Estimation of nested and zero-inflated ordered probit models

David Dale
Yandex
Moscow, Russia
dale.david@yandex.ru

Andrei Sirchenko
Higher School of Economics
Moscow, Russia
andrei.sirchenko@gmail.com

Abstract

We introduce three new STATA commands, `nop`, `ziop2` and `ziop3`, for the estimation of a three-part nested ordered probit model, the two-part zero-inflated ordered probit models of Harris and Zhao (2007, *Journal of Econometrics* 141: 1073–1099) and Brooks, Harris and Spencer (2012, *Economics Letters* 117: 683–686), and a three-part zero-inflated ordered probit model, with both exogenous and endogenous switching. The three-part models allow the probabilities of positive, neutral (zero) and negative outcomes to be generated by distinct processes. The zero-inflated models address the preponderance of zero responses and allows the zeros to emerge in two or three latent regimes. We provide the postestimation commands to compute the predicted probabilities and outcomes, the expected values of dependent variable, the marginal effects on the probabilities, the classification tables, and to perform model comparison using the Vuong test (1989, *Econometrica* 57: 307–333). We investigate the finite-sample performance of proposed maximum likelihood estimators by Monte Carlo simulations, discuss the relations among the models, and illustrate them with an empirical application to the U.S. federal funds rate target.

Keywords: ordinal outcomes, zero inflation, nested ordered probit, endogenous switching, two-part model, three-part model, Monte Carlo simulation, maximum likelihood, federal funds rate target.

1 Introduction

We introduce the STATA commands, `nop`, `ziop2` and `ziop3`, which estimate the two-level nested and zero-inflated ordered probit models including the zero- and middle-inflated models of Harris and Zhao (2007), Bagozzi and Mukherjee (2012), Brooks, Harris and Spencer (2012) and Sirchenko (2013). The rationale behind the two-level nested decision process is standard in the discrete-choice modeling when the set of alternatives faced by a decision-maker can be partitioned into subsets (or nests) with similar alternatives correlated due to the common unobserved factors. A choice among the nests and a choice among the alternatives within each nest can be driven by different sets of observed and unobserved factors (and common factors can have different weights).

In the case of unordered categorical data, in which choices can be grouped into the nests of similar options, the nested logit model is a popular method. The nested models for ordinal data are rare although the rationale behind them is similar: choosing among a negative response (decrease), a neutral response (no change) or a positive response (increase) is quite different from choosing the magnitude of negative response; and choosing the magnitude of negative response can be driven by quite different determinants than choosing the magnitude of positive response. This leads to three implicit decisions: an upper-level regime decision — a choice among the nests, and two lower-level outcome decisions — the choices of the magnitude of negative and positive responses (see the top left panel of Figure 1).

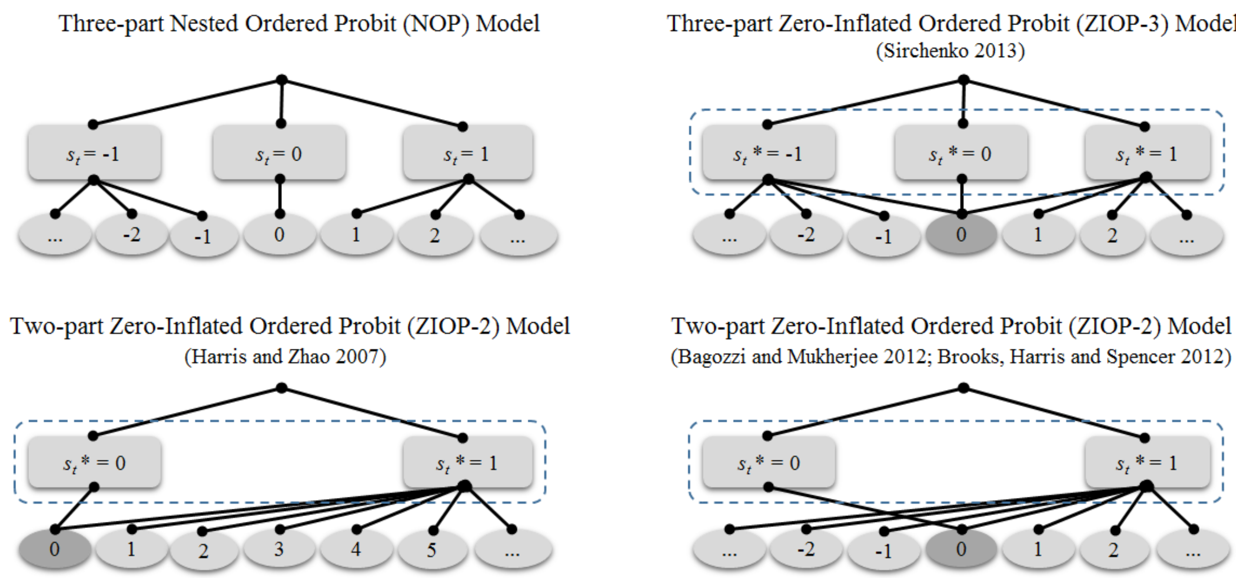
Furthermore, it would be reasonable for the zero (no-change) alternative to be in three nests: its own one, one with negative responses and one with positive responses; so some zeros can be driven by similar factors as negative or positive responses. This leads to a three-part cross-nested model with the nests overlapping at a zero response; hence, the probability of zeros is 'inflated'. Since the regime decision is not observable, the zeros are observationally equivalent — it is never known to which of the three nests the observed zero belongs. While several types of models with overlapping nests for unordered categorical responses are developed (Vovsha 1997; Wen and Koppelman 2001), the cross-nested models for ordinal outcomes are very scarce.[1]

The prevalence of status quo, neutral or zero outcomes is observed in many fields, including economics, sociology, technometrics, psychology and biology. The heterogeneity of zeros is widely recognized — see Winkelmann (2008) and Greene and Hensher (2010) for a review. Studies identify different types of zeros such as: no visits to doctor due to good health, iatrophobia, or medical costs; no illness due to strong immunity or lack of infec-

[1] Small (1987) introduced an ordered-choice model with overlapping nests, which contain two adjacent choices.

tion; no children due to infertility or choice. In the studies of survey responses using an odd-point Likert-type scale, where the respondents must indicate the negative, neutral or positive attitude or opinion, the heterogeneity of indifferent responses (a true neutral option versus an undecided, or ambivalent, or uninformed one, commonly reported as neutral) is also well-recognized and sometimes labeled as the middle category endorsement or inflation (Bagozzi and Mukhetjee 2012; Hernández, Drasgow and Gonzáles-Romá 2004; Kulas and Stachowski 2009).

Figure 1. Decision trees of nested and zero-inflated ordered probit models



Notes: Decisionmakers are not assumed to choose sequentially. The tree diagrams simply represent a nesting structure of the system of ordered probit models.

The two-part zero-inflated models, developed to address the unobserved heterogeneity of zeros, combines a binary choice model for the probability of crossing the hurdle (to participate or not to participate; to consume or not to consume) with a count or ordered-choice model for nonnegative outcomes above the hurdle: the two parts are estimated jointly, and the zero observations can emerge in both parts. The two-part zero-inflated models include the zero-inflated Poisson (Lambert 1992), negative binomial (Greene 1994), binomial (Hall 2002) and generalized Poisson (Famoye and Singh 2003) models for count outcomes, and the zero-inflated ordered probit model (Harris and Zhao 2007) and zero-inflated proportional odds model (Kelley and Anderson 2008) for non-negative ordinal responses.[2]

The model of Harris and Zhao (2007) is suitable for explaining decisions such as the levels of consumption, when the upper hurdle is naturally binary (to consume or not to consume), the responses are non-negative and the inflated zeros are situated at one end of the ordered scale (see the bottom left panel of Figure 1). Bagozzi and Mukherjee (2012) and Brooks, Harris and Spencer (2012) modified the model of Harris and Zhao (2007) and developed the middle-inflated ordered probit model for an ordinal outcome, which ranges from negative to positive responses, and where an abundant outcome is situated in the middle of the choice spectrum (see the bottom right panel of Figure 1).

