IES 2022 Innovation & Society 5.0: Statistical and Economic Methodologies for Quality Assessment

BOOK OF SHORT PAPERS

Editors: Rosaria Lombardo, Ida Camminatiello and Violetta Simonacci

Book of Short papers
10th International Conference **IES 2022**
Innovation and Society 5.0: Statistical and Economic Methodologies for Quality Assessment

Department of Economics, University of Campania "L. Vanvitelli",
January 27th - 28th 2022

Società Italiana di Statistica

PROFESSIONAL KNOWLEDGE EMPOWERMENT

Scientific Committee of the group of the Italian Statistical Society on Statistics for the Evaluation and Quality of Services – SVQS

Pietro Amenta -University of Sannio
Matilde Bini -European University of Roma
Luigi D'Ambra -University of Naples "Federico II"
Maurizio Carpita -University of Brescia
Paolo Mariani -University of Milan "Bicocca"
Marica Manisera -University of Brescia
Monica Palma -University of Salento
Pasquale Sarnacchiaro -University of Rome Unitelma Sapienza

Program Committee of the conference IES 2022

Chair: Rosaria Lombardo -University of Campania "L. Vanvitelli"
Fabio Bacchini -ISTAT
Laura Baraldi -University of Campania "L. Vanvitelli"
Eric Beh -University of Newcastle, Australia
Wicher Bergsma -The London School of Economic and Political Science, UK
Enrico Bonetti -University of Campania "L. Vanvitelli"
Eugenio Brentari -University of Brescia
Clelia Buccico -University of Campania "L. Vanvitelli"
Rosalia Castellano -University of Naples "Parthenope"
Ida Camminatiello -University of Campania "L. Vanvitelli"
Carlo Cavicchia -University of Rotterdam, The Netherlands
Enrico Ciavolino -University of Salento
Corrado Crocetta -University of Foggia
Claudio Conversano -University of Cagliari
Antonello D'Ambra -University of Campania "L. Vanvitelli"
Antonio D'Ambrosio -University of Naples "Federico II"
Alfonso Iodice D'Enza -University of Naples "Federico II"
Tonio Di Battista -University of Chieti "G. D'Annunzio"
Michele Gallo -University of Naples "L'Orientale"
Francesco Gangi -University of Campania "L. Vanvitelli"
Michele La Rocca -University of Salerno
Amedeo Lepore -University of Campania "L. Vanvitelli"
Riccardo Macchioni -University of Campania "L. Vanvitelli"
Filomena Maggino -University of Rome "La Sapienza"
Angelos Markos -Demokritus University of Thrace, Greece
Lucio Masserini -University of Pisa
Stefania Mignani -University of Bologna
Nicola Moscariello -University of Campania "L. Vanvitelli"
Francesco Palumbo -University of Naples "Federico II"
Alessandra Petrucci -University of Florence
Alessio Pollice -University of Bari

Editors

Rosaria Lombardo - University of Campania "L. Vanvitelli", Italy
Ida Camminatiello - University of Campania "L. Vanvitelli", Italy
Violetta Simonacci - University of Naples "L'Orientale", Italy

PKE - Professional Knowledge Empowerment s.r.l.
Sede legale: Villa Marelli - Viale Thomas Alva Edison, 45 - 20099 Sesto San Giovanni (MI)
Sede operativa: Villa Marelli - Viale Thomas Alva Edison, 45 - 20099 Sesto San Giovanni (MI)
Ufficio Di Rappresentanza: Via Giacomo Peroni, 400 - 00131 Roma (RM)
CF / P.I. 03167830920 — www.pke.it; e-mail info@pke.it — Privacy

February 2022 PKE s.r.l.
ISBN 978-88-94593-35-8 on print
ISBN 978-88-94593-36-5 online

Migrant Integration Policy Index (MIPEX): an analysis of countries via Gaussian mixture model-based clustering

Migrant Integration Policy Index (MIPEX): un'analisi internazionale attraverso un modello a mistura di gaussiane

Emiliano Seri, Leonardo Salvatore Alaimo, Enrico Di Bella, Rosanna Cataldo, Alfonso Piscitelli

