# Latent Class Analysis of Socio-Economic and Political Dynamics in Italy Before and During the COVID-19 Pandemic

Gaudenzia Genoni
gaudenzia.genoni@studenti.unitn.it
https://github.com/Ggenoni/CSS_project

***Abstract*— This study employs Latent Class Analysis (LCA) to uncover latent socio-economic and political classes within the Italian population, both before and during the COVID-19 pandemic. By analyzing data from the European Social Survey (ESS) and focusing on variables such as education, income, political trust, and civic engagement, the research identifies two distinct classes with varying profiles and observes shifts over time. While the approach reveals important insights, it is limited by potential biases from self-reported data and the oversimplification of using only two classes. To enhance understanding, future research should integrate additional datasets and employ longitudinal methods, which may provide more comprehensive insights for policy-making in a dynamically evolving socio-economic landscape.**

***Keywords*—Latent Class Analysis (LCA), European Social Survey (ESS), Socio-Economic Dynamics, Political Trust, COVID-19 Pandemic, Italy**

## I. Introduction

Understanding the dynamics of socio-economic and political change is increasingly vital as societies undergo rapid and profound transformations. By leveraging data from the European Social Survey (ESS) and examining variables such as education, income, political trust and civic engagement, this study proposes to explore the latent socio-economic and political classes within the Italian population both before and during the coronavirus outbreak using Latent Class Analysis (LCA), a powerful statistical technique for revealing concealed categories within multifaceted datasets.

Such an analysis holds significant potential benefits. By uncovering hidden profiles within Italian society, it provides a more nuanced and detailed understanding that aggregated data might easily overlook. This insight may enable policymakers and researchers to more effectively comprehend how different segments of the population have been impacted by events like the COVID-19 pandemic, allowing them to design targeted interventions such as tailored economic relief programs or comprehensive public health initiatives. Furthermore, tracking shifts in these profiles over time provides crucial information for adapting policies to better address ongoing societal challenges and support vulnerable groups.

The paper is structured as follows: Section II provides an overview of existing studies that have utilized LCA in the analysis of socio-economic and political variables, setting the context for this research and highlighting the relevance of LCA in examining pandemic-related changes. Section III details the European Social Survey data utilized for this study, including the selection of datasets and an examination of three key demographic factors within the Italian sample. Section IV describes the methodological approach of the analysis, covering model selection criteria and the interpretation of the results concerning the identified latent classes. Finally, Section V discusses the findings, acknowledges the study's strengths and limitations, and proposes directions for future research.

## II. Literature Review

Latent Class Analysis has proven to be a powerful tool in computational social sciences for uncovering hidden subgroups within complex datasets. A comprehensive reference on the methodologies and applications of LCA is provided by Andersen, Hagenaars, and McCutcheon in their volume *Applied Latent Class Analysis* (2004), which explores various innovations and techniques in LCA, offering valuable insights into its application across a range of social science research contexts [1].

Recent studies have applied LCA to investigate various social variables in the context of the last global health crisis. Strating et al. (2020) utilized LCA to analyze changes in alcohol consumption patterns during the initial lockdown period in Australia, New Zealand, and the United Kingdom. Their research highlights how pandemic restrictions influenced drinking behaviors across different social contexts [2].

Similarly, Wong et al. (2022) employed LCA to examine health disparities among young children in socially disadvantaged families during the pandemic. Their study identified distinct patterns of family hardship and correlated these with variations in parenting behavior and child well-being, illuminating how pandemic-related stressors differentially impact vulnerable populations [3].

Additionally, Malkowski et al. (2024) investigated socio-economic inequalities in internet use among older adults in

England before and during the pandemic. By using LCA, they revealed persistent digital exclusion and variations in internet use across socio-economic groups, underscoring the digital divide exacerbated by the pandemic [4].
Although not centered on LCA, the study by Satherley et al. (2024) is noteworthy for its examination of changes in political attitudes in New Zealand following COVID-19 lockdowns. Their research highlights shifts in government satisfaction and trust in institutions, which provides context for understanding political attitude changes — a topic that aligns with the focus of this paper [5].
These studies highlight the versatility of LCA in analyzing social variables and disparities, providing a valuable context for this research, which specifically aims to identify latent classes within the Italian population and analyze socio-economic and political dimensions both preceding and throughout the COVID-19 pandemic.

