Testing for state dependence in the fixed-effects ordered logit model

Highlights

Testing for state dependence in the fixed-effects ordered logit model

Francesco Bartolucci, Claudia Pigini, Francesco Valentini

- Persistence in self-reported health/well-being entails lasting effects of policies
- A test for state dependence in the fixed-effects ordered logit model is proposed
- The test is able to detect persistence either in the latent or observed variable
- It can be readily computed using existing software

Testing for state dependence in the fixed-effects ordered logit model

Francesco Bartolucci[a], Claudia Pigini[b,*], Francesco Valentini[b]

[a]*Department of Economics, University of Perugia, Via A. Pascoli 20, Perugia, 06123, Italy*
[b]*Department of Economic and Social Sciences, Marche Polytechnic University, Piazzale R. Martelli 8, Ancona, 60121, Italy*

Abstract

We propose a test for state dependence in the fixed-effects ordered logit, based on the Quadratic Exponential model. The test can be applied to models where persistence lies either in the latent or observed response variable.

Keywords: Conditional Maximum Likelihood, Fixed effects, Ordered panel data, Quadratic Exponential model, State dependence
JEL: C12, C23, C25

1. Introduction and motivation

Ordered choice models are widely employed to analyze survey data on self-rated measures, such as those describing physical and mental health, subjective well-being, and life satisfaction. These surveys often comprise a panel component, which makes it possible to specify dynamic models depicting persistence in the response variable. Being able to ascertain the presence of state dependence, which arises very naturally in these measures, is key to understand whether short-term policy interventions will have long-run effects on health or individual well-being (Contoyannis et al., 2004). However, there is disagreement in the literature on whether persistence lies in the unobserved perception or in the ordinal-scale self-reported measure.

In modeling subjective well-being, Pudney (2008) introduces the Latent

*Corresponding Author. E-mail address: c.pigini@univpm.it

Autoregressive model (LAR henceforth), according to which the latent perception y_{it}^* for individual i at time t is generated by a process of the type

$$y_{it}^* = \rho y_{i,t-1}^* + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \tag{1}$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where ρ is the autoregressive coefficient, $\boldsymbol{\beta}$ the vector of k regression parameters related to the $\boldsymbol{x}_{it}$ vector of covariates, α_i is the individual unobserved heterogeneity, and ε_{it} an idiosyncratic error term. Self-reported well-being y_{it} is then observed according to a standard set of threshold-crossing rules:

$$y_{it} = j \quad \text{if} \quad c_{j-1} < y_{it}^* \leq c_j, \ j = 1, \ldots, J,$$

with $c_0 = -\infty$ and $c_J = +\infty$. In the same vein, Heiss (2011) considers an ordered choice model where intertemporal dependence arises in a time varying unobserved heterogeneity component.

A different stream of literature focuses on dynamic models where state dependence is associated to dichotomizations of the lagged dependent variable. Muris et al. (2020) dichotomize $y_{i,t-1}$ at a fixed and known value τ, such that

$$y_{it}^* = \gamma \mathbb{1}(y_{i,t-1} \geq \tau) + \boldsymbol{x}_{it}'\boldsymbol{\theta} + \eta_i + \upsilon_{it}, \tag{2}$$

with $2 \leq \tau \leq J$, where $\boldsymbol{\theta}$ is now the vector of regression parameters and υ_{it} is the idiosyncratic error term. The above model is characterized by a single state dependence (SSD henceforth) parameter γ, and it can be seen as a special case of the mainstream dynamic ordered choice models (see Contoyannis et al., 2004; Carro and Traferri, 2014; Honoré et al., 2021), where there are multiple state dependence (MSD) parameters λ_j that are heterogeneous across thresholds. The latter formulation assumes that

$$y_{it}^* = \sum_{j=1}^{J} \lambda_j \mathbb{1}(y_{i,t-1} = j) + \boldsymbol{x}_{it}'\boldsymbol{\pi} + \omega_i + u_{it}, \tag{3}$$

where $\boldsymbol{\pi}$ are the regression parameters and u_{it} is the idiosyncratic error term.

