

# A mixture of ordered probit models with endogenous assignment to two latent classes

Jochem Huismans

Amsterdam School of Economics

University of Amsterdam

Amsterdam, the Netherlands

jochemhuismans@gmail.com

Jan Willem Nijenhuis

Amsterdam School of Economics

University of Amsterdam

Amsterdam, the Netherlands

janwillemnijenhuis@gmail.com

Andrei Sirchenko

Amsterdam School of Economics

University of Amsterdam

Amsterdam, the Netherlands

andrei.sirchenko@gmail.com

## Abstract

Ordinal responses can be generated, in a time-series context, by different latent regimes or, in a cross-sectional context, by different unobserved classes of population. We introduce a new command `swopit` that fits a switching ordered probit model with either exogenous or endogenous class assignment to two latent regimes (classes). Switching is endogenous if the unobservables in the regime-switching (class-assignment) model are correlated with the unobservables in the outcome models. We provide a battery of postestimation commands, access by Monte Carlo experiments the small-sample performance of the maximum likelihood estimator of the parameters, choice and regime probabilities and their standard errors (both the asymptotic and bootstrap ones), and apply the new command to model the policy interest rates and health status responses.

**Keywords:** `swopit`, ordinal responses, ordered probit, mixture models, latent class, regime switching, endogenous switching

# 1 Introduction

Ordinal responses can be generated, in a cross-sectional context, by different latent classes (unobserved segments of population) or, in a time-series context, by different latent regimes. We introduce a new command `swopit` that fits a mixture of ordered probit (OP) models for ordinal outcomes with either exogenous and endogenous assignment to two latent classes (or regimes). The decision-making process, which determines an outcome in each class (regime), is represented by a separate OP model. The class assignment (regime switching) mechanism is represented by a binary probit model. Endogenous regime switching implies that the unobservables in the class assignment model are correlated with the unobservables in both outcome models. The three latent equations from the class assignment model and two outcome models, each with its own set of observables (control variables) and unobservables (disturbance terms), are estimated simultaneously by full information maximum likelihood (ML), providing the probabilities of both the discrete choices and the latent classes for each observation. In this way, observed explanatory variables can have different marginal effects on the choice probabilities in different classes.

The idea of endogenous switching can be traced to Roy's discussion of earnings distribution and self-selection between two professions (Roy 1951). Numerous variants of endogenous switching models are often concerned with sample selection or treatment effects. In the sample selection models (the first econometric implementations belong to Gronau 1974; Heckman 1974; Lewis 1974; and Maddala and Nelson 1975), which are similar in structure to the regime-switching and mixture models, the outcomes from one regime are never observed (not selected to the sample), whereas in the regime-switching modes the observed outcomes can be generated by any regime but the regimes may not be known. A two-regime switching model can be thus viewed as two selection models merged together.

The typical treatment-effects (or program-evaluation) models contain a selection-into-treatment equation and a single outcome equation, which in addition to the control variables also contains one or sometimes several dummy variables (treatment indicators). The rest of parameters in the outcome equation is the same for both treated and untreated individuals, who are observed. In this literature, endogenous switching refers to a possible correlation of treatment dummies with the unobservables in the outcome equation. For example, Geweke *et al.* (2003) consider unordered multiple treatment and binary outcomes; Munkin and Trivedi (2008) consider unordered multiple treatment and multiple ordered outcomes; Miranda and Rabe-Hesketh (2006) provide a "wrapper" program, `ssm`, that calls the Stata's command `gllamm` to fit the binary treatment (or selection) model for multiple ordered (or count) outcomes; and the Stata command `eoprobit` fits an ordered probit regression model with endogenous treatment assignment (or sample selection).

In contrast to the typical treatment-effects regression models, in the switching regression models (also known as finite mixture models) the regimes may not be observed, and the outcomes in each regime are handled separately with the regime-specific values of all parameters in each outcome model (see McLachlan (2019) for a recent survey). In this literature, endogenous switching refers to a possible correlation between the unobservables in the class assignment (switching) model and the unobservables in the outcome models. The existing endogenous switching mixture models deal mainly with continuous or binary outcomes. Models with ordinal or count outcomes have received considerably less attention and are mostly limited to observed regimes or known sample separation (e.g., Carneiro *et al.* 2003; Chib and Hamilton 2000; Gregory 2015; Hasebe 2018; Hill 1990). In particular, Gregory (2015) and Hasebe (2020) develop the Stata commands,

`switchprobitsim` and `escount`, which fit, respectively, a two-regime endogenous switching OP model and a Poisson (or negative binomial) model separately for two observed classes of treated and untreated individuals.

The model for ordinal outcomes with endogenous assignment to two *latent* classes is proposed by Greene *et al.* (2008). It can be seen as a generalization of the zero-inflated OP model of Harris and Zhao (2007), in which the outcomes in one latent class are modeled by the OP model, and the outcomes in the other latent class have a density mass at a single point (a zero), generating “inflated” zeros. The endogenous switching OP model in Greene *et al.* (2008) allows all outcomes, not only the zeros, to be generated in two regimes.

A mixture of OP models with exogenous assignment to latent classes, when the unobservables in the class assignment and output models are independent of each other, can be also fitted by the Stata command for finite mixture models `fmm: oprobit`. In contrast to the commands with the `fmm` prefix, which use the multinomial logistic distribution to model the probabilities of the latent classes, the `swopit` command uses the probit model. The `fmm` commands can fit mixtures of models for different types of outcomes (censored, ordinal, count, categorical, zero-inflated, etc.), but do not allow for interdependence among the unobservables in the class assignment and output models as the `swopit` command does.

In the next section we describe the mixture of two OP models with endogenous assignment to latent classes and its estimation via full information ML, and discuss the identification issues. In Section 3 we present the new Stata command. In Section 4 we describe and report the results of the Monte Carlo experiments to assess the small-sample properties of the ML estimator of parameters and probabilities, and to compare the asymptotic and bootstrap estimators of standard errors. We illustrate the new command in Section 5 by modeling the policy interest rates and health status responses. Section 6 concludes.

## 2 Model

We let  $i$  ( $t = 1, 2, \dots, N$ ) be one of the available  $N$  observations, and  $y_i$  be an observed dependent variable that can take on a finite number  $J$  of ordinal values coded by index  $j$  ( $j = 1, 2, \dots, J$ ). We let the latent unobserved (or only partially observed) variables be denoted by “\*”. The observed outcome  $y_i$  can be generated in one of two states, coded by an index  $r_i^*$  ( $r_i^* = 1, 2$ ) and interpreted as the latent classes of population in the cross-sectional context or as the latent regimes in the time-series context. The realized states are not observed. The latent state  $r_i^*$  is determined by the observed data  $\mathbf{z}_i$  and unobservables  $\nu_i$  according to the class assignment model in the usual binary probit fashion with an unobserved threshold  $\mu$ . For each  $i$ , only one out of two potential realizations of  $y_i$  is observed. The observed outcome  $y_i$  is determined conditionally on the regime  $r_i^*$  by the observed data  $\mathbf{x}_{r_i^*, i}$  and unobservables  $\varepsilon_{r_i^*, i}$  according to the OP outcome models in the usual ordered-response fashion with the unobserved thresholds  $\alpha_{r_i^*}$ . The unobservables in the class assignment model can be correlated with those in each outcome model.

We let:  $\mathbf{z}_i$ ,  $\mathbf{x}_{1,i}$  and  $\mathbf{x}_{2,i}$  be the  $i$ th rows of the observed data matrices;  $\gamma$ ,  $\beta_1$  and  $\beta_2$  be the column vectors of unknown slope parameters;  $\mu$ ,  $\alpha_1$  and  $\alpha_2$  be the unknown thresholds;  $\rho_1$  and  $\rho_2$  be the correlation coefficients;  $\nu_i$ ,  $\varepsilon_{1,i}$  and  $\varepsilon_{2,i}$  be the error terms that are independently and identically distributed (*iid*) across  $i$  according to the normal distributions with the zero means and the variances  $\sigma^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$ , respectively;  $\Phi(g_i; \sigma_i^2)$

be the normal cumulative distribution function (CDF) of the random variable  $g_i$  with zero mean and variance  $\sigma_i^2$ ; and  $\Phi_2(g_1; g_2; \zeta)$  be the CDF of the bivariate normal distribution of the two random variables  $g_1$  and  $g_2$  with the zero means, variances  $\sigma_1^2$  and  $\sigma_2^2$ , and the correlation coefficient  $\zeta$ .