## III. The Dataset

The data source for this research was the European Social Survey (ESS), a cross-national research initiative launched in 2001 to collect and analyze data on social, political, and moral attitudes across Europe. The ESS is managed by ESS ERIC, a research infrastructure endorsed by the European Commission, which upholds high standards of data quality and methodological rigor.
Two datasets were chosen, each filtered to include only Italian respondents. The first dataset, from the European Social Survey Round 9 (referred to as ESS9), includes interviews with 2,745 individuals conducted between August 2018 and January 2020, representing the period before the coronavirus crisis [6]. The second dataset, from the European Social Survey Round 10 (referred to as ESS10), comprises interviews with 2,640 individuals conducted between September 2020 and September 2022, reflecting the period during and in the aftermath of the pandemic [7].

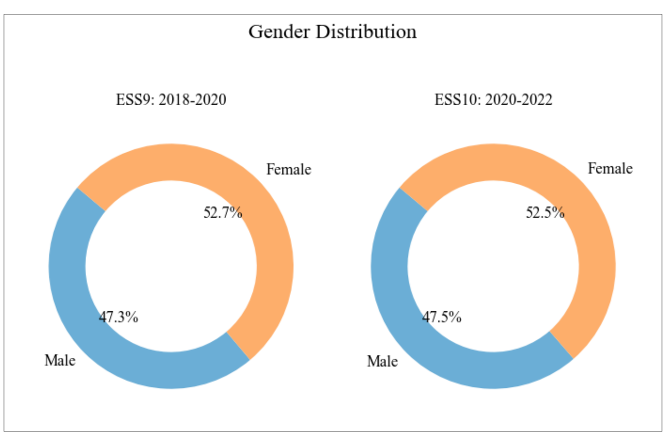


Fig. 1. Gender distribution in ESS9 and ESS10.

To understand the representativeness of the ESS data, it is important to conduct a brief preliminary examination of three demographic factors present in both surveys: gender distribution, age distribution, and geographical distribution.
Gender distribution (Figure 1) is balanced and remains stable across the two surveys, with females representing 52.7% in ESS9 and 52.5% in ESS10, while males constitute 47.3% and 47.5%, respectively.
The age distribution (Figure 2) follows a bell curve in both surveys, peaking in the 51-55 age group in ESS9 with 253 respondents and in the 61-65 age group in ESS10 with 252 respondents. This strong middle-aged and early senior representation reflects the population structure, while the youngest age group (16-20) shows a decline in participation in ESS10 compared to ESS9, indicating a possible reduction in engagement among younger individuals over time.

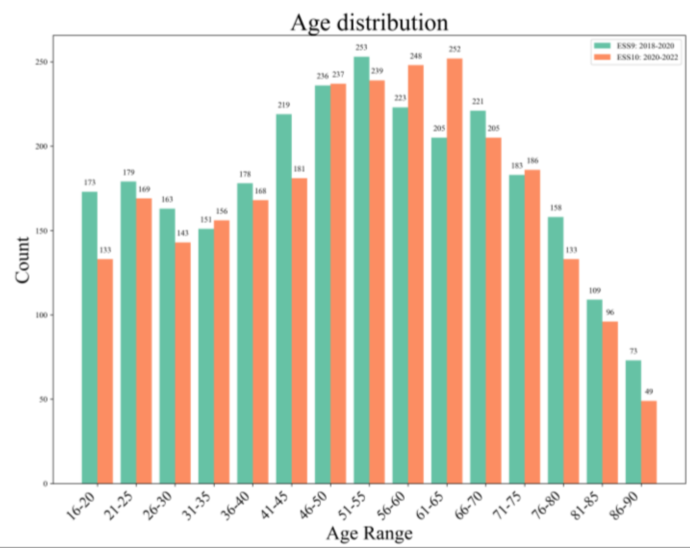


Fig. 2. Age distribution in ESS9 and ESS10.