The three-part cross-nested zero-inflated ordered probit model (see the top right panel of Figure 1) introduced in Sirchenko (2013) is a natural generalization of the models of Harris and Zhao (2007), Bagozzi and Mukherjee (2012) and Brooks, Harris and Spencer (2012). A trichotomous regime decision is more realistic and flexible than a binary decision (change or no change) if applied to ordinal data with negative, zero and positive values.

2 Models

2.1 Notation and assumptions

The observed dependent variable y_t, $t = 1, 2, ..., T$ is assumed to take on a finite number of ordinal values j coded as $\{-J^-, ..., -1, 0, 1, ..., J^+\}$, where a potentially heterogeneous (and typically predominant) response is coded as zero. The latent unobserved (or only partially observed) variables are denoted by $*$. Each model assumes an ordered-choice regime decision and the ordered-choice outcome decisions conditional on the regime. The regime decision is allowed to be correlated with each outcome decision. We denote by $\mathbf{x}_t$, $\mathbf{x}_t^-$, $\mathbf{x}_t^+$ and $\mathbf{z}_t$ the t^{th} rows of the observed data matrices (which in addition to predetermined explanatory variables may also include the lags of y_t); by $\boldsymbol{\beta}$, $\boldsymbol{\beta}^-$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the vectors of unknown slope parameters; by $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^-$, $\boldsymbol{\alpha}^+$ and $\boldsymbol{\mu}$ the vectors of unknown threshold parameters; by ρ, ρ^- and ρ^+ the vectors of unknown correlation coefficients; by ε_t, ε_t^-, ε_t^+ and ν_t the error terms that are independently and identically distributed (*iid*) across t with normal cumulative distribution function (CDF) Φ, the zero mean and the variances σ^2, σ_-^2, σ_+^2 and σ_ν^2, respectively; and by $\Phi_2(g_1;g_2;\sigma_1^2;\sigma_2^2;\rho)$ the CDF of the bivariate normal distribution of the two random variables g_1 and g_2 with the zero means, the variances σ_1^2 and σ_2^2 and the correlation coefficient ρ:

[2] The zero-inflated models, estimation of which is currently implemented in STATA, include: the zero-inflated Poisson model (the zip command), the negative binomial model (the zinb command), and the binomial model (the zib command) and the beta-binomial model (the zibbin command) developed by Hardin and Hilbe (2014).

$$\Phi_2(g_1;g_2;\sigma_1^2;\sigma_2^2;\rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{g_1}\int_{-\infty}^{g_2} \exp\left(-\frac{u^2/\sigma_1^2 - 2\rho uw/\sigma_1\sigma_2 + w^2/\sigma_2^2}{2(1-\rho^2)}\right) du dw.$$

2.2 Three-part nested ordered probit (NOP) model

Despite the wide-spread use of nested logit models for unordered categorical responses we are not aware of any example of the nested ordered probit/logit model in the literature. The two-level NOP model can be described as

Upper-level decision: $r_t^* = \mathbf{z}_t\boldsymbol{\gamma} + \nu_t, \quad s_t = \begin{cases} 1 & \text{if} \quad \mu_2 < r_t^*, \\ 0 & \text{if} \quad \mu_1 < r_t^* \le \mu_2, \\ -1 & \text{if} \quad r_t^* \le \mu_1. \end{cases}$

Lower-level decisions: $y_t^{-*} = \mathbf{x}_t^-\boldsymbol{\beta}^- + \varepsilon_t^-, \quad y_t^{+*} = \mathbf{x}_t^+\boldsymbol{\beta}^+ + \varepsilon_t^+,$

$$y_t = \begin{cases} j (j > 0) & \text{if} \quad s_t = 1 \text{ and } \alpha_{j-1}^+ < y_t^{+*} \le \alpha_j^+, \\ 0 & \text{if} \quad s_t = 0, \\ j (j < 0) & \text{if} \quad s_t = -1 \quad \text{and } \alpha_j^- < y_t^{-*} \le \alpha_{j+1}^-, \end{cases}$$

where $-\infty = \alpha_0^+ \le \alpha_1^+ \le ... \le \alpha_{J^+}^+ = \infty$

and $-\infty = \alpha_{-J^-}^- \le \alpha_{-J^-+1}^- \le ... \le \alpha_0^- = \infty.$

Correlation among decisions: $\begin{bmatrix} \nu_t \\ \varepsilon_t^i \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \sigma_\nu^2 & \rho^i\sigma_\nu\sigma_i \\ \rho^i\sigma_\nu\sigma_i & \sigma_i^2 \end{bmatrix}\right), i \in \{-,+\}.$

The probabilities of the outcome j in the NOP model are given by

$$\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t^-, \mathbf{x}_t^+) = I_{j<0}\Pr(r_t^* \le \mu_1 \text{ and } \alpha_j^- < y_t^{-*} \le \alpha_{j+1}^-|\mathbf{z}_t, \mathbf{x}_t^-) \\
&+I_{j=0}\Pr(\mu_1 < r_t^* \le \mu_2|\mathbf{z}_t) + I_{j>0}\Pr(\mu_2 < r_t^* \text{ and } \alpha_{j-1}^+ < y_t^{+*} \le \alpha_j^+|\mathbf{z}_t, \mathbf{x}_t^+) \\
&= I_{j<0}\Pr(\nu_t \le \mu_1 - \mathbf{z}_t\boldsymbol{\gamma} \text{ and } \alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^- < \varepsilon_t^- \le \alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-) \\
&+I_{j=0}\Pr(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma} < \nu_t \le \mu_2 - \mathbf{z}_t\boldsymbol{\gamma}) \\
&+I_{j>0}\Pr(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma} < \nu_t \text{ and } \alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+ < \varepsilon_t^+ \le \alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+) \\
&= I_{j<0}[\Phi_2(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-;\sigma_\nu^2;\sigma_-^2;\rho^-) - \Phi_2(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^-;\sigma_\nu^2;\sigma_-^2;\rho^-)] \\
&+I_{j=0}[\Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma};\sigma_\nu^2) - \Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma};\sigma_\nu^2)] \\
&+I_{j>0}[\Phi_2(-\mu_2 + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_\nu^2; \sigma_+^2; -\rho^+) \\
&-\Phi_2(-\mu_2 + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_\nu^2; \sigma_+^2; -\rho^+)],
\end{aligned} \tag{1}$$

where $I_{j<0}$ is an indicator function such that $I_{j<0} = 1$ if $j < 0$, and $I_{j<0} = 0$ if $j \ge 0$ (analogously for $I_{j=0}$ and $I_{j>0}$).

In the case of exogenous switching (when $\rho^- = \rho^+ = 0$), the probabilities of the outcome j in the NOP can be computed as

$$
\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t^-, \mathbf{x}_t^+, \rho^- = \rho^+ = 0)\\
&= I_{j<0}\Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)[\Phi(\alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_-^2) - \Phi(\alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_-^2)]\\
&+I_{j=0}[\Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma}) - \Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma})]\\
&+I_{j>0}[1 - \Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)][\Phi(\alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_+^2) - \Phi(\alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_+^2)].
\end{aligned}
$$

In the case of two or three outcome choices the NOP model degenerates to the conventional single-equation ordered probit model.

