

A mixture of ordered probit models with endogenous switching between 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

Department of Quantitative 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 mixture of ordered probit models with either exogenous or endogenous switching between two latent classes (or regimes). Switching is endogenous if the unobservables in the class-assignment model are correlated with the unobservables in the outcome models. We provide a battery of postestimation commands, assess by Monte Carlo experiments the finite-sample performance of the maximum likelihood estimator of the parameters, probabilities and their standard errors (both the asymptotic and bootstrap ones), and apply the new command to model the policy interest rates.

Keywords: `swopit`, ordinal responses, ordered probit, finite mixture model, latent class, regime switching, endogenous switching.

1 Introduction

Ordinal responses can be generated, in a cross-sectional context, by different latent classes of the population or, in a time-series context, by different latent regimes (states) of the underlying process. We introduce a new command **swopit** that fits a mixture of ordered probit (OP) models for ordinal outcomes with either exogenous or endogenous switching between two latent classes (or regimes). The class-membership (regime-switching) mechanism is represented by a binary probit model. The decision-making process, which determines an outcome in each latent class, is represented by a separate OP model. Endogenous regime switching implies that the unobservables in the class-membership model are correlated with the unobservables in the outcome models. The three latent equations from the class-membership 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. 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 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 models, 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-effect (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 the parameters in the outcome equation are 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 an unordered multiple treatment and binary outcomes; Munkin and Trivedi (2008) consider an 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 OP regression model with an endogenous treatment assignment (or sample selection).

In contrast to the typical treatment-effect regression models, in the switching regression models (also known as the 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 possible correlations between the unobservables in the class-membership 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 **switchoprobisim** and **escount** that 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 switching between two *latent* classes is proposed by Greene *et al.* (2008, 2014). It can be seen as a generalization of the zero-inflated OP model of Harris and Zhao (2007), implemented in the Stata command `ziop2` by Dale and Sirchenko (2021), in which the latent class membership is modeled by a binary probit model, the outcomes in one latent class are modeled by an 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 switching between latent classes, when the unobservables in the class-membership and outcome 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 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-membership and outcome models as the `swopit` command does.

In the next section we describe the mixture of two OP models with endogenous switching between two 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 finite-sample properties of the ML estimator of the parameters and probabilities, and to compare the asymptotic and bootstrap estimators of the standard errors. We illustrate the new command in Section 5 by modeling the policy interest rates. Section 6 concludes.

2 Model

We let i ($i = 1, 2, \dots, N$) 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 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 $r_i^* = 1, 2$ 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 ν_i according to the class-membership model in the usual binary probit fashion with an unobserved threshold μ . For each i , only one out of two potential realizations (states) of y_i is observed. The observed outcome y_i is determined conditionally on the regime $r_i^* = 1$ (or $r_i^* = 2$) by the observed data $\mathbf{x}_{1,i}$ (or $\mathbf{x}_{2,i}$) and unobservables $\varepsilon_{1,i}$ (or $\varepsilon_{2,i}$) according to the OP outcome models in the usual ordered-response fashion with the unobserved thresholds α_1 (or α_2). The unobservables in the class-membership model can be correlated with those in each outcome model.

The mixture of OP models with endogenous switching between two latent classes can be summarized by the following system:

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

where $j = 1, 2, \dots, J$, $-\infty = \alpha_{1,0} \leq \alpha_{1,1} \leq \dots \leq \alpha_{1,J} = \infty$; $-\infty = \alpha_{2,0} \leq \alpha_{2,1} \leq \dots \leq \alpha_{2,J} = \infty$; $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the column vectors of unknown slope parameters; ρ_1 and ρ_2 are the correlation coefficients; ν_i , $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ are the error terms that are independently and identically distributed (*iid*) across i according to the normal distributions with the zero means and the variances σ^2 , σ_1^2 and σ_2^2 , respectively.

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

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

where $\Phi_2(\cdot)$ is the cumulative distribution function (CDF) of the bivariate normal distribution.

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

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

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

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

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

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

where $\Phi(\cdot)$ is the CDF of the normal distribution.

2.1 ML estimation and parameter identification

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

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

where θ is a vector of all parameters, including γ , μ , β_1 , α_1 , β_2 , α_2 , ρ_1 and ρ_2 ; Θ 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 $\hat{\theta}$ are computed from the Hessian matrix.

In mixture models, the likelihood function may have multiple local maxima and large flat regions. To avoid the locally optimal solutions the developed command performs several estimation attempts with different initializations by randomly assigning observations to each class. After each random initialization, the command obtains the starting values for the slope and threshold parameters using the independent estimations of binary probit class-membership model and OP output models. Further, in the case of endogenous switching, the command obtains the starting values for ρ_1 and ρ_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). Using these starting values the command then performs the ML estimation employing at each attempt four optimization techniques, modified Newton-Raphson (NR), Berndt-Hall-Hall-Hausman (BHHH), Davidson-Fletcher-Powell (DFP), and Broyden-Fletcher-Goldfarb-Shanno (BFGS), one after another, until convergence is achieved or all four of them are used. 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.

Occasionally, when the likelihood function is too flat with respect to the changes of certain covariates (so the Hessian matrix is nearly singular), the function `optimize()`, which is employed for the numerical optimization of the likelihood, reports the missing standard errors for the estimated coefficients on those covariates (as is common for other Stata commands due to, for example, collinearity or poorly scaled covariates). In such cases, the implemented bootstrap estimator of the standard errors can be applied. We provide an option of the nonparametric stratified bootstrap estimator of the standard errors for all parameters and probabilities. The bootstrap estimator can also be beneficial in small samples. We illustrate the performance of both the asymptotic and bootstrap estimators in Section 4.

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 simultaneous equations models, and the third issue is typical in the estimation of finite mixture models.

First of all, the parameters are identified only up to scale and location, that is, the slope parameters in γ , β_1 and β_2 are identified only jointly with the variances σ^2 , σ_1^2 and σ_2^2 , respectively, and the intercept components in γ , β_1 and β_2 are identified only jointly with the corresponding threshold parameters μ , α_1 and α_2 , and variances σ^2 , σ_1^2 and σ_2^2 . Following the standard approach in the identification of discrete-choice models (see Long (1997), pp. 122-123), we fix the variances σ^2 , σ_1^2 and σ_2^2 to 1, and the intercept components of γ , β_1 and β_2 to 0. Importantly, the probabilities in (2) are absolutely identifiable and invariant to these (arbitrary) identifying assumptions on the parameters.