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), that measures policies to integrate migrants in 52 countries, has established itself as a solid reference on the subject over the years. In this work, we improve and facilitate the comparison between the treated countries by the application a Gaussian mixture model-based cluster analysis on the 8 MIPEX dimensions.
Abstract *Negli ultimi decenni, c'è stato un crescente interesse della ricerca per gli studi comparativi sull'integrazione e l'assimilazione dei migranti e la valutazione delle politiche attuate per questi scopi. Con questo obiettivo, il Migrant Integration Policy Index (MIPEX), che misura le politiche di integrazione dei migranti in 52 paesi, si è affermato negli anni come un solido riferimento sull'argomento. In questo lavoro, miglioriamo e facilitiamo il confronto tra i paesi trattati attraverso l'applicazione di una cluster analysis basata sul modello di mistura di Gaussiane sulle 8 dimensioni del MIPEX.*

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

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

Rosanna Cataldo
Department of Social Sciences, University of Naples "Federico II", e-mail: rosanna.cataldo2@unina.it.

Enrico di Bella
Department of Political Sciences, University of Genoa, e-mail: enrico.dibella@unige.it.

Alfonso Piscitelli
Department of Agricultural Sciences, University of Naples "Federico II", e-mail: alfonso.piscitelli@unina.it.

Key words: Migrant Integration Policy Index, Model based clustering, Finite mixture models

1 Introduction

Immigration regulation and immigrant assimilation have been a salient political issue in all industrialised countries since many decades. The growing interest in comparative analyses of immigration has led to a variety of attempts to quantify immigration policies and to build indices. 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 the lack of uniformity in the migration policy of many countries, which is based on multiple criteria. In a cross-country setting, to evaluate and compare what governments are doing to promote the integration of migrants, the Migrant Integration Policy Index (MIPEX) [1] has become a solid and useful tool . The project informs and engages key policy actors about how to use indicators to improve integration governance and policy effectiveness. For this purpose, the project identifies and measures integration policies and identifies the links between the latter, outcomes and public opinion, drawing on international scientific studies. Its aim is to measure policies that promote integration in both social and civic terms. In socio-economic terms, migrants must have equal opportunities to lead just as dignified, independent and active lives as the rest of the population. In civic terms, all residents can commit themselves to mutual rights and responsibilities on the basis of equality. The MIPEX includes 52 countries and collects data from 2007 to 2020, 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] using and conducting comparative research in their area of expertise. The policy areas of integration covered by the MIPEX are the following:

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

For each area, a synthetic measure (dimensional) is calculated as an arithmetic mean of the elementary indicators, i.e. those selected for measuring each policy area. Each dimensional synthetic indicator is bounded $[0, 100]$, in which the maximum of 100 is awarded when policies meet the highest standards for equal treatment. These

[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)

[2] Health data are only available for years 2014 and 2019

values are chosen by experts from each country, by means of a questionnaire. Although not without its critics, MIPEX has become a reference for comparative studies on migrant integration over the last decade. 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 countries considered, is it possible to identify homogeneous groups among them?.*

To answer this question, we applied a *Gaussian mixture model-based clustering* on the 52 considered countries for the eight dimensions of the MIPEX.

2 Methods

Given n independent observations identically distributed, $\mathbf{x} \equiv \{x_1, x_2, \ldots, x_n\}$, the distribution of each of them can be specified by a probability density function by means of a finite mixture model of G components as follows:

$$f(\mathbf{x}|\boldsymbol{\Psi}) = \sum_{k=1}^{G} \omega_k f_k(\mathbf{x}|\theta_{\mathbf{k}}) \tag{1}$$

where

$$\boldsymbol{\Psi} = \{\omega_1, \omega_2, \ldots, \omega_{G-1}; \theta_1, \ldots, \theta_G\}$$

are the parameters of the mixture model; $f_k(\mathbf{x_i}|\theta_{\mathbf{k}})$ is the k^{th} component density for observation x_i with parameter vector $\theta_{\mathbf{k}}$, $\{\omega_1, \omega_2, \ldots, \omega_{G-1}\}$ are the mixing weights, under the constraints:

$$\omega_j > 0 \quad \text{and} \quad \sum_{j=1}^{G} \omega_j = 1 \quad j = 1, \ldots, k$$

and G is the number of mixture components. Assuming that G is fixed, the mixture model parameters $\boldsymbol{\Psi}$ are usually unknown and must be estimated. Most applications assume that all component densities arise from the same parametric distribution family. A popular model is the Gaussian mixture model (GMM), which assumes a (multivariate) Gaussian distribution for each component ($f_k(\mathbf{x}|\theta_{\mathbf{k}}) \sim N(\mu_{\mathbf{k}}, \boldsymbol{\Sigma_k})$). Thus, a GMM is a weighted sum of G Gaussian component densities [2]:

$$f(\mathbf{x}|\boldsymbol{\Lambda}) = \sum_{i=1}^{G} \omega_i f_i(\mathbf{x}|\mu_i, \Sigma_i) \tag{2}$$

where $\boldsymbol{\Lambda} = \{\omega_i, \mu_i, \Sigma_i\}, i = 1 \ldots M$ are the model parameters, $\omega_i : i = 1, \ldots, M$ are the mixture weights, and $f_i(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i : i = 1, \ldots, M$, is the i^{th} Gaussian component density. A generic component density f_i is a D-variate Gaussian function of the form:

$$f_i(\mathbf{x}|\mu_i|\Sigma_i) = \frac{1}{(2\pi)^{D/2}|\Sigma_i|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)'\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right\} \qquad (3)$$

where μ_i is the mean and Σ_i the covariance matrix. The complete GMM is parameterized by the mean vector, the covariance matrices and the mixture weights for all component densities. Model parameters are estimated by using the iterative Expectation-Maximization (EM) algorithm [3]. A GMM based clustering [4] allows to use a mixture of Gaussians to represent a population formed by G groups with weights $\omega_1, \omega_2 \dots \omega_G$. An observation x can be classified into one of the G groups by computing the posterior probabilities:

$$p(g|x) = \frac{p_g f_g(x)}{\sum_h p_h f_h(x)}, \quad g = 1, \dots, G \qquad (4)$$

where $f(x)$ is the Gaussian density function. In GMM clustering approach, clusters are ellipsoidal, centered at the mean vector $\mu_\mathbf{i}$, and with other geometric features, such as volume, shape and orientation, determined by the covariance matrices $\boldsymbol{\Sigma_i}$. The choice of the optimal model and the optimal number of clusters is unsupervised and it is made according to the Bayesian information criterion (BIC) [6].

3 Application and results

As mentioned in Section 1, we proceed to analyse and cluster the data of the 8 MIPEX dimensional indicators. Figure 1 shows some useful descriptive statistics: above the main diagonal the Pearson's linear correlation coefficients (the correlation font is scaled by the size of the absolute correlation) and their significance level with confidence level $\alpha = 0.05$ are reported; on the main diagonal histograms and densities plot; below the main diagonal scatter plots and correlation ellipses[3]. The correlations show the relationships between the eight dimensions considered; the histograms and density lines give a graphical view of the form of the distributions of each indicator, while the scatterplots show graphically the relationships between each pair of indicators and their dispersion. We proceed to cluster the 52 countries considered by using a model-based approach via Gaussian mixture models. We choose a model-based approach to clustering, because we prefer an unsupervised approach, i.e. where choices such as the number of clusters, their shape and size and how clusters are assumed to differ, are made through inferential statistical methods (BIC coefficient). Among the possible distributional assumptions on the data, we focus on mixture of multivariate Gaussian densities for its ability to approximate the density function of any unknown distribution [7]. The clustering is computed via the *Mclust* package [5] of the **R** statistical software: the model selected is that with variable volume, equal shape and equal orientation (for details,

[3] The ellipse represents a level curve of the density of a bivariate normal with the matching correlation

Migrant Integration Policy Index

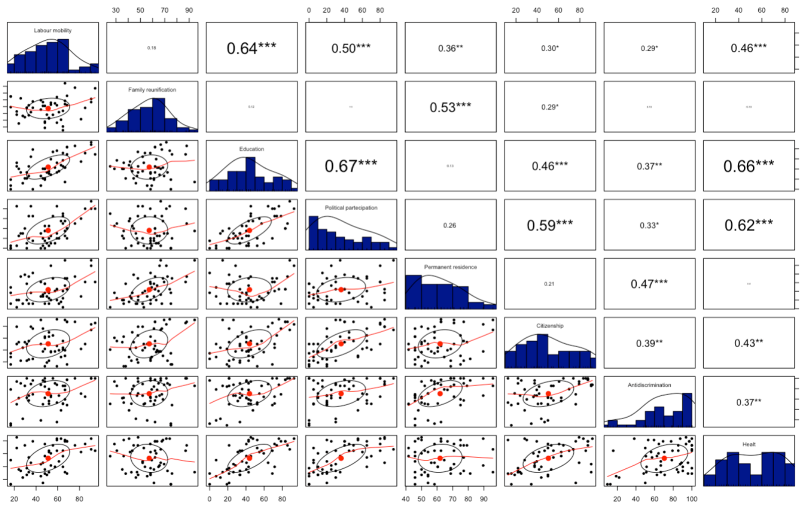

Fig. 1 Scatterplots, correlation ellipses, histograms, density plots and Pearson's linear correlation coefficients of the 8 dimensions of MIPEX.

