A longitudinal cross country comparison of migrant integration policies via Mixture of Matrix-Normals

Salvatore Leonardo Alaimo, Francesco Amato, Emiliano Seri

A longitudinal cross country comparison of migrant integration policies via Mixture of Matrix-Normals

Un confronto longitudinale tra paesi sulle politiche di integrazione degli immigrati attraverso Mixture of Matrix-Normals

Leonardo Salvatore Alaimo, Francesco Amato and Emiliano Seri

Abstract In recent decades, there has been a growing research interest in comparative studies of migrant integration, assimilation and the evaluation of policies implemented for these purposes. With this aim, the Migrant Integration Policy Index (MIPEX) measures policies to integrate migrants in 52 countries all over the world, over time. However, the comparison of very different countries on complex and multidimensional phenomena can lead to misleading interpretations and evaluations of the results. In this paper, we improve and facilitate the comparison between the treated countries on 7 MIPEX dimensions, applying a Mixture of Matrix-Normals classification model for longitudinal data. Trough the analysis, 5 clusters of countries have been discovered, allowing us to add new levels of interpretation of the data.

Abstract *Negli ultimi decenni, c'è stato un crescente interesse di ricerca negli studi comparativi sull'integrazione dei migranti, l'assimilazione e la valutazione delle politiche attuate per questi scopi. Con questo obiettivo, il Migrant Integration Policy Index (MIPEX) misura le politiche di integrazione dei migranti in 52 paesi di tutto il mondo nel tempo. Tuttavia, il confronto tra paesi molto diversi su fenomeni complessi e multidimensionali può portare a interpretazioni e valutazioni fuorvianti dei risultati. In questo articolo miglioriamo e facilitiamo il confronto tra i paesi trattati su 7 dimensioni MIPEX, applicando un modello di classificazione Mixture of Matrix-Normals per dati longitudinali. Attraverso l'analisi sono stati individuati 5 cluster di paesi, permettendoci di aggiungere nuovi livelli di interpretazione dei dati.*

Leonardo Salvatore Alaimo
Department of Social Sciences and Economics, Sapienza University of Rome, e-mail: leonardo.alaimo@uniroma1.it

Francesco Amato
Univ Lyon, Univ Lyon 2, ERIC, Lyon, e-mail: francesco.amato@univ-lyon2.fr

Emiliano Seri
Department of Statistic, Sapienza University of Rome, e-mail: emiliano.seri@uniroma1.it

Key words: Mixture of matrix-normals, MIPEX, Model-based classification, Migration policies

1 Introduction

Immigration regulation and immigrant assimilation have been a salient political issue in all industrialised countries for many decades [1]. The growing interest in comparative analyses of immigration has led to a variety of attempts to quantify immigration policies, i.e. to assess and put into numerical form what countries are doing to foster the integration and assimilation of immigrants. However, the study of these phenomena from a quantitative point of view is rather recent, due to the previous lack of data. Moreover, quantifying migrant integration is a difficult challenge, due to its complex nature and lack of uniformity in migration policies of many countries, which is based on multiple criteria. In the present work, we focus on the Migrant Integration Policy Index (MIPEX) [1]. The project informs and engages key policy actors about how to use indicators to improve integration governance and policy effectiveness. Its aim is to measure policies that promote integration in both social and civic terms, evaluating trough a survey the migration policies of each considered country, to construct a multi indicator system, first aggregated in 8 dimensions, each aggregated in one single composite indicator. The aim of this paper is to add new perspectives on the MIPEX data while respecting the complexity of the phenomenon under consideration by discovering structures and patterns in the behaviour of the considered countries. The research question from which this paper starts is:

- *Given the complexity of the phenomenon under consideration, in order to improve the comparison between the surveyed countries, is it possible to identify homogeneous groups over time among them, i.e. groups of countries which behave similarly across and within time?*

