Noname manuscript No.
(will be inserted by the editor)

Analysing the course of public trust via Hidden Markov Models: A focus on the Polish society

Fulvia Pennoni · Ewa Genge

the date of receipt and acceptance should be inserted later

Abstract We investigate public trust using measurements from individual items recorded through a long-term survey. We account for repeated and missing item responses by a hidden Markov model. Since trust may be conceived as an unobservable psychological process of each person that fluctuates over time, we allow for time-varying and time-fixed individual covariates affecting the latent process. We estimate the model parameters by a weighted log-likelihood through the Expectation-Maximization algorithm using longitudinal sampling weights and data collected in an East-Central European country like Poland. The latter is a country where the level of support to the national and international institutions is one of the lowest among the European member states. We apply a suitable algorithm based on the posterior probabilities to predict the best allocation to each latent typology. The proposed model is validated by generating out-of-sample responses, and we find good predictive values. We disentangle four hidden groups of Poles: discouraged, with no opinion, with selective trust and with full public trust. We reveal an increasing number of people that are going to trust only some selected institutions over time.

Keywords: Expectation-Maximization algorithm, missing responses, panel data, sampling weights, trust-building policy discussion.

Fulvia Pennoni
Department of Statistics and Quantitive Methods
University of Milano-Bicocca, Milano, Italy
Present address: via degli Arcimboldi 8, 20126 Milano
Tel.: +39-02-64485856
E-mail: fulvia.pennoni@unimib.it
ORCID: 0000-0002-6331-7211

Ewa Genge
Department of Economic and Financial Analysis
University of Economics, Katowice, Poland
Tel.: +48-32-2577091
E-mail: ewa.genge@ue.katowice.pl
ORCID: 0000-0002-8899-3697

1 Introduction

Public trust is a human attitude that express something which can be valuable for the individual dimension. It states the personal confidence towards the society, governments as well as public, economic, social and financial institutions and it is often conceived as a public good or *social capital* (Gille et al., 2016) and it is recently advocated in the ethics guidelines for trustworthy artificial intelligence (European Commission, 2019, p.4): "trust remains the bedrock of societies, communities, economies and sustainable development". It is related to the confidence on politicians, officials as well as on the main public and private and religious organizations. It is theorized to contribute to democratization according to the reasoning proposed by the following authors: Fukuyma (1995), Sztompka (1999), Putnam (2000). It is considered essential for the success of a wide range of public policies which surely depend on the behavioural responses of the targeted individuals (see, among others Paldam, 2008; Brehm and Rahn, 1997; van Oorschot et al., 2006).

We remark the dynamic nature of trust: people's views about an issue can develop and change over time and can be influenced by social media. People can be converted from indifference or poorly informed reactions to more thoughtful conclusions and to a settled public judgment and vice-versa. Mass media are also judged by individuals that can assign a certain level of trust to each communication fora. In the Western societies we assist to a more or less accentuated decline of public trust as also recently remarked in the trust barometer findings (Barometer, Edelman Trust, 2019). The Organisation for Economic Co-operation and Development (OECD) reported that in 2016 only 42% of citizens was confident in the national government, compared to 45% before 2007 (OECD, 2017). Dissatisfaction and lack of confidence in the functioning of the democratic institutions in many developed countries as a widespread phenomenon was first stressed by Crozier et al. (1975). The recent observed distrust on financial institutions is often explained with diametrically different reasons. These suspicious perceptions may be the result of the lack of experience in using the services provided by the institutions. On the other hand, there is also the greed of private institutions which are looking for solid profits from their activities.

Poland belongs to one of East-Central European (ECE) countries showing the lowest level of trust according to the survey on public opinions carried out by the European Commission (Eurobarometer, 2014). Mishler and Rose (1997) claimed that people in this group of countries evaluate political institutions according to a general frame, which is strongly determined by the economic situation of the country they lived in. Due to the 2007 financial crisis citizens might be generally distrustful on institutions, and they tend to perceive governors as corrupt (Marien, 2011). It is also worth to be noted that, in the course of democratic development, people could have become more aware of the differences between political, financial and international institutions.

Other two Western European countries like Italy and France are characterised by lower level of public trust than that of Poland (see, among others

Eurobarometer, 2014; Cautrès, 2017; Fazio et al., 2018). It is interesting to note that in France, this lack of public trust determined an impetus of social change due to discontentment about the government's method. In 2018 dissatisfaction with the system has been translated in factual claims manifested by the yellow vests movement[1], in a way to undermine legitimacy needed for governance (Van Prooijen and van Lange, 2014).

At individual level public trust is generally measured through items provided in the questionnaires concerning opinions (beliefs) and attitudinal survey questions (see among others, Albanese et al., 2013; Groenewegen et al., 2018). The data collected during the last decades have been analyzed through research mainly devoted to link institutional trust to good governance. Among the British society public trust is considered a social attitude such as neighbourhood attachment and civic participation (Li et al., 2005). Ersh et al. (2009) analyse trustworthiness of the Britishs with a newly designed experiment using real monetary rewards and Sapienza et al. (2007) consider the economic effects of distrust. Kuovo (2011) study public trust using data of the European Social Survey (ESS Round 4, 2008) and for selected European countries (Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine and United Kingdom) differing in the welfare system. Wang and Gordon (2011) using data of the World Value Survey (Inglehart et al., 2014) analyse trust at country level. As far as the Polish society is concerned Genge (2014) and Trzesiok (2016) consider data of Social Diagnosis (Czapiński, 2013) to evaluate attitudes towards the euro adoption in Poland and to explain the correspondence between trust in financial institutions and government, respectively.

Although there is considerable consensus about the importance of monitoring public trust, there is little consensus about its measurement. Mainly, basic frequencies of the items questionnaires on trust are reported in the literature, or composite indicators at country level are proposed (see, among others Gambetta, 1988; Glaeser et al., 2000; JRC-OECD, 2008). Boda (2014) use the European Social Survey data (ESS Round 4, 2008) to construct a composite indicator according to the mean value of the valid responses and show that Eastern Europeans do not demonstrate greater "materialistic" trust (related to income) compared to the Western peers. By using the same data Ydersbond (2015) propose an additive index of social and political trust created selecting specific items. Marozzi (2015) proposes a composite index to compare trust in public institutions and rank European countries on the basis of the European Social Survey (ESS Round 6, 2012) data. An additive index to measure trust in the executive, courts, police, armed forces, electoral commissions and government-run media is proposed by Hutchison and Johnson (2005) for 16 countries with data of the Afrobarometer. An index of general political trust and corruption for 103 countries is proposed by Clausen et al. (2011) made according to the Gallup World Poll, summing responses to questions on con-

[1] `https://en.wikipedia.org/wiki/Yellow_vests_movement`

fidence in the military, judicial system and courts, national government and honesty of elections. Similarly, other indices of political trust are developed for Latin America (Stoyan et al., 2016), Asia (Wong et al., 2011), Sub-Saharan Africa (Cho and Matthew, 2007; Lavallèe et al., 2008) and other countries (Catterberg and Moreno, 2005; Hakhverdian and Mayne, 2012). The use of such indexes especially when they are considered proxy for general institutional trust is criticized by many authors such as Fisher et al. (2010) claiming that citizens develop different forms of trust judgements that may vary both in application and significance depending on the given institution. Reeskens and Hooghe (2008) argue the uselessness of the comparisons of country specific means made by the generalized social trust index and more recently Schneider (2016) highlights some limitations in the measurement of political trust.

In accordance to Rosanvallon (2008) that describes the "latent" nature of trust as an "invisible institution" we conceive public trust as an hypothetical construct (concept) not directly measurable and only assessed through items. Therefore, a latent variable model (Skrondal and Rabe-Hesketh, 2004; Pennoni, 2014) is a valid tool to analyze responses provided to the items concerning this psychological process, which is multidimensional and dynamic over time. With respect to Fazio et al. (2018) that employ a latent class model (Lazarsfeld, 1950) for the item responses to the questions on trust towards the Italian provincial government, we propose a hidden Markov model (HMM, Bartolucci et al., 2013) to account for the chronological order of responses provided at different time points by each individual.

The novelty of our proposal is a multivariate HMM that employs the longitudinal sampling weights in order to be suitable to analyze data arising from complex social surveys with unequal sampling probabilities among respondents. Within the proposed HMM we account for missing responses due to refusal to respond, deleted data or absence of contact with the respondent, in order to avoid possible bias due to systematic unit non-response. This model-based approach allow us to use "raw" individual response categories provided to the question "Do you trust in..." administered throughout a questionnaire repeated over time. Most importantly, we model the category of "no opinion" among " "yes" and "no" since abstention may be due to uncertainty or mainly to profound discomfort and it is a very important aspect to account for. In fact, individuals may abstain to respond for many different reasons: they lack of opinions, they do not like to express opinions on the topic or to make them public. This absence of expression or indecision concerning public trust may also vary from a political refusal to poor interest, lack of information, lack of self-confidence, or perhaps it can be due to a more deep critical vision of the individual which cannot be expressed by a mark on a question, or it can be due to manifested disengagement generated by deep delusion.