The mixture of OP models with endogenous assignment to two latent classes (regimes) can be summarized by the following system:

$$\begin{aligned} \text{Class assignment} \\ \text{(regime switching):} \quad r_i^* &= \begin{cases} 1 & \text{if } \mathbf{z}_i\boldsymbol{\gamma} + \nu_i \leq \mu, \\ 2 & \text{if } \mu < \mathbf{z}_i\boldsymbol{\gamma} + \nu_i. \end{cases} \\ \\ \text{Outcome} \\ \text{models:} \quad y_i &= \begin{cases} j & \text{if } r_i^* = 1 \text{ and } \alpha_{1,j-1} < \mathbf{x}_{1,i}\boldsymbol{\beta}_1 + \varepsilon_{1,i} \leq \alpha_{1,j}, \\ j & \text{if } r_i^* = 2 \text{ and } \alpha_{2,j-1} < \mathbf{x}_{2,i}\boldsymbol{\beta}_2 + \varepsilon_{2,i} \leq \alpha_{2,j}, \end{cases} \quad (1) \\ \\ \text{Endo/exogeneity of} \\ \text{class assignment:} \quad \begin{bmatrix} \nu_i \\ \varepsilon_{s,i} \end{bmatrix} &\stackrel{iid}{\sim} \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho_s \sigma \sigma_s \\ \rho_s \sigma \sigma_s & \sigma_s^2 \end{bmatrix} \right), \quad s = 1, 2, \end{aligned}$$

where  $j = 1, 2, \dots, J$ ,  $-\infty = \alpha_{1,0} \leq \alpha_{1,1} \leq \dots \leq \alpha_{1,J} = \infty$  and  $-\infty = \alpha_{2,0} \leq \alpha_{2,1} \leq \dots \leq \alpha_{2,J} = \infty$ .

Conditional on the values of all independent variables in the model  $\mathbf{x}_i^{all} = (\mathbf{z}_i, \mathbf{x}_{1,i}, \mathbf{x}_{2,i})$ , the probability of the choice  $j$  is given by

$$\begin{aligned} \Pr(y_i = j | \mathbf{x}_i^{all}) &= \sum_{s=1}^{s=2} \Pr(r_i^* = s \text{ and } \alpha_{s,j-1} < \mathbf{x}_{s,i}\boldsymbol{\beta}_s + \varepsilon_{s,i} \leq \alpha_{s,j} | \mathbf{x}_i^{all}) \\ &= \Pr(\nu_i \leq \mu - \mathbf{z}_i\boldsymbol{\gamma} \text{ and } \alpha_{1,j-1} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1 < \varepsilon_{1,i} \leq \alpha_{1,j} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1 | \mathbf{x}_i^{all}) \\ &\quad + \Pr(\mu - \mathbf{z}_i\boldsymbol{\gamma} < \nu_i \text{ and } \alpha_{2,j-1} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2 < \varepsilon_{2,i} \leq \alpha_{2,j} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2 | \mathbf{x}_i^{all}) \\ &= \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{1,j} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \rho_1) - \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{1,j-1} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \rho_1) \\ &\quad + \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{2,j} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2) - \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{2,j-1} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2). \end{aligned} \quad (2)$$

More specifically, these probabilities can be computed for each choice as

$$\Pr(y_i = 1 | \mathbf{x}_i^{all}) = \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{1,1} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \rho_1) + \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{2,1} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2);$$

$$\Pr(y_i = j | \mathbf{x}_i^{all}) = \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{1,j} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \rho_1) - \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{1,j-1} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \rho_1) \\ + \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{2,j} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2) - \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; \alpha_{2,j-1} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2) \text{ for } 1 < j < J;$$

$$\Pr(y_i = J | \mathbf{x}_i^{all}) = \Phi_2(\mu - \mathbf{z}_i\boldsymbol{\gamma}; -\alpha_{1,J-1} + \mathbf{x}_{1,i}\boldsymbol{\beta}_1; -\rho_1) + \Phi_2(-\mu + \mathbf{z}_i\boldsymbol{\gamma}; -\alpha_{2,J-1} + \mathbf{x}_{2,i}\boldsymbol{\beta}_2; -\rho_2).$$

In the case of exogenous class assignment (when  $\rho_1 = \rho_2 = 0$ ), these probabilities simplify to

$$\Pr(y_i = j | \mathbf{x}_i^{all}, \rho_1 = \rho_2 = 0) = \Phi(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \sigma_\nu^2) [\Phi(\alpha_{1,j} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \sigma_1^2) - \Phi(\alpha_{1,j-1} - \mathbf{x}_{1,i}\boldsymbol{\beta}_1; \sigma_1^2)] \\ + [1 - \Phi(\mu - \mathbf{z}_i\boldsymbol{\gamma}; \sigma_\nu^2)] [\Phi(\alpha_{2,j} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; \sigma_2^2) - \Phi(\alpha_{2,j-1} - \mathbf{x}_{2,i}\boldsymbol{\beta}_2; \sigma_2^2)].$$

## 2.1 ML estimation and identification

The simultaneous estimation of the switching model and both outcome models can be performed using a full information ML estimator that maximizes the log-likelihood function  $l(\boldsymbol{\theta})$ :

$$\max_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta} | \mathbf{x}_i^{all}) = \max_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{i=N} \sum_{j=1}^{j=J} I_{ij} \ln[\Pr(y_i = j | \mathbf{x}_i^{all}, \boldsymbol{\theta})],$$

where  $\boldsymbol{\theta}$  is a vector of all parameters, including  $\boldsymbol{\gamma}$ ,  $\mu$ ,  $\boldsymbol{\beta}_1$ ,  $\boldsymbol{\beta}_2$ ,  $\boldsymbol{\alpha}_1$ ,  $\boldsymbol{\alpha}_2$ ,  $\rho_1$  and  $\rho_2$ ;  $\Theta$  is a parameter space;  $I_{ij}$  is an indicator function such that  $I_{ij} = 1$  if  $y_i = j$ , and  $I_{ij} = 0$  otherwise. The asymptotic standard errors of  $\hat{\boldsymbol{\theta}}$  can be computed from the Hessian matrix. There are three issues with the identification of the parameters: the first issue is typical for the models with limited dependent variables, the second is typical for the multiple-equation models, and the third issue is typical in the estimation of the mixture models.

Firstly, the intercept components of the slope parameters  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  are identified up to scale and location, that is, only jointly with the corresponding threshold parameters  $\mu$ ,  $\boldsymbol{\alpha}_1$  and  $\boldsymbol{\alpha}_2$ , and variances  $\sigma^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$ . Following the standard approach in the identification of discrete-choice models, we fix the variances  $\sigma^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$  to 1, and the intercept components of  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\beta}_2$  to 0. Importantly, the probabilities in (2) are absolutely invariant to these (arbitrary) identifying assumptions.

Secondly, we can actually identify all parameters in  $\boldsymbol{\theta}$  (again, up to scale and location) because of the nonlinearity of OP models, i.e. via the functional form (Heckman 1978; Wilde 2000). However, the ML estimation may be subject to a weak identification problem (a low precision of the estimates) if the class assignment model and outcome models contain the same or similar sets of covariates.

The third issue is a well known label-switching phenomenon, caused by the invariance of the likelihood function with respect to the switching of the latent regime labels (Redner and Walker 1984; Jasra *et al.* 2005). Since labeling of latent regimes is arbitrary, if both outcome models contain the same set of regressors ( $\mathbf{x}_{1,i} = \mathbf{x}_{2,i}$ ) we can say neither that  $\hat{\boldsymbol{\beta}}_1$  and  $\hat{\boldsymbol{\alpha}}_1$  indeed estimate  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\alpha}_1$  (but not  $\boldsymbol{\beta}_2$  and  $\boldsymbol{\alpha}_2$ ) nor that  $\hat{\boldsymbol{\beta}}_2$  and  $\hat{\boldsymbol{\alpha}}_2$  estimate  $\boldsymbol{\beta}_2$  and  $\boldsymbol{\alpha}_2$  (but not  $\boldsymbol{\beta}_1$  and  $\boldsymbol{\alpha}_1$ ). The model has two identical likelihood maxima because

$$\begin{aligned} & \Pr(y_i = j | \mathbf{x}_i^{all}, -\boldsymbol{\gamma}, -\mu, \boldsymbol{\beta}_2, \boldsymbol{\alpha}_2, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, -\rho_2, -\rho_1) \\ &= \Phi_2(-\mu + \mathbf{z}_i \boldsymbol{\gamma}; \boldsymbol{\alpha}_{2,j} - \mathbf{x}_{2,i} \boldsymbol{\beta}_2; -\rho_2) - \Phi_2(-\mu + \mathbf{z}_i \boldsymbol{\gamma}; \boldsymbol{\alpha}_{2,j-1} - \mathbf{x}_{2,i} \boldsymbol{\beta}_2; -\rho_2) \\ &+ \Phi_2(\mu - \mathbf{z}_i \boldsymbol{\gamma}; \boldsymbol{\alpha}_{1,j} - \mathbf{x}_{1,i} \boldsymbol{\beta}_1; \rho_1) - \Phi_2(\mu - \mathbf{z}_i \boldsymbol{\gamma}; \boldsymbol{\alpha}_{1,j-1} - \mathbf{x}_{1,i} \boldsymbol{\beta}_1; \rho_1) \\ &= \Pr(y_i = j | \mathbf{x}_i^{all}, \boldsymbol{\gamma}, \mu, \boldsymbol{\beta}_1, \boldsymbol{\alpha}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}_2, \rho_1, \rho_2). \end{aligned}$$