2.3 Two-part zero-inflated ordered probit (ZIOP-2) model

The ZIOP-2 model, which represents the two part zero-inflated ordered probit models of Bagozzi and Mukherjee (2012) and Brooks, Harris and Spencer (2012), can be described by the following system

$$
\begin{array}{ll}
\text{Regime decision:} & r_t^* = \mathbf{z}_t\boldsymbol{\gamma} + \nu_t, \quad s_t^* = \begin{cases} 1 & \text{if} \quad \mu < r_t^*, \\ 0 & \text{if} \quad r_t^* \leq \mu. \end{cases} \\
\text{Outcome decision:} & y_t^* = \mathbf{x}_t\boldsymbol{\beta} + \varepsilon_t, \\
& y_t = \begin{cases} j & \text{if} \quad s_t^* = 1 \text{ and } \alpha_{j-1} < y_t^* \leq \alpha_j, \\ 0 & \text{if} \quad s_t^* = 0, \end{cases} \\
& \text{where } -\infty = \alpha_{-J^- -1} \leq \alpha_{-J^-} \leq ... \leq \alpha_{J^+} = \infty. \\
\begin{array}{l}\text{Correlation among}\\ \text{decisions:}\end{array} & \begin{bmatrix} \nu_t \\ \varepsilon_t \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \sigma_\nu^2 & \rho\sigma_\nu\sigma \\ \rho\sigma_\nu\sigma & \sigma^2 \end{bmatrix} \right).
\end{array}
$$

The probabilities of the outcome j in the ZIOP-2 model are given by

$$
\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t) = I_{j=0}\Pr(r_t^* \leq \mu|\mathbf{z}_t) + I_{j\geq 0}\Pr(\mu < r_t^* \text{ and } \alpha_{j-1} < y_t^* \leq \alpha_j|\mathbf{z}_t, \mathbf{x}_t)\\
&= I_{j=0}\Pr(\nu_t \leq \mu - \mathbf{z}_t\boldsymbol{\gamma}) + I_{j\geq 0}\Pr(\mu - \mathbf{z}_t\boldsymbol{\gamma} < \nu_t \text{ and } \alpha_{j-1} - \mathbf{x}_t\boldsymbol{\beta} < \varepsilon_t \leq \alpha_j - \mathbf{x}_t\boldsymbol{\beta})\\
&= I_{j=0}\Phi(\mu - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2) + \Phi_2(-\mu + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_j - \mathbf{x}_t\boldsymbol{\beta}; \sigma_\nu^2; \sigma^2; -\rho)\\
&-\Phi_2(-\mu + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_{j-1} - \mathbf{x}_t\boldsymbol{\beta}; \sigma_\nu^2; \sigma^2; -\rho).
\end{aligned} \tag{2}
$$

In the case of exogenous switching (when $\rho = 0$), the probabilities of the outcome j in the ZIOP-2 model can be computed as

$$
\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t, \rho = 0) = I_{j=0}\Phi(\mu - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)\\
&+[1 - \Phi(\mu - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)][\Phi(\alpha_j - \mathbf{x}_t\boldsymbol{\beta}; \sigma^2) - \Phi(\alpha_{j-1} - \mathbf{x}_t\boldsymbol{\beta}; \sigma^2)].
\end{aligned}
$$

If $y_t \geq 0$ for $\forall t$, the ZIOP-2 model represents the model of Harris and Zhao (2007).

2.4 Three-part zero-inflated ordered probit (ZIOP-3) model

The ZIOP-3 model developed by Sirchenko (2013) is a three-part generalization of the ZIOP-2 model, and can be described by the following system

Regime decision: $$r_t^* = \mathbf{z}_t\boldsymbol{\gamma} + \nu_t, \quad s_t^* = \begin{cases} 1 & \text{if} \quad \mu_2 < r_t^*, \\ 0 & \text{if} \quad \mu_1 < r_t^* \le \mu_2, \\ -1 & \text{if} \quad r_t^* \le \mu_1. \end{cases}$$

Outcome decisions: $$y_t^{-*} = \mathbf{x}_t^-\boldsymbol{\beta}^- + \varepsilon_t^-, \quad y_t^{+*} = \mathbf{x}_t^+\boldsymbol{\beta}^+ + \varepsilon_t^+,$$

$$y_t = \begin{cases} j(j \ge 0) & \text{if} \quad s_t^* = 1 \quad \text{and } \alpha_{j-1}^+ < y_t^{+*} \le \alpha_j^+, \\ 0 & \text{if} \quad s_t^* = 0, \\ j(j \le 0) & \text{if} \quad s_t^* = -1 \text{ and } \alpha_j^- < y_t^{-*} \le \alpha_{j+1}^-, \end{cases}$$

where $-\infty = \alpha_{-1}^+ \le \alpha_0^+ \le ... \le \alpha_{J^+}^+ = \infty$

and $-\infty = \alpha_{-J^-}^- \le \alpha_{-J^-+1}^- \le ... \le \alpha_1^- = \infty.$

Correlation among decisions: $$\begin{bmatrix} \nu_t \\ \varepsilon_t^i \end{bmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{matrix} 0 \\ 0 \end{matrix}, \begin{bmatrix} \sigma_\nu^2 & \rho^i\sigma_\nu\sigma_i \\ \rho^i\sigma_\nu\sigma_i & \sigma_i^2 \end{bmatrix} \right), \; i \in \{-,+\}.$$

The probabilities of the outcome j in the ZIOP-3 model are given by

$$\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t^-, \mathbf{x}_t^+) = I_{j\le 0}\Pr(r_t^* \le \mu_1 \text{ and } \alpha_j^- < y_t^{-*} \le \alpha_{j+1}^-|\mathbf{z}_t, \mathbf{x}_t^-) \\
&+I_{j=0}\Pr(\mu_1 < r_t^* \le \mu_2|\mathbf{z}_t) + I_{j\ge 0}\Pr(\mu_2 < r_t^* \text{ and } \alpha_{j-1}^+ < y_t^{+*} \le \alpha_j^+|\mathbf{z}_t, \mathbf{x}_t^+) \\
&= I_{j\le 0}\Pr(\nu_t \le \mu_1 - \mathbf{z}_t\boldsymbol{\gamma} \text{ and } \alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^- < \varepsilon_t^- \le \alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-) \\
&+I_{j=0}\Pr(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma} < \nu_t \le \mu_2 - \mathbf{z}_t\boldsymbol{\gamma}) \\
&+I_{j\ge 0}\Pr(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma} < \nu_t \text{ and } \alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+ < \varepsilon_t^+ \le \alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+) \\
&= I_{j\le 0}[\Phi_2(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_\nu^2; \sigma_-^2; \rho^-) - \Phi_2(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_\nu^2; \sigma_-^2; \rho^-)] \\
&+I_{j=0}[\Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2) - \Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)] \\
&+I_{j\ge 0}[\Phi_2(-\mu_2 + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_\nu^2; \sigma_+^2; -\rho^+) \\
&-\Phi_2(-\mu_2 + \mathbf{z}_t\boldsymbol{\gamma}; \alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_\nu^2; \sigma_+^2; -\rho^+)],
\end{aligned} \tag{3}$$

where $I_{j\le 0}$ is an indicator function such that $I_{j\le 0} = 1$ if $j \le 0$, and $I_{j\le 0} = 0$ if $j > 0$ (analogously for $I_{j\ge 0}$).