Data on geographical distribution was recorded for groups of regions (see Table I for the correspondence between group codes and the regions included). As depicted by the choropleth map (Figure 3), a high number of respondents are present in northern and southern regions, with participation in the North-West exceeding 700 individuals in ESS9. Central regions show moderate participation, particularly in the second survey, while Sicily and Sardinia together record fewer than 300 participants. Although the overall geographical pattern remains consistent, there is a slight decrease in respondent numbers in ESS10, possibly due to external factors such as the pandemic.

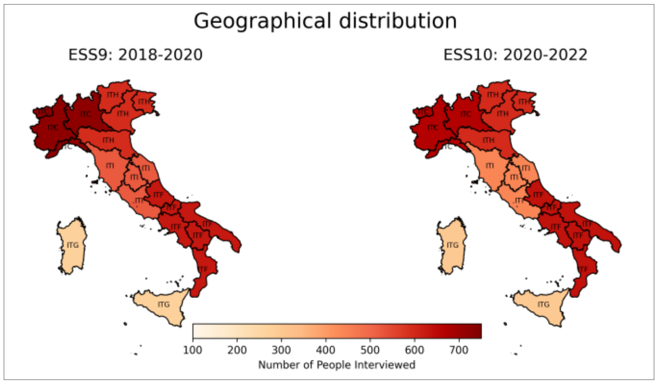


Fig. 3. Geographical distribution in ESS9 and ESS10.

TABLE I
REGIONS BY GROUP CODE

| Code | Group | Regions |
|---|---|---|
| ITC | Nord-Ovest | Piemonte, Valle d'Aosta/Vallée d'Aoste, Liguria, Lombardia |
| ITF | Sud | Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria |
| ITH | Nord-Est | Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria |
| ITI | Centro | Toscana, Umbria, Marche, Lazio |
| ITG | Isole | Sicilia, Sardegna |

Overall, it emerges that the ESS data exhibits a high degree of reliability and considerable representativeness of the Italian demographic.

To perform a Latent Class Analysis of the Italian population, eight categorical variables were selected from the available features in the datasets, covering a good range of socio-economic and political factors. Specifically, they are:

- **edlveit**: Highest level of education completed (ranging from 1 = 'none' to 21 = 'PhD')
- **hincfel**: Perception of household income (ranging from 1 = 'living comfortably' to 4 = 'very difficult')
- **stflife**: Life satisfaction (ranging from 0 = 'extremely dissatisfied' to 10 = 'extremely satisfied')
- **lrscale**: Political inclination (ranging from 0 = 'left' to 10 = 'right')
- **vote**: Participation in the last national election (1 = 'yes' or 2 = 'no')
- **trstplt**: Trust in politicians (ranging from 0 = 'no trust at all' to 10 = 'complete trust')
- **psppsgva**: Degree to which the political system is seen as allowing people to have a say in government actions (ranging from 1 = 'not at all' to 5 = 'a great deal')
- **polintr**: Interest in politics (ranging from 1 = 'very interested' to 4 = 'not at all').

These variables provide a multidimensional profile of respondents, capturing key aspects of their educational background, economic status, life satisfaction, political views and civic engagement: such a comprehensive approach should be able to reveal significant hidden subgroups within Italian society through the lenses of LCA.

Before proceeding, it is particularly interesting to observe a graph showing the distribution of trust level in politicians, a key component in the following LCA analysis. As is visible in Figure 4, a significant portion of the population harbors low trust in politicians, with 21.4% of respondents in the ESS9 dataset rating their trust at 0. In both datasets, more than 60% of the population scores between the negative range of 0 to 5, while only a very small fraction of respondents expresses high levels of trust (0.75% at level 9 in ESS9 and 0.50% at level 9 in ESS10; 0.38% at level 10 in ESS9 and 0.30% at level 10 in ESS10). Interestingly, however, in the ESS10 dataset, which reflects the period following the health crisis, there is a slight decrease in extreme distrust (with level 0 declining to 18.2%) accompanied by a slight increase in moderate trust levels (with level 6 reaching 10.8%, compared to 8.5% in ESS9), which may suggest a cautious shift towards more balanced views in the aftermath of the pandemic.