Secondly, we can actually identify all parameters in θ (again, up to scale and location) because of the nonlinearity of OP models, i.e. via the functional form (Heckman 1978;

Wilde 2000).¹ However, the ML estimation may be subject to a weak identification problem (a low precision of the estimates) if the class-membership model and outcome models contain the same or similar sets of covariates. In this case, especially if the sample size is small, the so-called exclusion restrictions (using different sets of covariates in the class-membership and outcome models) may be necessary to improve the precision of the estimates.

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 class labels (Redner and Walker 1984; Jasra *et al.* 2005). Since labeling 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 $\hat{\beta}_1$ and $\hat{\alpha}_1$ indeed estimate β_1 and α_1 (but not β_2 and α_2), nor that $\hat{\beta}_2$ and $\hat{\alpha}_2$ estimate β_2 and α_2 (but not β_1 and α_1). The model has two identical likelihood maxima because

$$\begin{aligned} & \Pr(y_i = j | \mathbf{x}_i^{all}, -\gamma, -\mu, \beta_2, \alpha_2, \beta_1, \alpha_1, -\rho_2, -\rho_1) \\ &= \Phi_2(-\mu + \mathbf{z}_i\gamma; \alpha_{2,j} - \mathbf{x}_{2,i}\beta_2; -\rho_2) - \Phi_2(-\mu + \mathbf{z}_i\gamma; \alpha_{2,j-1} - \mathbf{x}_{2,i}\beta_2; -\rho_2) \\ &+ \Phi_2(\mu - \mathbf{z}_i\gamma; \alpha_{1,j} - \mathbf{x}_{1,i}\beta_1; \rho_1) - \Phi_2(\mu - \mathbf{z}_i\gamma; \alpha_{1,j-1} - \mathbf{x}_{1,i}\beta_1; \rho_1) \\ &= \Pr(y_i = j | \mathbf{x}_i^{all}, \gamma, \mu, \beta_1, \alpha_1, \beta_2, \alpha_2, \rho_1, \rho_2). \end{aligned}$$

Hence, changing the signs of the parameters in the class-membership 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-membership model are identified up to sign only, and the latent class labels are interchangeable.

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:

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

where f is the probability density function of the standard normal distribution, and γ_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-membership model, the outcome model conditional on $r_i^* = 1$ and the outcome model conditional on

¹We don't make the nonlinear assumptions for the sake of identification: our original outcome models are already non-linear.

$r_i^* = 2$, respectively (γ_k^{all} , $\beta_{1,k}^{all}$ or $\beta_{2,k}^{all}$ is zero if the k th covariate in \mathbf{x}_t^{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 others are held fixed.

The asymptotic standard error of $\mathbf{PE}_{k,j,i}$ is computed using the delta method as a square root of $\nabla_{\theta}(\mathbf{PE}_{k,j,t})\mathbf{Var}(\boldsymbol{\theta})\nabla_{\theta}(\mathbf{PE}_{k,j,i})'$. The nonparametric stratified bootstrap estimator of the standard errors is also implemented.

3 The swopit command

The accompanying software includes the `swopit` command, postestimation commands and supporting help files available at <https://github.com/janwillemnienhuis/Swopit>.

3.1 Syntax

The following command fits a mixture of OP models with either exogenous or endogenous switching between 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 *indepvar* are, by default, included in each model. The alternative (and possibly not the same) lists of independent variables to be included in the class-membership model and each outcome model can be specified in *options*. The following options are available.

Options

regindepvars(*varlist*) specifies the list of independent variables included in the class-membership model. By default, the list is the same as in *indepvars*.

outoneindepvars(*varlist*) specifies the list of independent variables included in the first outcome model. By default, the list is the same as in *indepvars*.

outtwoindepvars(*varlist*) specifies the list of independent variables included in the second outcome model. By default, the list is the same as in *indepvars*.

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

guesses(*scalar*) specifies the number of estimation attempts to be performed by the command `optimize()` with different initializations by randomly assigning observations to each class, and obtaining starting values for the slope and threshold parameters using the independent estimations of binary probit regime-switching model and two OP outcome models. Further, in the case of endogenous switching, the `swopit` command obtains the starting values for ρ_1 and ρ_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. At each attempt, 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 output with the highest likelihood is reported. The default is `guesses(5)`.

limit(numlist) specifies a space-delimited list of the limits for the maximum absolute value of each parameter in the following order: γ , μ , β_1 , α_1 , β_2 , α_2 , ρ_1 and ρ_2 . If only one value is specified, this limit applies to all parameters. By default, no constraints on the parameters' values 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(numlist) specifies a space-delimited list of the starting values for the parameters in the following order: γ , μ , β_1 , α_1 , β_2 , α_2 , ρ_1 and ρ_2 . The elements of α_1 and α_2 should be provided in 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 is stopped if 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: $SV = SV + \text{change} * U(-|SV|, |SV|)$, where $U()$ represents a uniformly distributed random variable. In the case of endogenous switching, the SV for the correlation coefficients ρ_1 and ρ_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 option is ignored if the **initial()** option is not used. However, it is always applied in the bootstrap estimations if **bootstrap()** option is used. The default is **change(0.5)**.

log shows the progress of the numerical optimization of the log likelihood: current estimation attempt, optimization method, and convergence status. By default, the log output is suppressed.

bootstrap(scalar) specifies the number of bootstrap replications to be performed to estimate the standard errors. The default is **bootstrap(0)**, and no bootstrapping is performed.

bootguesses(scalar) specifies the number of attempts with different starting values of parameters to be performed with each bootstrap sample. At the first attempt, the starting values are the values of the parameters estimated with the original sample. At each new attempt, the starting values are selected as described in **change()**. At each attempt, 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 output with the highest likelihood is reported. 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(100)**.