We propose a test for state dependence in fixed-effects ordered logit models. The test is based on a modified Quadratic Exponential (QE) formulation (Bartolucci and Nigro, 2010; Bartolucci et al., 2018) for each possible dichotomization of the response variable. Under the null hypothesis of absence of state dependence, the QE probabilities for each dichotomization correspond[1] to those of the LAR, SSD, and MSD, when ρ, γ, and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$ are equal to zero in (1), (2), and (3), respectively, and provided that the error terms in the three equations are standard logistically distributed.

The QE model accommodates state dependence and admits sufficient statistics for the individual intercepts, which can therefore be treated as fixed effects and allowed to be correlated with the model covariates. As a consequence, the Conditional Maximum Likelihood (CML) estimator of the QE model parameters is fixed-T consistent and can be seen as a generalization of the popular Blow-Up and Cluster (BUC) procedure for the estimation of the fixed-effects static ordered logit model by Baetschmann et al. (2015).[2]

Finally, the test can be readily computed by the existing software package `cquad` for `R`[3] (Bartolucci and Pigini, 2017) and `CQUADR` for `Stata` (Bartolucci et al., 2020).

2. Proposed test

The test is based on the QE formulation as a dynamic binary choice model (Bartolucci and Nigro, 2010), with the modification introduced by Bartolucci et al. (2018) in testing for state dependence in the fixed-effects binary logit. Let $d_{it}^j = \mathbb{1}(y_{it} > j)$, $j = 1, \ldots, J-1$, be the binary dependent variable arising from the dichotomization of y_{it} at the cutoff j and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT})$.

The modified QE model is defined as the joint probability of the response configuration $\boldsymbol{d}_i^j = (d_{i1}^j, \ldots, d_{iT}^j)'$

$$p(\boldsymbol{d}_i^j | \alpha_i, \boldsymbol{X}_i, d_{i0}^j) = \frac{\exp\left(d_{i+}^j \delta_i + \sum_{t=1}^T d_{it}^j \boldsymbol{x}_{it}' \boldsymbol{\phi} + d_{i*}^j \psi\right)}{\sum_{\boldsymbol{z}} \exp\left(z_+ \delta_i + \sum_{t=1}^T z_t \boldsymbol{x}_{it}' \boldsymbol{\phi} + z_{i*} \psi\right)}, \tag{4}$$

where $d_{i+}^j = \sum_{i=1}^T d_{it}^j$ is the *total score*, δ_i is the individual unobserved heterogeneity, $\boldsymbol{\phi}$ is the vector of regression parameters, $d_{i*}^j = \sum_{t=1}^T \mathbb{1}(d_{it}^j = d_{i,t-1}^j)$, and ψ is the state dependence parameter; in the denominator, $\sum_{\boldsymbol{z}}$ ranges

[1] The correspondence holds under the assumption of regression parameters homogeneity across the different dichotomizations.

[2] An extension of the BUC using time-varying dichotomizations is proposed by Muris (2017).

[3] `https://CRAN.R-project.org/package=cquad`

over all binary response vectors $\boldsymbol{z} = (z_1, \ldots, z_T)'$, $z_+ = d^j_{i+}$, and $z_{i*} = \mathbb{1}(z_1 = d^*_{i0}) + \sum_{t=2}^T \mathbb{1}(z_t = z_{t-1})$. Bartolucci et al. (2018) show that the t-test for $H_0 : \psi = 0$ for the dynamic binary logit model has superior power properties compared to that based on the original QE model, in which d^j_{i*} would have to be replaced by $d^j_{it} d^j_{i,t-1}$.