In the case of exogenous switching (when $\rho^- = \rho^+ = 0$), the probabilities of the outcome j in the ZIOP-3 model can be computed as

$$\begin{aligned}
&\Pr(y_t = j|\mathbf{z}_t, \mathbf{x}_t^-, \mathbf{x}_t^+, \rho^- = \rho^+ = 0) = I_{j\le 0}\Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \boldsymbol{\sigma_\nu^2})[\Phi(\alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_-^2) \\
&-\Phi(\alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^-; \sigma_-^2)] + I_{j=0}[\Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2) - \Phi(\mu_1 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)] \\
&+I_{j\ge 0}[1 - \Phi(\mu_2 - \mathbf{z}_t\boldsymbol{\gamma}; \sigma_\nu^2)][\Phi(\alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_+^2) - \Phi(\alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+; \sigma_+^2)].
\end{aligned}$$

The inflated outcome does not have to be in the *very* middle of the ordered choices. If it is located at the *end* of the ordered scale, i.e. if $y_t \geq 0$ for $\forall t$, the ZIOP-3 model reduces to the ZIOP-2 model of Harris and Zhao (2007).

2.5 Maximum likelihood (ML) estimation

The probabilities in each ordered probit equation can be consistently estimated under fairly general conditions by an asymptotically normal ML estimator (Basu and de Jong 2007). The simultaneous estimation of the ordered probit equations in the NOP, ZIOP-2 and ZIOP-3 models can be also performed using an ML estimator of the vector of the parameters $\boldsymbol{\theta}$ that solves

$$\max_{\boldsymbol{\theta}\epsilon\boldsymbol{\Theta}} \sum_{t=1}^{T} \sum_{j=-J^-}^{J^+} I_{tj} \ln[\Pr(y_t = j|\mathbf{x}_t^{all}, \boldsymbol{\theta})], \tag{4}$$

where I_{tj} is an indicator function such that $I_{tj} = 1$ if $y_t = j$ and $I_{tj} = 0$ otherwise; $\boldsymbol{\theta}$ includes $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$, $\boldsymbol{\beta}^-$, $\boldsymbol{\beta}^+$, $\boldsymbol{\alpha}^-$, $\boldsymbol{\alpha}^+$, ρ^- and ρ^+ for the NOP and ZIOP-3 models, and $\boldsymbol{\gamma}$, μ, $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ρ for the ZIOP-2 model; Θ is a parameters' space; $\mathbf{x}_t^{all}$ is a vector that contains the values of all covariates in the model; and $\Pr(y_t = j|\mathbf{x}_t^{all}, \boldsymbol{\theta})$ are the probabilities from either (1) or (2) or (3). The asymptotic standard errors of $\widehat{\boldsymbol{\theta}}$ can be computed from the Hessian matrix.

The intercept components of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^-$, $\boldsymbol{\beta}^+$ and $\boldsymbol{\gamma}$ are identified up to scale and location, that is only jointly with the corresponding threshold parameters $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^-$, $\boldsymbol{\alpha}^+$ and $\boldsymbol{\mu}$ and variances σ^2, σ_-^2, σ_+^2, and σ_ν^2. As is common in the identification of discrete choice models, the variances σ^2, σ_-^2, σ_+^2, and σ_ν^2 are fixed to one, and the intercept components of $\boldsymbol{\beta}$, $\boldsymbol{\beta}^-$, $\boldsymbol{\beta}^+$ and $\boldsymbol{\gamma}$ are fixed to zero. The probabilities in (1), (2) and (3) are invariant to these (arbitrary) identifying assumptions: up to scale and location, we can identify all parameters in $\boldsymbol{\theta}$ because of the nonlinearity of ordered probit equations, i.e. via the functional form (Heckman 1978; Wilde 2000). However, since the normal CDF is approximately linear in the middle of its support, the simultaneous estimation of two or three equations may experience a weak identification problem if regime and outcome equations contain the same set of covariates. To enhance the precision of parameter estimates we may impose exclusion restrictions on the specification of covariates in each equation.

The three regimes (nests) in the NOP model are fully observable, contrary to the latent (only partially observed) regimes in the ZIOP-2 and ZIOP-3 models. The likelihood function of the NOP model — again in contrast with the ZIOP-2 and ZIOP-3 models — is separable with respect to the parameters in the three equations. Thus, solving (4) for the NOP model is equivalent to maximizing separately the likelihoods of the three ordered probit models representing the upper- and lower-level decisions.[3]

2.6 Marginal effects (ME)

The marginal effect of a continuous covariate k (the k^{th} element of $\mathbf{x}_t^{all}$) on the probability of each discrete outcome j are computed for the ZIOP-3 model as

$$
\begin{aligned}
\mathrm{ME}_{k,j,t} &= \tfrac{\partial \Pr(y_t=j|\boldsymbol{\theta})}{\partial \mathbf{x}_{t,k}^{all}} = I_{j\le 0}\Bigg\{\Bigg[\Phi\left(\tfrac{\mu_1-\mathbf{z}_t\boldsymbol{\gamma}-\rho^-(\alpha_j^- -\mathbf{x}_t^-\boldsymbol{\beta}^-)}{\sqrt{1-(\rho^-)^2}}\right) f(\alpha_j^- - \mathbf{x}_t^-\boldsymbol{\beta}^-) \\
&-\Phi\left(\tfrac{\mu_1-\mathbf{z}_t\boldsymbol{\gamma}-\rho^-(\alpha_{j+1}^- -\mathbf{x}_t^-\boldsymbol{\beta}^-)}{\sqrt{1-(\rho^-)^2}}\right) f(\alpha_{j+1}^- - \mathbf{x}_t^-\boldsymbol{\beta}^-)\Bigg]\boldsymbol{\beta}_k^{-all} \\
&-\left[\Phi\left(\tfrac{\alpha_{j+1}^- -\mathbf{x}_t^-\boldsymbol{\beta}^- -\rho^-(\mu_1-\mathbf{z}_t\boldsymbol{\gamma})}{\sqrt{1-(\rho^-)^2}}\right) - \Phi\left(\tfrac{\alpha_j^- -\mathbf{x}_t^-\boldsymbol{\beta}^- -\rho^-(\mu_1-\mathbf{z}_t\boldsymbol{\gamma})}{\sqrt{1-(\rho^-)^2}}\right)\right] f(\mu_1-\mathbf{z}_t\boldsymbol{\gamma})\boldsymbol{\gamma}_k^{all}\Bigg\} \\
&-I_{j=0}[f(\mu_2-\mathbf{z}_t\boldsymbol{\gamma}) - f(\mu_1-\mathbf{z}_t\boldsymbol{\gamma})]\boldsymbol{\gamma}_k^{all} \\
&+I_{j\ge 0}\Bigg\{\Bigg[\Phi\left(\tfrac{\mathbf{z}_t\boldsymbol{\gamma}-\mu_2+\rho^+(\alpha_{j-1}^+ -\mathbf{x}_t^+\boldsymbol{\beta}^+)}{\sqrt{1-(\rho^+)^2}}\right) f(\alpha_{j-1}^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+) \\
&-\Phi\left(\tfrac{\mathbf{z}_t\boldsymbol{\gamma}-\mu_2+\rho^+(\alpha_j^+ -\mathbf{x}_t^+\boldsymbol{\beta}^+)}{\sqrt{1-(\rho^+)^2}}\right) f(\alpha_j^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+)\Bigg]\boldsymbol{\beta}_k^{+all} \\
&+\left[\Phi\left(\tfrac{\alpha_j^+ -\mathbf{x}_t^+\boldsymbol{\beta}^+ +\rho^+(\mathbf{z}_t\boldsymbol{\gamma}-\mu_2)}{\sqrt{1-(\rho^+)^2}}\right) - \Phi\left(\tfrac{\alpha_{j-1}^+ -\mathbf{x}_t^+\boldsymbol{\beta}^+ +\rho^+(\mathbf{z}_t\boldsymbol{\gamma}-\mu_2)}{\sqrt{1-(\rho^+)^2}}\right)\right] f(\mathbf{z}_t\boldsymbol{\gamma}-\mu_2)\boldsymbol{\gamma}_k^{all}\Bigg\},
\end{aligned}
$$