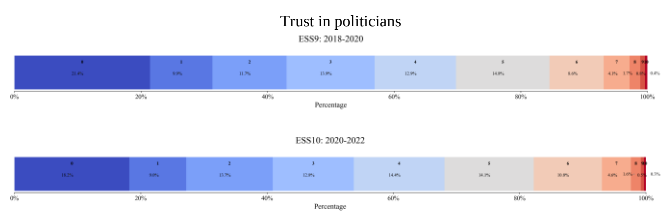


Fig. 4. Trust in politicians in ESS9 and ESS10.

Given the quality of the datasets, only minimal pre-processing was required. Missing values for specific questions were already coded as shown in Table II, and these values were retained during the analysis.

TABLE II
CODING FOR MISSING VALUES

| Code | Alternative code | Meaning |
|---|---|---|
| 7 | 77 (or 7777) | 'refusal' |
| 8 | 88 (or 8888) | 'don't know' |
| 9 | 99 (or 9999) | 'no answer' |

## IV. Latent Class Analysis

### A. Methodological overview

Latent Class Analysis (LCA) is a statistical technique used to identify unobserved subgroups within a population based on the patterns of responses or characteristics observed in the data [8]. As a finite mixture model, LCA assumes that the data are generated from a mixture of latent classes, each characterized by unique response patterns on observed variables.

In LCA, the primary goal is to estimate class membership probabilities, $\pi_k$, which represent the likelihood of an individual being in each latent class $k$. Additionally, the model estimates item-response probabilities, $P(X_i = x_j \mid C_k)$, indicating the likelihood of observing specific responses $x_j$ on variable $X_i$ given class membership $C_k$.

In this study, the model fitting process was performed in R, employing the package poLCA [9]. The process involves specifying the number of latent classes $K$ and estimating parameters through maximum likelihood estimation (MLE).

Once the model is fitted, latent classes can be interpreted based on the estimated item-response probabilities, providing insights into the underlying structure of the data.

It is important to note, however, that LCA is an unsupervised method where the researcher specifies the number of latent classes, but the resulting groups are not predetermined. This means the identified classes might not always match theoretical expectations. Additionally, LCA assumes local independence, meaning that within each latent class, observed variables are considered independent given the latent variable. This assumption can sometimes oversimplify complex dependencies among variables. Despite these limitations, LCA remains a powerful technique that provides meaningful findings into distinct population segments, with flexibility in the number of classes allowing for adaptation to various contexts.

### B. Model selection

In Latent Class Analysis, evaluating several models with varying numbers of classes is essential for identifying the most appropriate structure for the data. Different class configurations can reveal diverse underlying patterns and relationships, potentially leading to more accurate and insightful results. To determine the optimal model, researchers often use information criteria such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). AIC measures model fit while penalizing for complexity, aiming to find a balance between fit and parsimony. BIC also assesses model fit but applies a stricter penalty for complexity, therefore helping avoid overfitting.

Table III presents the AIC and BIC values for three models fitted to the ESS9 dataset. Model I has 2 classes, Model II has 3 classes, and Model III has 4 classes. According to the AIC, Model III achieves the lowest value, indicating the best performance in balancing model fit and complexity. However, the BIC, which applies a more stringent penalty for complexity, suggests that Model I is preferable, as it offers the best trade-off between model fit and parsimony.

TABLE III
ESS9: Model comparison on AIC and BIC

| **Model** | **Number of classes** | **AIC** | **BIC** |
|---|---|---|---|
| **I** | 2 | 71480.92 | 72421.80 |
| **II** | 3 | 71065.57 | 72479.86 |
| **III** | 4 | 70833.77 | 72421.80 |

Similarly, Table IV shows the AIC and BIC values for the three models fitted to the ESS10 dataset. Here, again, Model III has the lowest AIC but Model I reaches the lowest BIC. Therefore, choosing Model I is more advantageous due to its simpler structure and reduced risk of overfitting.