We aim to identify similar typologies of individuals sharing common perceptions according to different trust's dimensions. We are interested to characterize how these typologies are evolving over time and to explain the resulting variability due to the available time-constant and time-varying socio-economic features of the respondents including respondents with missing responses. A secondary aim is to classify individuals in each latent component and to make individual predictions according to the estimated posterior probabilities to belong to a latent state at any moment in time. This prediction is highly informative to detect the main course of trust over the whole population. We also show the correctness of the predictive power of the proposed HMM by producing out-of-sample realisations to verify if they are able to maintain the interpretability.

In summary the methodological contribution of the current proposal can be stated in the following aspects which we cover jointly: the categorical nature of the response variables which are made by the responses to several items referred to trust towards different institutions, the multidimensional nature of the phenomena, the time order of the responses, the probabilities of selection of the units in the population due to the survey design, the missing responses, the time-varying and time fixed individual covariates.

The rest of the paper is structured in four sections. Section 2 describes the data motivating the current research collected within the Social Diagnosis survey. Section 3 illustrates the methodological proposal and focuses on the steps for the maximum likelihood estimation of the model parameters for the basic HMM and for the HMM with covariates. Section 4 reports the main results using data collected from 2009 to 2015, as well as predictions. It also shows the results of the HMM estimated on data collected on previous waves. Section 5 concludes the paper with some highlights on the current course of trust in Poland and some general remarks.

2 Social Diagnosis Survey data

The available data refer to the long-term longitudinal Social Diagnosis survey conducted in Poland over a large sample of households aged 16 and above. The survey is aimed at highlighting information on the labour, education and other features such as public trust at the household and individual level (Social Diagnosis, 2015). The sample is stratified according to a two-stage sampling design where census areas are sampled with probabilities proportional to the number of dwellings. Urban strata are divided into large towns (with more than 100k residents), medium-sized towns (with a number of resident between 20k and 100k) and small towns (with a number of residents less than 20k). Furthermore, in five largest cities the strata covered the household districts. At the second stage, the dwelling are sampled according to the five census areas. The survey is carried out every two years since 2000 mainly to support the decision makers with data derived from indicators concerning attitudes, mind-sets and behaviours of the households and their members.

The research interest concerns individual responses provided by the households over the last four waves of the survey carried out from 2009 to 2015. In fact, the questionnaires contain a complete set of questions able to detect the individual perceptions related to different trust's dimensions . For this period of time, we dispose of a representative sample of $n = 10,728$ individuals and longitudinal sample weights which allow us to preserve the representative characteristics of the sample and to consider its deficiencies in covering the target population (see, among others, Ernst, 1989; Paas et al., 2007).

Table 1 illustrates the weighted percentages of responses for each item referred to 11 public and financial institutions addressed in the survey and reports the weighted percentages of missing responses. It is interesting to note that the question *Do you trust in . . . ?* is referred mainly to national institutions except one concerning European Parliament. Uncertainty is accounted in the questionnaire by allowing the category *"I have no opinion" along "yes I trust"*, *"no I don't trust"*. In 2009 the questions related to *Stock Exchange*, *Court* and *Insurance Companies* have not been administered. The largest percentage of missing responses (close to 4%) is observed in the sixth wave (2011) of the survey. For sake of comparison in Table 2 we show the unweighted response frequencies of two items referred to the 2013 survey.

We notice that for each year the majority of individuals express public trust in *Police*, *National Bank of Poland* as well as in *Commercial Banks* except for 2011 (in this year the question about trust in *National Bank of Poland* was asked for the first time). In 2015 there is a higher percentage of people feeling that they can trust the *Insurance Companies* and the *Social Insurance Institutions* with respect to the previous years. In 2013 and 2015, there are more people expressing their support towards the *President*, and *Court*. Trust in *Government* is increasing over time. In 2015 people with distrust towards the National Parliament are about 47%. According to the constitutions the system of government of Poland is based on the separation of and balance between the legislative, executive and judicial powers. Parliament in Poland (the Sejm and the Senate) is the determining body (legislative power). The government (the President and the Council of Ministers) is approved by the parliament by the vote of confidence, it is the executive body (executive power). Judicial power is vested in courts and tribunals.

As far as the third response category ("no opinion") is concerned, we note an increased number of people not expressing their opinion towards *Stock Exchange*. We observe that there is less indecision towards in this institution during years 2011-2015 since the percentage of "yes" ranges from 7.42% to 15.6% . In 2009 to 2015 a steady decrease from 23.96% up to 9.89% is observed for "no opinion" towards Police. Concerning the *EU Parliament*, we notice a marked decline of undecided people (with "no opinion") from 2009 to 2011 and more confident respondents even if in 2015 the percentage of those discouraged towards this international institution remains high (34%).

To characterize trustworthiness and the its dynamics we consider many personal features selected among the available covariates because they are supposed to have an effect on trusting. We account for the following socio-economic characteristics of the respondents: gender, marital status, education, place of living, socio-professional status. From the observed frequencies depicted in Table 3 weighted with sampling weights we notice that the majority of respondents are married, have secondary education and live mainly in the cities below 20,000 inhabitants or in rural areas. In 2009 respondents are on average 45 years old, age is analyzed in years as a continuous time-varying covariate.

3 Multivariate hidden Markov model

The HMM is an extension of the latent class model (Lazarsfeld and Henry, 1968) initially proposed in its basic version by Wiggins (1955). In his book (Wiggins, 1973) published in 1973 first noticed that the model based on latent variables would be suitable for a wide range of panel data especially for data referred to the illustrated political studies on vote intention and on winner expectations to political elections referred to the year 1940. Since that time, research on latent variable models as well as on finite mixture models to which this class of models belongs to, has been developed (see, among others McLachlan and Peel, 2000; McLachlan et al., 2018). In particular, the extensions concerns latent variable models for longitudinal data using mixtures (see, among others Vermunt, 2010). More recently, many theoretical advances on the HMM have been made and it has been extensively employed as advanced statistical methods to analyse several different panel data. We refer to Bartolucci et al. (2014) for a synthetic overview of the HMMs. The model estimation has been developed in many ways by considering a full likelihood or a Bayesian approach. We mention a recent extension proposed by Bartolucci et al. (2016) tailored to draw causal inference when the effect of policies on the responses is of interest. In the recent literature, the HMM model has been employed as an advanced statistical method to analyse survey data (Pennoni, 2016) along with the latent class model (see, among others, Magidson and Vermunt, 2000; Pennoni and Nakai, 2018). The HMM has been compared with the competitive model named latent growth mixture model (Muthen, 2002) by Pennoni and Romeo (2017) where the authors employ an application based on survey data on self-rated heath status. It has been proposed to analyse material deprivation (Dotto et al., 2019) and the level of satisfaction on the primary work (Bartolucci et al., 2017). Note that both deprivation and work dissatisfaction have been found as negatively associated with levels of trust (see Fazio et al., 2018, and the references therein).

The multivariate formulation of the HHM deals with the responses recorded at several time occasions and it accounts for covariates in the latent model or in the measurement model (see Bartolucci et al., 2013). In the following, by assuming that the missing pattern is non-informative, as it depends on the survey structure we propose a specific HMM to account for the missing responses and for the longitudinal survey weights which are derived by raw data from subsequent panel waves. By considering the longitudinal weights we avoid the bias due to representativeness of each unit in the population. By considering the missing responses we avoid the bias due the complete data solution. We assume that the mechanism inducing the missing responses generates missing-at-random responses: the missing responses and the latent variables which characterize the HMM are conditionally independent from the missing data mechanism given the observed variables. In the following, we illustrate the basic HMM since it is fitted as a preliminary step to recover the parameters of the manifest model and then we show its formulation in order to allow the covariates to influence the latent part of the model.

3.1 Basic multivariate hidden Markov model with longitudinal sampling weights

We assume that the observed sequence of responses provided at each time occasion indirectly measure a latent trait representing perceptions that are tempered at individual level. Let $\boldsymbol{Y}_{it}$ be the observed response vector for individual i, $i = 1, \dots, n$, at each time occasion t, $t = 1, \dots, T$. We denote as Y_{ijt} the single response variable provided to item $j, j = 1, \dots, r$ by individual i, $i = 1, \dots, n$ at time occasion t. The basic HMM assumes that the observed responses depend on the unobserved trait which is time-varying denoted as $\boldsymbol{U} = (U_1, \dots, U_T)$. This hidden stochastic process has a distribution with k support points assuming finite discrete values. For individual i and time occasion t it is denoted as U_{it} with $i = 1, \dots, n$, $t = 1, \dots, T$. Specifically, we make the the first-order Markov assumption under which the unobserved perceptions are dependent only on those at the previous time point that is the trait held at time 3 depends only on the trait held at time 2 and not on that at time 1. We assume that the responses observed at a given time point $\boldsymbol{Y}_t$, only depend on the latent variable U_t referred to the same time occasion. We make the local independence assumption that is that the responses collected in the vectors $\boldsymbol{Y}_{i1}, \dots, \boldsymbol{Y}_{iT}$ are independent one another conditionally to the latent process $\boldsymbol{U}_i$. Moreover, for each individual i at time occasion t each response collected in the vector $\boldsymbol{Y}_{it}$ is conditionally independent given U_{it}.

