NOVA IMS Information Management School

MEGI

Master Degree Program in

Statistics and Information Management

RETHINKING CHILD POVERTY IN THE EU

A cluster-based alternative to AROPE

Maria João Gonçalves Dutra Serra

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Statistics and Information Management

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

RETHINKING CHILD POVERTY IN THE EU

A cluster-based alternative to AROPE

by

Maria João Gonçalves Dutra Serra

Master Thesis presented as partial requirement for obtaining the Master’s degree in Statistics and Information Management, with a specialization in Information Analysis and Management

Supervised by

Maria Helena Baptista, PhD, NOVA Information Management School

July, 2025

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism, any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisboa, 20th of February 2026
Maria Serra

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Maria Helena Baptista for her guidance, constructive feedback, and continuous support throughout the development of this thesis. Her insights and high standards were fundamental in strengthening both the methodological rigour and the clarity of this work.

I am deeply grateful to my family for their constant encouragement, patience, and unconditional support. Their confidence in me made this journey possible.

I would also like to thank my friends and colleagues for their support, understanding, and for the many conversations that helped me reflect on and refine my ideas along the way.

Finally, I acknowledge the importance of the publicly available Eurostat data used in this research, which made this comparative analysis across EU Member States possible.

ABSTRACT

This thesis develops a multidimensional, child-specific typology of poverty and social exclusion across European Union Member States using harmonised national-level indicators from Eurostat. While the EU's official measure, AROPE, plays a central role in social monitoring, it relies on household-level income and employment metrics that overlook key disadvantages experienced by children. To address this limitation, the study constructs a dataset of nine child-focused indicators capturing housing deprivation, economic insecurity, and early education access. After standardisation and targeted imputation, two clustering techniques, Ward's hierarchical method and k-means, are applied to identify cross-country deprivation profiles. The analysis yields a four-cluster structure that distinguishes low, moderate, and high levels of child deprivation, including two extreme profiles corresponding to Greece and the Bulgaria–Romania group. In parallel, the AROPE components are disaggregated into seven mutually exclusive intersections and subjected to the same analytical pipeline, producing a simpler three-cluster typology characterised by a large low-risk group and a sharply isolated high-risk cluster. PaCMAP embeddings provide additional insight into the geometric structure of the data, revealing that the child-specific indicators form a dispersed and multidimensional landscape, whereas the AROPE components generate a compressed configuration with limited internal variation. Overall, the findings demonstrate that child-focused indicators capture structural patterns of disadvantage that remain invisible under income-based measures alone, underscoring the importance of multidimensional approaches for monitoring child well-being and informing EU social policy.

KEYWORDS

AROPE; Child Poverty; Cluster Analysis; European Union; Multidimensional Deprivation; PaCMAP;

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ADAM	Adaptive Moment Estimation
AF	Alkire-Foster method
AROP	At-Risk-Of-Poverty rate
AROPE	At Risk Of Poverty or Social Exclusion
CRC	Convention on the Rights of the Child
EAPN	European Anti-Poverty Network
EPSR	European Pillar of Social Rights
EU	European Union
EU-LFS	EU Labour Force Survey
EuroVoc	Multilingual thesaurus of the Publications Office of the European Union
EU-SILC	EU Statistics on Income and Living Conditions
HCA	Hierarchical Cluster Analysis
MODA	Multiple Overlapping Deprivation Analysis
MPI	Multidimensional Poverty Index
NHCA	Non-Hierarchical Cluster Analysis
PaCMAP	Pairwise Controlled Manifold Approximation Projection
PCA	Principal Component Analysis
SDG	Sustainable Development Goal
SMSD	Severe Material and Social Deprivation rate
t-SNE	t-distributed Stochastic Neighbour Embedding
UMAP	Uniform Manifold Approximation and Projection
UNICEF	United Nations International Children’s Emergency Fund
VLWI	Very Low Work Intensity indicator

1. INTRODUCTION

Child poverty and social exclusion remain pressing challenges across the European Union, despite long-standing policy commitments to reduce disadvantage and promote equal opportunities. Children experience poverty differently from adults, as their well-being depends not only on household resources but also on access to adequate housing, education, care, safety, and other fundamental environments for development. These characteristics make child poverty inherently multidimensional and distinct in both its causes and its consequences.

The European Union currently monitors poverty and social exclusion through the At Risk of Poverty or Social Exclusion (AROPE) indicator, which combines income poverty, severe material and social deprivation, and very low work intensity at the household level. While AROPE has an important institutional role, concerns have been raised about its limited capacity to capture child-specific disadvantages. Many relevant aspects of children's experiences, such as overcrowding, quality of housing, energy poverty, or participation in early education, lie outside the scope of income-based or household-level indicators. As a result, there is a growing need for complementary approaches that reflect a more comprehensive and child-centred understanding of deprivation.