TABLE IV
ESS10: Model comparison on AIC and BIC

| **Model** | **Number of classes** | **AIC** | **BIC** |
|---|---|---|---|
| **I** | 2 | 68592.50 | 69515.43 |
| **II** | 3 | 68202.57 | 69589.91 |
| **III** | 4 | 68064.37 | 69916.11 |

Based on these considerations, both datasets were analyzed using a model with only two latent classes. The relevant findings are discussed in the following section.

### C. Results

1) ESS9

The two latent classes identified by the Latent Class Analysis of the ESS9 dataset exhibit significant differences in socio-economic and political characteristics (see Figure 5).

Class 1, representing 43.69% of the population, is characterized by moderate to low educational attainment. A substantial portion of this group has completed middle school (39.28% at level 4), with a notable percentage at lower educational levels (19.49% at level 2). Financially, Class 1 appears moderately stable yet constrained, as nearly half (44.83%) feel they are getting by, and 29.62% experience financial difficulties. Their life satisfaction is generally moderate, with a significant portion expressing dissatisfaction.

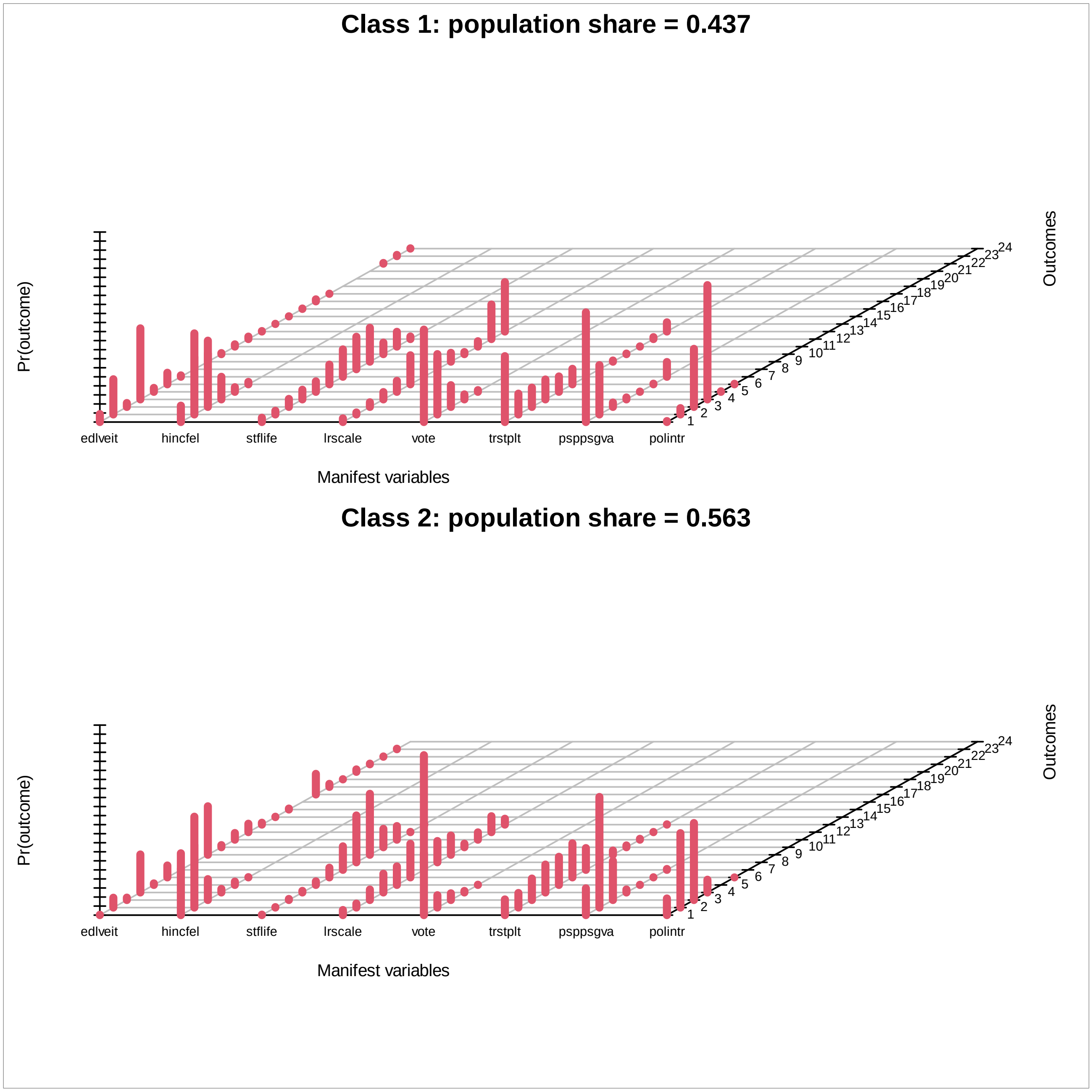


Fig. 5. ESS9 Latent Class Profile Plot.

Politically, Class 1 shows a centrist tendency but also significant disengagement, with 27.31% not positioning themselves on the political spectrum. Civic engagement is moderate, with 50.84% having voted in the last national election. Trust in politicians is low, with 36.42% expressing no trust at all, reflecting a broader skepticism about the political system's responsiveness, as 60.60% believe it offers virtually no opportunities for public input. Additionally, the overwhelming majority (95.28%) reports little to no interest in politics.

Class 2, comprising 56.31% of the sample, exhibits slightly higher educational attainment, with a notable concentration at level 9 (26.81%), which corresponds to a high school diploma. Financially, this group is relatively more stable, with 50.23% reporting comfort and 34.18% feeling secure. Their life satisfaction is higher than in Class 1, with 33.71% rating it as 8 and 25.91% as 7. Politically, Class 2 shows more engagement, as indicated by 88.40% having voted in the last national election. Despite this, trust in politicians remains moderate, although the perception of the political system is slightly more positive, with 21.64% believing it provides some opportunities for public input. Political interest is more balanced, though a notable minority (7.08%) remains completely disinterested.