A set of parameters is made by the conditional probabilities of the responses given the latent variables and it is referred to the manifest model:

$$\phi_{jy|u} = p(Y_{ijt} = y | U_{it} = u), \tag{1}$$

for $i = 1, \dots, n,\ j = 1, \dots, r,\ y = 0, 1, 2, t = 1, \dots, T,\ u = 1, \dots, k$. The other set of parameters concerns the initial and the transition probabilities and it is referred to the latent model. They are denoted as

$$\pi_u = p(U_{i1} = u) \quad u = 1, \dots, k$$

$$\pi_{u|\boldsymbol{u}} = p(U_{it} = u | U_{i,t-1} = \bar{u}) \quad u, \bar{u} = 1, \dots, k,\ t = 2, \dots, T.$$

The transition probabilities are generally stored in the transition matrix denoted as $\boldsymbol{\Pi}$ which along with the initial probability vector denoted as $\boldsymbol{\pi}$ defines the latent Markov chain (Cappe et al., 2009). Note that we are assuming time homogeneity according to which the transition probabilities do not depend on the time occasion. This is a reasonable assumption for the context at hand since in this way we avoid overfitting preferring to estimate the covariates effects on these probabilities.

We account for individual longitudinal sampling weights denoted as w_i, $i = 1, \dots, n$ by considering a weighted log-likelihood (a pseudo-likelihood) that given a sample of n independent individuals providing the responses $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ is written as

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

where $\boldsymbol{\theta}$ represents the vector of all free parameters arranged in a suitable way. The *manifest probability* of the responses $p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT})$ is computed by suitable recursions developed in the hidden Markov literature (Baum et al., 1970; Welch, 2003; Zucchini and MacDonald, 2009).

The Expectation-Maximization (EM) algorithm (Baum et al., 1970; Dempster et al., 1977) represents the main tool to estimate the HMM models. It is based on the *complete data likelihood* that for the proposed model is given by

$$\begin{aligned} \ell_1^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg[& \sum_{u=1}^{k} \sum_{j=1}^{r} \sum_{t=1}^{T} \sum_{y=0}^{2} a_{iujty} \log \phi_{jy|u} + \sum_{u=1}^{k} \sum_{t=1}^{T} w_i b_{iu1} \log p(U_{i1} = u) \\ & + \sum_{\bar{u}=1}^{k} \sum_{u=1}^{k} \sum_{t=2}^{T} w_i b_{i\bar{u}ut} \log p(U_{it} = u | U_{i,t-1} = \bar{u}) \Bigg], \end{aligned} \tag{2}$$

where a_{ujty} corresponds to the (weighted) frequency of people responding to the j-th item and belonging to latent state u at occasion t, b_{iu1} is an indicator variable equal to 1 if individual i belongs to latent state u at the beginning of the period, with $p(U_{i1} = u)$ being the initial probabilities and $b_{i\bar{u},t} = b_{i\bar{u},t-1} b_{iut}$ is an indicator variable equal to 1 if the same respondent moves from state $\bar{u}$ to state u at occasion t, with $p(U_{it} = u | U_{i,t-1} = \bar{u})$ being the transition probabilities. Details concerning the steps of the EM algorithm can be found in Bartolucci et al. (2013, 2016).

The HMM model needs to be estimated several times since it is important to explore the entire parameters space of each model with a different number of hidden states due to the fact that the log-likelihood function may be multi-modal. The appropriate number of latent states, when not known in advance, is recovered through penalized-likelihood criteria such as the Bayesian Information Criterion (BIC, Schwarz, 1978) as measures of relative goodness of fit. A comparative study conducted by Bacci et al. (2014) shows that the BIC is to be preferred for the HMM. The model with the lowest values of the index is suggested to be the best one. However, it is possible that its values are strictly decreasing when the number of components is increased. In such cases, to follow the parsimony principle the literature suggests to consider the smallest number of components able to represent the main typologies of the latent phenomena under study. To respect the parsimony principle it is also a common practice to plot the BIC values against the increasing number of latent states and to choose the number of components according to the point corresponding to the angle in the plot (elbow criterion). The latter is the point where the max decrease is reached. The other principle is to check for a reasonable interpretation of the resulting states according with the subject matter knowledge. This suggestion is always valid for all the models in the class of finite mixture models (see, among others, Magidson and Vermunt, 2000). Standard errors for the parameters are used as a measure of precision mainly for the covariates effects and are computed as the square root of the elements in the main diagonal of the observed or expected information matrix at the maximum likelihood estimate. Otherwise, they are computed by applying the parametric or non-parametric bootstrap, see Bartolucci et al. (2013) for more details.

3.2 Hidden Markov model with covariates in the latent model

The time-fixed and time-varying covariates are denoted by $\boldsymbol{X_t}$, $t = 1, \ldots, T$ for the t-th time occasion. The covariates are supposed to influence the initial hidden states of the Markov chain as well as the probability to transit between states. To explore how they characterize the latent trait giving rise to the expressed level of trust on the observed items we consider a parameterization on the transition probabilities by using a generalized linear model for each row of the transition matrix. First, we constrain the parameters of the measurement model to be fixed at the estimates of the manifest probabilities $\hat{\phi}_{jy|u}$ obtained with the basic HMM presented in Section 3.1 since this reduce the bias due to covariates. Then, we consider simple logit models for the initial probabilities with the first latent state as reference category and multinomial logit models for the transition probabilities with $\bar{u}$ as reference category as the followings

$$\log \frac{\pi_{u|\boldsymbol{x}}}{\pi_{1|\boldsymbol{x}}} = \beta_{0u} + \boldsymbol{x}'\boldsymbol{\beta}_{1u}, \quad u = 2, \ldots, k, \tag{3}$$

$$\log \frac{\pi_{u|\bar{u}\boldsymbol{x}}}{\pi_{\bar{u}|\bar{u}\boldsymbol{x}}} = \delta_{\bar{u}u} + \boldsymbol{x}'\boldsymbol{\delta}_{1u\bar{u}},\ \bar{u} \neq u \tag{4}$$

where $t \geq 2$, and $\boldsymbol{\delta}_{11} = \boldsymbol{0}$ for the model identifiability. In the above equations, $\boldsymbol{\beta}'_{1u}$ and $\boldsymbol{\delta}'_{1\bar{u}u}$ define the influence of the covariates.

The *complete data log-likelihood* is given by the following two terms

$$\begin{aligned}\ell_2^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \Bigg[& \sum_{u=1}^{k}\sum_{t=1}^{T} w_i b_{iu1} \log p(U_{i1} = u|\boldsymbol{x}_{it}) \\ & + \sum_{\bar{u}=1}^{k}\sum_{u=1}^{k}\sum_{t=2}^{T} w_i b_{i\bar{u}ut} \log p(U_{it} = u|U_{i,t-1} = \bar{u}, \boldsymbol{x}_{it}) \Bigg],\end{aligned}$$

where b_{iu1} and $b_{i\bar{u}u,t} = b_{i\bar{u},t-1} b_{iut}$ are computed on the basis of the parameters of the measurement model ($\hat{\phi}_{jy|u}$) estimated within the basic HMM as illustrated in the previous section.

We predict the allocation of each individual to each latent state according to the maximum a-posteriori probability known as MAP-based approach. We use an adapted version of the Viterbi algorithm (Viterbi, 1967; Juang and Rabiner, 1991) as proposed by Bartolucci et al. (2013). This allocation called global decoding is considered as optimal since the algorithm maximizes the overall a-posteriori probabilities provided by the estimated HMM. We consider the joint posterior estimated probability $\hat{p}_1(u, \boldsymbol{y})$ concerning the first time occasion and the same probabilities $\hat{p}_t(u, \boldsymbol{y})$ for $t = 2, \ldots, T$ and we predict the optimal states $\hat{\hat{u}}_T(\boldsymbol{y})$ and $\hat{\hat{u}}_t(\boldsymbol{y})$ first by a forward recursion and then by the following backward recursions

$$\hat{\hat{u}}_T(\boldsymbol{y}) = \underset{u=1,\ldots,k}{\operatorname{argmax}}\, \hat{p}_T(u, \boldsymbol{y}),$$

and

$$\hat{\hat{u}}_t(\boldsymbol{y}) = \underset{u=1,\ldots,k}{\operatorname{argmax}}\, \hat{p}_t(u, \boldsymbol{y}) \hat{p}_{(t+1)}(\hat{\hat{u}}_{(t+1)}(\boldsymbol{y})|u), \quad t = T-1, \ldots, 1.$$

4 Results

We show the results of the basic HMM and then those of the HMM estimated with covariates on the data illustrated in Section 2. Then, we focus on predictions. Finally, we discuss the results obtained by estimating the HMM for waves from 2003 to 2007 of the same survey.

4.1 Results of the basic HMM and of the HMM with covariates

The results of the model estimated without covariates in terms of likelihood and information criteria are reported in Table 4 for a number of latent states ranging from 1 to 11. We choose four hidden states as the suitable number to represent the subpopulations forming distinctive clusters of perceptions towards the institutions since the values of the BIC index are slightly decreasing when the number of latent states is greater than four. In this way, as stated in Section 3.1, we apply the parsimony principle and we favour interpretability since we noticed that the typologies of the latent phenomena are well defined. The basic HMM model proposed in Section 3.1 with $k = 4$ states has a maximum log-likelihood equal to $\hat{\ell} = -295,923.7$, BIC = 592,988.8 with 127 free parameters. Suitable `R` code and functions to prepare the data and to estimate the model parameters have been adapted from the `R` package `LMest` Bartolucci et al. (2017). The complete results are available from the authors upon request.