Hence, changing the signs of the parameters in the class assignment model and the signs of the correlation coefficients switches the regime labels, and results in two identical likelihood values. Thus, the correlation coefficients and the slope and threshold parameters in the class assignment model are identified up to sign only, and the latent class labels are interchangeable.

In mixture models, the likelihood function may have multiple local maxima. To avoid the locally optimal solutions the developed command performs several estimation attempts (five by default) with different initializations by randomly assigning observations to each class (regime). After each random initialization, the command obtains the starting values for the slope and threshold parameters using the independent estimations of binary probit class assignment model and OP output models. Further, in the case of endogenous class assignment, the command obtains the starting values for  $\rho_1$  and  $\rho_2$  by maximizing the likelihood functions over a grid search from  $-0.95$  to  $0.95$  in increments of  $0.05$  holding the other parameters fixed at their estimates in the exogenous switching case, following Sirchenko (2020) and Dale and Sirchenko (2021)<sup>1</sup>. Using these starting values the command then performs ML estimation. The estimation results with the highest obtained likelihood are reported. The Monte Carlo simulations show that the ML estimator arrives at the global maximum and is consistent.

---

<sup>1</sup>We are thankful to David Dale, whose Stata command `ziop2` for the zero-inflated OP model from Dale and Sirchenko (2021) we adopted and modified.

## 2.2 Partial effects

We assemble the partial effects (PE) of each covariate on the probability of each choice into a matrix  $\mathbf{PE}_i$ , in which rows are linked with covariates and columns are linked with choices. The PE of a continuous-valued covariate on the choice  $j$  is computed as follows:

$$\begin{aligned} \mathbf{PE}_{k,j,i} = & + \left[ \Phi \left( \frac{\mu - \mathbf{z}_i \gamma - \rho_1 (\alpha_{1,j-1} - \mathbf{x}_{1,i} \beta_1)}{\sqrt{1 - \rho_1^2}} \right) f(\alpha_{1,j-1} - \mathbf{x}_{1,i} \beta_1) \right. \\ & - \Phi \left( \frac{\mu - \mathbf{z}_i \gamma - \rho_1 (\alpha_{1,j} - \mathbf{x}_{1,i} \beta_1)}{\sqrt{1 - \rho_1^2}} \right) f(\alpha_{1,j} - \mathbf{x}_{1,i} \beta_1) \left. \right] \beta_{1,k}^{all} \\ & + \left[ \Phi \left( \frac{\alpha_{1,j} - \mathbf{x}_{1,i} \beta_1 - \rho_1 (\mu - \mathbf{z}_i \gamma)}{\sqrt{1 - \rho_1^2}} \right) - \Phi \left( \frac{\alpha_{1,j-1} - \mathbf{x}_{1,i} \beta_1 - \rho_1 (\mu - \mathbf{z}_i \gamma)}{\sqrt{1 - \rho_1^2}} \right) \right] f(\mu - \mathbf{z}_i \gamma) \gamma_k^{all} \\ & + \left[ \Phi \left( \frac{\mathbf{z}_i \gamma - \mu + \rho_2 (\alpha_{2,j-1} - \mathbf{x}_{2,i} \beta_2)}{\sqrt{1 - \rho_2^2}} \right) f(\alpha_{2,j-1} - \mathbf{x}_{2,i} \beta_2) \right. \\ & - \Phi \left( \frac{\mathbf{z}_i \gamma - \mu + \rho_2 (\alpha_{2,j} - \mathbf{x}_{2,i} \beta_2)}{\sqrt{1 - \rho_2^2}} \right) f(\alpha_{2,j} - \mathbf{x}_{2,i} \beta_2) \left. \right] \beta_{2,k}^{all} \\ & + \left[ \Phi \left( \frac{\alpha_{2,j} - \mathbf{x}_{2,i} \beta_2 + \rho_2 (\mathbf{z}_i \gamma - \mu)}{\sqrt{1 - \rho_2^2}} \right) - \Phi \left( \frac{\alpha_{2,j-1} - \mathbf{x}_{2,i} \beta_2 + \rho_2 (\mathbf{z}_i \gamma - \mu)}{\sqrt{1 - \rho_2^2}} \right) \right] f(\mathbf{z}_i \gamma - \mu) \gamma_k^{all}, \end{aligned}$$

where  $f$  is the probability density function of the standard normal distribution, and  $\gamma_k^{all}$ ,  $\beta_{1,k}^{all}$  and  $\beta_{2,k}^{all}$  are the coefficients on the  $k$ th covariate in  $\mathbf{x}_i^{all}$  in the class assignment model, the outcome model conditional on  $r_i^* = 1$  and the outcome model conditional on  $r_i^* = 2$ , respectively ( $\gamma_k^{all}$ ,  $\beta_{1,k}^{all}$  or  $\beta_{2,k}^{all}$  is zero if the  $k$ th covariate in  $\mathbf{x}_i^{all}$  is not included into the corresponding model). For a discrete-valued covariate, the PE can be computed as the change in the probabilities when this variable changes by one increment and all other variables are held fixed. The asymptotic standard error of  $\mathbf{PE}_{k,j,i}$  is computed using the delta method as a square root of  $\nabla_{\theta}(\mathbf{PE}_{k,j,i}) \mathbf{Var}(\boldsymbol{\theta}) \nabla_{\theta}(\mathbf{PE}_{k,j,i})'$ .

## 3 The swopit command

The accompanying software includes the `swopit` command, postestimation commands and supporting help files.

### 3.1 Syntax

The following command fits a mixture of OP models with either exogenous or endogenous assignment to two latent classes (regimes):

```
swopit depvar [indepvars] [if] [in] [, options]
```

The dependent variable *depvar* may take on two or more discrete ordered values. The independent variables listed in *indepvars* will be, by default, included in each model. The alternative (and possibly not the same) lists of independent variables to be included in the class assignment model and each outcome model can be specified in *options*. The following options are available.

### Options

The underlying denotes the shortest acceptable abbreviation.

**regindepvars**(*varlist*) specifies the list of independent variables included in the class assignment model. By default, it is equal to all independent variables listed in *indepvars*.

**outoneindepvars**(*varlist*) specifies the list of independent variables included in the first outcome model. By default, it is equal to all independent variables listed in *indepvars*.

**outtwoindepvars**(*varlist*) specifies the list of independent variables included in the second outcome model. By default, it is equal to all independent variables listed in *indepvars*.

**endogenous** specifies that the endogenous class assignment (regime switching) is to be used instead of the default exogenous switching.

**guesses**(*scalar*) specifies the number of estimation attempts with different random starting values using the command **optimize()**. The default is **guesses(5)**. At each attempt, the following optimization techniques are applied one after another until convergence is achieved or all four of them are used: modified Newton-Raphson (NR), Berndt-Hall-Hall-Hausman (BHHH), Davidson-Fletcher-Powell (DFP) and Broyden-Fletcher-Goldfarb-Shanno (BFGS). The estimation output with the highest likelihood is reported. If starting values are specified in **initial()** then the estimation attempts are stopped after the first converged one (if any).

**limit**(*scalar*) specifies the limit for the maximum absolute value of each parameter in the ML estimation. The default is **limit(0)**, and no constraints are applied.

**maxiter**(*scalar*) specifies the maximum number of iterations before the optimization algorithm quits. The default is **maxiter(500)**.

