A multivariate hidden Markov model: prospects for the course of public trust in Poland

Ewa Genge* and Fulvia Pennoni◇

*University of Economics,Katowice, Poland (ewa.genge@ue.katowice.pl)
◇University of Milano-Bicocca, Milano, Italy (fulvia.pennoni@unimib.it)

Presented at the 34th International Workshop on Statistical Modelling (IWSM), University of Minho, Portugal, 7-12 July 2019

Abstract

We propose a multivariate Hidden Markov Model (HMM) to analyse longitudinal survey data able to account for repeated responses over time along with longitudinal survey weights and missing responses. Since trust may be conceived as a psychological unobservable process of each person that fluctuates over time we consider observed time-varying and time-fixed individual covariates influencing the initial and transition probabilities. We employ the joint estimated posterior probabilities to make predictions on the latent trajectories of the course of public trust of the Polish society.

Introduction and Aims

❖ We formulate a HHM to account for longitudinal sampling weights and missing responses to measure public trust in one of the country with the lowest level of support to national government and parliament among the European member states (Eurobarometer survey [4]).

❖ We use a weighted log-likelihood to include the probabilities of sampling units in the likelihood function and consistently estimate the models parameters.

❖ We aim to find latent subgroups of Poles sharing the same perceived trust and we explain the resulting variability according to the available time-constant and time-varying socio-economic features.

❖ We predict the course of trust through the sequence of the predicted individual allocation to each latent state by means of the Viterbi algorithm.

Data and Methods

❖ Data are collected through the Polish Social Diagnosis survey [6] and the items are designed to monitor trust in **public institutions** over time. The opinions of 10,728 respondents are expressed by items having categories: "yes", "no" and "no opinion" during the years 2009, 2011, 2013 and 2015.

❖ The majority of respondents express public trust in *Police* and *National Bank of Poland*. The highest level of distrust (about 50%) is observed for *Government* and *National Parliament*. The highest percentage of "no opinion" in 2013 is towards *Stock Exchange* (see Table 1).

❖ The majority of respondents are married, have secondary education and live mainly in cities below 20,000 inhabitants or in rural areas. They are predominantly employed in private sector, with an average age close to 50 years old in 2013 (see Table 2).

Table 1. Weighted frequency distributions of each response variable and missing responses in 2013.

Trust in **2013** (%)	No	Yes	No opinion	Missing
Commercial banks	28.89	**36.09**	31.82	3.20
National Bank of Poland	11.56	**62.47**	22.83	3.14
National Parliament	51.89	27.76	17.38	2.97
President	31.27	**49.21**	16.54	2.98
European Parliament	33.44	36.32	27.18	3.06
Police	22.75	**64.33**	**9.90**	3.02
National Government	53.22	28.84	14.99	2.95
Social Insurance Institution	40.33	38.17	18.57	2.93
Stock exchange	32.39	14.80	**49.73**	3.08
Court	28.46	**50.37**	18.00	3.17
Insurance companies	31.86	26.53	38.47	3.14

Table 2. Weighted frequencies of individual covariates in 2013 and 2015: married, unmarried; education: primary, grammar, post-secondary; profession: public, private, other.

Covariate in **2013** (%)			
marital status	**61.36**	38.64	
education	16.40	**62.02**	21.57
socio-professional status	12.95	29.07	**57.97**
age	*mean=49.31*	*s.d.=17.45*	
Covariate in **2015** (%)			
marital status	**60.69**	39.31	
education	16.33	**60.03**	23.64
socio-professional status	12.66	30.97	**56.36**
age	*mean=51.31*	*s.d.=17.45*	

Latent Markov model

❖ Trust is conceived as a psychological construct fluctuating over time and conceptualized according to a latent process $\boldsymbol{U} = (U_1, \dots, U_T)$ influencing the distribution of the response variables and assumed as a stochastic Markov process of first-order with K discrete states (see [1]).

❖ The model parameters are the following:

1. The conditional distribution of the response variables given the latent states – measurement model:

$$\phi_{jy|u} = p(Y_{ijt} = y | U_{it} = u), \quad j = 1, \dots, r, \quad y = 0, 1, 2.$$

where Y_{ijt} is the j-th response variable provided by the i-th individual, $i = 1, \dots, n$, at the t-th occasion $t = 1, \dots, T$.