Saved results

swopit saves the following in `e()`:

Scalars

<code>e(N)</code>	number of observations	<code>e(r2_p)</code>	McFadden pseudo-R-squared
<code>e(k_cat)</code>	number of choices	<code>e(ll)</code>	log likelihood
<code>e(k)</code>	number of parameters	<code>e(ll_0)</code>	log likelihood, constant-only model
<code>e(df_m)</code>	model degrees of freedom	<code>e(aic)</code>	Akaike Information Criterion
<code>e(chi2)</code>	Chi-square test statistic	<code>e(bic)</code>	Bayesian Information Criterion
<code>e(p)</code>	p-value of Chi-square test		

Macros

<code>e(cmd)</code>	swopit program	<code>e(predict)</code>	swopitpredict program
<code>e(opt)</code>	optimization method	<code>e(depvar)</code>	name of dependent variable
<code>e(properties)</code>	b V	<code>e(switching)</code>	type of regime switching
<code>e(vce)</code>	standard error method	<code>e(vctype)</code>	title to label standard errors

Matrices

<code>e(ll_obs)</code>	vector of observation-wise log-likelihood	<code>e(b)</code>	coefficient vector
<code>e(V)</code>	variance-covariance matrix of the estimators		
<code>e(boot)</code>	coefficient vectors in the bootstrap samples		

Functions

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

3.2 Postestimation commands

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

The **swopitpredict** command

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

This command provides the predicted probabilities of the observed choices (by default) or latent classes for each observation. The following options are available:

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 $Pr(y_i \leq 1)$, $Pr(y_i \leq 2)$, ..., $Pr(y_i \leq J)$. If **output**() is not specified, the usual choice probabilities such as $Pr(y_i = 1)$, $Pr(y_i = 2)$, ..., $Pr(y_i = J)$ are predicted.

name(*varname*) creates the variables named *varname_**, where *** 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. If **name**() is not specified, either the variables named *swopit_r_** are created if the **regimes** option is used, or the variables named *swopit_pr_** are created if the **output**() option is used.

`tabstat` displays a compact table of summary statistics for the variables created by `swopitpredict`. If option `name(varname)` is used, the statistics are provided for all available variables `varname_*`. If `name()` is not specified, the statistics are provided for all available variables `swopit_r_*` and `swopit_pr_*`.

The `swopitprobabilities` command

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

This command provides the predicted probabilities of the observed choices (by default) or latent classes and 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 not included, the probabilities are estimated at the median value of this variable.

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

Saved results `swopitprobabilities` saves the following in `r()`:

Matrices

<code>r(at)</code>	vector of the covariates values, at which the probabilities are computed
<code>r(pr)</code>	vector of the probabilities
<code>r(se)</code>	vector of the standard errors of the probabilities
<code>r(t)</code>	vector of the t-statistics of the probabilities
<code>r(pval)</code>	vector of the p-values of the t-tests for the probabilities

3.2.1 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 and 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 partial effects. By default, the partial 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 not included, the partial effects are estimated at the median value of this variable.

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

Saved results `swopitmargins` saves the following in `r()`:

Matrices

<code>r(at)</code>	vector of the covariates values, at which the marginal effects are computed
<code>r(me)</code>	matrix of the marginal effects on the probabilities
<code>r(se)</code>	matrix of the standard errors of the marginal effects on the probabilities
<code>r(t)</code>	matrix of the t-statistics of the marginal effects on the probabilities
<code>r(pval)</code>	matrix of the p-values of the t-tests for the marginal effects

3.2.2 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 Monte Carlo experiments to assess the finite-sample performance of the proposed full information 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).

In brief, the results of the experiments show that as the sample size grows, the biases and the RMSEs decrease, and the coverage rates approach the nominal level, suggesting the consistency of the proposed ML estimator. The estimates of choice and class probabilities in the exogenous switching case are quite reliable in small samples: at the 95% nominal level, the coverage rates are 85%–92% with 250 observations, and 90%–94% with 500 observations (with 10–11 parameters to estimate). In the case of endogenous switching (with two extra parameters to estimate), the coverage rates are in the same range for the regime probabilities, but are only 79%–84% for the choice probabilities. With 2,000 observations, the coverage rates for the regime probabilities for both types of switching are 94%–95%; however, for the choice probabilities they are 94%–95% in the exogenous switching case but less than 86% with the endogenous switching. The bootstrap estimator of the standard errors for the correlation coefficients and the probabilities can provide less biased estimates and better coverage rates than the asymptotic estimator (based on its performance with 250 and 500 observations). However, at the same time, the bootstrap estimator of the standard errors for the parameters are much more biased.

Of course, all Monte Carlo results are subject to the particular design of the experiment: they may be better or worse under different specifications of the model and different data sets. It seems clear, however, that in order to be reliable the estimation of endogenous switching models requires much larger samples.

4.1 Monte Carlo design

We generated five independent variables $g_m \stackrel{iid}{\sim} 4 * \mathcal{N}(0, 1)$, $m \in \{1, \dots, 5\}$ 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-membership model	g_1	g_1, g_3	g_1
Class 1: Outcome model	g_2, g_3	g_2, g_3	g_2, g_3
Class 2: Outcome model	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 membership: 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 the Appendix. These values were calibrated in such a way that both latent classes have the same theoretical probability, and each of the three choices is equally likely. The variances of the error terms were fixed to one. The probabilities were computed at the median values of the covariates. 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 switching, and estimated the standard errors of the parameters in each process using both the asymptotic and bootstrap estimators. We generated 2,000 replications independently with 250 and 500 observations. For each bootstrap sample, the observations were re-sampled separately for each choice, so the number of observations of each choice in every bootstrap sample was the same as in the original one. To compute the nonparametric stratified bootstrap estimates of the standard errors, we drew with replacement 300 bootstrap samples for each Monte Carlo iteration, and computed the standard deviations of the estimates of parameters, correlations coefficients, and choice and class probabilities. The “true” values of the standard errors were calculated as the standard deviations of the corresponding estimates in the 2,000 Monte Carlo replications.

4.2 Monte Carlo results

Table 1 summarizes the finite-sample performance of the estimator of the parameters. We divided all parameters into three groups (slope parameters, threshold parameters and correlation coefficients), and computed the average statistics for each. The slope parameters γ , β_1 and β_2 are estimated better than the threshold parameters μ , α_1 and α_2 and much better than the correlation coefficients ρ_1 and ρ_2 . As the sample sizes increase eight times from 250 to 2,000:

- 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;
- for the slope parameters the biases decrease 14–16 times and the RMSEs decrease 9–13 times;
- 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; and

– the RMSEs 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.

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

Sample size	Class membership: Covariates' overlap:	Exogenous			Endogenous		
		none	partial	complete	none	partial	complete
Slope parameters γ , β_1 and β_2							
250	Absolute 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
Threshold parameters μ , α_1 and α_2							
250	Absolute 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
Correlation coefficients ρ_1 and ρ_2							
250	Absolute 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

Table 2 reports the finite-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 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:

- 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;
- 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;
- the RMSEs 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;

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

Sample size	Class membership:	Exogenous			Endogenous		
	Covariates' overlap:	none	partial	complete	none	partial	complete
Probabilities of choices							
250	Absolute 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
Probabilities of latent classes							
250	Absolute 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

- 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;
- the RMSEs 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
- the delta-method estimator of standard errors is downward biased for all probabilities, but also consistent: the biases become smaller than 4% in the exogenous switching models, and smaller than 3% for the class probabilities but are up to 40% for the choice probabilities in the endogenous switching models.