**ptol**(*scalar*) specifies the tolerance for parameters. The default is **ptol(1e - 6)**.

**vtol**(*scalar*) specifies the tolerance for log likelihood. The default is **vtol(1e - 7)**.

**nrtol**(*scalar*) specifies the tolerance for scaled gradient. The default is **nrtol(1e - 5)**.

**initial**(*string*) specifies a space-delimited list of the starting values for the parameters in the following order:  $\gamma$ ,  $\mu$ ,  $\beta_1$ ,  $\alpha_1$ ,  $\beta_2$ ,  $\alpha_2$ ,  $\rho_1$  and  $\rho_2$ . The elements of  $\alpha_1$  and  $\alpha_2$  should be provided in the ascending order.

**change**(*scalar*) specifies the interval for randomly selecting new starting values (SV) for the next estimation attempt if the user has specified the starting values in **initial()**. The estimation attempts are stopped after the first converged one or until all attempts specified in **guesses()** are performed. The SV for all coefficients with the exception of the correlation coefficients are adjusted for each estimation attempt according to the formula:  $SV = SV + change * U(-|SV|, |SV|)$ , where  $U()$  represents a uniformly distributed random variable. In the case of endogenous switching, the SV for the correlation coefficients  $\rho_1$  and  $\rho_2$  are determined by maximizing the likelihood function over a grid search from  $-0.95$  to  $0.95$  in increments of  $0.05$  holding the other parameters fixed. The default is **change(0.5)**. The option is ignored if the **initial()** option is not used. However, it is always applied for the bootstrap.

**bootstrap**(*scalar*) specifies the number of bootstrap replications to be performed to estimate the standard errors. Bootstrapping uses the initial values of estimated parameters as the starting ones. The default is **bootstrap**(0), and no bootstrapping is performed.

**bootguesses**(*scalar*) specifies the maximum number of attempts with different random starting values in the bootstrap estimations. At each new attempt, the starting values are selected as described in **change**() and the following optimization techniques are applied one after another until convergence is achieved or all four of them are used: NR, BHHH, DFP and BFGS. The estimation attempts are stopped after the first converged one or until all attempts specified in **bootguesses**() are performed. The default is **bootguesses**(3).

**bootiter**(*scalar*) specifies the maximum number of iterations in the bootstrap estimations before the optimization algorithm quits. The default is **bootiter**(50).

## Stored results

The descriptions of the stored results can be found in the help files.

## 3.2 Postestimation commands

The following postestimation commands are available after **swopit**.

### The **swopitpredict** command

**swopitpredict** *varname* [, **regimes** **output**(*string*)]

This command provides the predicted probabilities of the observed choices (by default) or latent classes for each observation. It creates the variables named *varname\_i* where *i* is the label of the observed choice or latent class. *varname* can only consist of letters and underscores. If an invalid name is given an error message is displayed. The options **regimes** and **output**(*string*) are defined as:

**regimes** calculates the predicted probabilities of the latent classes (regimes) instead of the choice probabilities (by default).

**output**(*string*) specifies the different types of predictions. The possible options for *string* are: *choice* for reporting the predicted outcome (the choice or regime with the largest predicted probability); *mean* for reporting the expected value of the dependent variable computed as a summation of  $j * Pr(y_i = j)$  across all choices  $j = 1, 2, \dots, J$ ; and *cum* for predicting the cumulative choice probabilities such as  $Pr(y_i \leq 1), Pr(y_i \leq 2), \dots$ . If **output**() is not specified, the usual choice probabilities such as  $Pr(y_i = 1), Pr(y_i = 2), \dots$  are predicted.

### The **swopitprobabilities** command

**swopitprobabilities** [, **at**(*string*) **regimes**]

This command provides the predicted probabilities of the observed choices (by default) or latent classes with their standard errors for the specified values of the independent variables. The options **at**(*string*) and **regimes** are defined as follows:



`at(string)` specifies the values of the independent variables, at which the probabilities are estimated. By default, the probabilities are computed at the median values of the independent variables. The syntax of this command is *varname = value* for each variable, separated by a blank space. *varname* is the name of the variable listed in *indepvars*. If an independent variable from *indepvars* is excluded from this option, the probabilities are estimated at the median value of this variable.

`regimes` calculates the predicted probabilities of the latent classes (regimes) instead of the choice probabilities (by default).

### The `swopitmargins` command

`swopitmargins` [, `at(string)` `regimes`]

This command provides the marginal (partial) effects on the predicted probabilities of the observed choices (by default) or latent classes with their standard errors for the specified values of the independent variables. The options `at(string)` and `regimes` are defined as follows:

`at(string)` specifies the values of the independent variables to estimate the marginal effects. By default, the marginal effects are computed at the median values of the independent variables. The syntax of this command is *varname = value* for each variable, separated by a blank space. *varname* is the name of the variable listed in *indepvars*. If an independent variable from *indepvars* is excluded from this option, the marginal effects are estimated at the median value of this variable.

`regimes` calculates the marginal effects on the probabilities of the latent classes (regimes) instead of the choice probabilities (by default).

### The `swopitclassification` command

`swopitclassification`

This command constructs a confusion matrix (classification table) for the dependent variable. The classification table shows the observed choices in the rows and the predicted ones (the choices with the highest predicted probability) in the columns. The diagonal elements give the numbers of correctly predicted choices. The command also reports the accuracy (the percentage of correct predictions), the Brier probability score (Brier 1950), the ranked probability score (Epstein 1969), the precisions, the recalls and the adjusted noise-to-signal ratios (Kaminsky and Reinhart 1999).

## 4 Monte Carlo experiments

We have performed the Monte Carlo experiments to assess the small sample performance of the proposed ML estimator of the parameters and the probabilities of choices and latent classes as well as of their standard errors (both the asymptotic and bootstrap ones). The performance is measured by the bias (the absolute difference between the estimated and true values), the root mean square error (RMSE) and the coverage rate (the percentage of times the estimated 95% confidence intervals cover the true values)<sup>2</sup>.

---

<sup>2</sup>The “true values” of the standard errors are computed as the standard deviations of the parameters’ estimates in Monte Carlo replications.

The results of the experiments show that as the sample size grows the biases and RMSE decrease, and the coverage rates approach the nominal level. The results suggest that the proposed ML estimator is consistent and quite reliable in small samples (with only 250–500 observations).

## 4.1 Monte Carlo design

We generated five independent variables  $g_j \stackrel{iid}{\sim} 4 * \mathcal{N}(0, 1)$ ,  $j \in \{1, \dots, 5\}$  and simulated the following three different scenarios for the overlap among the covariates in the two outcome models, “no overlap”, “partial overlap”, and “complete overlap”:

	no overlap	partial overlap	complete overlap
Class assignment model	$g_1$	$g_1, g_3$	$g_1$
Outcome model 1	$g_2, g_3$	$g_2, g_3$	$g_2, g_3$
Outcome model 2	$g_4, g_5$	$g_3, g_4$	$g_2, g_3$

In the “partial overlap” case, there is a covariate that belongs to all three latent equations. We simulated six data-generating processes according to system (1), using three overlap scenarios and two types of class assignment: exogenous (when  $\rho_1 = \rho_2 = 0$ ) and endogenous (when  $\rho_1 = 0.3$  and  $\rho_2 = 0.5$ ). The dependent variable  $y$  was generated with three ordinal choices. The values of the parameters used in the simulations are shown in Table A1 of Appendix. These values were calibrated in such a way that both latent classes have the same theoretical probability, and each of three choices is equally likely. The variances of the error terms were fixed to one. 10,000 repeated samples with 250, 500, 1,000 and 2,000 observations were independently generated for each process, and estimated using its true specification.

In addition, to study the performance of the bootstrap estimator of the standard errors, we simulated two processes with “no overlap” specification, one with exogenous and the other with endogenous class assignment, and estimated the parameters of each process using both the asymptotic and bootstrap estimators of the standard errors. We generated 3,000 replications independently with 250 and 500 observations. To compute a nonparametric stratified bootstrap estimator of standard errors, we drew with replacement 300 bootstrap samples for each Monte Carlo iteration, and computed the standard deviations of the replicated statistics. For each bootstrap sample, the observations were re-sampled separately for each choice, so the number of observations in each choice in every bootstrap sample was the same as in the original one.