From Table 5 we disentangle the four latent subpopulations corresponding to the hidden states on the basis of the estimated manifest probabilities

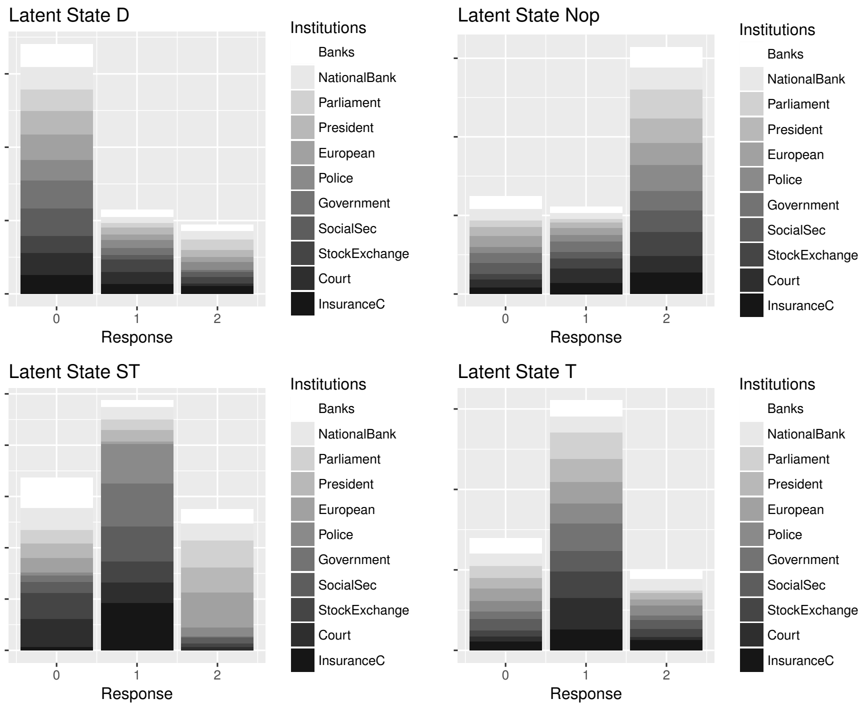

Fig. 1: *Estimated conditional probabilities showed according to the latent states labelled as Discouraged, (D, no Trust), No Opinion (Nop), Selected Trust (ST), Trust (T) and to the response categories 0 "no I don't trust", 1 "yes I trust", 2 "I have no opinion" of the question* Do you trust in ...? *referred to each institution on the right.*

referred to the joint responses to all items as in Equation 1. The estimated probabilities are also depicted in Figure 1 to have a graphical inspection of the heterogeneity of each latent group. Interestingly, the cluster labelled as U_D (discouraged or distrust) collects individuals with the higher estimated probability for the response category referred to absence of public trust. It represents Poles not supporting institutions, especially *National Government* and *Social Insurance Institutions* (over 75%) as well as *European Parliament*. The cluster labeled as U_{Nop} (no opinion) is mainly referred to the latent subgroup of Poles reluctant to manifest interest towards institutions, or as explained in Section 1, not prompt to share opinions. Primarily they do not judge *National Parliament* (75%), *Police* (66%) and *Stock Exchange* (60%). However, they manifest some support to *Court* (37%). The cluster labeled as U_T (trust) assemble Poles showing predominant confidence in both public and financial institutions. The highest level of trust is towards *Court* (close to 80 per cent) and the smallest is towards *Commercial Banks* (close to 40%). The cluster labeled as U_{ST} (selected trust) represents the most heterogeneous latent subpopulation: people belonging to this cluster are mainly oriented to support *Insurance Companies* (92 per cent), *Government* (84 per cent), *Police* (76 per cent) and *Social Insurance Institutions* (68 per cent). Instead, they do not like to express their opinions on *National Parliament*, *EU Parliament* and *President* and more than half show a lack of trust towards *Court* and *Stock Exchange*. From Figure 1 it is evident that they show the highest probability to trust European Parliament which is the only international institution represented in the questionnaire, compared to those in the other clusters.

As explained in Section 3.2 the parameters of the latent model are estimated after that the measurement model's parameters are fixed at the estimates obtained with the basic HMM. The latent HMM includes the effects of the individual covariates through a multinomial logit parameterization as in Equations 3 and 4. The full set of parameters is referred to the covariates illustrated in Table 3 as time-fixed and time-varying covariates. For the initial probabilities of the latent chain we consider the covariates collected in 2009. In estimating the HMM with covariates we get a BIC value at convergence equal to -298,837.400 with 238 parameters. We have 39 $\boldsymbol{\beta}$'s and 156 $\boldsymbol{\delta}$'s parameters referred to the initial and transition probabilities respectively. The estimated coefficient for the multinomial logit on the initial probability of the U_T latent state versus U_D referred to the higher and post-secondary education is significant and is equal to 0.26 indicating that those with an higher education tend to belong to the cluster of those showing public trust compared to those with lower levels of education (the odds ratio for higher vs less educated individuals is equal to exp(0.260) =1.29) fixing the values of the other covariates. The same coefficient referred to females is significant and is equal to 0.315 indicating that the odds ratio for females vs males to belong to the cluster of those trusting in institutions is 1.37. We are not reporting all the estimated coefficients since we prefer to show the most significant effects, according to the estimated standard errors, comparing the averaged initial and transition probabilities across covariates.

Table 6 shows the averaged initial probabilities to belong to each cluster for the whole population. We observe that at the beginning of the period the clusters U_D distrust, U_{Nop} no opinion and U_T fully trust are equally represented in the population (each has a percentage of about 30) whereas only 1% is located in U_{ST} the cluster of Poles with selected trust.

Table 7 shows the estimated averaged transition matrix for the period 2013-2015. The probabilities among the four groups are higher in the main diagonal indicating persistence in each cluster. In particular, we notice a transition towards the cluster of selected trust: the discouraged Poles U_D and the Poles with no opinion U_{Nop} are especially prone to switch to that of Poles who trust in selected institutions U_{ST} ($\hat{\pi}_{3|1} = 0.14$ and $\hat{\pi}_{3|2} = 0.14$). Interestingly, about 20 per cent of Poles which are confident in institutions U_T are switching towards the cluster of those with selective confidence U_{ST} ($\hat{\pi}_{3|4} = 0.20$).

Table 8 reports the averaged initial and transition probabilities for people differing with respect to the level of education. At the first year of the survey, Poles with less numbers of years of education are more prone to belong to the cluster of Poles with "no opinion". They also show a lower probability to belong to the confident group U_T compared to Poles holding a post secondary education. 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} compared to those with only primary education. Less educated individuals also show higher probability to remain in the subpopulation of Poles not supporting the institutions U_D or to stay in the group of those with no opinions U_{Nop} compared to the higher-educated Poles.

Table 9 reports the same probabilities referred to people working in the public sector and those that are professional inactive. Referring to the initial period we notice that Poles employed in the public sector have the highest probability of positive feeling towards the institutions U_T as opposed to the Poles which are professionally inactive characterized instead by the highest probabilities to belong to discouraged Poles U_D or to Poles with no opinion U_{Nop}. Government workers show higher probability to belong to clusters of Poles who select the institutions U_{ST} compared to inactive Poles. Workers employed in the public sector also show an higher probability to switch from discouraged people U_D and from people with no opinion U_{Nop} to Poles confident in selected institutions U_{ST} compared to inactive Poles.

Tables 10–12 show the same probabilities comparing males and females, married and unmarried people and people living in big and small cities respectively. We notice that at the beginning, females, unmarried people living in small towns or rural areas are more prone to belong to the latent subpopulation without opinions U_{Nop}. At the initial period, males show higher probability to belong to cluster of distrusted Poles U_D compared to females. According to the transitions males supporting all the institutions U_T are persistent with their support compared to women. Married Poles as well as Poles living in big cities have a higher probability to belong to cluster of confident people U_T. The probability to move from no confidence U_D towards selected trust (U_{ST}) is higher for females, married individuals, living in small cities and rural areas of Poland.

4.2 Prediction of the cluster membership

We predict the allocation of each individual to the four latent states through the estimated HMM a-posteriori probabilities as introduced in Section 3.1. For sake of illustration in Table 13 we show the sequence of the observed responses $\boldsymbol{y}_t$ provided at each time occasion by a married man of middle age (50 years old), with secondary education, living in a small town, working in the private sector, along with his predicted profile estimated according to the Viterbi algorithm (Viterbi, 1967) denoted as $\hat{\hat{\boldsymbol{u}}}(\boldsymbol{y})$.

The global decoded sequence illustrated in the last column of Table 13 indicates that this person would probably change from supporting all the institutions to no opinion. This prediction is highly informative to detect the main course of trust of the whole population. According to the estimated posterior probabilities only 25% of Poles is going to be allocated from a different cluster of the initial one. The majority of Poles (24%) is predicted to remain in the cluster of those with no opinion U_{Nop}, 20% is predicted to remain in the group of skeptical people U_D and only 14% is predicted to remain in the cluster of confident people U_T.