Table 3. The finite-sample performance of the bootstrap estimators of standard errors

Class membership:		Exogenous		Endogenous	
Sample size:		250	500	250	500
Class-membership parameters γ and μ					
Coverage rate	Asymptotic	96.5	95.9	93.8	94.8
(at 95% level), %	Bootstrap	99.5	98.3	98.9	98.5
Bootstrap s.e. / true s.e.		13.2	6.2	13.0	2.9
Outcome parameters β_1 , α_1 , β_2 and α_2					
Coverage rate	Asymptotic	97.2	96.4	84.4	86.8
(at 95% level), %	Bootstrap	98.8	99.3	98.8	98.9
Bootstrap s.e. / true s.e.		1.6	1.6	1.2	1.3
Correlation coefficients ρ_1 and ρ_2					
Coverage rate	Asymptotic			28.0	42.0
(at 95% level), %	Bootstrap			88.6	88.1
Bootstrap s.e. / true s.e.				0.86	0.86
Latent class probabilities					
Coverage rate	Asymptotic	91.2	93.3	90.5	92.0
(at 95% level), %	Bootstrap	93.2	95.3	91.3	94.6
Bootstrap s.e. / true s.e.		1.05	1.03	0.94	0.99
Choice probabilities					
Coverage rate	Asymptotic	85.5	90.2	82.8	80.3
(at 95% level), %	Bootstrap	89.9	92.0	86.4	90.5
Bootstrap s.e. / true s.e.		0.99	0.98	0.84	0.87

Table 3 reports the finite-sample performance of the bootstrap estimator of the standard errors in comparison with the estimator of the asymptotic ones. The coverage rates with asymptotic errors for the slope and threshold parameters in the exogenous switching case are above the nominal level (96%–97%), whereas in the endogenous switching case they are 85%–90%, below the nominal level. The estimated errors for the slope and threshold parameters do not benefit from bootstrapping, at least in the exogenous switching case. The bootstrap estimator overestimates the parameters' standard errors with both types of switching (the coverage rates are 98.3%–99.5%). The bootstrap standard errors for the class parameters are overestimated far more than those of the outcome

model parameters: for example, with 250 observations the bootstrap errors for the class model parameters are 13 times as large as the true values of the errors for both types of switching, whereas the bootstrap errors for the outcome parameters are larger by only 20%–60%. The bootstrap standard errors for the class parameters are inflated by a very small fraction of very large estimates of the parameters in the bootstrap samples. The very large values of the parameters in the regime-switching model imply the more certain estimations of the latent classes: all estimated class probabilities become very close to either 1 or 0. The variability of parameters’ estimates in the outcome models in the bootstrap samples is less affected, and their bootstrap standard errors are inflated far less.

However, the estimates of correlation coefficients do benefit from bootstrapping: the coverage rates move toward the nominal level (from 28% and 42% to 88% for both sample sizes); the bootstrap standard errors are still underestimated but not that much as the asymptotic ones. More importantly, the probabilities’ estimates seem to clearly benefit from bootstrapping: 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. The bootstrap standard error estimates differ from the true ones by no more than 5% in the exogenous case, and no more than 16% in the endogenous case.

The simulations suggest that with 250 and 500 observations the bootstrap estimator of the standard errors for the correlation coefficients, the choice and class probabilities can provide less biased estimates and better coverage rates than the asymptotic estimator despite the fact that at the same time the latter can demonstrate less biased estimates and better coverage rates for the parameters. However, the simulation results depend on the particular design: we do not know whether our conclusions are valid for other sample sizes, other specifications of the covariates in the regime-switching and outcome models, and other true values of the parameters.

5 Empirical example

Estimations of monetary policy rules have been historically implemented using models for a continuous dependent variable. During the last three decades, most major central banks began changing their policy rates by discrete amounts, typically by multiples of 25 basis points (bp). It motivated the use of discrete-choice approaches such as the OP models (e.g., Gerlach, 2007; Sirchenko, 2008; Vanderhart, 2000). However, a single-equation OP model does not allow for oscillating switches in monetary policy — following a widely used in macroeconomic modeling regime-switching approach, built on the seminal work of Hamilton (1989), policy rates are better modeled as stochastic and repeated fluctuations among two or more regimes (see, among others, Assenmacher-Wesche, 2006).

We apply both the standard and switching OP models to the decisions of the U.S. Federal Open Market Committee (FOMC) on the federal funds rate target, made at both scheduled and unscheduled meetings from 2/1989 to 9/2019. The FOMC used to change its target in multiples of 6.25 bp, but since 10/1989 it changed the rate in multiples of 25 bp only. The frequencies of original changes to the target rate are shown below (data source: ALFRED (ArchivaL Federal Reserve Economic Data), available at <https://alfred.stlouisfed.org>):

Change, bp	-75	-50	-31.25	-25	0	6.25	12.5	18.75	25	43.75	50	75
Frequency	3	16	1	35	180	1	1	1	35	1	4	1

For further analysis, we consolidated these FOMC decisions into three categories,

“decrease”, “no change” and “increase”:²

y	decrease (-1)	no change (0)	increase (1)
Frequency	55	181	43

In the sample of 279 observations, the FOMC decisions are matched with the values of the following independent variables as they were observed at the previous day:

house (the Greenbook projection of the total number of new privately-owned housing units started for the current quarter; data source: RTDSM (Real-Time Data Set for Macroeconomists), available at <https://www.philadelphiafed.org>);

gdp (the Greenbook projection of quarterly growth in the nominal gross domestic (before 1992: national) product for the current quarter, annualized percentage points; data source: RTDSM);

spread (the difference between the one-year treasury constant maturity rate and the effective federal funds rate, three-business-day moving average; data source: ALFRED);

bias (the indicator that we constructed from the “policy bias” and “balance-of-risks” statements made 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; data source: the FOMC statements and minutes, available at <https://www.federalreserve.gov>).