2. The initial and the transition probabilities of the latent process are parameterized according to time-fixed and time-varying covariates through a multinomial logit parameterizarion:

$$\log \frac{p(U^{(1)} = u | \boldsymbol{X}^{(1)} = \boldsymbol{x})}{p(U^{(1)} = 1 | \boldsymbol{X}^{(1)} = \boldsymbol{x})} = \log \frac{\pi_{u|\boldsymbol{x}}}{\pi_{1|\boldsymbol{x}}} = \beta_{0u} + \boldsymbol{x}'\boldsymbol{\beta}_{1u}, \quad u \geq 2,$$

$$\log \frac{p(U^{(t)} = u | U^{(t-1)} = \bar{u}, \boldsymbol{X}^{(t)} = \boldsymbol{x})}{p(U^{(t)} = \bar{u} | U^{(t-1)} = \bar{u}, \boldsymbol{X}^{(t)} = \boldsymbol{x})} = \log \frac{\pi_{u|\bar{u}\boldsymbol{x}}}{\pi_{\bar{u}|\bar{u}\boldsymbol{x}}} = \delta_{\bar{u}u} + \boldsymbol{x}'\boldsymbol{\delta}_{1\bar{u}u}, \quad t = 1, \dots, T, \quad \bar{u} \neq u.$$

where $\boldsymbol{X}_t$ is the vector of individual covariates and $\boldsymbol{\beta}'_{1u}$ and $\boldsymbol{\delta}'_{1\bar{u}u}$ are the vectors of regression parameters defining the influence of the covariates.

❖ Given a sample of n independent individuals providing the responses $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ we account for individual sampling weights w_i maximising the log-likelihood through the Expectation-Maximization (EM) algorithm [3]:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{t=1}^{T} w_i \ell_i(\boldsymbol{\theta}), \quad \ell_i(\boldsymbol{\theta}) = \log p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}).$$

❖ A suitable version of the Viterbi algorithm [7] is employed to predict the allocation of each individual to a latent state at each time occasion (global decoding) on the basis of the maximum a-posteriori probability.

Results

❖ The model is estimated by adapting suitable functions of the R package LMest [2]. Bayesian Information Criterion (BIC) lead us to select a model with $k = 4$ latent states.

❖ We label the latent states by looking at the estimated parameters of the measurement model:

- U_D people discouraged towards all the institutions,
- U_{Nop} people reluctant to manifest interest,
- U_T people confident in both public and financial institutions,
- U_{ST} people mainly oriented to support *Insurance Companies* (92%), *Government* (84%), *Police* (76%) and *Social Insurance Institutions* (68%), reluctant to express their opinions on *National Parliament*, *EU Parliament*, and showing lack of trust towards *Court* and *Stock Exchange*.

❖ Individuals with few years of educations are initially more prone to belong to the cluster of Poles with "no opinion" and show a lower probability to belong to the confident group U_T compared to Poles holding a post secondary education (see Table 3).

❖ After on time, higher-educated Poles show higher probability of supporting all the institutions U_T or of remaining in the cluster of those with selective confidence U_{ST} (see Table 3).

❖ The probability to move from no opinion U_{Nop} towards selective trust U_{ST} is higher for males, married people living in big cities of Poland.

Table 3. Averaged initial and transition probabilities (from $\bar{u}$ row, to u column) across educational levels.

	Primary education				Post secondary education			
	U_D	U_{Nop}	U_{ST}	U_T	U_D	U_{Nop}	U_{ST}	U_T
	0.2928	**0.4420**	0.0004	0.2648	0.3411	0.1757	0.0004	**0.4828**
U_D	**0.8525**	0.0156	0.1305	0.0013	0.8013	0.0004	**0.1961**	0.0022
U_{Nop}	0.0005	**0.8683**	0.1301	0.0012	0.0000	0.8107	**0.1832**	0.0060
U_{ST}	0.0000	**0.0911**	0.9087	0.0002	0.0052	0.0001	**0.9946**	0.0001
U_T	0.0049	0.0503	0.2058	0.7390	0.0006	0.0074	0.2092	**0.7828**

Prediction

❖ Table 4 shows the observed responses at each time occasion provided by a 50 years old married man, with secondary education, living in a small town, and working in the private sector, and his predicted pattern $\hat{\tilde{\boldsymbol{u}}}(\boldsymbol{y})$ estimated according to the Viterbi algorithm [7] (see Table 4)

❖ The majority of Poles (24%) is predicted to remain in the cluster of those with no opinion U_{Nop}, some of them (20%) remain in the group of skeptical people U_D and only few (14%) to remain in the cluster of confident people U_T.