Under $H_0 : \psi = 0$, the probability in (4) is the same as the joint probability that arises for the j-th dichotomization from the LAR model in (1), with $\rho = 0$, $\boldsymbol{\beta} = \boldsymbol{\phi}$, and $\alpha_i = \delta_i$, from the SSD model in (2), with $\gamma = 0$, $\boldsymbol{\theta} = \boldsymbol{\phi}$, and $\eta_i = \delta_i$, and from the MSD model in (3), with $\boldsymbol{\lambda} = \boldsymbol{0}$, $\boldsymbol{\pi} = \boldsymbol{\phi}$, and $\omega_i = \delta_i$, provided that ε_{it}, υ_{it}, and v_{it} are standard logistic random variables.

As in the static logit model, in the QE model the total scores are sufficient statistics for the individual-specific intercepts, so that conditioning the probability in (4) on d^j_{i+} eliminates δ_i:

$$p(\boldsymbol{d}^j_i | \boldsymbol{X}_i, d^j_{i+}, d^j_{i0}) = \frac{\exp\left(\sum_{t=1}^T d^j_{it} \boldsymbol{x}'_{it} \boldsymbol{\phi} + d^j_{i*} \psi\right)}{\sum_{\boldsymbol{z}: z_+ = d^j_{i+}} \exp\left(\sum_{t=1}^T z_t \boldsymbol{x}'_{it} \boldsymbol{\phi} + z_{i*} \psi\right)}.$$

With a static formulation, a single dichotomization to estimate the model parameters could be used, while Baetschmann et al. (2015) devised the BUC procedure, where the information brought by each dichotomization can be combined in a single quasi-likelihood function. Following the latter, the quasi-loglikelihood function can be written as

$$\ell(\boldsymbol{\phi}, \psi) = \sum_{j=1}^{J-1} \sum_{i=1}^{n} \mathbb{1}(1 < d^j_{i+} < T) \log p(\boldsymbol{d}^j_i | \boldsymbol{X}_i, d^j_{i+}, d^j_{i0})$$

and can be maximized by a standard Newton-Raphson algorithm, so as to obtain the CML estimator $(\hat{\boldsymbol{\phi}}', \hat{\psi})'$. It is then straightforward to compute the t-statistic for testing $H_0 : \psi = 0$ as

$$W = \frac{\hat{\psi}}{\mathrm{se}(\hat{\psi})},$$

where $\mathrm{se}(\cdot)$ is the standard error derived form a cluster–robust variance estimator allowing for correlation within subjects, as they contribute more than once to the log-likelihood (Baetschmann et al., 2015).

3. Simulation study

The design mirrors that in Honoré et al. (2021). Based on (1), for $i = 1, \dots, n$ and $t = 1, \dots, T$, it is assumed that

$$y_{it}^* = \rho y_{i,t-1}^* + \sum_{l=1}^{3} x_{it}^l \beta_l + \alpha_i + \varepsilon_{it},$$

with $y_{i0}^* = \rho u_i + \sum_{l=1}^3 x_{i0}^l \beta_l + \alpha_i + \varepsilon_{i0}$, where u_i and $\varepsilon_{i0}, \dots, \varepsilon_{iT}$ are standard logistic random variables; covariates are $x_{it}^1 = \sqrt{3}(Z_{it}^1 + Z_i)/\sqrt{2}$ and $x_{it}^s = \sqrt{3}(Z_{it}^s + Z_{it}^1)/\sqrt{2}$ with $s = 2, 3$, where Z_i and Z_{it}^l, with $l = 1, 2, 3$, are standard normals, and $\alpha_i = \sqrt{3} Z_i$. The outcome takes $J = 4$ possible values with thresholds $\{-\infty, -2, 0, 2, \infty\}$. The parameters $\beta_1, \beta_2, \beta_3$ are set to $1, -0.5, 0$, respectively, while ρ ranges between -0.9 and 0.9 with step of 0.1. For the SSD model in (2), we have

$$y_{it}^* = \gamma \mathbb{1}(y_{i,t-1} \geq 3) + \sum_{l=1}^{3} x_{it}^l \beta_l + \alpha_i + \varepsilon_{it},$$

with $y_{i0}^* = \gamma \mathbb{1}(u_i \geq 3) + \sum_{l=1}^3 x_{i0}^l \beta_l + \alpha_i + \varepsilon_{i0}$, where here γ takes values between -1.6 and 1.6 with step 0.2. For the MSD model in (3), we use the same design and let $\boldsymbol{\lambda}$ take values in $(0, 0, 0, 0)$, $(-1, 0, 0, 1)$, and $(-1, -0.5, 0.5, 1)$.