The summary statistics of the variables used are shown below:

```
. use "policy_rate.dta", clear
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
y	279	-.0430108	.5921667	-1	1
bias	279	-.1397849	.8036526	-1	1
house	279	1.293154	.3728651	.4	2.15
gdp	279	4.36905	1.977448	-4.3	9.4
spread	279	.026129	.4995737	-1.48	1.713333

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

```
. set seed 3
. set more off
. 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
```

²A single 6.25-bp change is included in the “no change” category: such a small change is viewed as a technical adjustment and typically combined with the no-change decisions in the literature (see, e.g., Hamilton and Jorda 2002).

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.

The OP model suggests that the larger/smaller the value of any explanatory variable, the more likely that the FOMC will hike/cut the rate (holding the other variables fixed). Next, we let the outcomes be generated in two latent regimes, and fit a mixture of two OP models with exogenous switching using the `swopit` command:

```
. set seed 3
. set more off
. swopit y house gdp bias spread, reg(house gdp) outone(bias spread) outtwo(bias spread) log
Finding regime starting values
Finding outcome starting values
Attempt #1 with method: nr
Trying again with different method: bhhh
Trying again with different method: dfp
convergence (32 iterations)
Finding regime starting values
Finding outcome starting values
Attempt #2 with method: nr
convergence with likelihood improvement (24 iterations)
Finding regime starting values
Finding outcome starting values
Attempt #3 with method: nr
convergence without likelihood improvement
Finding regime starting values
Finding outcome starting values
Attempt #4 with method: nr
convergence without likelihood improvement
Finding regime starting values
Finding outcome starting values
Attempt #5 with method: nr
convergence without likelihood improvement
Printing converged estimation with highest likelihood:
```

A mixture of ordered probit models with two latent classes

```
Latent class switching = Exogenous
SE method = OIM
Optimization method = Newton Raphson
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	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Class membership model							
	house	4.911729	1.386313	3.54	0.000	2.194606	7.628852
	gdp	.6790068	.202835	3.35	0.001	.2814574	1.076556
	/cut1	11.96884	2.980525	4.02	0.000	6.127117	17.81056
Outcome model (Class 1)							
	bias	.6798781	.1525716	4.46	0.000	.3808433	.9789129
	spread	1.819521	.240331	7.57	0.000	1.348481	2.290561
	/cut1	-1.251123	.1463508	-8.55	0.000	-1.537965	-.9642805
	/cut2	2.318764	.2520935	9.20	0.000	1.82467	2.812858
Outcome model (Class 2)							
	bias	3.954417	1.183389	3.34	0.001	1.635017	6.273817
	spread	14.12324	3.432332	4.11	0.000	7.395996	20.85049
	/cut1	-16.28826	4.240216	-3.84	0.000	-24.59893	-7.977591
	/cut2	1.448369	.6057136	2.39	0.017	.2611923	2.635546

An identical value of the likelihood can be obtained with the opposite signs of the parameters in the regime-switching model and the switched labels of the latent regimes. The variances of error terms are fixed to one, and the intercept parameters are fixed to zero since the parameters are identified only up to scale and location. Nevertheless, the regime and choice probabilities are absolutely identified regardless of the above assumptions.

By default, five estimation attempts are executed. At each attempt, the observations are randomly assigned to one of the latent classes. The reported value of the log likelihood is obtained at the second attempt by the NR method (after 24 iterations). At the first attempt, the NR method does not converge whereas the DFP method converges but with a lower value of the likelihood than that obtained at the second attempt by the NR method. At the next three attempts, the NR method converges with the same value of the likelihood as at the second attempt.

The switching OP model suggests that the FOMC decisions are generated in two different regimes. In both regimes, the larger/smaller the value of **spread** or **bias**, the more likely that the FOMC will hike/cut the rate (all other things equal). However, in one of the two regimes, the FOMC reactions to the values of **spread** and **bias** are much stronger than in the other one: the coefficients on **spread** and **bias** in the outcome model of the regime 2 are, respectively, eight and six times as large as in the outcome model of the regime 1. Besides, the larger/smaller the value of **house** or **gdp**, the higher/smaller the likelihood of the more aggressive regime 2. The empirical evidence in favor of regime switching is convincing: the switching OP model substantially improves the likelihood, and is clearly preferred to the standard OP model according to both AIC and BIC.

Given our sample size of 279, as suggested by the MC experiments, we could suspect that the parameters' estimates are not accurate enough and the coverage rate of the 95% confidence intervals could be higher than 97%. The simulations suggest that increasing sample size from 250 to 500 could lead to a threefold decrease of the biases of the slope parameters' estimates, and a twofold decrease of those of the threshold estimates as well as to a sevenfold reduction of the biases of the standard errors' estimates for all parameters.

Next, we report the selected output of the postestimation commands. The predicted choice probabilities at the specified values of the independent variables can be estimated using the **swopitprobabilities** command:

```
. swopitprobabilities, at(house=1.5 gdp=8.9 bias=1 spread=-0.0633333)

Evaluated at:
      house      gdp      bias      spread
      1.5000  8.9000  1.0000  -0.0633

Predicted probabilities of different outcomes
      Pr(y=-1)      Pr(y=0)      Pr(y=1)
      0.0026      0.1186      0.8788

Delta-method standard errors of probabilities
      Pr(y=-1)      Pr(y=0)      Pr(y=1)
      0.0034      0.1241      0.1260
```

In contrast, $\Pr(y = -1)$, $\Pr(y = 0)$ and $\Pr(y = 1)$ from the standard OP model are rather different, 0.0028, 0.6610 and 0.3362, respectively:

```
. quietly oprobit y house gdp bias spread

. predict pr*, p
. display pr1[2], pr2[2], pr3[2]
.00281791 .66099417 .3361879
```

According to the standard OP model, the predicted choice (that with the largest

predicted probability) for the second observation (`house` = 1.5, `gdp` = 8.9, `bias` = 1, `spread` = -0.0633333) is 0, that is “no change”.

