Markov-switching autoregressive latent variable models for longitudinal data

Silvia Bacci University of Perugia (Italy)

Francesco Bartolucci University of Perugia (Italy)

Fulvia Pennoni University of Milano Bicocca (Italy)

silvia.bacci@stat.unipg.it

Summary

- Introduction
- Different approaches for the treatment of longitudinal data
- Proposed model: the Markov-switching LAR (SW-LAR) model
- SW-LAR model: special cases
- SW-LAR model: estimation
- Application
- Future developments
- References

Introduction

- Context:

  analysis of longitudinal data (we refer to the case of ordinal response variables y_{it} depending on covariates $\mathbf{x}_{it}$)

- Problem:

  taking into account the effect that unobservable factors have on the occasion-specific response variables

- Different approaches:

  1. Individual-specific random intercept model
  2. Latent autoregressive (LAR) model (Chi and Reinsel, 1989)
  3. Latent Markov (LM) regression model (Wiggins, 1973)

- Aim:

  We propose a generalization of the LAR model based on assuming a latent Markov-switching AR(1) process with correlation coefficient depending on the regime of the chain

Individual-specific random intercept model

The unobserved heterogeneity is taken into account through individual-specific random intercepts

Ordinal response variable for subject *i* at occasion *t* with *j = 1, ..., I* categories

Covariates

$$\log \frac{p(y_{it} > j \mid u_{i0}, \mathbf{x}_{it})}{p(y_{it} \leq j \mid u_{i0}, \mathbf{x}_{it})} = \mu_j + u_{i0} + \mathbf{x}_{it}'\boldsymbol{\beta}$$

Random part of the intercept

$$u_{i0} \sim N(0, \sigma^2) \quad \forall t$$

↑ Parsimony

↓ The effect of unobservable factors is assumed to be time constant

LAR model

The unobserved heterogeneity is taken into account through the inclusion, within subjects, of occasion-specific random effects which follow an AR(1) process

$y_{i1},\ldots, y_{iT}$ are conditionally independent given the latent variables u_{it} and the covariates $\mathbf{x}_{it}$

$$\log \frac{p(y_{it} > j \mid u_{it}, \mathbf{x}_{it})}{p(y_{it} \le j \mid u_{it}, \mathbf{x}_{it})} = \mu_j + u_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}$$

Occasion-specific continuous latent variable

$$u_{i1} \sim N(0, \sigma^2) \quad \text{for } t = 1$$

$$u_{it} \mid u_{i,t-1} = \rho u_{i,t-1} + \varepsilon_{it} \quad \text{for } t > 1$$

$$\varepsilon_{it} \sim N\left(0, \sigma^2 / \sqrt{1-\rho^2}\right)$$

LAR model

- ↑ Parsimony
- ↑ The effect of unobservable factors is time varying
- ↑ In many applications, error terms are naturally represented by continuous random variables

- ↓ Estimation may be problematic from the computational point of view (Heiss, 2008)

LM regression model

The unobserved heterogeneity is taken into account through the inclusion of a sequence of discrete latent variables which follow a first-order Markov chain

$y_{i1},\dots, y_{iT}$ are conditionally independent given the latent variables u_{it} and the covariates $\mathbf{x}_{it}$

$$\log \frac{p(y_{it} > j \mid u_{it}, \mathbf{x}_{it})}{p(y_{it} \le j \mid u_{it}, \mathbf{x}_{it})} = \mu_j + u_{it} + \mathbf{x}_{it}'\boldsymbol{\beta}$$

Occasion-specific discrete latent variable

⇓

1. Any latent variable u_{it} is conditionally independent of $u_{i1}, \dots, u_{i,t-2}$ given $u_{i,t-1}$
2. The latent variable u_{it} can assume k different regimes (or states)

LM regression model

- ↑ It may reach a better fit than the LAR model
- ↑ It is easier to estimate than the LAR model
- ↑ It provides a classification of subjects in a reduced number of groups
- ↑ It may be seen as a semi-parametric version of the LAR model

- ↓ It is less parsimonious than the LAR model: the LM model is based on k-1 initial probabilities and $k(k$-1$)$ transition probabilities, whereas the LAR model is based on only 2 parameters for the latent process.

We formulate a model for longitudinal data based on the assumption that the error terms follow a

Markov-switching AR(1) process (Hamilton, 1989)

- Main characteristics:
  1. The latent process is continous as in the LAR model, but the correlation coefficient is not restricted to be constant.
  2. A set of different regimes are possible, with each regime corresponding to a different value of the correlation coefficient
  3. How a subject moves between regimes is governed by a time-homogenous latent Markov chain

We expect that the resulting model has a fit comparable to that of a LM model, but it is more parsimonious

Proposed model:

Markov-Switching LAR model (SW-LAR)

Assumptions of LAR model are substituted by:

$$u_{it} \,\middle|\, u_{i,t-1}, v_{it} = \rho_{v_{it}} u_{i,t-1} + \varepsilon_{it} \quad \text{for } t > 1$$

$$\varepsilon_{it} \,\middle|\, v_{it} \sim N\left(0, \sigma^2 \Big/ \sqrt{1-\rho_{v_{it}}^2}\right)$$

Note that every latent variable u_{it} has marginal distribution $N(0,\sigma^2)$ as in the LAR model.

