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

J. L. Huismans[1] J. W. Nijenhuis[2] A. Sirchenko[3]

[1,2,3]Faculty of Economics and Business Economics
University of Amsterdam

August 6, 2021

Overview

We introduce the new **swopit** command that fits a switching ordered probit model with exogenous or endogenous class assignment to two latent regimes.

- produced as a software package for STATA
- exogenous and endogenous switching
- model the decisions of the FOMC to change the federal funds rate
- model (ordinal) health status responses

Previous work

Currently, the following is observed with regard to advancements in the field of endogenous switching:

- existing endogenous switching models deal mainly with continuous or binary data
- zero-inflated model for endogenous class-assignment does not allow for all outcomes in each regime

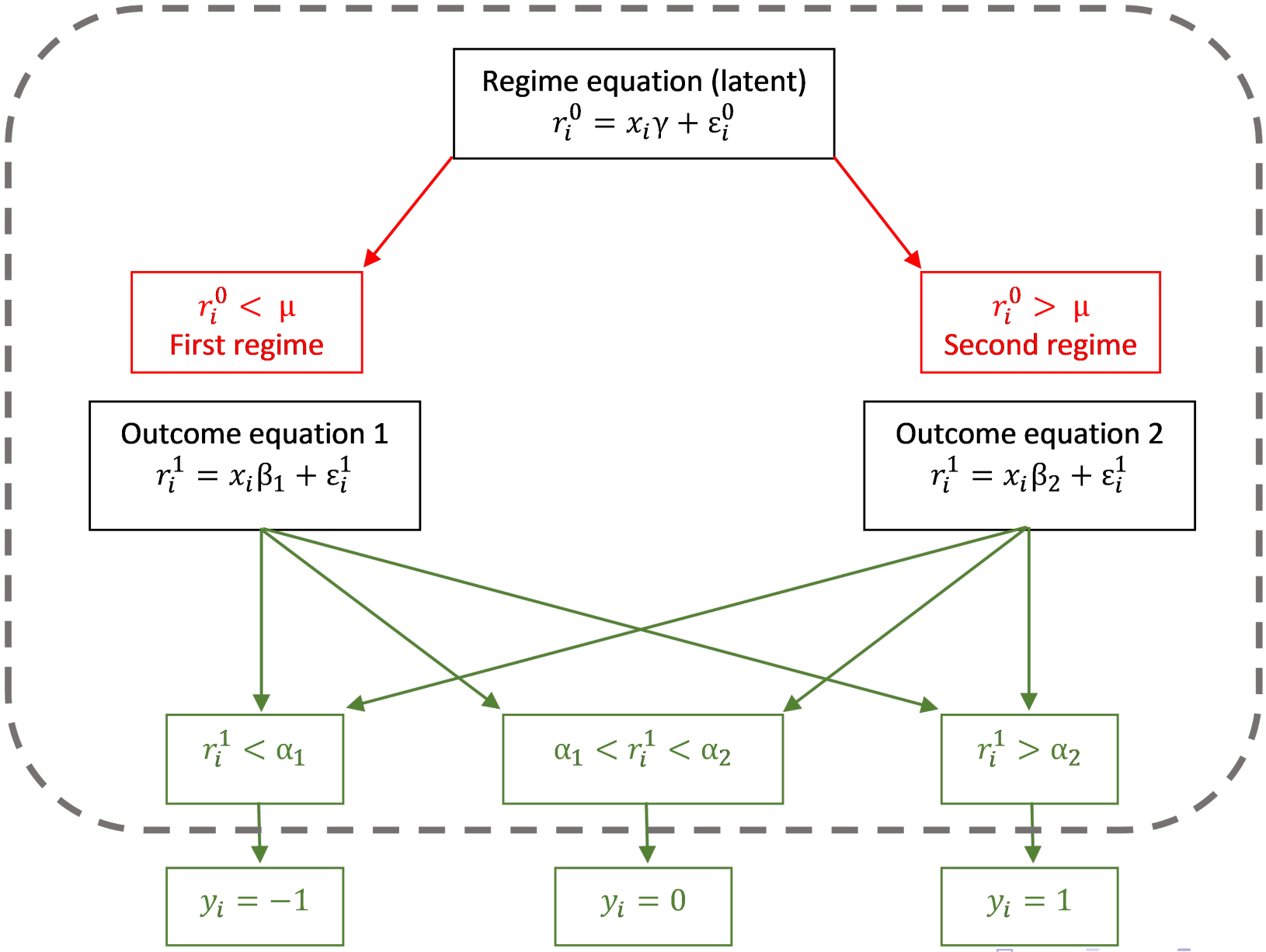
The **swopit** model fits a switching ordered probit model with either exogenous or endogenous switching. There are no restrictions on outcomes in the two regimes.