As a further piece of information, referring to people predicted to be in cluster of selected trust U_{ST} at the end of the period we show their trajectories in Figure 2. Looking at the figure we notice that a majority of them transit to this cluster at the third period. They arrive to the cluster U_T from all the other clusters. We notice that all the individuals in the cluster of fully trust U_T at the first occasion transit to the cluster of selected trust U_{ST}.

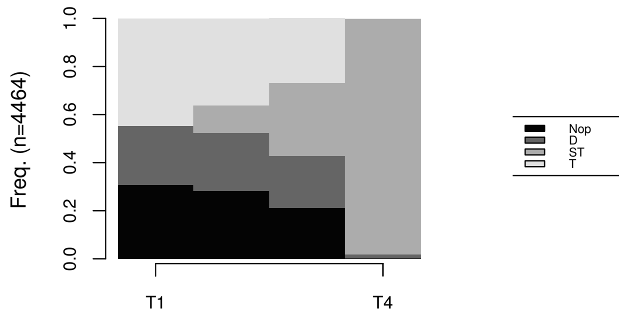

Fig. 2: *Predicted trajectories across clusters on the subset of people changing latent states within the four waves (eight years from 2009 to 2015): D: Discouraged, Nop: No Opinion, ST: Selected Trust, T: Trust.*

The estimated values can also be described according to the covariates. For example, the group of Poles which is predicted to change from no trust U_D to selected trust U_{ST} is mainly made by married Poles, holding secondary education, living in small towns and not employed in the public or private sector.

We show the performance of the estimated HMM by predicting the responses provided to each item at each occasion by a number of individuals equal to three times that of the observed sample (32,184). The realisations depicted in Figures 3 and 4 are generated according to the estimated parameters by assuming the same values of the covariates observed in the sample data. Concerning the predicted responses to the 11 questions we note that the only items showing decreasing frequencies of trust over time are those on confidence in the European and National Parliament and in the President. Commercial banks show the highest probability of distrust.

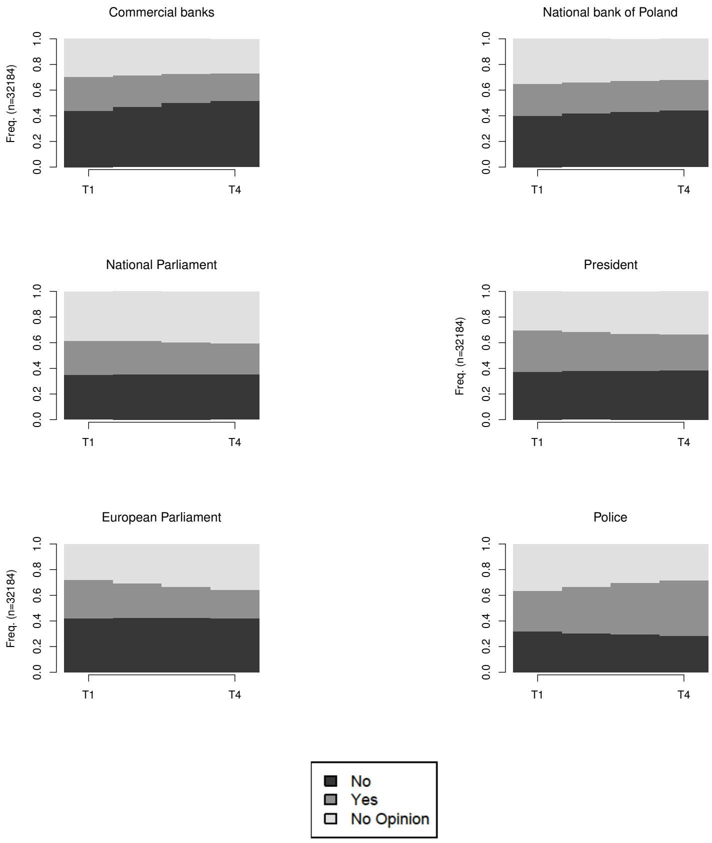

Fig. 3: *Predicted probabilities of 32,184 out of sample Poles for each response category No, Yes, No opinion of the question "Do you trust in . . . ?" referred to each institution, obtained through the estimated HMM model with covariates.*

4.3 Results for data collected on the previous waves

Data collected from 2003 to 2007 refer to responses provided by the 5,001 respondents interviewed from the second to the fourth wave of the survey. As explained in Section 2 only few questions on trust where included in the questionnaires of that period. For this reason, it is not possible to directly compare the responses with those of the previous analysis. The questions are about trust towards *Commercial banks*, *Social Insurance Institutions* and *Stock*

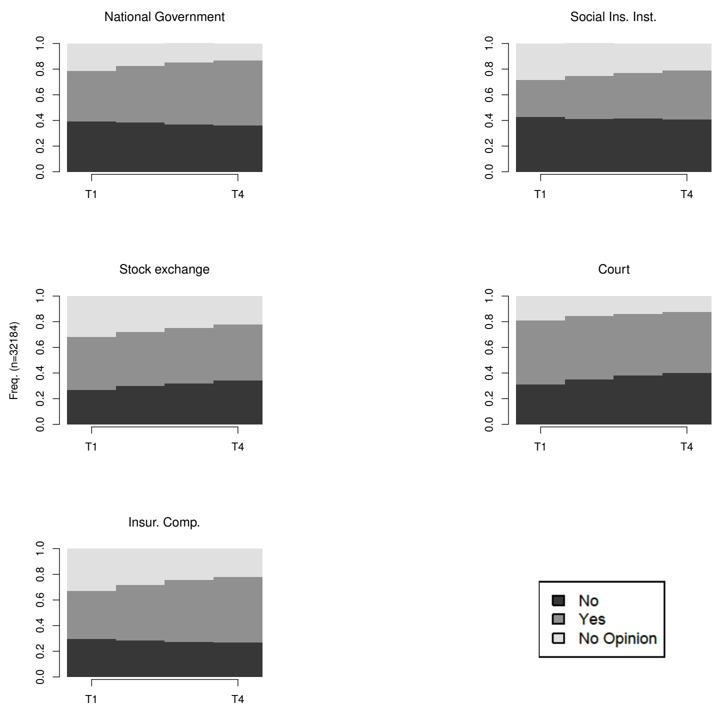

Fig. 4: *Predicted probabilities of 32,184 out of sample Poles for each response category No, Yes, No opinion of the question "Do you trust in . . . ?" referred to each institution, obtained through the estimated HMM model with covariates.*

Exchange. The response frequencies are not reporting for space limitations and we notice that in 2003 by using the corresponding sample weights the level of support towards *Commercial banks* was over 37 per cent and the lack of trust was 17 per cent whereas 32 per cent of the Poles had no opinion. The percentage of missing responses was 13 per cent. By considering the covariates listed in Table 3 we notice that the majority of respondents are married, have secondary education and they mainly live in cities below 20,000 inhabitants or in rural areas. In 2013 respondents are mainly adults with an average age close to 50 years old.

We estimated the HMM presented in Section 3 with $k = 4$ latent states accounting for t covariates, missing responses and longitudinal weights to compare the results with those obtained in Section 4 for the data collected more recently. The estimated model has a log-likelihood equal to $\hat{\ell}$= -26,489.21, BIC index equal to 54,844.65 with 219 free parameters. The clusters are labelled according to the estimated conditional probabilities as the following: "no opinion" U_{Nop}, supporting the institutions "trust" U_T, not supporting the institutions Discouraged, U_D and not supporting *Stock Exchange* and having no opinions on *Commercial banks* and *Social Insurance Institutions* denoted as $U_{D_N op}$.

In 2003 the majority of Poles (about 68%) belongs to the cluster of confident people U_T and only 11 per cent of the Poles are in the group of people not supporting the institutions, 21 per cent are in the cluster with no opinion and none in the cluster denoted as U_{DNop}. According to the estimated transition probabilities 14 per cent of individuals with no opinion U_{Nop} switch to the cluster of those without confidence or with no opinion U_{DNop}. Interestingly, 47% of Poles with trust U_T switch to the cluster of distrust and no opinion U_{DNop}; 39% of the Poles in the cluster denoted as U_{DNop} switch to the cluster of distrust U_D; 14% of the Poles in the cluster denoted as U_D, switch to the mixed group of no trust and no opinion U_{DNop}; 89 per cent of the Poles with no opinion remain in the same group, showing the highest persistence.

The transition of those in the confident group U_T to the undecided or discouraged people U_{DNop} is more evident for females with respect to males, for those having a primary education compared with higher educated individuals, for who is not working in the public or private sector, for who lives in big cities compared with other places and married individuals compared to unmarried Poles.

The transition from discouraged people U_{DNop} to confident people U_T is considerably more frequent for individuals which are females, married, with primary education living in small cities and rural areas. Showing that the opinions tends to fluctuate the more the level of education is low. The probability to switch from undecided or discouraged people U_{DNop} to those discouraged towards all the institutions U_D is higher for men, married and higher educated individuals, living in big cities and mainly working in the public sector. Considering these results and comparing them with those showed in the previous section we argue that the effort made by the governors from the 2007 have been effective to raise the rate of trust in many national institutions.