The predicted choices for each observation can be computed for the switching OP model using the command `swopitpredict` with the option `output(choice)`. In particular, the predicted choice for the second observation (as well as the observed choice) is 1, that is an “increase”:

```
. swopitpredict, output(choice)
. display swopit_pr[2]
1
```

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

```
. swopitpredict, regimes tabstat
```

Variable	N	Mean	SD	Variance	Max	Min
swopit_r_0	279	.8363	.2967	.08801	1	.0000455
swopit_r_1	279	.1637	.2967	.08801	1	0

The average predicted probabilities of two latent regimes in the sample are 0.84 and 0.16. For the second observation, the predicted probabilities of two latent regimes are:

```
. display swopit_r_0[2], swopit_r_1[2]
.07466315 .92533685
```

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=1.56 gdp=5.9 bias=1 spread=-0.41)
```

Evaluated at:

house	gdp	bias	spread
1.5600	5.9000	1.0000	-0.4100

Marginal effects of all variables on the probabilities of different outcomes

	Pr(y=-1)	Pr(y=0)	Pr(y=1)
house	-0.2210	0.2361	-0.0150
gdp	-0.0306	0.0326	-0.0021
bias	-0.0831	0.0706	0.0125
spread	-0.2223	0.1864	0.0359

Delta-method standard errors of marginal effects

	Pr(y=-1)	Pr(y=0)	Pr(y=1)
house	0.1140	0.1137	0.0143
gdp	0.0163	0.0163	0.0020
bias	0.0198	0.0228	0.0130
spread	0.0899	0.0994	0.0390

The marginal effects computed for the same values of the covariates (the 120th observation) in the standard OP model are also rather different:

```
. quietly oprobit y house gdp bias spread
. margins, dydx(house gdp bias spread) at(house=(1.56) gdp=(5.9) bias=(1) spread=(-0.41))
```

Conditional marginal effects
Model VCE: OIM

Number of obs = 279

```

dy/dx wrt: house gdp bias spread

1._predict: Pr(y=-1), predict(pr outcome(-1))
2._predict: Pr(y=0), predict(pr outcome(0))
3._predict: Pr(y=1), predict(pr outcome(1))

At: house = 1.56
    gdp   = 5.9
    bias  = 1
    spread = -.41

```

		Delta-method		z	P> z	[95% conf. interval]	
		dy/dx	std. err.				
house	_predict						
	1	-.0748088	.0375775	-1.99	0.047	-.1484593	-.0011584
	2	-.008311	.047125	-0.18	0.860	-.1006743	.0840524
	3	.0831198	.0368925	2.25	0.024	.0108119	.1554277
gdp	_predict						
	1	-.0183997	.0076835	-2.39	0.017	-.033459	-.0033404
	2	-.0020441	.0116896	-0.17	0.861	-.0249553	.0208671
	3	.0204439	.0087203	2.34	0.019	.0033523	.0375354
bias	_predict						
	1	-.0545135	.014821	-3.68	0.000	-.0835621	-.025465
	2	-.0060563	.0353416	-0.17	0.864	-.0753245	.063212
	3	.0605698	.0288366	2.10	0.036	.004051	.1170885
spread	_predict						
	1	-.2019291	.0842712	-2.40	0.017	-.3670976	-.0367607
	2	-.0224336	.1259947	-0.18	0.859	-.2693786	.2245115
	3	.2243627	.0575838	3.90	0.000	.1115005	.337225

In particular, the marginal effects on the probabilities of no change ($\Pr(y = 0)$) for all four covariates are negative in the OP model, but positive in the switching OP model. Besides, the marginal effects of **house** and **gdp** on the probabilities of an increase ($\Pr(y = 1)$) also have the opposite signs in the two models. A flexible two-regime structure of the switching OP model, in which the regimes are driven by the covariates, allows the sign of the marginal effect on the choice probabilities to change more than once from the lowest choice to the highest one. Thus, the marginal effects on the probabilities of an increase and a decrease may have the same sign as we can see above for **house** and **gdp**. In contrast, the marginal effects in the standard OP model are subject to the single crossing property — the sign of the marginal effects changes once (and only once), so the marginal effects on the probabilities of an increase and a decrease always have the opposite signs.

The `swopitclassification` command computes the 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

```

The bootstrap standard errors can be computed by the `swopit` command using option `boot`:

```

. set seed 3
. set more off
. swopit y house gdp bias spread, reg(house gdp) outone(bias spread) outtwo(bias spread) ///
> boot(900) bootguesses(2) bootiter(100) change(0.25)
Printing converged estimation with highest likelihood:

```

A mixture of ordered probit models with two latent classes

```

Latent class switching =      Exogenous
SE method              =      Bootstrap
Optimization method    =      Newton Raphson
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

```

		Bootstrap					
	y	Coefficient	std. err.	z	P> z	[95% conf. interval]	
Class membership model							
	house	4.911729	81.77341	0.06	0.952	-155.3612	165.1847
	gdp	.6790066	5.751994	0.12	0.906	-10.59469	11.95271
	/cut1	11.96884	155.371	0.08	0.939	-292.5527	316.4903
Outcome model (Class 1)							
	bias	.6798781	.6030137	1.13	0.260	-.5020071	1.861763
	spread	1.819521	.6087107	2.99	0.003	.6264699	3.012572
	/cut1	-1.251123	.1590352	-7.87	0.000	-1.562826	-.9394194
	/cut2	2.318764	1.409538	1.65	0.100	-.4438799	5.081408
Outcome model (Class 2)							
	bias	3.954373	2.159801	1.83	0.067	-.2787601	8.187506
	spread	14.12308	11.27749	1.25	0.210	-7.980387	36.22656
	/cut1	-16.28807	10.94617	-1.49	0.137	-37.74217	5.166038
	/cut2	1.448379	.9848017	1.47	0.141	-.481797	3.378555

The bootstrap standard errors for the parameters are larger than the asymptotic ones: for the regime-switching parameters — by 47 times on average, and for the outcome parameters — by 2.8 times on average. These results are in accordance with those in the Monte Carlo simulations with 250 observations: the bootstrap errors for the regime-switching parameters are larger than the asymptotic ones by 61 times, and for the outcome parameters — by 2.7 times. In the simulations, the coverage rates with the asymptotic standard errors are above the nominal 95% level, so with the bootstrap errors they are even higher (around 99%) than with the asymptotic ones.

Table 4 reports the summary statistics of the estimated parameters in the bootstrap replications and the bootstrap estimates of the standard errors along with the values of the parameters estimated in the original sample and their asymptotic standard errors. Interestingly, the median values of the bootstrap parameters' estimates are almost the same as the values of the parameters observed in the sample. However, the distributions of the slope parameters' estimates in the bootstrap replications are highly skewed: the coefficients of skewness are from 7 to 27. The distributions of the threshold parameters are also severely skewed. Therefore, in compliance with the Monte Carlo experiments, the bootstrap estimator of standard errors of the parameters are likely to be much more biased

than the asymptotic one, especially for the regime-switching parameters (in the simulations the bias is 60% for the outcome parameters but 1200% for the regime-switching ones).