please see: [5]) and 4 components (clusters). Table 1 reports the indicators' means of each component and the number of units.

Table 1 Means of components and number of units.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Labour market mobility	26.39	65.60	53.33	39.41
Family reunification	62.98	65.39	52.42	57.43
Education	9.50	76.08	42.12	25.17
Political participation	7.48	70.62	34.95	11.19
Permanent residence	51.04	68.51	59.57	64.27
Citizenship	35.93	78.80	48.80	28.35
Antidiscrimination	15.62	91.61	62.96	84.96
Healt	18.22	73.90	55.88	35.47
Number of units	4	12	25	11

Figure 2 shows the subdivision of the countries according to the cluster to which they belong. Cluster 1 comprises 4 countries: China, India, Indonesia and Russia. This cluster represents countries with the lowest level of integration policies for migrants. Cluster, including 12 countries, is the one of the "best integration", i.e. that present the higheest values in all the indicators. 25 countries are in the Cluster 3. This cluster group up the countries with average performances in all the indicators. The 11 remaining countries (Bulgaria, Croatia, Hungary, Latvia, Moldova, North Macedonia, Poland, Romania, Serbia, Slovakia, Slovenia) are classified in Cluster

4, characterised by low values in Political participation and Citizenship, but high in Permanent residence and the highest in Anti-discrimination.

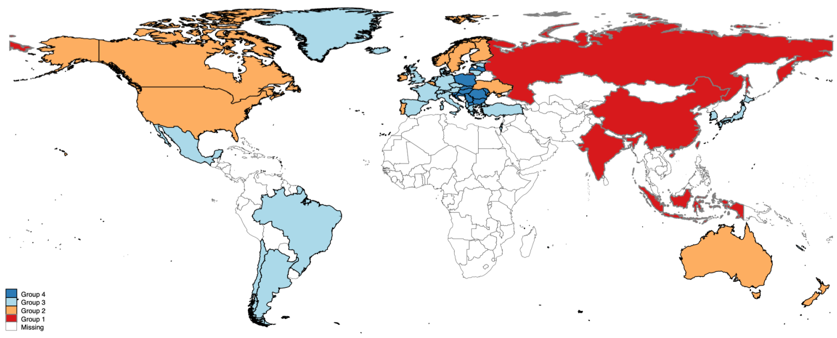

Fig. 2 MIPEX dimensional indices: clusters' composition of countries. Year 2019.

4 Conclusion

MIPEX aims to allow comparison of migration policies between countries. However, given the complex nature of the phenomenon analysed, the classification by means of Gaussian mixture model-based clustering, has made it possible to improve the reading of the results and therefore a better comparison and evaluation of the performance of the countries considered, for the 8 dimensions of MIPEX.

References

1. G. Solano, and T. Huddleston: "Migrant Integration Policy Index 2020.": 259. https://www.mipex.eu/. (2020).
2. D. A. Reynolds: "Gaussian mixture models." Encyclopedia of biometrics **741**: 659-663. (2009).
3. A. P. Dempster, N. M. Laird, and D. B. Rubin: "Maximum likelihood from incomplete data via the EM algorithm." Journal of the Royal Statistical Society: Series B (Methodological) 39.1: 1-22. (1977).
4. G. J. McLachlan, and K. E. Basford: "Mixture models: Inference and applications to clustering". Vol. **38**. New York: M. Dekker, (1988).
5. L. Scrucca, et al.: "mclust 5: clustering, classification and density estimation using Gaussian finite mixture models." The R journal 8.1: 289. (2016).
6. G.Schwarz: "Estimating the dimension of a model." The annals of statistics: 461-464. (1978).
7. D. M. Titterington, et al.: "Statistical analysis of finite mixture distributions". Vol. **198**. John Wiley & Sons Incorporated, (1985).