According to the predictions made on the basis of the Viterbi algorithm (as illustrated in Section 3) Figure 5 shows the decoded sequences for each individual estimated to change cluster over the years. At the end of the period only few people are predicted to stay in the cluster of those not supporting stock exchange and with no confidence in some institutions U_{DNop}. People not confident in the institutions U_D transit to U_{DNop} and at the end of the period half of them transit to U_T and half return to U_{DNop}.

5 Discussion and conclusions

The general model-based clustering approach proposed with this research to analyse multivariate categorical responses collected through surveys provides the possibility to identify and cluster similar response patterns according to different latent dimensions. The motivating example concerns trust that like other human features is a latent psychological dimension that can be mea-

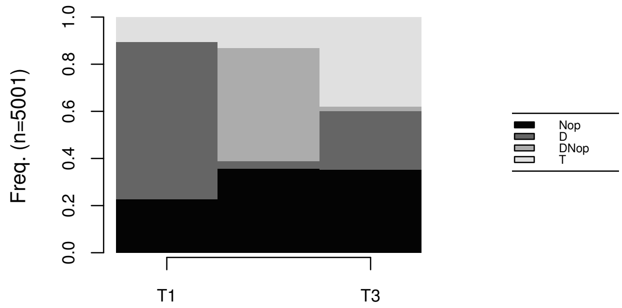

Fig. 5: *Predicted trajectories across clusters on the subset of people changing latent states within the three waves (six years) for the data collected from 2003 to 2007:* U_{Nop} *No Opinion,* U_T *Trust,* U_{DNop} *Discouraged and No opinion.*

sured only by proxies. Considering the Social Diagnosis Survey we detect the perceptions towards institutions of the Polish society and we provide a picture of the segments of the entire population to show possible changes over time.

Nowadays, the analysis of this phenomena is considered of main importance because for example politics need trust by their followers and the perceived trust is monitored frequently by politicians to assess their popularity. The level of trust towards a national institution could also be used to monitor the expected people's behaviour or the level of what is known as sovereign and nationalism both affecting many European countries. Understanding who and when manifest high public distrust permits to target specific individuals belonging to one or more clusters to raise the level of trust.

With respect to the analyses made in the literature and illustrated in the introduction that are based on the construction of composite indicators like that of Boda (2014) and Marozzi (2015) within our proposal the covariates enter into the model in an active way in order to contribute to the classification process. We are able to consider many covariates such as the professional status that relates with the perceived trust and to characterize the latent states according to them. Disposing of data from other countries within our proposal it would be possible to rank them according to the estimated posterior probabilities of belonging to each cluster in other to produce similar rankings. This issue will be addressed with further research.

With respect to the latent class model employed by Fazio et al. (2018) to analyse trust with Italian survey data the proposed multivariate hidden Markov model with respect to the latent class model allows us to estimate the regularities in perceptions and trends over time accounting for the multiple dimensions of this phenomena. In addition, the current proposal allows us to make predictions by predicting trends also for individuals with missing responses under the missing-at-random assumption.

Comparing the results obtained with the data of the initial waves with those of the last waves we notice that the Polish society is more trusted oriented. This may be due to the constant raise of gross domestic product per capita from 2000 since 2015. It may be also due to some changes made by the government, for example concerning the immigration policy aimed to restrict considerably the number of immigrants in the country over the last years.

The hidden Markov model is able to detect the trend that Polish people are becoming more selective, that they are less trustworthy towards the elites in general and they have the tendency to perform a selection of the reliable institutions. By considering the predicting probabilities the model allow to perform targeted policies to be addressed to some of the clusters according with the predicted trajectory. The estimated initial and transition probabilities showed in Tables from 8 to 12 permit us to measure the influence of the personal attributes. We notice that in the Polish society trust is more likely if people are married, work in the public sector and live in big cities. Males are more prone to retain trustworthiness with respect to females. We also notice that the main differences on the initial probabilities are related to people with primary and post secondary education. Higher educated people are less prompt to be disengagement, more prompt to trust. They are also more selective over time compared to their counter part while remaining less discouraged. Therefore, 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 at each level for all the citizens of the country independently by their age.

Special actions could be delivered to Poles belonging to the cluster of discouraged people since this group is predicted as a quarter of the population and is rather stable within this position. For example, they could be interviewed face to face to better understand the reasons of their abstention and/or they could be addressed with policies of civil cohesion. Other reforms could be addressed to unmarried people living in small cities and rural areas characterized by the highest probability of remaining distrustful. The obtained results may become the driving force for actions taken by both financial and state institutions: initiatives increasing the knowledge, and the financial awareness of Poles, strengthening the society's education regarding the provided services, defining more clear institutional procedures.

Another interesting feature of the results is related to the differences between the cluster of Poles with stable trust and those with selective trust. The second is characterized by higher probability of missing opinions towards European Parliament with respect to the first. According with the predicted values this cluster is going to enlarge in the future. Therefore, it would be desired for the European politicians to be much more close to the needs of this part of the society.

References

Albanese, G., De Blasio, G., and Sestito, P. (2013). Trust and preferences: Evidence from survey data. Temi di Discussione, Working paper, 911, Bank of Italy.

Bacci, S., Pandolfi, S., and Pennoni, F. (2014). A comparison of some criteria for states selection in the latent Markov model for longitudinal data. *Advances in Data Analysis and Classification*, 8:125–145.

Barometer, Edelman Trust (2019). Edelman trust barometer: Executive summary. *Report. Retrieved from http://www. Scribd. Com/Doc/79026497/2012-Edelman-Trust-Barometer-Executive-Summary. Accessed September.*

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

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2014). Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates (with discussion). *TEST*, 23:433–465.

Bartolucci, F., Pandolfi, S., and Pennoni, F. (2017). `LMest`: An `R` package for latent Markov models for longitudinal categorical data. *Journal of Statistical Software*, 81:1–38.

Bartolucci, F., Pennoni, F., and Vittadini, G. (2016). Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. *J. Educ. Behav. Stat/*, 41:146–179.

Baum, L., Petrie, T., Soules, 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.

Boda, Z. (2014). Does institutional trust in east central Europe differ from Western Europe? *Quarterly of Political Attitudes and Mentalities*, 3:1–17.

Brehm, J. and Rahn, W. (1997). Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science*, 41:999–1023.

Cappe, O., Moulines, O., and Ryden, T. (2009). *Inference in Hidden Markov Models*. Springer, New York.

Catterberg, G. and Moreno, A. (2005). Modelling a multivariate hidden Markov process on survey data. *International Journal of Public Opinion Research*, 18:31–48.

Cautrès, B. (2017). 'jusqu'ici tout va bien? la dèmocratie de la dèfiance en annèe èlectorale'. In *Wave 8*, pages 1–10. Paris: CEVIPOF.

Cho, W. and Matthew, F. (2007). A vicious circle of corruption and mistrust in institutions in sub-saharan africa: A micro-level analysis. *Cape Town: AfroBarometer Working Paper*, 71:1–27.

Clausen, B., Kraay, A., and Nyiri, Z. (2011). Corruption and confidence in public institutions: Evidence from a global survey. *The World Bank Economic Review*, 25:212–249.

Crozier, M., Huntington, S., and Watanuki, J. (1975). *The crisis of democracy. Report on the governability of democracies to the Trilateral Commission.*

New York University Press, New York.

Czapiński, J. (2013). Social diagnosis 2013 objective and subjective quality of life in poland-full report. *Contemporary Economics*, 7.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39:1–38.

Dotto, F., Farcomeni, A., Pittau, M., and Zelli, R. (2019). A dynamic inhomogeneous latent state model for measuring material deprivation. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 182:495–516.

Ernst, L. (1989). *Weighting Issues for Longitudinal Household and Family Estimates*. Wiley, New York.

Ersh, J., Gambetta, D., Laurie, H., Siedler, T., and Uhrig, S. (2009). Measuring people's trust. *Journal of the Royal Statistical Society A (Statistics in Society)*, 172:749–769.

ESS Round 4, E. S. S. (2008). Data file edition 4.5, norwegian centre for research data, norway. Data Archive and distributor of ESS data for ESS ERIC.

ESS Round 6, E. S. S. R. . D. (2012). Data file edition 2.3. norwegian centre for research data, norway. Archive and distributor of ESS data for ESS ERIC.

Eurobarometer, E. C. (2014). Public opinion in the european union. Report, Standard Eurobarometer Spring 81.

European Commission, D.-G. f. C. (2019). Ethics guidelines for trustworthy ai. *Report*. High-Level Expert Group on Artificial Intelligence.

Fazio, G., Giambona, F., Vassallo, E., and Vassiliadis, E. (2018). A measure of trust: The italian regional divide in a latent class approach. *Social Indicators Research*, 140:209–242.

Fisher, J., Heerde, J., and Tucker, A. (2010). Does one trust judgement fit all? linking theory and empirics. *The British Journal of Politics & International Relations*, 12:161–168.

Fukuyma, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity*. Free Press, New York.

Gambetta, D. (1988). *Trust: Making and breaking cooperative relations*. Blackwell, Oxford.

Genge, E. (2014). A latent class analysis of the public attitude towards the euro adoption in Poland. *Advances in Data Analysis and Classification*, 8:427–442.

