A mover-stayer model with covariates for installment loans repayment process

HALINA FRYDMAN* AND ANNA MATUSZYK**

*STERN SCHOOL OF BUSINESS, NEW YORK UNIVERSITY

**WARSAW SCHOOL OF ECONOMICS

Credit Scoring and Credit Control XIV
Edinburgh, August 26-28, 2015

Agenda

- Mover-stayer (M-S) model
- Data and EM algorithm description
- Algorithm application
- Patterns’ classification consistency assessment
- Summary

Mover-stayer (M-S) model

- extension of a discrete time Markov chain
- population heterogeneity
- two types of individuals in a population:

- stayers - never leave their initial states

- movers - move among the states according to a discrete time Markov chain.

Literature review

- Blumen et al (1955) – introduced the M-S to study industrial labour mobility in US
- Frydman (1984) obtained the maximum likelihood estimators of the M-S model's parameters by direct maximization of the likelihood function
- Fuchs and Greenhouse (1988) – provided expectation-maximization (EM) algorithm

Data description

- Loans granted by a Polish bank to 605 customers
- Loans period: April 2003-September 2011
- Observation period: 24 months
- Repayment history of each customer observed at 3-month intervals for 24 months
- Every customer is characterized by K time fixed covariates
- Repayment status recorded as:

state 1: customer repays loan instalment on time

state 2: customer repays with delay

state 3: customer defaults on a loan.

- All 605 loans initially in state 1, of those:

560 never left state 1,

Of 45 that left state 1, 14 defaulted

Probability that an r'th individual is a stayer in state i

$$\pi_{ri} = \frac{exp(\alpha_i x_r)}{1+exp(\alpha_i x_r)}$$

where:

- x_r is a vector of K covariates for the r'th individual,
- α_i is the state dependent vector of coefficients.

The EM algorithm

Consists of two steps: E-step and M-step

E-step: At the p'th iteration, p ≥ 1 computes:

$n_{si}^{(p)}$ – expected number of stayers in state i and

$e_{ki}^{(p)}$ – expected value of the sum of the k'th covariate for stayers

T – # of times a customer is observed

$$n_{si}^{(p)} = \sum_{r \in S_i^{(p)}} \left[\frac{\pi_{ri}^{(p)}}{\pi_{ri}^{(p)} + (1 - \pi_{ri}^{(p)}) m_{ii}^{(p)T}} \right]$$

$$e_{ki}^{(p)} = \sum_{r \in S_i^{(p)}} \left[\frac{x_{kr}\pi_{ri}^{(p)}}{\pi_{ri}^{(p)} + (1 - \pi_{ri}^{(p)}) m_{ii}^{(p)T}} \right] (1 \le k \le K)$$

The EM algorithm – cont.

M-step: solves the following equations for $\alpha_i^{(p+1)}$:

$$n_{si}^{(p)} = \sum_{r=1}^{n} \pi_{ri}^{(p+1)} \quad (1 \le i \le w)$$

$$e_{ki}^{(p)} = \sum_{r=1}^{n} x_{kr} \pi_{ri}^{(p+1)} \quad (1 \le i \le w)(1 \le k \le K)$$

And computes $m_{ij}^{(p+1)}$, $1 \le j \le w$:

$$m_{ij}^{(p+1)} = \frac{n_{ii} - J n_{si}^{(p)}}{n_i^* - J n_{si}^{(p)}}, \; m_{ij}^{(p+1)} = \frac{n_{ij}}{n_i^* - T n_{si}^{(p)}} \quad (j \ne i)$$

The algorithm is implemented in SAS.

Steps E and M repeat until a convergence criterion is satisfied.

Algorithm application

- fitting the Markov chain model
- fitting the M-S model without covariates

$$\hat{P} = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \begin{pmatrix} 0.9901 & 0.0084 & 0.0015 \\ 0.07 & 0.79 & 0.14 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\hat{M} = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \begin{pmatrix} 0.8701 & 0.1102 & 0.0197 \\ 0.07 & 0.79 & 0.14 \\ 0 & 0 & 1 \end{pmatrix}$$

$\hat{\pi}$=0.8892 (proportion of stayers in state 1)

$\hat{n}$=538 (#of stayers in state 1)

- fitting the M-S model with covariates

$$\hat{M}_c = \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \begin{pmatrix} 0.874 & 0.107 & 0.019 \\ 0.07 & 0.79 & 0.14 \\ 0 & 0 & 1 \end{pmatrix}$$

$\hat{\pi}$=0.8869 (proportion of stayers in state 1)

$\hat{n}$=537 (#of stayers in state 1)

Final logistic regression model

Covariate	Coeff (std error)	p-value	odds ratio	95% CI for odds ratio	
Intercept	-1.3931 (0.5732)	0.0151			
Score	0.8624 (0.4124)	0.0365	2.369	1.056	5.316
Residence type	2.1904 (0.4551)	<.0001	8.939	3.664	21.809
Supports a family member(s)	0.8103 (0.3466)	0.0194	2.249	1.140	4.436
Education	1.2432 (0.3181)	<.0001	3.467	1.858	6.467

Initial patterns' classification

Types of patterns		# of customers	# of patterns
Only stayers' patterns		10	4
Only movers' patterns		23	5
Shared patterns	stayers	527	5
	movers	45	
Total		605	14

Analysis of shared patterns

Pattern	# of stayers	# of movers	Proportion of stayers with a given pattern
0100	13	14	0.48
1100	151	18	0.89
1101	100	5	0.95
1110	171	6	0.97
1111	92	2	0.98

Final assignment of patterns

Types of patterns		# of customers	# of patterns
High probability stayers' patterns		386	7
Only movers' patterns		23	5
Shared patterns	stayers	164	2
	movers	32	
Total		605	14

Estimation of the inconsistency rate

Proportion of customers in V with inconsistently classified patterns or patterns inconsistency rate for V

$$IR_V = \frac{1}{|V| - N_V} \sum_{p \in C} c_p$$

N_V - total number of customers in V sample with new patterns compared to B sample
C - set of inconsistently classified patterns in V sample
c_p- number of misclassified customers with a p pattern

Replication number	IR	# of customers with inconsistently classified patterns/($\|V\| - N_V$)
1	0.0182	4/220
2	0	0/235
3	0.0137	3/219
4	0	0/217
5	0.03125	7/224
mean (std dev)	0.012626 (0.01321)	

Q AND A

Thank you.

Contacts:

hfrydman@stern.nyu.edu

amatuszyk@matuszyk.com