## 4.2 Monte Carlo results

Table 1 reports the small-sample performance of the estimator of the parameters. The slope parameters  $\gamma$ ,  $\beta_1$  and  $\beta_2$  are estimated better than the threshold parameters  $\mu$ ,  $\alpha_1$  and  $\alpha_2$  and much better than the correlation coefficients  $\rho_1$  and  $\rho_2$ . As the sample sizes increase eight times from 250 to 2,000:

(i) the coverage rates move toward the nominal 95% level: for the slope and threshold parameters from 87% to 94% in the endogenous switching models and from 97% to 95% in the exogenous switching ones, whereas for the correlation coefficients from 28%–30% to 75%–79% only;

(ii) for the slope parameters the biases decrease 14–16 times, the RMSE decrease 9–13 times, and the biases of the standard error estimates decrease 143–257 times;

Table 1. The small-sample performance of the estimator of parameters

Sample size	Class assignment: Covariates' overlap:	Exogenous			Endogenous		
		none	partial	complete	none	partial	complete
Slope parameters $\gamma$ , $\beta_1$ and $\beta_2$							
250	Bias, x10	5.5	5.6	5.4	4.8	4.5	4.7
500		1.7	1.7	1.7	1.6	1.6	1.6
1000		0.8	0.7	0.8	0.7	0.7	0.7
2000		0.3	0.3	0.3	0.3	0.3	0.3
250	RMSE, x10	18.2	18.6	16.6	13.3	12.0	12.9
500		4.6	4.3	4.6	4.3	4.1	4.4
1000		2.3	2.2	2.3	2.3	2.2	2.3
2000		1.5	1.4	1.5	1.4	1.4	1.5
250	Coverage rate (at 95% level), %	97.3	97.3	97.4	87.6	86.1	87.2
500		96.3	96.5	96.4	89.9	87.9	89.8
1000		95.5	95.6	95.4	92.3	90.9	92.6
2000		95.3	95.3	95.4	94.3	93.7	94.1
250	Bias of standard error estimator, x100	95.6	111.8	69.5	46.7	55.9	59.7
500		10.5	8.6	9.7	8.2	8.5	9.9
1000		1.7	1.5	1.6	1.9	1.8	1.5
2000		0.4	0.4	0.4	0.3	0.3	0.4
Threshold parameters $\mu$ , $\alpha_1$ and $\alpha_2$							
250	Bias, x10	26.7	25.0	27.1	22.4	21.5	23.0
500		13.3	12.7	13.2	12.6	11.7	12.3
1000		8.4	8.2	8.3	8.2	7.9	8.2
2000		5.7	5.5	5.6	5.7	5.5	5.7
250	RMSE, x10	27.7	27.0	26.6	24.2	22.8	25.5
500		9.3	8.4	9.0	8.8	8.3	8.9
1000		4.9	4.7	4.8	4.8	4.7	4.8
2000		3.1	3.0	3.0	3.0	3.0	3.0
250	Coverage rate (at 95% level), %	97.2	97.0	97.4	86.0	87.5	85.3
500		96.2	95.9	96.3	88.2	88.4	88.1
1000		95.6	95.3	95.4	91.4	91.0	91.6
2000		95.2	95.1	95.3	94.0	93.5	93.6
250	Bias of standard error estimator, x100	120.2	123.5	106.5	108.4	100.3	119.1
500		18.2	12.5	15.5	17.1	17.4	20.0
1000		3.1	2.7	3.0	4.0	4.3	3.5
2000		0.9	0.7	0.7	0.6	0.8	0.8
Correlation coefficients $\rho_1$ and $\rho_2$							
250	Bias, x10				8.4	9.3	8.8
500					6.9	6.7	7.1
1000					6.5	6.5	6.9
2000					5.4	5.4	5.7
250	RMSE, x10				8.8	8.9	8.7
500					7.4	7.5	7.3
1000					5.6	5.8	5.6
2000					3.9	4.0	3.9
250	Coverage rate (at 95% level), %				28.2	28.1	32.4
500					43.4	40.2	46.1
1000					60.9	58.9	64.5
2000					75.0	72.8	78.6
250	Bias of standard error estimator, x100				41.8	36.6	39.1
500					37.8	39.7	36.2
1000					22.7	24.3	20.9
2000					11.3	12.6	10.0

(iii) the biases of the slope parameter estimator relative to those of the threshold and correlation estimators decrease, respectively, around three and a half times and nine times;

(iv) the RMSE of the slope parameter estimator relative to those of the threshold estimator decrease by 23%–33% in the exogenous switching models and by 6%–13% in the endogenous switching ones, and relative to those of the correlation coefficient estimator decrease around four times; and

(v) the biases of the standard error estimator for the slope parameters relative to those for the thresholds decrease by 10%–59% (but increase by 26% for “no overlap” case with endogenous switching), and relative to those for the correlation coefficients decrease 39–58 times.

The standard errors are underestimated for all parameters, all models and all sample sizes. The coverage rates for all parameters are below the nominal level in the endogenous regime-switching case, but nevertheless above the nominal level in the exogenous case due to the right-skewed distribution of the standard error estimates.

Table 2 reports the small-sample performance of the estimator of the probabilities of choices and latent classes. In the endogenous switching models, the estimator of class probabilities demonstrates better accuracy and certainty than the estimator of choice probabilities, while in the exogenous switching models the estimator of the class probabilities provides better accuracy but slightly less certainty than the estimator of choice probabilities. As the sample sizes grow from 250 to 2,000:

(i) the coverage rates of the class probability estimator move toward the nominal 95% level from 86%–92% to 94%–95%, whereas those of the choice probability estimator move from 86%–90% to 94%–95% in the exogenous switching models, but only slightly move in the interval 80%–85% in the endogenous switching ones;

(ii) the biases of the class probability estimator are smaller by 44%–99% than those of the choice probabilities for all overlap scenarios, all sample sizes and both switching types;

(iii) the RMSE of the class probability estimator are smaller by 23%–56% than those of the choice probability estimator in the endogenous switching models, but are more or less similar in the exogenous switching models (smaller by 17% in the “no overlap” case, larger by 28% in the “partial overlap” case and larger by 15% in the “complete overlap” case than those of the choice probability estimator for all sample sizes);

(iv) the biases of the class probability estimator relative to those of the choice probability estimator change in different directions for different overlap scenarios, sample sizes and types of switching, and no clear patterns are observed;

(v) the RMSE of the class probability estimator relative to those of the choice probability estimator decrease slightly by 1%–15% in the exogenous switching models, and more substantially by 34%–37% in the endogenous switching ones; and

(vi) the biases of the standard error estimator for the class probabilities are on average larger by more than 60% in the exogenous switching models but lower by about 95% (with 1000 and 2000 observations) in the exogenous switching models; in the exogenous switching models the biases of the standard error estimator for the class probabilities relative to those for the choice probabilities change in different directions for different overlap scenarios and different sample sizes, and no clear patterns are observed; in the endogenous switching models the relative biases of the standard error estimator for the class probabilities decrease 37–47 times.

Table 2. The small-sample performance of the estimator of probabilities

Sample	Class assignment:	Exogenous			Endogenous		
size	Covariates' overlap:	none	partial	complete	none	partial	complete
Probabilities of choices							
250	Bias, x1000	9.79	3.25	11.12	36.20	47.56	35.95
500		4.57	3.06	6.04	13.13	24.71	13.74
1000		3.15	1.78	2.76	5.77	6.58	5.31
2000		1.19	0.53	1.34	9.65	3.49	8.01
250	RMSE, x10	1.57	1.63	1.25	2.01	2.06	2.00
500		1.04	1.04	0.77	1.64	1.65	1.67
1000		0.73	0.69	0.51	1.27	1.24	1.30
2000		0.50	0.47	0.35	0.88	0.85	0.90
250	Coverage rate (at 95% level), %	85.6	86.9	89.7	82.5	82.9	80.9
500		90.5	91.5	92.9	80.4	84.2	79.4
1000		92.5	93.6	94.0	77.3	83.2	78.7
2000		94.0	94.3	94.8	80.7	85.3	84.0
250	Bias of standard error estimator, x100	1.68	1.86	1.49	1.63	1.20	1.33
500		0.52	0.64	0.43	3.70	3.26	3.88
1000		0.21	0.14	0.11	3.69	3.00	3.42
2000		0.03	0.07	0.03	2.29	1.77	2.01
Probabilities of latent classes (regimes)							
250	Bias, x1000	2.93	0.97	1.68	1.29	8.60	6.58
500		0.76	0.68	0.27	0.30	5.80	2.07
1000		0.21	0.83	0.28	0.48	3.05	1.01
2000		0.67	0.18	0.04	0.02	0.86	0.34
250	RMSE, x10	1.45	2.10	1.45	1.43	1.59	1.41
500		0.87	1.34	0.87	0.87	0.97	0.87
1000		0.58	0.88	0.58	0.58	0.63	0.58
2000		0.39	0.61	0.40	0.39	0.43	0.40
250	Coverage rate (at 95% level), %	91.3	85.6	91.8	90.2	88.7	89.9
500		94.2	91.3	94.1	93.2	92.7	93.1
1000		94.6	93.4	94.3	94.1	93.9	94.1
2000		94.9	94.0	94.8	94.8	94.1	94.2
250	Bias of standard error estimator, x100	2.48	3.80	2.39	1.99	2.47	1.88
500		0.61	1.26	0.56	0.51	0.72	0.52
1000		0.17	0.31	0.18	0.18	0.27	0.17
2000		0.05	0.20	0.06	0.06	0.10	0.08

Table 3 reports the small-sample performance of the asymptotic and bootstrap estimators of standard errors. The slope and threshold parameters' estimates do not benefit from bootstrapping: the biases of the bootstrap estimator are 1.5–13 times larger than those of the asymptotic estimator, and the coverage rates of the bootstrap estimator are close to 99%. However, the estimates of correlation coefficients do benefit from bootstrapping: the coverage rates move drastically towards the nominal level (from 28% and 42% to 88% for both sample sizes), and the biases of standard error estimator decrease by 75%.