Gille, F., Smith, S., and Mays, N. (2016). *Theory and conceptualisation of public trust in the health care system. Three English case studies: care data, biobanks and 100,000 Genomes Project*. University of London, London.

Glaeser, E., Laibson, D., Scheinkman, J., and Soutter, C. (2000). Measuring trust. *The Quarterly Journal of Economics*, 3:811–846.

Groenewegen, P., Hansen, J., and de Jond, J. (2018). Trust in times of health reform. In *Health Policy*, pages 1–10.

Hakhverdian, A. and Mayne, Q. (2012). Institutional trust, education, and corruption: a micro-macro interactive approach. *Journal of Politics*, 74:739–

750.

Hutchison, M. and Johnson, K. (2000–2005). Capacity to trust? institutional capacity, conflict, and political trust in africa. *Journal of Peace Research*, 48:737–752.

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J .and Lagos, M., Norris, P., Ponarin, E., Puranen, B., et al. (2014). World values survey: All rounds-country-pooled datafile version. *Madrid: JD Systems Institute. Retrieved August*, 1:2018.

JRC-OECD (2008). Handbook on copages. OECD Paris.

Juang, B. and Rabiner, L. (1991). Hidden Markov models for speech recognition. *Technometrics*, 33:251–272.

Kuovo, A. (2011). The source of generalized trust and institutional confidence in europe. *Research on Finnish Society*, 4:9–40.

Lavallèe, E., Razafindrakoto, M., and Roubaud, F. (2008). Corruption and trust in political institutions in sub-saharan africa. *Afrobarometer Working Paper*, 102:1–25.

Lazarsfeld, P. (1950). *The logical and mathematical foundation of latent structure analysis*. Princeton University Press.

Lazarsfeld, P. F. and Henry, N. (1968). *Latent structure analysis*. Houghton Mifflin, Boston, MA.

Li, Y., Pickles, A., and Savage, M. (2005). Social capital and social trust in britain. *European Sociological Review*, 21:109–123.

Magidson, J. and Vermunt, J. (2000). *Latent class analysis*. Cambridge University press, Cambridge.

Marien, S. (2011). *Measuring Political Trust Across Time and Space*. ECPR Press, Colchester.

Marozzi, M. (2015). Measuring trust in european public institutions. *Social Indicators Research*, 3:879–895.

McLachlan, G., Lee, S., and Rathnayake, S. (2018). Finite mixture models. *Annual Review of Statistics and Its Application*, 6:11–24.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

Mishler, W. and Rose, R. (1997). Trust, distrust and skepticism:popular evaluations of civil and political institutions in post-communist societies. *The Journal of Politics*, 59:418–451.

Muthen, B. (2002). Beyond sem: general latent variable modeling. *Behaviormetrika*, 29:81–118.

OECD (2017). Government at a glance. OECD Publishing, Paris.

Paas, L. J., Vermunt, J. K., and Bilmolt, T. H. A. (2007). Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Series A*, 170:955–974.

Paldam, M. (2008). Social capital: One or many? definition and measurement. *Journal of Economic Surveys*, 14:629–265.

Pennoni, F. (2014). *Issues on the estimation of latent variable and latent class models*. Scholars' Press, Saarbuken.

Pennoni, F. (2016). Modelling a multivariate hidden Markov process on survey data. In *Proceedings of the 48th Scientific Meeting of the Italian Statistical Society*, pages 1–10. Università degli studi di Salerno.

Pennoni, F. and Nakai, M. (2018). A latent class analysis towards stability and changes in breadwinning patterns among coupled households.

Pennoni, F. and Romeo, I. (2017). Latent Markov and growth mixture models for ordinal individual responses with covariates: a comparison. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10:29–39.

Putnam, R. (2000). *Bowling Alone-The Collapse and Revival of American Community*. Simon & Schuster, Cambridge.

Reeskens, T. and Hooghe, M. (2008). Cross-cultural measurement equivalence of generalized trust. evidence from the european social survey (2002 and 2004). *Social Indicators Research*, 85:515–532.

Rosanvallon, P. (2008). *Counter-democracy: Politics in an age of distrust*. Cambridge University Press, Cambridge.

Sapienza, P., Toldra, A., and Zingales, L. (2007). Understanding trust. In *National Bureau of Economic Research*, number 13387 in Working Paper 13387, pages 1–10, Cambridge.

Schneider, I. (2016). Can we trust measures of political trust? assessing measurement equivalence in diverse regime types. *Social Indicators Research*, 133:963–984.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. Chapman and Hall/CRC, London.

Social Diagnosis (2015). Objective and subjective quality of live in poland. Czapinski, J. and Panek,s T. (eds. Warszaa, Social Monitoring Council (15.10.2016).

Stoyan, A., Sara, N., Jana, M., Jonathan, H., and Rosario, E. (2016). Trust in government institutions: the effects of performance and participation in the dominican republic and haiti. *International Political Science Review*, 37:18–35.

Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press, Cambridge.

Trzesiok, J. (2016). Badanie zaufania do instytucji finansowych w polsce z wykorzystaniem analizy korespondencji. *Studia Ekonomiczne*, 265:80–94.

van Oorschot, W., Arts, W., and Gelissen, J. (2006). Social capital in europe. *Acta Sociologica*, 49:149–167.

Van Prooijen, J. and van Lange, P. (2014). *Power, politics, and Paranoia. Why people are suspicious of their leaders*. Cambridge University Press, Cambridge.

Vermunt, J. (2010). *Longitudinal research using mixture models*. Springer, Verlag, Berlin and Heidelberg.

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

Wang, L. and Gordon, P. (2011). The bases of political trust in six asian societies: Institutional and cultural explanations compared. *International Political Science Review*, 32:263–281.

Welch, L. (2003). Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, 53:1–13.

Wiggins, L. (1955). *Mathematical Models for the Analysis of Multi-wave Panels*. PhD thesis, Columbia University, Arbor A, MI.

Wiggins, L. (1973). *Panel Analysis: Latent probability models for attitude and behavious processes*. Elsevier, Amsterdam.

Wong, T. K., Wan, P., and Hsiao, H.-H. M. (2011). The bases of political trust in six asian societies: Institutional and cultural explanations compared. *International Political Science Review*, 3:263–281.

Ydersbond, I. (2015). Social and political trust in europe. *Report*.

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov Models for time series: an introduction using R*. Springer-Verlag, New York.

Table 1: *Weighted frequency for each item and related frequencies of missing responses over years.*

Trust in **2009** (%)	No	Yes	No opinion	Missing
Commercial banks	24.80	**44.88**	28.61	1.71
National Bank of Poland	-	-	-	100.00
National Parliament	**59.76**	8.40	30.23	1.61
President	56.20	13.84	28.27	1.69
European Parliament	31.63	15.30	**51.28**	1.79
Police	30.96	**43.33**	**23.96**	1.75
National Government	**55.84**	13.98	28.48	1.70
Social Insurance Institution	42.80	22.69	32.91	1.60
Stock exchange	-	-	-	100.00
Court	-	-	-	100.00
Insurance companies	-	-	-	100.00
Trust in **2011** (%)	No	Yes	No opinion	Missing
Commercial banks	29.01	19.95	47.14	3.90
National Bank of Poland	14.59	**46.21**	35.49	3.71
National Parliament	55.16	14.26	26.93	3.64
President	39.42	30.05	26.95	3.58
European Parliament	31.32	22.67	**42.27**	3.74
Police	28.22	**46.93**	**21.26**	3.59
National Government	54.23	17.80	24.30	3.67
Social Insurance Institution	41.41	25.79	29.14	3.68
Stock exchange	32.54	7.42	**56.35**	3.69
Court	32.44	32.84	31.13	3.59
Insurance companies	36.57	13.99	45.67	3.77
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
Trust in **2015** (%)	No	Yes	No opinion	Missing
Commercial banks	25.30	**40.74**	30.85	3.11
National Bank of Poland	11.35	**61.99**	23.69	2.97
National Parliament	**47.28**	33.18	16.64	2.90
President	32.13	**50.17**	14.76	2.94
European Parliament	34.08	**37.69**	25.28	2.95
Police	18.94	**68.32**	**9.89**	2.85
National Government	49.84	32.83	14.46	2.87
Social Insurance Institution	38.16	**41.69**	17.33	2.82
Stock exchange	26.26	15.06	**55.74**	2.94
Court	29.79	35.88	31.32	3.00
Insurance companies	9.00	**82.37**	5.67	2.96

Table 2: *Unweighted frequency distributions for the responses provided in 2013 about trust towards National Parliament and Insurance companies.*

Trust in **2013** (%)	No	Yes	No opinion	Missing
National Parliament	48.74	25.48	17.60	**8.18**
Insurance companies	28.87	23.31	39.47	**8.35**

Table 3: *Weighted frequencies of the respondents'socio-economic features according to the time occasion. The categories are the following:* gender*: 1 male, 2 female;* marital status*: 1: married, 2: other;* education*: 1 primary/no education, 2 vocational/grammar or secondary, 3 higher and post-secondary;* place of living*: 1 cities with more than 500,000 inhabitants, 2 cities with 20,000 to 500,00 inhabitants; 3 inhabitants, cities below 20,000 inhabitants and rural areas;* socio-professional status*: 1: employees in public sector, 2: employees in private sector and entrepreneur/self-employed, 3: other.*