where f is the probability density function of the standard normal distribution, and $\boldsymbol{\gamma}_k^{all}$, $\boldsymbol{\beta}_k^{-all}$ and $\boldsymbol{\beta}_k^{+all}$ are the coefficients on the k^{th} covariate in $\mathbf{x}_t^{all}$ in the regime equation, outcome equation conditional on $s_t^* = 1$ and outcome equation conditional on $s_t^* = -1$, respectively ($\boldsymbol{\gamma}_k^{all}$, $\boldsymbol{\beta}_k^{-all}$ or $\boldsymbol{\beta}_k^{+all}$ is zero if the k^{th} covariate in $\mathbf{x}_t^{all}$ is not included into the corresponding equation). For a discrete-valued covariate, the ME can be computed as the change in the probabilities when this covariate changes by one increment and all other covariates are fixed.

The MEs for the NOP model are computed by replacing $I_{j\ge 0}$ with $I_{j>0}$ and $I_{j\le 0}$ with $I_{j<0}$.

The MEs for the ZIOP-2 model are computed as

$$
\begin{aligned}
\mathrm{ME}_{k,j,t} &= \tfrac{\partial \Pr(y_t=j|\boldsymbol{\theta})}{\partial \mathbf{x}_{t,k}^{all}} = -I_{j=0}[f(\mu-\mathbf{z}_t\boldsymbol{\gamma})]\boldsymbol{\gamma}_k^{all} \\
&+\left[\Phi\left(\tfrac{\mathbf{z}_t\boldsymbol{\gamma}-\mu+\rho(\alpha_{j-1}-\mathbf{x}_t\boldsymbol{\beta})}{\sqrt{1-\rho^2}}\right) f(\alpha_{j-1}-\mathbf{x}_t\boldsymbol{\beta}) - \Phi\left(\tfrac{\mathbf{z}_t\boldsymbol{\gamma}-\mu+\rho(\alpha_j-\mathbf{x}_t\boldsymbol{\beta})}{\sqrt{1-\rho^2}}\right) f(\alpha_j-\mathbf{x}_t\boldsymbol{\beta})\right]\boldsymbol{\beta}_k^{all} \\
&+\left[\Phi\left(\tfrac{\alpha_j-\mathbf{x}_t\boldsymbol{\beta}+\rho(\mathbf{z}_t\boldsymbol{\gamma}-\mu)}{\sqrt{1-\rho^2}}\right) - \Phi\left(\tfrac{\alpha_{j-1}-\mathbf{x}_t\boldsymbol{\beta}+\rho(\mathbf{z}_t\boldsymbol{\gamma}-\mu)}{\sqrt{1-\rho^2}}\right)\right] f(\mathbf{z}_t\boldsymbol{\gamma}-\mu)\boldsymbol{\gamma}_k^{all},
\end{aligned}
$$

[3] The data matrices in the lower-level decisions should be truncated to contain only those rows of $\mathbf{x}_t^-$ or $\mathbf{x}_t^+$ for which $y_t < 0$ or $y_t > 0$, respectively.

where $\boldsymbol{\beta}_k^{all}$ is the coefficient on the k^{th} covariate in $\mathbf{x}_t^{all}$ in the outcome equation ($\boldsymbol{\beta}_k^{all}$ is zero if the k^{th} covariate in $\mathbf{x}_t^{all}$ is not included into the outcome equation).

The asymptotic standard errors of the MEs are computed using the Delta method as the square roots of the diagonal elements of

$$\widehat{Var(\widehat{\mathbf{ME}}_{k,j,t})} = \nabla_\theta \widehat{\mathbf{ME}}_{k,j,t} \widehat{Var(\widehat{\boldsymbol{\theta}})} \nabla_\theta \widehat{\mathbf{ME}}_{k,j,t})'.$$

2.7 Relations among the models and their comparison

We discuss now the choice of a formal model-selection test, which depends on whether the models are nested in each other.

The exogenous-switching version of each model is nested in its endogenous-switching version as its uncorrelated special case; their comparison can be performed using any classical lilkelihood-based test for nested hypotheses, such as the likelihood ratio (LR) test.

The NOP model is nested in the ZIOP-3 model. The latter becomes a NOP model if $\alpha_{-1}^- \to \infty$ and $\alpha_1^+ \to -\infty$; therefore, $\Pr(y_t = 0|\mathbf{x}_t^+, s_t^* = 1) \to 0$ and $\Pr(y_t = 0|\mathbf{x}_t^-, s_t^* = -1) \to 0$. Thus, the comparison of the NOP and ZIOP-3 models can also be performed with the LR test; however, the critical values of the classical LR test are invalid since some standard regularity conditions of the classical LR test fail to hold. In particular, the values of α_{-1}^- and α_1^+ in the null hypothesis are not the interior points of the parameter space; hence, the asymptotic distribution of the LR statistics is not standard. Instead, one may use the simulated critical values provided in Andrews (2001).

Generally, the ZIOP-2 model is not a special case of the ZIOP-3 model, and vice versa. However, they are not strictly non-nested and overlap if all their slope parameters are fixed to zeros. We can compare them using a likelihood-based test for non-nested overlapping models, such as the Vuong test (Vuong 1989). A special case when the ZIOP-3 model nests the ZIOP-2 model emerges under some restrictions on the parameters as explained below. In this case, the selection between the ZIOP-3 and ZIOP-2 models can be performed using any classical likelihood-based test for nested hypotheses.

The special case emerges if y_t takes on only three discrete values $j \in \{-1, 0, 1\}$, the regressors in $\mathbf{x}_t^-$ and $\mathbf{x}_t^+$ in the outcome equations of the ZIOP-3 model contain all regressors in the ZIOP-2 regime equation (denoted below by $\mathbf{z}_{2t}$ with the parameter vector $\boldsymbol{\gamma}_2$), and the regressors in the regime equation of the ZIOP-3 model (denoted below by $\mathbf{z}_{3t}$ with the parameter vector $\boldsymbol{\gamma}_3$) include all regressors in the $\mathbf{x}_t$ in the ZIOP-2 outcome equation. According to (2) the probabilities of the outcome j in the ZIOP-2 model are given by

$$
\begin{aligned}
&\Pr(y_t=-1|\mathbf{z}_{2t},\mathbf{x}_t)=\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};-\rho);\\
&\Pr(y_t=0|\mathbf{z}_{2t},\mathbf{x}_t)=\Phi(\mu-\mathbf{z}_{2t}\boldsymbol{\gamma}_2)+\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\alpha_0-\mathbf{x}_t\boldsymbol{\beta};-\rho)\\
&-\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};-\rho)=1-\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;-\alpha_0+\mathbf{x}_t\boldsymbol{\beta};\rho)\\
&-\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};-\rho);\\
&\Pr(y_t=1|\mathbf{z}_{2t},\mathbf{x}_t)=\Phi(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma})-\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\alpha_0-\mathbf{x}_t\boldsymbol{\beta};-\rho)\\
&=\Phi_2(-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;-\alpha_0+\mathbf{x}_t\boldsymbol{\beta};\rho),
\end{aligned}
\tag{5}
$$

since $\Phi_2(x;y;\rho)=\Phi(x)-\Phi_2(x;-y;-\rho)$.