We considered scenarios with $n = 500, 1000$ and $T = 3, 7$, for 1000 Monte Carlo replications. Simulation results are reported in Figure 1 and Table 1 for a 5% nominal size. For the LAR, the nominal size is attained in every scenario considered, while the value of the empirical power, for values of ρ close to zero, increases in T. The power curves depicted for the SSD are wider, although rejection rates increase sharply in both the sample size and number of time occasions. This is a result of the proposed test using all the possible dichotomizations of the response variable, whereas in the SSD model state dependence is associated only with a specific one for the lagged response variable. The proposed procedure, however, is robust to such dichotomization happening at any threshold and that power properties are nonetheless satisfactory in scenarios that are likely to occur in practice. Finally, the test attains the nominal size when data are generated according to a MSD model and achieves high power under departures from the static model.

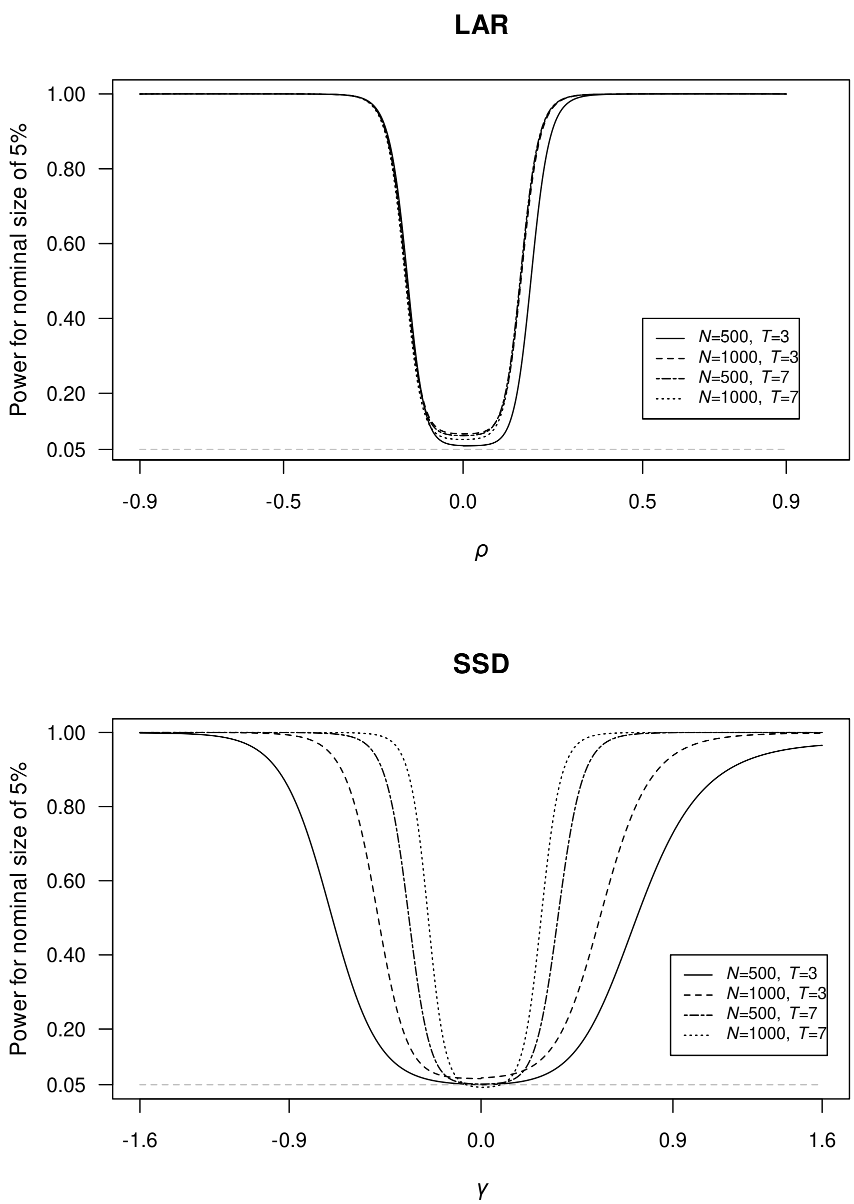

Figure 1: LAR and SSD models: power curves

$\boldsymbol{\lambda}$	$n = 500$		$n = 1000$	
	$T = 3$	$T = 7$	$T = 3$	$T = 7$
$(0, 0, 0, 0)$	0.050	0.048	0.046	0.046
$(-1, 0, 0, 1)$	0.972	1.000	1.000	1.000
$(-1, -0.5, 0.5, 1)$	0.986	1.000	1.000	1.000

Table 1: MSD model: power for nominal size of 5%

4. Conclusions

Persistence in ordered measures such as self-assessed health or subjective well-being is an important element in understanding whether interventions may have lasting effects. Yet research is still currently in pursuit of a fixed-T consistent estimator for the parameters of a dynamic fixed-effects ordered choice model (Honoré et al., 2021). In addition, different model formulations have been proposed, according to whether persistence lies in the latent or observed variable.

We partially fill this gap by proposing a test for state dependence, which is robust against alternative specifications of the dynamic ordered logit. We also show that the test finite-sample performance is satisfactory in realistic scenarios. Finally, the test could be particularly appealing to practitioners as it can be readily computed using existing software.