This thesis addresses this gap by developing a multidimensional, child-specific typology of deprivation across EU Member States using harmonised national-level data from Eurostat. It adopts a quantitative, cross-sectional design and applies a set of unsupervised learning techniques, hierarchical clustering, k-means, and Pairwise Controlled Manifold Approximation Projection (PaCMAP) embedding, to identify patterns of disadvantage based on a selection of child-focused indicators. To assess how these patterns relate to existing EU monitoring practices, the same analytical framework is applied to the disaggregated components of AROPE, enabling a direct comparison between child-specific and household-based approaches.

The central research question guiding this work is: How do multidimensional, child-specific deprivation patterns across EU Member States compare with those identified by the AROPE framework, and what do these differences imply for the measurement of child poverty? Addressing this question contributes to a more nuanced understanding of child deprivation, highlights the implications of indicator selection for cross-country classification, and informs ongoing debates on the adequacy of EU poverty monitoring tools.

The thesis makes two main contributions. First, it develops a child-centred analytical framework grounded in domains that are directly relevant to children's well-being. Second, it provides a systematic comparison between this framework and the existing AROPE structure, offering insights into the alignment and misalignment between child-specific and income-based approaches to measuring disadvantage.

The remainder of the thesis is organised as follows. Chapter 2 reviews the theoretical foundations of multidimensional poverty, critiques of AROPE, and approaches to child-specific measurement. Chapter 3 outlines the methodological design, including data sources, indicator selection, preprocessing, and analytical procedures. Chapter 4 presents the empirical analysis for both sets of indicators. Chapter 5 discusses the typologies and the structural patterns revealed by the embedding technique. Finally, Chapter 6 summarises the study's contributions, policy relevance, and paths for future research.

2. LITERATURE REVIEW

2.1 THEORETICAL FOUNDATIONS

Poverty and social exclusion are increasingly recognised as multidimensional and relational phenomena, especially concerning children. Traditional income-based definitions, typically using thresholds like 60% of median income, have long dominated academic and policy frameworks. However, critics argue that such approaches fail to capture non-monetary deprivations such as inadequate housing, limited access to quality education or healthcare, and restricted opportunities for social participation (Clerici, 2022; UNICEF, 2024a). These narrow definitions risk obscuring structural disadvantages and the lived experiences of children, particularly those from marginalised groups or single-parent households. In response, social exclusion has emerged as a critical lens, shifting the focus from static income poverty to dynamic disconnection from essential institutions, relationships, and opportunities (Bessell, 2021; Clerici, 2022).

To address these limitations, several theoretical frameworks have been developed. The capability approach, pioneered by Amartya Sen and further expanded by scholars such as Alkire et al. (2022) and Robeyns & Byskov (2025), reframe poverty as a deprivation of capabilities — that is, the fundamental freedoms individuals have to lead lives they have reason to value. In the case of children, this approach shifts the focus to real opportunities for well-being and development, including access to education, play, care, and social relationships (D'Agostino et al., 2018; Domínguez-Serrano & Del Moral-Espín, 2022). Parallel to this, rights-based approaches, grounded in the United Nations Convention on the Rights of the Child (CRC), highlight the legal and moral obligations of states to guarantee children's rights to an adequate standard of living, education, healthcare, and protection from discrimination (Biggeri & Cuesta, 2020; UNICEF, 2024a). These frameworks are further embedded in international policy commitments, such as Sustainable Development Goal (SDG) 1.2 and the European Pillar of Social Rights (EPSR), which emphasise the need to eradicate poverty and promote social inclusion in all its forms (De La Rasilla et al., 2024).

These theoretical perspectives critique adult-centric, income-focused measurement tools, arguing that they fail to account for children's agency, relational needs, and intra-household inequalities. Children may experience deprivation even within households that are not classified as income poor due to unmet specific needs or unequal allocation of resources (Bessell, 2021; Karagiannaki & Burchardt, 2024). Consequently, there is growing consensus in the literature that child poverty should be measured using child-specific, multidimensional indicators. These should reflect material deficits and relational, participatory, and subjective dimensions of well-being (Chzhen et al., 2018; Dirksen & Alkire, 2021). This conceptual shift underpins the development of more inclusive and ethically grounded frameworks for measuring child poverty across the European Union (EU).

2.2 CRITIQUE OF AROPE AND CURRENT EU INDICATORS

The European Union monitors poverty and social exclusion through a composite measure of intersections of the at-risk-of-poverty rate (AROP), the severe material and social deprivation rate (SMSD), and the very low work intensity indicator (VLWI) at the household level (Eurostat, n.d.). This indicator, At Risk of Poverty or Social Exclusion (AROPE), has been central to EU (European Union) social policy monitoring since 2010. Despite being revised under the Europe 2030 strategy, AROPE faces substantial critique. Researchers argue that its binary thresholds, reliance on household-level data, and limited dimensional scope render it inadequate for addressing the complex realities of child poverty (Clerici, 2022; UNICEF, 2024b). Furthermore, critics contend that AROPE's structure remains misaligned with the EU's commitments under SDG 1.2 and the European Pillar of Social Rights. This misalignment is particularly problematic considering the EU's stated aim to "leave no child behind," which requires more granular and child-centred monitoring tools. Together, these limitations underscore the need for more comprehensive, child-specific indicators to guide EU social policy and improve monitoring outcomes (Copeland, 2023; De La Rasilla et al., 2024; UNICEF, 2024a).