More importantly, the probabilities' estimates in general benefit from bootstrapping as well: the coverage rates move closer to the nominal level (by 1–3 percentage points for class probabilities and by 2–10 percentage points for choice probabilities), and the

biases of standard error estimator decrease by 53%–95% for class probabilities for both types of switching, and by 40%–47% for choice probabilities in the exogenous switching model. In the endogenous switching model, the bias of standard error estimator of choice probabilities decreases by 46% with 500 observations but increases by 43% with 250 observations.

Table 3. Comparison of the asymptotic and bootstrap estimators of standard errors

Class assignment:		Exogenous		Endogenous	
Sample size:		250	500	250	500
Slope parameters $\gamma$ , $\beta_1$ and $\beta_2$					
Coverage rate (at 95% level), %	Asymptotic	97.3	96.6	87.3	89.2
	Bootstrap	98.9	99.1	98.8	98.9
Bias of standard error estimator	Asymptotic	1.44	0.07	0.34	0.09
	Bootstrap	16.26	1.03	4.18	0.34
Threshold parametrs $\mu$ , $\alpha_1$ and $\alpha_2$					
Coverage rate (at 95% level), %	Asymptotic	96.9	96.0	85.3	87.6
	Bootstrap	98.9	99.2	98.8	98.7
Bias of standard error estimator	Asymptotic	1.01	0.15	0.65	0.19
	Bootstrap	5.19	0.60	1.10	0.27
Correlation coefficients $\rho_1$ and $\rho_2$					
Coverage rate (at 95% level), %	Asymptotic			28.0	42.0
	Bootstrap			88.6	88.1
Bias of standard error estimator	Asymptotic			0.47	0.40
	Bootstrap			0.12	0.10
Class (regime) probabilities					
Coverage rate (at 95% level), %	Asymptotic	91.2	93.3	90.5	92.0
	Bootstrap	93.2	95.3	91.3	94.6
Bias of standard error estimator, x100	Asymptotic	2.34	0.99	1.70	0.99
	Bootstrap	0.73	0.25	0.79	0.05
Choice probabilities					
Coverage rate (at 95% level), %	Asymptotic	85.5	90.2	82.8	80.3
	Bootstrap	89.9	92.0	86.4	90.5
Bias of standard error estimator, x100	Asymptotic	1.58	0.72	2.14	4.09
	Bootstrap	0.95	0.38	3.05	2.20

## 5 Examples

### 5.1 Policy interest rate

We apply the new command to model the decisions of the U.S. Federal Open Market Committee (FOMC) to decrease, leave unchanged or increase the federal funds rate target. In the sample of 279 observations from 1989 to 2019, the FOMC decisions (both scheduled and unscheduled ones) are matched with the values of the explanatory variables as they were observed one day before each FOMC decision. We employ the follow-

ing explanatory variables: **house** (the Greenbook projection of the total number of new privately owned housing units started for the current quarter)<sup>3</sup>; **gdp** (the Greenbook projection of quarterly growth in the nominal gross domestic (before 1992: national) product for the current quarter, annualized percentage points)<sup>4</sup>; **spread** (the difference between the one-year treasury constant maturity rate and the effective federal funds rate, three-business-day moving average)<sup>5</sup>; and **bias** (the indicator that we constructed from the “policy bias” statements at the previous FOMC meeting: it equals 1 if the statement was asymmetric toward tightening, 0 if the statement was symmetric, and -1 if the statement was asymmetric toward easing)<sup>6</sup>.

First, we fit the standard OP model using the `oprobit` command:

```
. use "policy_rate.dta", clear
. oprobit y house gdp bias spread, nolog

Ordered probit regression               Number of obs   =       279
                                      LR chi2(4)         =      189.38
                                      Prob > chi2        =       0.0000
Log likelihood = -153.35319             Pseudo R2        =       0.3818
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
house		.7062818	.2644774	2.67	0.008	.1879157	1.224648
gdp		.1737146	.0457331	3.80	0.000	.0840793	.2633499
bias		.5146708	.1300754	3.96	0.000	.2597277	.7696138
spread		1.906444	.2110997	9.03	0.000	1.492697	2.320192
/cut1		.2311623	.3433395			-.4417708	.9040954
/cut2		3.422302	.4426521			2.55472	4.289884

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	279	-248.0452	-153.3532	6	318.7064	340.4937

Note: N=Obs used in calculating BIC; see [R] BIC note.

Next, we let the outcomes be generated in two latent regimes, and we fit the mixture of OP models with exogenous switching. The `swopit` command yields the following results<sup>7</sup>:

```
. swopit y house gdp bias spread, reg(house gdp) outone(spread bias) outtwo(spread bias)
(output omitted)
Two-regime switching ordered probit regression
Regime switching:      Exogenous
Number of observations =       279
Log likelihood         = -132.6306
McFadden pseudo R2    =    0.4653
LR chi2( 9)           =   230.8291
Prob > chi2           =    0.0000
AIC                   =   287.2612
BIC                   =   327.2046
```

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
--	---	-------	-----------	---	------	----------------------	--

<sup>3</sup>Data source: RTDSM (Real-Time Data Set for Macroeconomists), available at <https://www.philadelphiafed.org>.

<sup>4</sup>Data source: RTDSM.

<sup>5</sup>Data source: ALFRED (Archival Federal Reserve Economic Data), available at <https://alfred.stlouisfed.org/>.

<sup>6</sup>Raw data source: the FOMC statements and minutes, available at [https://www.federalreserve.gov/monetarypolicy/fomc\\_historical.htm](https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm).

<sup>7</sup>An identical value of the likelihood can be obtained with the opposite signs of the parameters in the class assignment model and switched labels of the latent regimes.

Regime equation							
house		4.911731	1.386307	3.54	0.000	2.19462	7.628842
gdp		.6790069	.2028345	3.35	0.001	.2814585	1.076555
/cut1		11.96884	2.980508	4.02	0.000	6.127154	17.81053
-----							
Outcome equation 1							
spread		1.819521	.2403311	7.57	0.000	1.348481	2.290561
bias		.6798781	.1525716	4.46	0.000	.3808432	.9789129
/cut1		-1.251123	.1463508	-8.55	0.000	-1.537965	-.9642805
/cut2		2.318764	.2520933	9.20	0.000	1.82467	2.812858
-----							
Outcome equation 2							
spread		14.12316	3.43147	4.12	0.000	7.397605	20.84872
bias		3.954395	1.183196	3.34	0.001	1.635374	6.273416
/cut1		-16.28817	4.239138	-3.84	0.000	-24.59672	-7.979607
/cut2		1.448363	.6056882	2.39	0.017	.2612362	2.63549
-----							

The empirical evidence in favor of regime switching is convincing: the switching OP model substantial improves the likelihood, and is preferred to the standard OP model according to both AIC and BIC.

Next, we let the unobservables in the regime switching model be correlated with the unobservables in the outcome models. The `swopit` command with endogenous switching yields the following results<sup>8</sup>:

```
. swopit y house gdp bias spread, reg(house gdp) outone(spread bias) outtwo(spread bias) endo
      (output omitted)
```

Two-regime switching ordered probit regression

Regime switching:	Endogenous
Number of observations =	279
Log likelihood	= -132.2678
McFadden pseudo R2	= 0.4668
LR chi2(11)	= 231.5549
Prob > chi2	= 0.0000
AIC	= 290.5355
BIC	= 337.7413

  

	y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
-----						
Regime equation						
house		-4.804476	1.412871	-3.40	0.001	-7.573652 -2.0353
gdp		-.6180352	.1823049	-3.39	0.001	-.9753462 -.2607243
/cut1		-11.48332	2.989303	-3.84	0.000	-17.34225 -5.624396
-----						
Outcome equation 1						
spread		12.40142	4.229464	2.93	0.003	4.111827 20.69102
bias		3.389226	1.50522	2.25	0.024	.4390492 6.339402
/cut1		-14.70935	4.594482	-3.20	0.001	-23.71437 -5.704335
/cut2		1.092356	1.04458	1.05	0.296	-.9549827 3.139694
-----						
Outcome equation 2						
spread		1.771607	.2478844	7.15	0.000	1.285763 2.257451
bias		.6495292	.1555852	4.17	0.000	.3445878 .9544707
/cut1		-1.199481	.1591716	-7.54	0.000	-1.511452 -.8875102
/cut2		2.316014	.2597303	8.92	0.000	1.806952 2.825076
-----						
Correlations						
rho1		.4467066	1.151927	0.39	0.698	-1.811028 2.704441
rho2		.4063907	.5008222	0.81	0.417	-.5752028 1.387984
-----						

Both AIC and BIC favor exogenous switching. We report the selected output of the postestimation commands, performed after the `swopit` command with exogenous switching. The predicted choice probabilities at the specified values of the independent variables can be estimated using the `swopitprobabilities` command:

```
. swopit y house gdp bias spread, reg(house gdp) outone(spread bias) outtwo(spread bias)
```

<sup>8</sup>An identical value of the likelihood can be obtained with the opposite signs of the correlation coefficients and of the parameters in the class assignment model, and switched labels of the latent regimes.



```
(output omitted)
. swopitprobabilities, at(house=1.65 gdp=4.1 bias=0 spread=0.55)
  Evaluated at:
    house    gdp    bias    spread
  1.6500  4.1000  0.0000  0.5500

  Predicted probabilities of different outcomes
    Pr(y=-1)  Pr(y=0)  Pr(y=1)
    0.0448    0.8257    0.1295

  Standard errors of the probabilities
    Pr(y=-1)  Pr(y=0)  Pr(y=1)
    0.0194    0.0705    0.0717
```

The predicted choices (those with the largest predicted probability) for each observation can be estimated using the command **swopitpredict** with the option **output(choice)**:

```
. swopitpredict choice, output(choice)
. tab choice
```

choice	Freq.	Percent	Cum.
-1	37	13.26	13.26
0	206	73.84	87.10
1	36	12.90	100.00
Total	279	100.00	

The predicted probabilities of two latent regimes for each observation can be estimated using the command **swopitpredict** with the option **regimes**:

```
. swopitpredict pregim, regimes
. tabstat pregim*, stat(mean)
```

stats	pregim_0	pregim_1
mean	.8362734	.1637266

The average predicted probabilities of two latent regimes in the sample are 0.84 and 0.16.

The marginal effects of the covariates on the choice probabilities at the specified values of the covariates can be estimated using the **swopitmargins** command:

```
. swopitmargins, at(house=1.65 gdp=4.1 bias=0 spread=0.55)
  Evaluated at:
    house    gdp    bias    spread
  1.6500  4.1000  0.0000  0.5500

  Marginal effects of all variables on the probabilities of different outcomes
    |    Pr(y=-1)    Pr(y=0)    Pr(y=1)
  ---+-----
    house |   -0.0569   -0.7523    0.8092
      gdp |   -0.0079   -0.1040    0.1119
      bias |   -0.1667   -0.5331    0.6998
    spread |   -0.0623   -0.1425    0.2048

  Standard errors of marginal effects
    |    Pr(y=-1)    Pr(y=0)    Pr(y=1)
  ---+-----
    house |    0.0384    0.3861    0.4096
      gdp |    0.0044    0.0383    0.0399
      bias |    0.0576    0.5999    0.5836
    spread |    0.0135    0.1583    0.1528
```

Finally, the `swopitclassification` command computes different measures of the accuracy of probabilistic predictions:

```
. swopitclassification
+-----+-----+-----+
|          | Precision | Recall | Adj. noise-to-signal |
+-----+-----+-----+
| y=-1 | 0.7568 | 0.5091 | 0.0789 |
| y=0 | 0.7961 | 0.9061 | 0.4730 |
| y=1 | 0.7778 | 0.6512 | 0.0521 |
+-----+-----+-----+
Accuracy = 0.7885
Brier score = 0.2952
Ranked probability score = 0.1483

Confusion matrix
+-----+-----+-----+
| True | y=-1 | y=0 | y=1 |
+-----+-----+-----+
| Predicted |
| y=-1 | 28 | 9 | 0 |
| y=0 | 27 | 164 | 15 |
| y=1 | 0 | 8 | 28 |
+-----+-----+-----+
```

## 5.2 Health status

We also apply the new command to model the health status responses using the cross-sectional data on 3203 individuals. The dependent variable `health` takes on five ordinal categories: “pure”, “fair”, “average”, “good”, and “excellent”. There are four explanatory variables: `area` (sq. km), `weight` (kg), `female` (1=female, 0=male) and `rural` (1=rural, 0=urban).

The standard OP model provides the following output:

```
. use "fmm_health.dta", clear
. oprobit health area weight female rural, nolog

Ordered probit regression               Number of obs   =      3,203
                                         LR chi2(4)       =      470.17
                                         Prob > chi2      =      0.0000
Log likelihood = -4641.5835              Pseudo R2       =      0.0482

+-----+-----+-----+-----+-----+-----+
| health | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval] |
+-----+-----+-----+-----+-----+-----+
| area | -.025212 | .0011947 | -21.10 | 0.000 | -.0275536 | -.0228704 |
| weight | -.0029607 | .0013401 | -2.21 | 0.027 | -.0055871 | -.0003342 |
| female | -.1144565 | .040577 | -2.82 | 0.005 | -.1939859 | -.0349271 |
| rural | -.101257 | .0387958 | -2.61 | 0.009 | -.1772954 | -.0252186 |
+-----+-----+-----+-----+-----+-----+
| /cut1 | -2.58716 | .1189811 | | | | -2.820358 | -2.353961 |
| /cut2 | -1.740583 | .1148413 | | | | -1.965668 | -1.515498 |
| /cut3 | -.9011639 | .1126798 | | | | -1.122012 | -.6803156 |
| /cut4 | -.1430569 | .1121749 | | | | -.3629157 | .0768018 |
+-----+-----+-----+-----+-----+-----+

. estimates stat

Akaike's information criterion and Bayesian information criterion

+-----+-----+-----+-----+-----+-----+-----+
| Model | Obs | ll(null) | ll(model) | df | AIC | BIC |
+-----+-----+-----+-----+-----+-----+-----+
| . | 3,203 | -4876.67 | -4641.584 | 8 | 9299.167 | 9347.742 |
+-----+-----+-----+-----+-----+-----+-----+
```

The mixture of OP models significantly improves the likelihood, and is preferred to the ordinary OP model according to AIC. The `swopit` command with exogenous class assignment yields the following results:

```

swopit health weight female rural, reg(area weight female rural) outone(area weight female rural)
> outtwo(area weight female rural)
Two-regime switching ordered probit regression
Regime switching:      Exogenous
Number of observations =      3203
Log likelihood         = -4603.413
McFadden pseudo R2    =      0.0560
LR chi2(17)           =   546.5138
Prob > chi2            =      0.0000
AIC                   =   9248.8256
BIC                   =   9376.3344

```

	health	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----							
Regime equation							
area		-.0451184	.0057608	-7.83	0.000	-.0564094	-.0338274
weight		-.0225067	.0036099	-6.23	0.000	-.029582	-.0154313
female		-1.19694	.2025826	-5.91	0.000	-1.593995	-.7998858
rural		1.059583	.2180968	4.86	0.000	.6321211	1.487045
/cut1		-3.088778	.	.	.	.	.
-----							
Outcome equation 1							
area		-.0294304	.0029473	-9.99	0.000	-.0352069	-.0236539
weight		-.0153952	.001554	-9.91	0.000	-.0184409	-.0123495
female		-.6963554	.094571	-7.36	0.000	-.8817112	-.5109996
rural		.2951696	.1488002	1.98	0.047	.0035265	.5868126
/cut1		-4.564172	.0975106	-46.81	0.000	-4.755289	-4.373055
/cut2		-3.400328	.	.	.	.	.
/cut3		-2.478094	.0500958	-49.47	0.000	-2.57628	-2.379908
/cut4		-1.710041	.06242	-27.40	0.000	-1.832382	-1.5877
-----							
Outcome equation 2							
area		-.0498046	.005135	-9.70	0.000	-.059869	-.0397402
weight		.0000911	.0031486	0.03	0.977	-.0060801	.0062623
female		-.1699151	.0992853	-1.71	0.087	-.3645108	.0246805
rural		.0904578	.0955062	0.95	0.344	-.0967309	.2776464
/cut1		-2.587944	.2633434	-9.83	0.000	-3.104088	-2.071801
/cut2		-1.733168	.2467896	-7.02	0.000	-2.216866	-1.249469
/cut3		-.817695	.2394527	-3.41	0.001	-1.287014	-.3483764
/cut4		.0332917	.2384071	0.14	0.889	-.4339776	.500561
-----							

The `swopit` command with endogenous class assignment yields the following results:

```

. swopit health weight female rural, reg(area weight female rural) outone(area weight female rural)
> outtwo(area weight female rural) endo
Two-regime switching ordered probit regression
Regime switching:      Endogenous
Number of observations =      3203
Log likelihood         = -4597.9364
McFadden pseudo R2    =      0.0572
LR chi2(19)           =   557.4667
Prob > chi2            =      0.0000
AIC                   =   9241.8727
BIC                   =   9381.5251

```

	health	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----							
Regime equation							
area		.0610783	.0053706	11.37	0.000	.050552	.0716046
weight		.0329797	.0025214	13.08	0.000	.0280378	.0379215
female		1.574811	.178264	8.83	0.000	1.22542	1.924202
rural		-1.293424	.1899111	-6.81	0.000	-1.665643	-.9212053
/cut1		3.942168	.	.	.	.	.
-----							
Outcome equation 1							
area		-.0354348	.0030747	-11.52	0.000	-.0414611	-.0294085
weight		.0055531	.0011537	4.81	0.000	.003292	.0078142
female		.1106933	.0651754	1.70	0.089	-.0170481	.2384347
rural		-.1204076	.065746	-1.83	0.067	-.2492674	.0084522
/cut1		-2.090245	.	.	.	.	.
/cut2		-1.307607	.0528805	-24.73	0.000	-1.411251	-1.203964
/cut3		-.4411496	.0631265	-6.99	0.000	-.5648753	-.3174239
/cut4		.3563662	.0763013	4.67	0.000	.2068183	.505914
-----							
Outcome equation 2							
area		-.0257581	.002044	-12.60	0.000	-.0297642	-.021752
weight		-.0096472	.0012137	-7.95	0.000	-.0120261	-.0072683

female		-.3782872	.0569361	-6.64	0.000	-.48988	-.2666944
rural		.0238924	.0772799	0.31	0.757	-.1275734	.1753581
/cut1		-3.591378	.	.	.	.	.
/cut2		-2.475482	.077368	-32.00	0.000	-2.62712	-2.323843
/cut3		-1.590479	.0829777	-19.17	0.000	-1.753113	-1.427846
/cut4		-.8499883	.085508	-9.94	0.000	-1.017581	-.6823957
-----							
Correlations							
rho1		.8853588	.0504253	17.56	0.000	.7865269	.9841906
rho2		.5874218	.1548505	3.79	0.000	.2839204	.8909231
-----							

The mixture of OP models with endogenous class assignment further improves the likelihood, and is preferred to the exogenous class assignment according to AIC.

## 6 Concluding remarks

Certain important classifications of the cross-sectional data or the states of a time-series process are often not observed. These latent classes or states can distort the inference in a traditional single-equation model. Finite mixture or regime switching models surmount the problem of unobserved heterogeneity or clustering through their flexible form. The available Stata command for finite mixture of ordered probit models, `fmm: oprobit`, does not allow for endogenous switching, when the unobservables in the switching model are correlated with the unobservables in the outcome models. This article describes the ML estimation of the switching ordered probit model with exogenous or endogenous switching between two latent regimes or classes using the new Stata command `swopit`.

## References

- Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78 (1): 1–3.
- Carneiro, P., K. Hansen, and J. Heckman. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* 44: 361–422.
- Chib, S., and B. H. Hamilton. 2000. Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics* 97 (1): 25–50.
- Clarke, K. A. 2003. Nonparametric model discrimination in international relations. *Journal of Conflict Resolution* 47: 72–93.
- Dale, D., and A. Sirchenko (2021). Estimation of nested and zero-inflated ordered probit models. *Stata Journal* 21 (1): 1–36.
- Epstein, E. S. 1969. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology* 8: 985–987.
- Geweke, J., G. Gowrisankaran, and R. J. Town. 2003. Bayesian inference for hospital quality in a selection model. *Econometrica* 71 (4): 1215–1238.
- Greene, W. H., M. Harris, B. Hollingsworth, and P. Maitra. 2008. A bivariate latent class correlated generalized ordered probit model with an application to modeling observed obesity levels. Working paper No. 08-18, New York University.
- Gregory, C. A. 2015. Estimating treatment effects for ordered outcomes using maximum simulated likelihood. *Stata Journal* 15 (3): 756–774.
- Gronau, R. 1974. Wage comparisons — a selectivity bias. *Journal of Political Economy* 82 (6): 1119–1143.

- Harris, M. N., and X. Zhao. 2007. A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics* 141 (2): 1073–1099.
- Hasebe, T. 2018. Treatment effect estimators for count data models. *Health Economics Letters* 27: 1868–1873.
- Hasebe, T. 2020. Endogenous switching regression model and treatment effects of count-data outcome. *Stata Journal* 20 (3): 627–646.
- Heckman, J. J. 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42: 679–694.
- Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959.
- Hill, D. H. 1990. An endogenously-switching ordered-response model of information, eligibility and participation in SSI. *Review of Economics and Statistics* 72 (2): 368–371.
- Jasra, A., C. Holmes, and D. Stephens. 2005. Markov Chain Monte Carlo Methods and the Label Switching Problem in Bayesian Mixture Modeling. *Statistical Science* 20: 50–67.
- Kaminsky, G. L., and C. M. Reinhart. 1999. The twin crises: the causes of banking and balance-of-payments problems. *American Economic Review* 89 (3): 473–500.
- Lewis, H. 1974. Comments on selectivity biases in wage comparisons. *Journal of Political Economy* 82 (6): 1145–1155.
- McLachlan, G. J., S. X. Lee, and S. I. Rathnayake. 2019. Finite mixture models. *Annual Review of Statistics and Its Application* 6: 355–378.
- Maddala, G. S. and F. Nelson. 1975. Switching regression models with exogenous and endogenous switching. *Proceedings of the American Statistical Association (Business and Economics Section)*: 423–426.
- Miranda, A., and S. Rabe-Hesketh. 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *Stata Journal* 6 (3): 285–308.
- Munkin, M. K., and P. K. Trivedi. 2008. Bayesian analysis of the ordered probit model with endogenous selection. *Journal of Econometrics* 143: 334–348.
- Redner, R., and H. Walker. 1984. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* 26: 195–239.
- Roy, A. D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3: 135–146.
- Sirchenko, A. 2020. A model for ordinal responses with heterogeneous status quo outcomes. *Studies in Nonlinear Dynamics & Econometrics* 24 (1).
- Wilde, J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69 (3): 309–312.

# Appendix

Table A1. The true values of parameters in the Monte Carlo experiments

Parameters	Overlap among covariates		
	no overlap	partial overlap	complete overlap
$\gamma_1$	2	2	2
$\gamma_3$		1	
$\mu$	0.2	0.2	0.2
$\beta_{1,2}$	2	2	2
$\beta_{1,3}$	1	1	1
$\alpha_{1,1}$	-3.83	-5.23	-3.83
$\alpha_{1,2}$	3.76	2.46	3.81
$\beta_{2,2}$			1
$\beta_{2,3}$		-2	-2
$\beta_{2,4}$	1	1	
$\beta_{2,5}$	-2		
$\alpha_{2,1}$	-3.97	-6.17	-3.83
$\alpha_{2,2}$	3.97	0.97	3.93
$\rho_1$	0.3	0.3	0.3
$\rho_2$	0.5	0.5	0.5

Notes: The variances  $\sigma^2$ ,  $\sigma_1^2$  and  $\sigma_2^2$  are fixed to one in all simulations.