Acknowledgments

We would like to thank the Editor Graham Elliott and an anonymous Referee for their comments. We are grateful to Chris Muris for his helpful remarks and to the audience of the 2022 meeting of the Challenges for Categorical Data Analysis group.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Baetschmann, G., Staub, K.E., Winkelmann, R., 2015. Consistent estimation of the fixed effects ordered logit model. Journal of the Royal Statistical Society: Series A 178, 685–703.

Bartolucci, F., Nigro, V., 2010. A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$-consistent conditional estimator. Econometrica 78, 719–733.

Bartolucci, F., Nigro, V., Pigini, C., 2018. Testing for state dependence in binary panel data with individual covariates by a modified quadratic exponential model. Econometric Reviews 37, 61–88.

Bartolucci, F., Pigini, C., 2017. cquad: An R and Stata package for conditional maximum likelihood estimation of dynamic binary panel data models. Journal of Statistical Software 78.

Bartolucci, F., Pigini, C., Valentini, F., 2020. CQUADR: Stata module to estimate Quadratic Exponential models running the cquad R package. Statistical Software Components, Boston College Department of Economics. URL: `https://ideas.repec.org/c/boc/bocode/s458852.html`.

Carro, J.M., Traferri, A., 2014. State dependence and heterogeneity in health using a bias-corrected fixed-effects estimator. Journal of Applied Econometrics 29, 181–207.

Contoyannis, P., Jones, A.M., Rice, N., 2004. The dynamics of health in the british household panel survey. Journal of Applied Econometrics 19, 473–503.

Heiss, F., 2011. Dynamics of self-rated health and selective mortality. Empirical economics 40, 119–140.

Honoré, B.E., Muris, C., Weidner, M., 2021. Dynamic ordered panel logit models. arXiv preprint arXiv:2107.03253 .

Muris, C., 2017. Estimation in the fixed-effects ordered logit model. Review of Economics and Statistics 99, 465–477.

Muris, C., Raposo, P., Vandoros, S., 2020. A dynamic ordered logit model with fixed effects. arXiv preprint arXiv:2008.05517 .

Pudney, S., 2008. The dynamics of perception: modelling subjective well-being in a short panel. Journal of the Royal Statistical Society: Series A 171, 21–40.