Similarly, according to (3) the probabilities of the outcome j in the ZIOP-3 model are given by

$$
\begin{aligned}
&\Pr(y_t=-1|\mathbf{z}_{3t},\mathbf{x}_t^-,\mathbf{x}_t^+)=\Phi_2(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3;\alpha_0^--\mathbf{x}_t^-\boldsymbol{\beta}^-;\rho^-);\\
&\Pr(y_t=0|\mathbf{z}_{3t},\mathbf{x}_t^-,\mathbf{x}_t^+)=\Phi(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3)-\Phi_2(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3;\alpha_0^--\mathbf{x}_t^-\boldsymbol{\beta}^-;\rho^-)\\
&+\Phi(\mu_2-\mathbf{z}_{3t}\boldsymbol{\gamma}_3)-\Phi(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3)+\Phi_2(-\mu_2+\mathbf{z}_{3t}\boldsymbol{\gamma}_3;\alpha_0^+-\mathbf{x}_t^+\boldsymbol{\beta}^+;-\rho^+)\\
&=\Phi_2(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3;-\alpha_0^-+\mathbf{x}_t^-\boldsymbol{\beta}^-;-\rho^-)+\Phi(\mu_2-\mathbf{z}_{3t}\boldsymbol{\gamma}_3)\\
&-\Phi(\mu_1-\mathbf{z}_{3t}\boldsymbol{\gamma}_3)+\Phi_2(-\mu_2+\mathbf{z}_{3t}\boldsymbol{\gamma}_3;\alpha_0^+-\mathbf{x}_t^+\boldsymbol{\beta}^+;-\rho^+);\\
&\Pr(y_t=1|\mathbf{z}_{3t},\mathbf{x}_t^-,\mathbf{x}_t^+)=\Phi(-\mu_2+\mathbf{z}_{3t}\boldsymbol{\gamma}_3)-\Phi_2(-\mu_2+\mathbf{z}_{3t}\boldsymbol{\gamma}_3;\alpha_0^+-\mathbf{x}_t^+\boldsymbol{\beta}^+;-\rho^+)\\
&=\Phi_2(-\mu_2+\mathbf{z}_{3t}\boldsymbol{\gamma}_3;-\alpha_0^++\mathbf{x}_t^+\boldsymbol{\beta}^+;\rho^+).
\end{aligned}
\tag{6}
$$

Suppose the covariates in $\mathbf{x}_t^-$ and $\mathbf{x}_t^+$ in the ZIOP-3 outcome equations are identical to the covariates in $\mathbf{z}_{2t}$ in the ZIOP-2 regime equation, the covariates in $\mathbf{z}_{3t}$ in the ZIOP-3 regime equation are identical to the covariates in the $\mathbf{x}_t$ in the ZIOP-2 outcome equation, and the parameters are restricted as follows: $-\boldsymbol{\beta}^-=\boldsymbol{\beta}^+=\boldsymbol{\gamma}_2$, $\boldsymbol{\beta}=\boldsymbol{\gamma}_3$, $\mu_1=\alpha_{-1}$, $\mu_2=\alpha_0$, $-\alpha_0^-=\alpha_0^+=\mu$ and $-\rho^-=\rho^+=\rho$. Then, since $\mathbf{x}_t^-=\mathbf{x}_t^+=\mathbf{z}_{2t}$, $\mathbf{z}_{3t}=\mathbf{x}_t$ and $\Phi(-x)=1-\Phi(x)$, the probabilities for the ZIOP-3 model in (6) can be written as

$$
\begin{aligned}
&\Pr(y_t=-1|\mathbf{x}_t,\mathbf{z}_{2t})=\Phi_2(\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;-\rho);\\
&\Pr(y_t=0|\mathbf{x}_t,\mathbf{z}_{2t})=\Phi_2(\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};\mu-\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\rho)+\Phi(\alpha_0-\mathbf{x}_t\boldsymbol{\beta})-\Phi(\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta})\\
&+\Phi_2(-\alpha_0+\mathbf{x}_t\boldsymbol{\beta};\mu-\mathbf{z}_{2t}\boldsymbol{\gamma}_2;-\rho)=-\Phi_2(\alpha_{-1}-\mathbf{x}_t\boldsymbol{\beta};-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;-\rho)+1\\
&-\Phi_2(-\alpha_0+\mathbf{x}_t\boldsymbol{\beta};-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\rho);\\
&\Pr(y_t=1|\mathbf{x}_t,\mathbf{z}_{2t})=\Phi_2(-\alpha_0+\mathbf{x}_t\boldsymbol{\beta};-\mu+\mathbf{z}_{2t}\boldsymbol{\gamma}_2;\rho),
\end{aligned}
$$

which are identical to the probabilities for the ZIOP-2 model in (5).

Notice that the restrictions $-\boldsymbol{\beta}^-=\boldsymbol{\beta}^+=\boldsymbol{\gamma}_2$ and $-\alpha_0^-=\alpha_0^+=\mu$ impose a sort of symmetry in the ZIOP-3 model, because they imply that the conditional probability of a positive response is equal to the conditional probability of a negative response:

$$\Pr(y_t = 1|\mathbf{z}_{3t}, \mathbf{x}_t^+, s_t^* = 1) = 1 - \Phi(\alpha_0^+ - \mathbf{x}_t^+\boldsymbol{\beta}^+) =$$
$$= \Phi(-\alpha_0^+ + \mathbf{x}_t^+\boldsymbol{\beta}^+) = \Phi(\alpha_0^- - \mathbf{x}_t^-\boldsymbol{\beta}^-) = \Pr(y_t = -1|\mathbf{z}_t, \mathbf{x}_t^-, s_t^* = -1).$$

In general, if $\mathbf{x}_t^-$ and $\mathbf{x}_t^+$ are not identical to $\mathbf{z}_{2t}$ but contain all covariates in $\mathbf{z}_{2t}$, and if $\mathbf{z}_{3t}$ is not identical to $\mathbf{x}_t$ but contains all covariates in $\mathbf{x}_t$, the ZIOP-2 model is stil nested in the ZIOP-3 model with the additional zero restrictions for the coefficients on all extra covariates in $\mathbf{x}_t^-$, $\mathbf{x}_t^+$ and $\mathbf{z}_{3t}$.

3 The nop, ziop2 and ziop3 commands

3.1 Syntax

`ziop3` *depvar indepvars* [*if*] [*in*] [, `xp(`*varlist*`)` `xn(`*varlist*`)` `infcat(`*integer* 0`)` `endoswitch` `cluster(`*varname*`)` `robust` `initial(`*string*`)`]

This command estimates by ML the three-part cross-nested zero-inflated ordered probit model with possibly different sets of covariates in the regime and outcome equations and possibly endogenous switching among three latent regimes.

`ziop2` *depvar indepvars* [*if*] [*in*] [, `x (`*varlist*`)` `infcat(`*integer* 0`)` `endoswitch` `cluster(`*varname*`)` `robust` `initial(`*string*`)`]

This command estimates by ML the two-part cross-nested zero-inflated ordered probit model with possibly different sets of covariates in the regime and outcome equations and possibly endogenous switching among two latent regimes.

`nop` *depvar indepvars* [*if*] [*in*] [, `xp(`*varlist*`)` `xn(`*varlist*`)` `infcat(`*integer* 0`)` `endoswitch` `cluster(`*varname*`)` `robust` `initial(`*string*`)`]

This command estimates by ML the three-part nested ordered probit model with possibly different sets of covariates in the regime and outcome equations and possibly endogenous switching among three latent regimes..