SW-LAR model

$v_{i1},\ldots,v_{iT}$ $\Rightarrow$ follow a Markov chain with k latent states:

$\Downarrow$

1. Each latent state corresponds to a correlation coefficient:

$$\rho_1,\ldots,\rho_k$$

2. The latent states are characterized by a vector of initial probabilities:

$$\boldsymbol{\lambda} = \{\lambda_v\},\;\; v = 1,\ldots,k$$

   and by a transition probability matrix:

$$\boldsymbol{\Pi} = \{\pi_{v_0 v}\},\;\; v_0, v = 1,\ldots,k$$

SW-LAR model: special cases

$$k = 1$$ $\Rightarrow$ Basic LAR model:

The correlation coefficient is the same for all subjects and occasions

$$\mathbf{\Pi} = \mathbf{I}$$ $\Rightarrow$ SW-LAR$_1$ model:

The correlation coefficient may be different between subjects belonging to different latent states, but not between occasions

$$\mathbf{\Pi} = \mathbf{1} \otimes \boldsymbol{\lambda}'$$ $\Rightarrow$ SW-LAR$_2$ model:

The correlation coefficient may change between subjects and occasions, since each subject randomly moves betwen different regimes

SW-LAR model: estimation

We maximize the log-likelihood:

$$\ell(\boldsymbol{\theta}) = \sum_i \log p(\mathbf{y}_i | \mathbf{X}_i)$$

Manifest probability

It is based on a *T*-dimensional integral

$\Downarrow$

Sequential numerical integration method

(Heiss, 2008 which is strictly related to Baum et al., 1970)

SW-LAR model: sequential numerical integration

$$\boxed{p(\mathbf{y}_i|\mathbf{X}_i) = \sum_v \int_{\Re} q_{iT}(u,v)du}$$

$$\boxed{q_{it}(u,v) = p(u_{it} = u, v_{it} = v, y_{i1},\ldots, y_{iT})}$$

We compute first

$$q_{i1}(u,v) = p(y_{i1}|u)\rho_v g(u)$$

and then for $t > 1$

$$q_{it}(u,v) = p(y_{it}|u)\sum_{v_0} \pi_{v_0 v} \int_{\Re} q_{i,t-1}(u_0,v_0)g(u|u_0,v)du_0$$

Each integral above is computed by a Gaussian Quadrature

Application

- Data from the Health and Retirement Study (University of Michigan)
- A set of 1000 American people who self-evaluated their health status over 8 occasions
- Health status is an ordinal qualitative response variable: poor, fair, good, very good, excellent
- Time-constant covariates: gender, race, education
- Time- varying covariate: age
- We consider three models: LAR, SW-LAR_1 with 2 latent states, SW-LAR_2 with 2 latent states
- Model selection criterion: BIC

Results: parameter estimates, maximum log-likelihood and BIC

	LAR	SW-LAR$_1$	SW-LAR$_2$
μ_1	7.327	9.152	7.645
μ_2	4.195	5.275	4.301
μ_3	1.023	1.248	0.908
μ_4	-2.376	-3.028	-2.692
female	-0.057	0.044	-0.059
non white	-1.852	-2.207	-1.876
education	1.588	1.940	1.675
age	-0.101	-0.121	-0.093
σ	2.916	3.997	3.241
ρ_1	0.955	0.489	0.441
ρ_2	--	0.976	1
λ_1	1	0.241	0.127
λ_2	--	0.759	0.873
log-likelihood	-8884.7	-8795.6	-8818.2
# parameters	10	12	12
BIC	17838	17674	17719

We have two different levels of persistence of the effect of the unobservable factors on the response variables

SW-LAR$_1$ model has only two more parameters than LAR model

SW-LAR$_1$ model has a better fit than LAR model

What's next?

- Simulation study to detect the differences among LAR, LM and SW-LAR models
- Implementation of a sequential numerical integration algorithm to estimate a general SW-LAR model and to obtain standard errors for the parameter estimates

References

- Baum L.E., Petrie, T., Solues, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics, 41, 164-171.
- Chi, E.M., and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. Journal of the American Statistical Association, 84, 452-459.
- Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Eocnometrica, 57, 357-384.
- Heiss, F. (2008). Sequential numerical integration in nonlinear state space models for microeconometric panel data. Journal of Applied Econometrics, 23, 373-389.
- Wiggins, L.M. (1973). Panel analysis: latent probability models for attitude and behavior processes. Amsterdam: Elsevier.