In conclusion, Class 1, characterized by moderate education, financial constraints, moderate life satisfaction, and low political trust and engagement, could be aptly described as "Disengaged Moderates". In contrast, Class 2, which features higher educational attainment, greater financial stability, and increased political engagement, along with higher life satisfaction but some lingering skepticism about the political system, could be labeled "Active Pragmatists".

2) ESS10

Compared to ESS9, the two classes identified by the Latent Class Analysis of the ESS10 dataset reveal both similarities and differences (see Figure 6).

Class 1 in ESS10 (48.82% of the population) mirrors Class 1 from ESS9 in terms of moderate educational attainment and political centrism. However, financial sentiment has improved slightly, with 51.56% reporting comfort, compared to greater financial constraints in ESS9. Life satisfaction is similarly moderate to high. Civic engagement remains moderate (only 49.94% voted), and trust in politicians is low (31.54% express no trust), consistent with ESS9. Skepticism towards the political system and low political interest also persist.

Class 2 in ESS10 (51.18% of the population) shows a broader range of educational attainment with a concentration at high secondary education. Financial sentiment is somewhat improved compared to Class 2 in ESS9, though a small portion (7.68%) still finds financial management challenging. Life satisfaction is higher than in ESS9, with increased satisfaction levels. Politically, this class is balanced with a central tendency, similar to Class 2 in ESS9, but with slightly lower civic engagement (85.51% voted). Trust in politicians is moderate, and the perception of the political system is less cynical than in ESS9, reflecting an overall positive shift.

In summary, contrary to an expectation that the COVID-19 pandemic would lead to a dramatic worsening of socio-economic and political conditions, the analysis of the two latent classes from ESS10 indicates a more nuanced reality, as financial sentiment for both classes has shown improvement, and life satisfaction levels have remained stable or even slightly increased. Factors that might have contributed to this unexpected resilience, such as social support systems, and adaptations to the economic challenges, should be further investigated.