Model

The **swopit** model allows for:

- two latent regimes which are estimated by binary probit
- a seperate ordered probit model for each regime
- custom starting values
- exogenous and endogenous switching

Switching Ordered Probit Model



The probabilities

$$
\begin{aligned}
Pr(y_i = j | x_i^{all}) &= \Phi_2(\mu - z_i\gamma; \alpha_{1,j} - x_{1,i}\beta_1; \rho_1) \\
&\quad - \Phi_2(\mu - z_i\gamma; \alpha_{1,j-1} - x_{1,i}\beta_1; \rho_1) \\
&\quad + \Phi_2(-\mu + z_i\gamma; \alpha_{2,j} - x_{2,i}\beta_2; -\rho_2) \\
&\quad - \Phi_2(-\mu + z_i\gamma; \alpha_{2,j-1} - x_{2,i}\beta_2; -\rho_2)
\end{aligned}
$$

Identification issue: By design, the model has two maxima in maximum likelihood estimation. Changing the signs of the correlation coefficients and the parameters in the class assignment models switches the regime labels and results in two identical likelihood values.

(Post)estimation command(s)

Within the **swopit** command the user has the option to choose which covariates should be used in the regime equation, as well as both outcome equations. The user can choose his/her own starting values, which type of switching and limitations on the optimization algorithm.
After estimation, the following postestimation commands are available:

- swopitpredict
- swopitprobabilities
- swopitmargins
- swopitclassification

Monte Carlo Simulations (1)

In order to assess the performance of the model, the following Monte Carlo setup was used for a situation with two *latent* regimes and three classes:

1. 5 independent variables $g_j \sim 4 * \mathcal{N}(0, 1)$ were generated and used as covariates.
2. We simulated three different scenarios: 'complete overlap', 'partial overlap' and 'no overlap' between the covariates.
3. Parameters were calibrated to ensure that both *latent* classes have the same theoretical probability. The same holds for the outcomes.
4. To study consistency of the estimates, the simulation is applied for 250, 500, 1000 and 2000 observations.
5. For endogenous switching the correlation between unobservables is set to a positive number, for exogenous switching it is fixed at 0.

Monte Carlo Simulations - Parameters

Sample size	Class assignment: Covariates' overlap:	Exogenous none	Exogenous partial	Exogenous complete	Endogenous none	Endogenous partial	Endogenous complete
		Slope parameters γ, β_1 and β_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

Monte Carlo Simulations - Thresholds

		Threshold parameters μ, α_1 and α_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

Application

The model is applied to two datasets. First, a standard OP model is fitted. Then, the swopit model with exogenous switching and consecutively the swopit model with endogenous switching. The best model is chosen based on the Akaike Information Criterion (AIC). For the FOMC decisions an swopit model with exogenous switching outperforms the other methods. For the health data the swopit model with endogenous switching performs best.

Application FOMC decisions - Setup

- y_t: decision of the FOMC to decrease (-1), increase (1) or leave the federal funds rate target unchanged (0). This is the target variable.
- $house_t$: the Greenbook projection of the total number of new privately owned housing units started for the current quarter. Used as a measure of the housing market.
- gdp_t: the Greenbook projection of quarterly growth in the nominal gross domestic (before 1992: national) product for the current quarter, annualized percentage points.
- $spread_t$: the difference between the one-year treasury constant maturity rate and the effective federal funds rate, three business-day moving average.
- $bias_t$: 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.

Application FOMC decisions - Results

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

Application Health Data - Setup

Consists of 3,202 observations on health status responses.

- $health_i$: reported status of health. Variable can take on the values "pure", "fair", "average", "good" and "excellent". This is the target variable.
- $area_i$: size of area a person is living.
- $weight_i$: weight of an individual in kg.
- $female_i$: indicator if the individual is female (1=female, 0=male).
- $rural_i$: indicator if the individual lives in an urban or rural area (1=rural, 0=urban).

Application health data - Results

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

Concluding remarks

- The currently available Stata commands for ordered probit model do not allow for endogenous switching.
- The currently available Stata commands that allow for endogenous switching do not provide the option for application to ordered probit models.
- The **swopit** command fills this gap in the literature.
- The **swopit** command developed in this paper shows to be consistent and reliable, even in small samples of 250-500 observations.

Contact & Questions

- janwillemnijenhuis@gmail.com
- https://github.com/janwillemnijenhuis/Swopit
- https://ase.uva.nl/content/research-programmes/uva-econometrics/discussion-papers/discussion-papers.html#2021