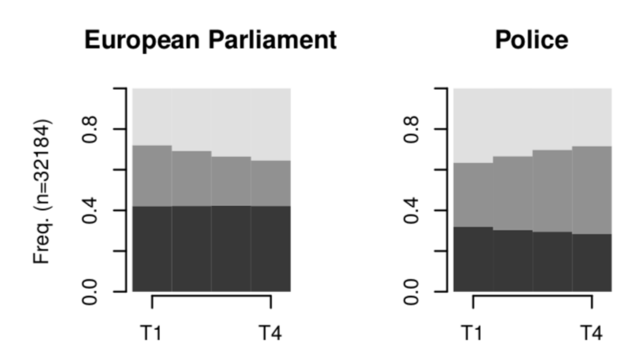
❖ Among the Poles predicted to change the initial latent state (25%) are allocated in the group of those with selective trust U_{ST} at the last time occasion.

❖ We checked the model tenability by predicting the responses provided by 32,184 individuals through an out-of-sample analysis. For these individuals in Figure 2 we observe a decreasing probability to trust *European Parliament* and an increasing probability to trust *Police*.

Table 4. Observed responses and predicted pattern.

$\boldsymbol{y}_{it}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	$\hat{\tilde{\boldsymbol{u}}}(\boldsymbol{y})$
$\boldsymbol{y}_{i1}$	1	NA	2	0	2	0	2	0	NA	NA	NA	U_T
$\boldsymbol{y}_{i2}$	1	1	2	2	2	1	2	1	2	1	2	U_{Nop}
$\boldsymbol{y}_{i3}$	2	1	0	1	0	0	0	2	0	0	1	U_{Nop}
$\boldsymbol{y}_{i4}$	2	2	2	1	2	1	1	1	0	1	1	U_{Nop}

Figure 2. Predicted response probabilities for EU and Police from an out-of-sample analysis. Black for No, grey for Yes, light grey for No opinion.



Conclusions

❖ We deal with a multivariate HMM for categorical longitudinal data.

❖ Polish people are less trustworthy towards the elites in general and they have the tendency to perform a selection of the reliable institutions.

❖ Special actions could be delivered for Poles predicted in the cluster of undecided people.

❖ To restore a positive engagement in politics and to increase social cohesion an effective policy would surely be to pursue free and good quality education. More details are illustrated in [5].

References

[1] Bartolucci, F., Farcomeni, A., Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman and Hall/CRC press, Boca Raton.

[2] Bartolucci, F., Pandolfi, S., Pennoni, F. (2017). LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software* **81**, 1–38. https://CRAN.R-project.org/package=LMest.

[3] Dempster, A. P., Laird, N. M., Rubin, D.B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1–38.

[4] Eurobarometer-European Commission (2014). *Public Opinion in the European Union*. Report, Standard Eurobarometer Spring **81**. http://ec.europa.eu/commfrontoffice/publicopinion/archives/eb/eb81/eb81_publ_en.pdf.

[5] Pennoni, F., Genge, E. (2019). Analysing the course of public trust towards via Hidden Markov Models: a focus on the Polish society, *to appear, Statistical methods and applications*, 1–27.

[6] Social Diagnosis (2015). Objective and subjective quality of live in Poland. Czapinski J., Panek T. (eds.), Warszawa, Social Monitoring Council.

[7] Viterbi, A. J. (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, **13**, 260–269.

Acknowledgements

The IWSM 2019 workshop attendance has been supported by the research project (SONATA 12, UMO-2016/23/D/HS4/00989, "Latent variable models in the identification of homogenous structures in socio-economic longitudinal data") of the National Science Centre, Poland.