In ESS10, the shift towards a more balanced distribution of the population between the two classes, that nearly reach an even split, is also notable. It may reflect shifts in class definitions or changes in public attitudes that were influenced by the pandemic, resulting in a more even distribution of socio-economic and political characteristics among the population.

## V. Discussion and Conclusion

This study explored the latent socio-economic and political classes within the Italian population before and during the COVID-19 pandemic using Latent Class Analysis (LCA) on data from the European Social Survey (ESS).

The chosen method facilitated a nuanced identification of hidden subgroups, revealing distinct profiles across two latent classes, and enabled a comparative analysis over a critical period in the lives of many.

However, the limitations of this study should not be overlooked. The reliance on self-reported survey data introduces potential biases, such as social desirability bias - where respondents may alter their answers to align with perceived social norms - and recall bias - where inaccuracies arise from difficulties in remembering past events. These biases can affect the accuracy of the findings, as individuals may not fully or accurately disclose their socio-economic conditions or political attitudes. Additionally, the decision to limit the analysis to two latent classes, guided by the Bayesian Information Criterion (BIC), while promoting model simplicity, might have oversimplified the diversity within the population. A more complex model could capture a broader spectrum of attitudes, though this might come at the cost of reduced interpretability.

These considerations suggest that future research could benefit from incorporating additional datasets or employing longitudinal methods to track changes in latent classes over a longer period. Expanding the analysis to include more latent classes could also enrich the understanding of the underlying factors driving socio-economic and political changes.

In conclusion, while the LCA approach provided valuable insights, the study’s limitations highlight the need for cautious interpretation and suggest avenues for further exploration. Future research on this topic may offer crucial information for shaping policies that address the diverse needs of different segments of the Italian population in the post-pandemic landscape.