To answer this question, a *Finite Mixture of Matrix-Normals model* has been applied to cluster the units, taking into account the time dimension. The MIPEX includes 52 countries and collects data from 2007 to 2019, in order to provide a view of integration policies across a broad range of differing environments. It considers a system of 58 indicators (for more information, please consult [1]) covering 8 policy areas that have been designed to benchmark current laws and policies against the highest standards through consultations with top scholars and institutions[1]. The policy areas of integration covered by the MIPEX are the following:

[1] The highest standards are drawn from Council of Europe Conventions, European Union Directives and international conventions (for more information see: http://mipex.eu/methodology)

- Labour Market Mobility
- Family Reunion
- Education
- Political Participation
- Long-term Residence
- Access to Nationality
- Anti-discrimination
- Health[2]

Each dimensional synthetic indicator is bounded between $[0,100]$, in which the maximum of 100 is awarded when policies meet the highest standards for equal treatment. These values are chosen by experts from each country, by means of a questionnaire. The analysis carried out in the present work uses the listed above dimensions[3] excluding health.

2 Mixture of Matrix-Normals

Finite Mixture of Matrix-Normals (MNN), as introduced in [2], can be a useful tool to cluster time-dependent data. Let $\mathbf{Y} = \{Y_i\}_{i=1}^{N}$ be a sample of $J \times T$-variate matrix observations (i.e. $Y_i \in R^{J\times T}$), arose from studies with J-variate vector observations measured repeatedly over T time points, as in a longitudinal study case[4]. Assume that each Y_i follows a matrix-normal distribution, $Y_i \sim \mathscr{MN}_{(J\times T)}(M,\Phi,\Omega)$,where $M \in R^{J\times T}$ is the matrix of means, $\Phi \in R^{T\times T}$ is a covariance matrix containing the variances and covariances between the T occasions or times and $\Omega \in R^{J\times J}$ is the covariance matrix containing the variance and covariances of the J variables. The matrix-normal probability density function (pdf) is given by

$$f(Y \mid M,\Phi,\Omega) = (2\pi)^{-\frac{TJ}{2}} \mid \Phi \mid^{-\frac{J}{2}} \mid \Omega \mid^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}[\Omega^{-1}(Y-M)\Phi^{-1}(Y-M)]\right\} \tag{1}$$

The matrix-normal distribution is a natural extension of the multivariate normal distribution, since if $Y \sim \mathscr{MN}_{(J\times T)}(M,\Phi,\Omega)$, then $\mathrm{vec}(Y) \sim \mathscr{MVN}_{TJ}(\mathrm{vec}(M), \Phi \otimes \Omega)$, where $\mathrm{vec}(.)$ is the vectorization operator and $\otimes$ denotes the Kronecker product. Being a special case of the multivariate normal distribution, the matrix-normal distribution shares the same various properties, like, for instance, closure under marginalization, conditioning and linear transformations [3]. The pdf of the MMN model is given by

$$f(Y \mid \pi,\Theta) = \sum_{k=1}^{K} \pi_k \phi^{(J\times T)}(Y \mid M_k,\Phi_k,\Omega_k) \tag{2}$$

[2] Health data are only available for years 2014 and 2019, therefore this dimension could not be used in the analysis

[3] An extensive explanation of the MIPEX dimensions is given in [1]

[4] The three-way data time arrays analysed are represented as: $\mathbf{Y} \equiv \{y_{ijt} : i = 1,\ldots,N; j = 1,\ldots,J; t = 1,\ldots,T\}$, where $i = 1,2,\ldots,52$ indicates the generic country, $j = 1,2,\ldots,7$ the generic MIPEX dimension and $t = 2014,2015,\ldots,2019$ the generic year.