Time-fixed covariates **2009-2015** (%)	1	2	3
gender	45.37	**54.63**	
place of living	10.17	35.51	**54.32**
Covariate in **2009** (%)			
marital status	**61.30**	38.70	
education	17.64	**63.99**	18.37
socio-professional status	13.25	25.26	**66.19**
age	*mean=45.31*	*s.d.=17.45*	
Covariate in **2011** (%)			
marital status	**61.22**	38.78	
education	16.67	**63.21**	20.13
socio-professional status	13.48	27.59	**58.93**
age	*mean=47.31*	*s.d.=17.45*	
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*	

Table 4: *Maximum log-likelihood, number of parameters, BIC index for the HMMs with latent states varying from 1 to 11.*

k	$\hat{\ell}$	#par	BIC
1	-355,609.315	22	711,416.358
2	-314,981.073	51	630,420.514
3	-303,133.099	86	607,039.133
4	-295,923.693	127	592,988.812
5	-290,625.743	174	582,815.331
6	-286,398.329	227	574,836.847
7	-282,631.280	286	567,833.017
8	-280,277.796	351	563,710.244
9	-278,191.210	422	560,175.192
10	-276,263.083	499	557,010.984
11	-274,692.256	582	554,615.302

Table 5: *Estimated conditional probabilities of the latent states given each item response category to the question* Do you trust in ...? *referred to the eleven institutions. The labels assigned to the clusters are:* U_D *distrust,* U_{Nop} *no opinion,* U_{ST} *selected trust, and* U_T *trust. In bold some values to be highlighted.*

		Latent states			
Institutions	Categories	U_D	U_{Nop}	U_{ST}	U_T
Commercial banks	No	0.6244	0.3267	**0.5902**	0.3844
	Yes	0.2001	0.1530	0.1285	0.3985
	No opinion	0.1755	0.5203	0.2813	0.2171
National Bank of Poland	No	0.6210	0.2846	**0.4285**	0.3097
	Yes	0.1558	0.1600	0.2429	0.4003
	No opinion	0.2232	0.5554	0.3286	0.2900
National Parliament	No	0.5755	0.1691	0.2717	0.2939
	Yes	0.1328	0.0820	0.2027	0.6500
	No opinion	0.2918	**0.7488**	**0.5257**	0.0561
President	No	0.6415	0.2384	0.2814	0.2599
	Yes	0.1779	0.1484	0.2258	0.5756
	No opinion	0.1806	0.6132	**0.4928**	0.1645
European Parliament	No	**0.6991**	0.2683	0.2768	0.3209
	Yes	0.1583	0.1715	0.0488	0.5359
	No opinion	0.1426	0.5602	**0.6744**	0.1432
Police	No	0.5602	0.1630	0.0594	0.2513
	Yes	0.2227	0.1732	**0.7647**	0.4923
	No opinion	0.2171	0.6638	0.1759	0.2564
National Government	No	**0.7575**	0.2540	0.1319	0.1962
	Yes	0.1816	0.2565	**0.8427**	0.6929
	No opinion	0.0609	0.4895	0.0254	0.1109
Social Insurance Institution	No	**0.7513**	0.2713	0.2103	0.2760
	Yes	0.1252	0.1780	**0.6797**	0.5080
	No opinion	0.1235	0.5507	0.1100	0.2160
Stock exchange	No	0.4742	0.1436	**0.5101**	0.1508
	Yes	0.3426	0.2481	0.4144	0.6503
	No opinion	0.1832	**0.6083**	0.0755	0.1989
Court	No	0.5943	0.2080	**0.5462**	0.1242
	Yes	0.3289	**0.3713**	0.3977	**0.7943**
	No opinion	0.0768	0.4207	0.0561	0.0815
Insurance companies	No	0.5178	0.1645	0.0664	0.2248
	Yes	0.2702	0.2815	**0.9239**	0.5156
	No opinion	0.2120	0.5540	0.0097	0.2596

Table 6: *Estimated initial probabilities* $\hat{\pi}_1$.

	U_D	U_{Nop}	U_{ST}	U_T
$\hat{\pi}_1$	0.333	0.319	0.001	0.347

Table 7: *Estimated transition probabilities from $\bar{u}$ (row) to u (column).*

$\hat{\pi}_{u\|\bar{u}}$	U_D	U_{Nop}	U_{ST}	U_T
U_D	0.8464	0.0107	**0.1397**	0.0031
U_{Nop}	0.0006	0.8599	**0.1382**	0.0013
U_{ST}	0.0009	0.0202	**0.9760**	0.0029
U_T	0.0012	0.0224	**0.2091**	0.7672

Table 8: *Estimated initial ($\hat{\pi}_1$, upper panel) and transition probabilities ($\hat{\pi}_{u|\bar{u}}$, bottom panel) from $\bar{u}$ (row) to u (column) according to the educational level (*primary vs post-secondary education*).*

	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**

Table 9: *Estimated initial ($\hat{\pi}_1$, upper panel) and transition probabilities ($\hat{\pi}_{u|\bar{u}}$, bottom panel) from $\bar{u}$ (row) to u (column) according to the professional status (*public sector vs inactive*).*

	Public sector				Professionally inactive			
	U_D	U_{Nop}	U_{ST}	U_T	U_D	U_{Nop}	U_{ST}	U_T
	0.3531	0.2275	0.0005	**0.4189**	0.3149	**0.3600**	0.0003	0.3248
U_D	0.8399	0.0019	**0.1571**	0.0011	**0.8511**	0.0118	0.1330	0.0042
U_{Nop}	0.0001	0.8513	**0.1427**	0.0059	0.0009	**0.8623**	0.1360	0.0008
U_{ST}	0.0003	0.0085	**0.9909**	0.0003	0.0013	0.0296	0.9648	0.0043
U_T	0.0001	0.0106	0.2094	**0.7798**	0.0015	**0.0309**	**0.2126**	0.7550

Table 10: *Estimated initial ($\hat{\pi}_1$, upper panel) and transition probabilities ($\hat{\pi}_{u|\bar{u}}$, bottom panel) from $\bar{u}$ (row) to u (column) according to gender (*females vs males*).*

	Females				Males			
	U_D	U_{Nop}	U_{ST}	U_T	U_D	U_{Nop}	U_{ST}	U_T
	0.2848	**0.3611**	0.0015	0.3526	**0.3900**	0.2691	0.0000	0.3409
U_D	0.8422	0.0132	**0.1446**	0.0000	0.8514	0.0078	0.1340	0.0068
U_{Nop}	0.0000	0.8662	0.1338	0.0000	0.0014	0.8525	**0.1433**	0.0028
U_{ST}	0.0000	**0.0361**	0.9585	0.0054	0.0020	**0.0017**	0.9963	0.0000
U_T	0.0006	0.0322	**0.2159**	**0.7513**	0.0020	0.0111	0.2011	**0.7858**

Table 11: *Estimated initial ($\hat{\pi}_1$, upper panel) and transition probabilities ($\hat{\pi}_{u|\bar{u}}$, bottom panel) from $\bar{u}$ (row) to u (column) according to the marital status (*married vs unmarried*).*

	Unmarried				Married			
	U_D	U_{Nop}	U_{ST}	U_T	U_D	U_{Nop}	U_{ST}	U_T
	0.3106	**0.3801**	0.0019	0.3074	0.3478	0.2796	0.0001	**0.3725**
U_D	0.8563	0.0090	1339	0.0008	0.8401	0.0118	**0.1435**	0.0046
U_{Nop}	0.0016	0.8753	0.1230	0.0001	0.0000	0.8501	**0.1479**	0.0021
U_{ST}	0.0000	**0.0488**	**0.9511**	0.0001	0.0015	**0.0020**	**0.9918**	0.0047
U_T	0.0000	0.0164	0.2047	0.7789	0.0020	0.0263	**0.2119**	0.7598

Table 12: *Estimated initial ($\hat{\pi}_1$, upper panel) and transition probabilities ($\hat{\pi}_{u|\bar{u}}$, bottom panel) from $\bar{u}$ (row) to u (column) according to the place of living covariate (*big cities vs small cities*).*

	Big cities				Small cities and rural areas			
	U_D	U_{Nop}	U_{ST}	U_T	U_D	U_{Nop}	U_{ST}	U_T
	0.3600	0.1351	0.0000	**0.5049**	0.3247	**0.3625**	0.0013	0.3115
U_D	0.8644	0.0000	0.1356	0.0000	0.8450	0.0110	**0.1435**	0.0005
U_{Nop}	0.0000	0.8086	**0.1914**	0.0000	0.0000	0.8696	0.1304	0.0000
U_{ST}	0.0000	0.0005	**0.9995**	0.0000	0.0014	0.0309	0.9631	0.0046
U_T	0.0000	0.0090	0.2214	0.7696	0.0019	0.0287	0.2112	0.7581

Table 13: *Observed responses provided by a married man of middle age (50 years old), with secondary education, living in a small town, and working in the private sector, with 0 "no I don't trust", 1 "yes I trust" , 2 "I have no opinion" to the eleven institutions in Table 5 and predicted cluster memberships* $\hat{\tilde{\boldsymbol{u}}}(\boldsymbol{y})$.

$\boldsymbol{y}_{it}$	**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}