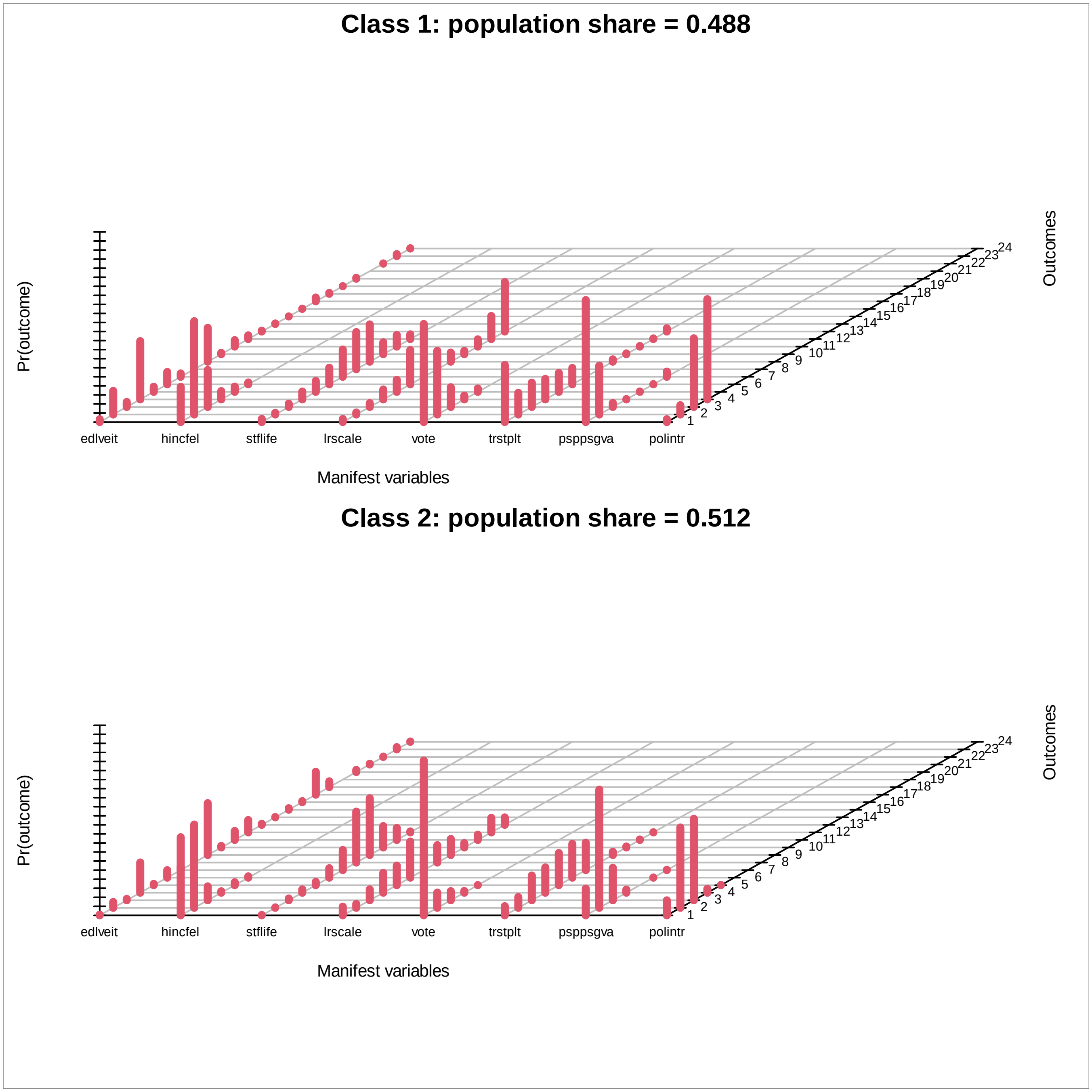


Fig. 6. ESS10 Latent Class Profile Plot.

## References


[1] Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied Latent Class Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511499531.

[2] Strating, T. R., Puljević, C., Davies, E., Barratt, M. J., Winstock, A., & Ferris, J. (2023). Not all home drinking is equal: A latent class analysis of drinking patterns and alcohol consumption levels following initial COVID-19 restrictions in Australia, New Zealand and the United Kingdom. *Drug and Alcohol Review, 42*(5), 1041-1053. https://doi.org/10.1111/dar.13686.

[3] Wong, R. S., Tung, K. T. S., Rao, N., Chan, K. L., Fu, K.-W., Yam, J. C., Tso, W. W. Y., Wong, W. H. S., Lum, T. Y. S., & Ip, P. (2022). Using latent class analyses to examine health disparities among young children in socially disadvantaged families during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health, 19*(13), 7893. https://doi.org/10.3390/ijerph19137893.

[4] Malkowski, O. S., Townsend, N. P., Kelson, M. J., Foster, C. E. M., & Western, M. J. (2024). Socio-economic inequalities in the breadth of internet use before and during the COVID-19 pandemic among older adults in England. *PLoS ONE*, *19*(5), e0303061. https://doi.org/10.1371/journal.pone.0303061.

[5] Satherley, N., Zubielevitch, E., Greaves, L. M., Barlow, F. K., Osborne, D., & Sibley, C. G. (2022). Political attitude change over time following COVID-19 lockdown: Rallying effects and differences between left and right voters. *Frontiers in Psychology, 13*, 1041957. https://doi.org/10.3389/fpsyg.2022.1041957.

[6] European Social Survey European Research Infrastructure (ESS ERIC) (2023) ESS9 - integrated file, edition 3.2 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess9e03_2.

[7] European Social Survey European Research Infrastructure (ESS ERIC) (2023) ESS10 - integrated file, edition 3.2 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10e03_2.

[8] Vermunt, J., & Magidson, J. (2004). *Latent class analysis*. In *The Sage encyclopedia of social science research methods* (pp. 549–553). Sage Publications. https://www.researchgate.net/publication/280940069_Latent_Class_Analysis_in_The_Sage_encyclopedia_of_social_science_research_methods.

[9] Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software*, *42*(10), 1–29. https://doi.org/10.18637/jss.v042.i10.