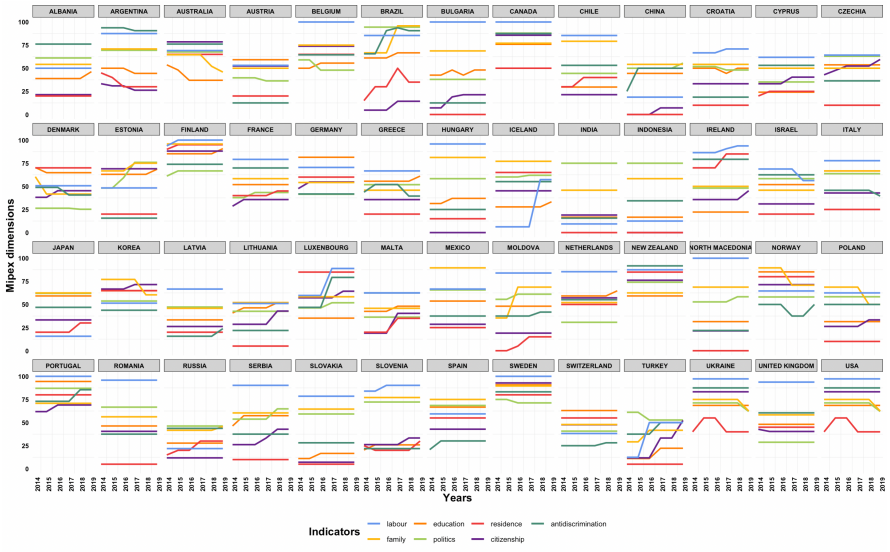
where K is the number of mixture components, $\pi = \{\pi_k\}_{k=1}^{K}$ is the vector of mixing proportions, subject to constraint $\sum_{k=1}^{K} \pi_k = 1$ and $\Theta = \{\Theta_k\}_{k=1}^{K}$ is the set of component-specific parameters with $\Theta_k = \{M_k, \Phi_k, \Omega_k\}$.

In [4], carrying forward from the previous papers, the over-parametrization issue is addressed. For identifiability issues of the model, the determinant of the time-covariance matrix must be restricted to be $\mid \Phi_k \mid = 1$.

3 Results

Figure 1 outline that most of the countries does not change much the values of their indicators through time. Following, the MNN will be used to model together the changes between and within time, grouping together the units which behave similarly across and within time[5].

Fig. 1 Country trajectories of the 7 MIPEX dimensions over time

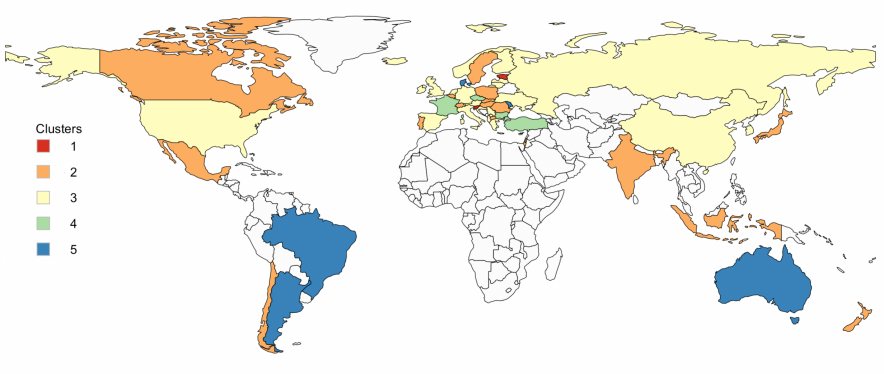


Since our dataset is composed by 52 units, we carried out the MMN model for K ranging from 1 to 8, and choose the best number of clusters by means of the BIC. The selected K is 5. According to BIC, the selected parametrization of the model is A-VEV-VV, which means that the means are better parsimoniously parametrized in additive way, Ω_k with varying volume, equal shape and varying orientation and

[5] The analyzes have been carried out using the package of the software R, MatTransMix [4]