Options

options	Description
`xp(`*varlist*`)`	list of covariates for positive response in NOP and ZIOP models; by default, it equals *indepvars*, the list of covariates for initial stage
`xn(`*varlist*`)`	list of covariates for negative response in NOP and ZIOP models; by default, it equals *indepvars*, the list of covariates for initial stage
`x(`*varlist*`)`	list of covariates for non-zero response in ZIOP models; by default, it equals *indepvars*, the list of covariates for initial stage
`infcat(`*integer*`)`	value of the response variable that should be modeled as inflated; by default, it equals 0
`endoswitch`	flag that errors in the first and second stages may be correlated, forcing estimation of endogenous switching models
`robust`	flag that variance-covariance estimator must be robust (based on "sandwich") estimate
`cluster(`*varname*`)`	clustering variable for robust variance estimator
`initial(`*string*`)`	whitespace-delimited list of initial parameter values for estimation, in the following order: β, α, γ^+, μ^+, γ^-, μ^-, ρ^-, ρ^+

Stored results

`nop`, `ziop2`, and `ziop3` store the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(ll)`	total log-likelihood of the model
Macros	
`e(cmd)`	`nop`, `ziop2`, or `ziop3`, respectively
`e(depvar)`	dependent variable of regression
Matrices	
`e(b)`	parameters vector
`e(V)`	variance-covariance matrix
`e(ll_obs)`	vector of observation-wise log-likelihood
Functions	
`e(sample)`	marks estimation sample

ß

3.2 Postestimation commands

The predict command

The `predict` command after the nop, ziop2 and ziop3 estimation commands produces either predicted probabilities or expected values of the responses.

`predict` *varname* [*if*] [*in*] [`, zeros regime output(`*string*`) at(`*string*`)`]

`name` is the name of predicted variable, if it is single, or prefix for names, if there are several predicted variables

`zeros` indicates that different types of zeros (i.e. "intrinsic zeros", or "positive zeros", or "negative zeros") must be predicted instead of different response values.

`regime` indicates that different groups of response (negative, positive or zero) must be predicted instead of different response values. This option is ignored if `zeros` option is on.

`output(string)` specifies type of aggregating predicted probabilities of different response. Possible values are: `mode` for reporting the outcome with the highest predicted probability, and `mean`, for predicting the expected outcome computed as $\sum_i Pr(y_t = i) \times i$, and `cum` for predicting cumulative response probabilities (i.e. $\Pr(y_t <= -2)$, $\Pr(y_t <= -1)$, $\Pr(y_t <= 0)$ etc.). If not specified, raw response probabilities are predicted ($\Pr(y_t = -2)$, $\Pr(y_t = -1)$, $\Pr(y_t = 0)$ etc.), and placed into multiple variables.

The ziopmargins command

`ziopmargins [, at(`*string*`) nominal(`*varlist*`) zeros regime]`

This command prints marginal effects for the last estimated model (either NOP, or ZIOP-2, or ZIOP-3), calculated at the specified point, along with confidence intervals.

`at(string)` specifies at which point predictions must be calculated. If at is specified, (as a list of `varname=value` expressions, separated by comma), prediction is calculated at this point and posted on the screen without saving to the dataset. If some covariate names are not specified, their mean value is taken instead.

`nominal` is a space-separated list of covariates which should be considered as nominal; marginal effect is then calculated as difference between values at 0 and at 1.

`zeros` and `regime` indicate that marginal effects should be calculated for different zeros or for groups of response variable, as in `predict` command.

The ziopprobabilities command

`ziopprobabilities [, at(`*string*`) zeros regime]`

This command prints predicted probabilities for the last estimated model (either NOP, or ZIOP-2, or ZIOP-3) , calculated at the specified point, along with confidence intervals. The point `at` is specified like in `ziopmargins`.

The ziopcontrasts command

`ziopcontrasts [, at(`*string*`) to(`*string*`) zeros regime]`

This command prints differences in predicted probabilities for the last estimated model (either NOP, or ZIOP-2, or ZIOP-3), calculated between the specified points, along with confidence intervals. The points `at` and `to` are specified like `at` in `ziopmargins`.

The ziopclassification command

`ziopclassification`

This command prints the classification table (confusion matrix). It displays the predicted (most probable) outcome in rows, actual outcome in columns, and number of (mis)classifications in each cell. The command also prints the percentage of correct predictions.

The zioppscores command

`zioppscores`

This command computes two strictly proper scoring rules: the probability, or Brier, score (Brier 1950) and ranked probability score (Epstein 1969). Brier probability score is computed as $\frac{1}{T}\sum_{t=1}^{T}\sum_{j=J^-}^{J^+}[\Pr(y_t = j) - I_{jt}]^2$, where indicator $I_{jt} = 1$ if $y_t = j$ and $I_{jt} = 0$ otherwise. Ranked probability score is computed as $\frac{1}{T}\sum_{t=1}^{T}\sum_{j=J^-}^{J^+}[Q_{jt} - D_{jt}]^2$, where $Q_{jt} = \sum_{i=J^-}^{j}\Pr(y_t = i)$ and $D_{jt} = \sum_{i=J^-}^{j} I_{it}$. The better the prediction, the smaller both score values. Both scores have a minimum value of zero when all actual outcomes are predicted with a unit probability.

The ziopvuong command

`ziopvuong` *modelspec*$_1$ *modelspec*$_2$

This command performs non-nested Vuong test which compares closeness of two models to the true data distribution. Arguments *modelspec*$_1$ and *modelspec*$_2$ are the names under which estimation results were saved using `estimates store` command. Any model that stores vector `e(ll_obs)` of observation-wise log-likelihood technically can be used to perform the test.

The command calculates and outputs a z-score for the difference of pointwise log likelihoods of the two models. It can be used to test the hypothesis that one of the models explains the data better than the other.

4 Monte Carlo simulations

We conducted extensive Monte Carlo experiments to illustrate the finite sample performance of the ML estimators of each model.

4.1 Monte Carlo design

We simulated six processes generated by the NOP, ZIOP-2 and ZIOP-3 models, each of them with both exogenous and endogenous switching. The repeated samples with 200, 500 and 1000 observations were independently generated and then estimated by the true model. The number of replications was 10,000 in each experiment.

Three covariates $\mathbf{w_1}$, $\mathbf{w_2}$ and $\mathbf{w_3}$ were drawn in each replication as $\mathbf{w_1} \overset{iid}{\sim} \mathcal{N}(0,1)+2$, $\mathbf{w_2} \overset{iid}{\sim} \mathcal{N}(0,1)$, and $\mathbf{w_3} = -1$ if $\mathbf{u} \leq 0.3$, 0 if $0.3 < \mathbf{u} \leq 0.7$, or 1 if $\mathbf{u} > 0.7$, where $\mathbf{u} \overset{iid}{\sim} \mathcal{U}[0,1]$. The repeated samples were generated for the NOP and ZIOP-3 models with $\mathbf{Z} = (\mathbf{w_1}, \mathbf{w_2})$, $\mathbf{X}^- = (\mathbf{w_1}, \mathbf{w_3})$, $\mathbf{X}^+ = (\mathbf{w_2}, \mathbf{w_3})$, and for the ZIOP-2 model with $\mathbf{Z} = (\mathbf{w}_1, \mathbf{w}_3)$, $\mathbf{X} = (\mathbf{w}_2, \mathbf{w}_3)$. The dependent variable y was generated with five values: -2, -1, 0, 1 and 2. The parameters were calibrated to yield on average the following frequencies of the above outcomes: 7%, 14%, 58%, 14% and 7%, respectively. To avoid the divergence of ML estimates due to the problem of complete separation (perfect prediction), which could happen if the actual number of observations in any outcome category is very low, the samples with any outcome category frequency lower than 6% were re-generated. The matrix of the MEs has $3 \times 5 = 15$ elements; their values, which depend on the values of the regressors, are computed at the population medians of the covariates. The variances of the errors in all equations were fixed to one. The true values of all other parameters in the simulations are shown in Table A1 in Appendix. The starting values for slope and threshold parameters were obtained using the independent ordered probit estimations of each equation. The starting values for ρ, ρ^- and ρ^+ were obtained by maximizing the likelihood functions of the endogenous-switching models holding the other parameters fixed at their estimates in the corresponding exogenous-switching model.