Table 4. Summary statistics of the bootstrap estimates of parameters

	Original sample		Bootstrapped samples					
	Coeff.	Std error	Mean	Median	Min	Max	Skewness	Std dev
Class-membership model								
house	4.9	1.4	12.7	5.0	-1.4	1896.7	17.6	81.8
gdp	0.7	0.2	1.3	0.7	-0.1	105.7	13.7	5.8
/cut1	12.0	3.0	27.3	12.1	-1.8	3496.4	16.6	155.4
Class 1: Outcome model								
bias	0.7	0.2	0.7	0.7	0.3	18.3	27.3	0.6
spread	1.8	0.2	2.0	1.9	0.8	15.3	12.3	0.6
/cut1	-1.3	0.1	-1.3	-1.3	-2.4	-0.8	-1.2	0.2
/cut2	2.3	0.3	2.7	2.4	1.8	15.5	5.9	1.4
Class 2: Outcome model								
bias	4.0	1.2	4.3	3.9	0.4	37.0	6.9	2.2
spread	14.1	3.4	16.5	15.1	-0.7	292.7	17.1	11.3
/cut1	-16.3	4.2	-18.5	-17.3	-279.2	-3.0	-15.4	10.9
/cut2	1.4	0.6	1.8	1.7	-1.8	16.6	5.9	1.0

The bootstrap standard errors for predicted choice probabilities, estimated using the `swopitprobabilities` command, are larger than the asymptotic ones by 11% on average:

```
. swopitprobabilities, at(house=1.5 gdp=8.9 bias=1 spread=-0.0633333)
```

```
Evaluated at:
```

```
house    gdp    bias    spread
1.5000  8.9000  1.0000  -0.0633
```

```
Predicted probabilities of different outcomes
```

```
Pr(y=-1)  Pr(y=0)  Pr(y=1)
0.0026    0.1186  0.8788
```

```
Bootstrap standard errors of the probabilities
```

```
Pr(y=-1)  Pr(y=0)  Pr(y=1)
0.0035    0.1922  0.1933
```

It is also in accordance to the simulations with 250 observations, in which the bootstrap standard errors for probabilities are larger than the asymptotic ones by 37% on average. The coverage rates with asymptotic standard errors in the simulations are below the nominal level with both types of errors, but the bootstrap coverage rates are closer to the nominal level. Therefore, the bootstrap estimator of standard errors of the probabilities are likely to be less biased than the asymptotic one with our sample size (in the simulations the bias is only 1%). The summary statistics of the estimated choice probabilities in the bootstrap replications, reported in Table 5, show that the bootstrap estimates of the probabilities are much less skewed than the estimates of parameters.

Table 5. Summary statistics of the bootstrap estimates of choice probabilities

	Original sample		Bootstrapped samples					
	Estimate	Std error	Mean	Median	Min	Max	Skewness	Std dev
Pr(y=-1)	0.00	0.00	0.00	0.00	0.00	0.05	4.32	0.00
Pr(y=0)	0.12	0.12	0.23	0.17	0.00	0.90	1.20	0.19
Pr(y=1)	0.88	0.13	0.77	0.82	0.09	1.00	-1.19	0.19

Following Hamilton (1989), switching macroeconomic regimes are usually modelled as latent states driven by a stochastic exogenous Markov-chain process, which has the constant probabilities of transition from one state to another and which does not depend on the endogenous economic variables. In the context of monetary policy, it is intuitive to assume that central banks' actions are systematically related to the economic conditions. Therefore, regime switches should be endogenous to the state of the economy. 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:³

```
. set seed 3
. set more off
. swopit y house gdp bias spread, reg(house gdp) outone(bias spread) outtwo(bias spread) endo ///
> guesses(25)
(output omitted)
Printing converged estimation with highest likelihood:
```

A mixture of ordered probit models with two latent classes

```
Latent class switching =      Endogenous
SE method              =      OIM
Optimization method    =      Newton Raphson
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	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
-----+-----							
Class membership model							
	house	-4.804481	1.411601	-3.40	0.001	-7.571168	-2.037794
	gdp	-.6180356	.1822873	-3.39	0.001	-.9753122	-.260759
	/cut1	-11.48333	2.986709	-3.84	0.000	-17.33717	-5.629489
-----+-----							
Outcome model (Class 1)							
	bias	3.389238	1.508778	2.25	0.025	.4320874	6.346389
	spread	12.40146	4.241626	2.92	0.003	4.088026	20.71489
	/cut1	-14.70939	4.604909	-3.19	0.001	-23.73485	-5.683934
	/cut2	1.092362	1.045621	1.04	0.296	-.9570166	3.141741
-----+-----							
Outcome model (Class 2)							
	bias	.6495292	.1555939	4.17	0.000	.3445707	.9544877

³To quickly replicate these results run:

```
. set seed 3
. set more off
. swopit y house gdp bias spread, reg(house gdp) outone(bias spread) outtwo(bias spread) endo initial(-
4.804481 -.6180356 -11.48333 3.389238 12.40146 -14.70939 1.092362 .6495292 1.771607 -1.199481 2.316013
.4467007 .4063919) guesses(1)
```


	spread		1.771607	.247858	7.15	0.000	1.285814	2.2574
	/cut1		-1.199481	.1591936	-7.53	0.000	-1.511495	-.8874674
	/cut2		2.316013	.2594048	8.93	0.000	1.807589	2.824437
-----+-----								
Correlations								
	rho1		.4467007	1.153861	0.39	0.699	-1.814825	2.708227
	rho2		.4063919	.5005391	0.81	0.417	-.5746467	1.38743

The estimates of the correlation coefficients have the expected positive signs, but are not significant.⁴ Both AIC and BIC favor exogenous switching. A simultaneous ML estimation of three latent equations with correlated disturbances requires more observations: in the Monte Carlo simulations, the biases of the choice and class probabilities' estimator in the endogenous switching case are 3–4 times as large as in the exogenous switching case. In the sample, there are on average less than 22 observations per parameter. With a larger sample in hand, it would be feasible to estimate the parameters and probabilities more accurately, to include more control variables, and to obtain a stronger empirical evidence in favor of the endogenous switching. In the simulations, when the sample size is increased from 250 to 500 (1,000), the biases of the choice and class probabilities' estimator are reduced threefold (six times), and the RMSE are smaller by 17%–35% (38%–59%).

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 a finite mixture of ordered probit models, `fmm: oprobit`, does not allow for endogenous switching, when the unobservables in the latent class switching model are correlated with the unobservables in the outcome models. This article described 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`.

7 Acknowledgements

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), the editor Stephen P. Jenkins and an anonymous referee for helpful comments and suggestions on both the manuscript and the routines, and StataCorp technical support staff for help.

⁴An identical value of the likelihood can be obtained with the opposite signs of the correlation coefficients and of the parameters in the regime-switching model, and the switched labels of the latent regimes.

References

- Assenmacher-Wesche, K. 2006. Estimating central banks' preferences from a time-varying empirical reaction function. *European Economic Review* 50 (8): 1951–1974.
- 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.
- Gerlach, S. 2007. Interest rate setting by the ECB, 1999–2006: Words and deeds. *International Journal of Central Banking* 3 (3): 1–46.
- 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. N. 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.
- Greene, W. H., M. N. Harris, B. Hollingsworth, and P. Maitra. 2014. A latent class model for obesity. *Economic Letters* 123 (1): 1–5.
- 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.
- Hamilton, J. D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57 (2): 357–384.
- Hamilton, J. D., and O. Jorda. 2002. A model for the federal funds rate target. *Journal of Political Economy* 110 (5): 1135–1167.
- 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.
- Long, J. S. 1997. Regression models for categorical and limited dependent variables. *Sage Publications, Inc.*
- 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. 2008. Modeling monetary policy in real time: Does discreteness matter? EERC Working Paper Series 08/07e, EERC Research Network, Russia and CIS.
- Sirchenko, A. 2020. A model for ordinal responses with heterogeneous status quo outcomes. *Studies in Nonlinear Dynamics & Econometrics* 24 (1).
- Vanderhart, P. G. 2000. The Federal Reserve’s reaction function under Greenspan: An ordinal probit analysis. *Journal of Macroeconomics* 22 (4): 631–644.
- Wilde, J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69 (3): 309–312.

About the authors

Jochem Huismans was a student of the Amsterdam School of Economics at the University of Amsterdam in the Netherlands.

Jan Willem Nijenhuis was a student of the Amsterdam School of Economics at the University of Amsterdam in the Netherlands.

Andrei Sirchenko was a lecturer in econometrics in the Department of Quantitative Economics at the University of Amsterdam in the Netherlands.

Appendix

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

Parameters	Overlap among covariates		
	no overlap	partial overlap	complete overlap
γ_1	2	2	2
γ_3		1	
μ	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
ρ_1	0.3	0.3	0.3
ρ_2	0.5	0.5	0.5

Note: The variances σ^2 , σ_1^2 and σ_2^2 are fixed to one in all simulations.

Table A2. The average estimated values of parameters in the Monte Carlo experiments with exogenous switching

Sample size	Overlap among covariates		
	no overlap	partial overlap	complete overlap
	250 500 1000 2000	250 500 1000 2000	250 500 1000 2000
γ_1	2.78 2.19 2.08 2.04	2.96 2.26 2.10 2.05	2.70 2.19 2.09 2.04
γ_3		1.48 1.13 1.05 1.02	
μ	0.29 0.22 0.21 0.20	0.33 0.22 0.21 0.20	0.27 0.22 0.21 0.20
$\beta_{1,2}$	2.63 2.22 2.10 2.04	2.64 2.21 2.10 2.04	2.64 2.22 2.10 2.04
$\beta_{1,3}$	1.31 1.11 1.05 1.02	1.32 1.10 1.05 1.02	1.32 1.11 1.05 1.02
$\alpha_{1,1}$	-5.03 -4.25 -4.01 -3.91	-6.89 -5.78 -5.48 -5.34	-5.06 -4.25 -4.01 -3.91
$\alpha_{1,2}$	4.93 4.17 3.94 3.84	3.24 2.71 2.58 2.51	5.02 4.23 3.99 3.89
$\beta_{2,2}$			1.34 1.12 1.05 1.02
$\beta_{2,3}$		-2.63 -2.21 -2.09 -2.04	-2.68 -2.23 -2.10 -2.05
$\beta_{2,4}$	1.34 1.12 1.05 1.02	1.31 1.11 1.05 1.02	
$\beta_{2,5}$	-2.69 -2.23 -2.10 -2.04		
$\alpha_{2,1}$	-5.32 -4.43 -4.17 -4.06	-8.09 -6.83 -6.46 -6.29	-5.13 -4.27 -4.02 -3.92
$\alpha_{2,2}$	5.32 4.42 4.17 4.06	1.27 1.07 1.02 0.99	5.26 4.39 4.12 4.02

Table A3. The average estimated values of parameters in the Monte Carlo experiments with endogenous switching

	Overlap among covariates		
	no overlap	partial overlap	complete overlap
Sample size	250 500 1000 2000	250 500 1000 2000	250 500 1000 2000
γ_1	2.59 2.16 2.07 2.03	2.67 2.23 2.09 2.05	2.53 2.17 2.07 2.03
γ_3		1.33 1.11 1.05 1.02	
μ	0.25 0.22 0.21 0.20	0.27 0.23 0.21 0.20	0.25 0.22 0.21 0.20
$\beta_{1,2}$	2.60 2.20 2.09 2.04	2.58 2.20 2.09 2.04	2.60 2.21 2.09 2.04
$\beta_{1,3}$	1.30 1.10 1.05 1.02	1.29 1.10 1.05 1.02	1.30 1.11 1.05 1.02
$\alpha_{1,1}$	-4.96 -4.21 -4.00 -3.91	-6.72 -5.74 -5.46 -5.34	-4.96 -4.22 -4.00 -3.91
$\alpha_{1,2}$	4.91 4.15 3.94 3.84	3.21 2.72 2.57 2.51	4.97 4.22 3.99 3.89
$\beta_{2,2}$			1.31 1.11 1.05 1.02
$\beta_{2,3}$		-2.56 -2.20 -2.09 -2.04	-2.62 -2.22 -2.09 -2.04
$\beta_{2,4}$	1.30 1.11 1.05 1.02	1.28 1.10 1.04 1.02	
$\beta_{2,5}$	-2.60 -2.21 -2.09 -2.04		
$\alpha_{2,1}$	-5.17 -4.41 -4.16 -4.06	-7.90 -6.81 -6.45 -6.29	-5.03 -4.26 -4.01 -3.92
$\alpha_{2,2}$	5.12 4.37 4.15 4.05	1.22 1.05 1.01 0.99	5.12 4.35 4.11 4.02
ρ_1	-0.03 0.14 0.25 0.29	-0.10 0.10 0.22 0.29	0.01 0.15 0.24 0.29
ρ_2	0.38 0.40 0.47 0.50	0.42 0.43 0.47 0.50	0.39 0.41 0.46 0.50