Φ_k has both varying shape and orientation. Because of the matrices Φ_k Ω_k, each MMN component models not only the conditional means, but also covariances of the response variables and the covariances among times. In this way, each cluster provides a broad profile of units belonging to it. We report graphically the countries that belongs to each cluster in Figure 2 and the interpretation of the results is as follows:

Fig. 2 MIPEX dimensional indices: MMN clusters' composition of MIPEX countries. Years 2014 – 2019



- **Cluster 1:** *Estonia and Slovenia*. This cluster presents lower correlations in time between the first three years (2014-2016) and the second ones (2017-2019). Moreover, it has negative correlation between labour and the other dimensions, except for family reunification policies. Countries in this cluster have the lowest score for the access to citizenship and rank low for political participation as well, while ranking average for labour mobility, educational policies and high for family reunification, long-term residence and anti-discrimination legislation.
- **Cluster 2:** *Belgium, Canada, Chile, Hungary, India, Indonesia, Israel, Japan, Mexico, New Zealand, North Macedonia, Poland, Portugal, Romania, Slovakia, Sweden, Switzerland*. During the study period, countries belonging to this cluster did not change much their policies, and countries that rank high in some areas tend to rank high in the others as well. The countries of this group tend to have good policies for long-term residency, family reunification and anti-discrimination, but rank low for education and political participation.
- **Cluster 3:** *Albania, Austria, China, Croatia, Cyprus, Finland, Germany, Greece, Iceland, Ireland, Italy, Korea, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Russia, Serbia, Spain, Ukraine, UK, USA*. Countries in this cluster represents the group that reformed less their immigration legislation during the study period. They tend to rank average in most of the policies areas, with the exception of residence and anti-discrimination laws, where they tend to rank

higher. However, low correlation among variables signals that countries do not move homogeneously among the policies areas.

- **Cluster 4:** *Bulgaria, Czech Republic, France, Turkey*. Despite ranking generally high for anti-discrimination policies, countries within this cluster tend to rank low for policies in education, access to citizenship and labour market mobility, while scoring average for long-residence legislation. Yet, low correlation among variables indicates that countries do not move homogeneously among the dimensions, with the exception of policies regarding access to long-term residence and anti-discrimination, that have high positive correlation. These countries have seen their score moderately changing in time, indicating that some changes in the legislation have happened.
- **Cluster 5:** *Argentina, Australia, Brazil, Denmark, Moldova*. Countries belonging to this cluster have high-correlation in time, but they tend to decrease faster with time, meaning that some changes in the policies have been made especially in the last years. They generally rank low in policies related to educational support for foreign pupils and political participation, average for legislation related to access to labour market and high in family reunion, residence, access to citizenship and anti-discrimination. The policies' dimensions have low correlation, meaning that the countries tend not to move homogeneously among them.

4 Conclusions

This paper has explored immigrant integration policies, analyzing 7 dimensions of the MIPEX from the year 2014 to 2019, to identify groups of units with similar behaviour, to improve the ease of reading of the phenomenon. We addressed this issue trough the application of an unsupervised approach to clustering for longitudinal data namely Mixture of Matrix-Normals model, that accounts simultaneously for the within and between time dependency structures. The analysis, allowed the comparison of clusters with each other and of the countries within each cluster. Also, the correlations in time shown the general trend of each indicator over time in each cluster, and the correlations between variables purified from time effect shown the behaviour of each indicator in relation to the others within each cluster.

References

1. Giacomo Solano and Thomas Huddleston. Migrant integration policy index 2020. *Barcelona Center for International Affairs (CIDOB)*, 2020.
2. Cinzia Viroli. Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, 21(4):511–522, 2011.
3. A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 2000.
4. Xuwen Zhu, Shuchismita Sarkar, and Volodymyr Melnykov. MatTransMix: an R Package for Matrix Model-Based Clustering and Parsimonious Mixture Modeling. *Journal of Classification*, pages 1–24, 2021.