4.2 Monte Carlo results

Table 1 reports the measures of accuracy for the estimates of the probabilities and MEs. For each model, the bias and RMSE decrease as sample size increases. RMSE decreases in most cases faster than asymptotic rate $\sqrt{n}$. This may be caused by a small number of large deviations in parameter estimation in small samples. For most of models and sample sizes, the bias and RMSE are slightly higher for the endogenous-switching version. This is expected from a more complex model estimated with the same sample size.

Standard error estimates for parameters on average correspond to the actual standard errors. Large deviations make standard errors estimates biased, especially in small samples, but this problem rapidly decreases as sample size grows. Anyway, rare large deviations do not prevent asymptotic coverage rates to be close to the nominal values even with only 200 obsevations. The simulations show that the ML estimators are consistent and reliable even in samples with only 200 observations: the biases of probability estimates are smaller than five percent and coverage rates differ from the nominal level by less than one percent. The ME estimates are less precise and approach the similar accuracy with about 1000 observations: however, with 500 observations the ME biases are smaller than ten percent and the ME coverage rates differ from the nominal level by less than four percent only. The accuracy in the NOP models is expectedly higher than in the cross-nested zero-inflated models.

Table 1. Monte Carlo results: The accuracy of ML estimators

Notes: Bias – the absolute difference between the estimated and true values, devided by the true value; RMSE – the absolute root mean square error of the estimates; Coverage rate – the percentage of times the estimated asymptotic 95% confidence intervals cover the true values; Bias of standard error estimates – the absolute difference between the average of the estimated asymptotic standard errors of the estimates and the standard deviation of the estimates in all replications. The above measures are averaged across five outcome categories, and for the estimates of the MEs are also averaged across all three covariates.

5 Application

6 Concluding remarks

This article describes the ML estimation of the nested and cross-nested zero-inflated ordered probit models using the new STATA commands `nop`, `ziop-2` and `ziop-3` . Such models can be applied to a variety of data sets in which the discrete ordinal outcomes can be divided into the groups (nests) of similar choices, for example, the decisions to reduce, or leave unchanged, or increase the choice variable (monetary policy interest rates, rankings, prices, consumption levels), or the negative, or neutral, or positive attitudes to the survey questions. The choice among the nests is driven by an ordered-choice switching mechanism that can be either exogenous or endogenous to the outcome decisions, which are also naturally ordered (large or small increase/decrease; disagree or strongly disagree; etc). The models allow the probabilities of choices from different nests (e.g., no change and an increase) to be driven by distinct mechanisms. Moreover, the zero-inflated cross-nested models allow the often abundant no-change or neutral outcomes to belong to all nests and be inflated by several different processes.

The results of Monte Carlo simulations indicate that the proposed ML estimators are consistent and perform well in small samples.

7 Acknowledgments

We gratefully acknowledge support from the Basic Research Program of the National Research University Higher School of Economics in Moscow.

References

Andrews, D. W. K. 2001. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69 (3): 683–734.

Bagozzi, B. E., and B. Mukherjee. 2012. A mixture model for middle category inflation in ordered survey responses. *Political Analysis* 20: 369–386.

Basu, D., and R. M. de Jong. 2007. Dynamic multinomial ordered choice with an application to the estimation of monetary policy rules. *Studies in Nonlinear Dynamics and Econometrics* 11 (4): 1–35.

Brier, G. W. 1950. Verification of forecasts expressed in terms of probability. *Monthly Weather Review* 78 (1): 1–3.

Brooks, R., M. N. Harris, and C. Spencer. 2012. Inflated ordered outcomes. *Economics Letters* 117 (3): 683–686.

Epstein, E. S. 1969. A scoring system for probability forecasts of ranked categories. *Journal of Applied Meteorology* 8: 985–987.

Famoye, F., and K. P. Singh. 2003. On inflated generalized Poisson regression models. *Advanced Applied Statistics* 3 (2): 145–158.

Greene, W. H. 1994. Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working Paper No. 94-10, Department of Economics, Stern School of Business, New York University.

Greene, W. H., and D. A. Hensher. 2010. *Modeling ordered choices: A primer*. Cambridge University Press.

Hardin, J. W., and J. M. Hilbe. 2014. Estimation and testing of binomial and beta-binomial regression models with and without zero inflation. *Stata Journal* 14(2): 292-303.

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.

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46: 931–959.

Hernández, A., F. Drasgow and V. Gonzáles-Romá. 2004. Investigating the functioning of a middle category by means of a mixed-measurement model. *Journal of Applied Psychology* 89 (4): 687-699.

Kelley, M. E., and S. J. Anderson. 2008. Zero inflation in ordinal data: incorporating susceptibility to response through the use of a mixture model. *Statistics in Medicine* 27: 3674–3688.

Kulas, J. T. and A. A. Stachowski. 2009. Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality* 43: 489-493.

Lambert, D. 1992. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* 34 (1): 1–14.

MacKinnon, J. G. 1996. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics* 11: 601–618.

Sirchenko, A. 2013. A model for ordinal responses with an application to policy interest rate. National Bank of Poland Working Paper No. 148.

Small, K. 1987. A discrete choice model for ordered alternatives. *Econometrica* 55: 409–424.

Vovsha, P. 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record* 1607: 6–15.

Vuong, Q. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57 (2): 307–333.

Wen, C.-H., and F. Koppelman. 2001. The generalized nested logit model. *Transportation Research B* 35: 627–641.

Wilde, J. 2000. Identification of multiple equation probit models with endogenous dummy regressors. *Economics Letters* 69 (3): 309–312.

Winkelmann, R. 2008. *Econometric analysis of count data.* 5th edition, Springer.

Appendix

Table A1. Monte Carlo simulations: The true values of parameters

	NOP (exog)	NOP	ZIOP-2 (exog)	ZIOP-2	ZIOP-3 (exog)	ZIOP-3
γ	(0.6, 0.4)'	(0.6, 0.4)'	(0.6, 0.8)'	(0.6, 0.8)'	(0.6, 0.4)'	(0.6, 0.4)'
μ	(0.21, 2.19)'	(0.21, 2.19)'	0.45	0.45	(0.9, 1.5)'	(0.9, 1.5)'
β			(0.5, 0.6)'	(0.5, 0.6)'		
β^-	(0.3, 0.9)'	(0.3, 0.9)'			(0.3, 0.9)'	(0.3, 0.9)'
β^+	(0.2, 0.3)'	(0.2, 0.3)'			(0.2, 0.3)'	(0.2, 0.3)'
α			(-1.45, -0.55, 0.75, 1.65)'	(-1.18,-0.33, 0.9, 1.76)'		
α^-	-0.17	-0.5			(-0.67, 0.36)'	(-0.88, 0.12)'
α^+	0.68	1.3			(0.02, 1.28)'	(0.49, 1.67)'
ρ			0	0.5		
ρ^-	0	0.3			0	0.3
ρ^+	0	0.6			0	0.6

Notes: (exog) – exogenous switching; the variances σ^2, σ^2_-, σ^2_+, and σ^2_ν are all fixed to one.