One of the most prominent criticisms of AROPE concerns its reliance on binary thresholds, which classify individuals as either "at risk" or "not at risk" based on arbitrary cut-offs without accounting for the depth or intensity of deprivation (Bárcena-Martín et al., 2020; Fabbris, 2024; Kumar et al., 2023). This approach oversimplifies complex realities and fails to reflect gradations of poverty, particularly among children who may experience multiple, overlapping deprivations that remain invisible under such dichotomous metrics. Additionally, AROPE's use of national median income thresholds for determining income poverty (AROP) undermines cross-country comparability within the EU, as wealthier nations set much higher poverty lines than poorer member states, leading to distorted perceptions of poverty incidence (Aguayo et al., 2016; Darvas, 2018).

As previously discussed in the theoretical literature, AROPE's reliance on household-level indicators makes it particularly ill-suited for capturing child-specific vulnerabilities. It assumes an even distribution of resources within families, ignoring the reality of intra-household inequalities that often leave children, particularly girls or children in large families, more vulnerable (Bessell, 2021; Karagiannaki & Burchardt, 2024; UNICEF, 2024a). Moreover, the composite measure lacks age-disaggregated indicators, masking critical differences in how poverty affects children compared to adults (Dirksen & Alkire, 2021; UNICEF, 2024b). The reliance on aggregated household-level data excludes crucial child-level indicators essential for policy targeting and evaluation. In addition, the indicator also neglects non-material aspects of child poverty, such as access to education, safe play spaces, and social participation, which are essential for a child's well-being but fall outside AROPE's current scope (Bessell, 2021; D'Agostino et al., 2018).

On top of this methodological shortcomings, AROPE has been critiqued for its role within the political economy of EU governance. Analysts argue that AROPE has become technicized, serving as a tool for statistical benchmarking rather than a meaningful instrument for addressing poverty's root causes (Fabbris, 2024; Zieleńska & Wnuk, 2024). This depoliticization of poverty measurement aligns with the EU's broader emphasis on fiscal discipline and market integration, often side lining structural inequalities and social rights in policy discourse (Copeland, 2023). Civil society organisations, such as the European Anti-Poverty Network (EAPN), have voiced concerns that AROPE legitimises minimalist social policies while obscuring the lived realities of marginalised groups, particularly children (Duffy, 2020). Despite commitments under the European Pillar of Social Rights, AROPE remains poorly equipped to support transformative social protection strategies, partly because of its embeddedness in institutional routines like the European Semester, which prioritises macroeconomic stability over social investment (Eurochild, 2023).

In conclusion, the limitations of AROPE provide both the conceptual and empirical justification for alternative, child-specific approaches to measuring poverty and social exclusion. This thesis contributes to that agenda by developing a multidimensional framework that draws on child-level deprivation indicators and applies statistical clustering techniques to uncover patterns of disadvantage across EU countries.

2.3 CHILD-SPECIFIC MULTIDIMENSIONAL POVERTY

The theoretical and measurement critiques outlined earlier have led to increasing calls in the literature for more complex, multidimensional, and child-specific poverty measures. Traditional indicators fail to reflect children's unique vulnerabilities, especially in education, housing, and social participation (Bessell, 2021; Chzhen et al., 2015; D'Agostino et al., 2018). Frameworks such as the United Nations International Children's Emergency Fund (UNICEF)'s Multiple Overlapping Deprivation Analysis (MODA) have gained prominence for capturing these deprivations through a rights-based lens aligned with the CRC and SDG 1.2 (Chzhen et al., 2018; Chzhen & Bruckauf, 2017).

While MODA and related models are typically applied to microdata sources like EU Statistics on Income and Living Conditions (EU-SILC), their conceptual principles remain relevant for national-level approaches. Studies using MODA with EU-SILC data have shown that multidimensional child poverty measures can identify deprived children missed by income-based tools such as AROPE (Chzhen & Bruckauf, 2017; UNICEF, 2024b). Although designed for individual-level data, these studies provide conceptual guidance for constructing multidimensional indicators using national aggregates, particularly when child-specific breakdowns are unavailable. Even though this thesis does not use individual-level information, it draws from these principles to select and group indicators to reflect children's multidimensional experiences of deprivation.

Empirical research across Europe underscores the persistent structural inequalities faced by children. In Austria, Premrov et al. (2022) used a modified Alkire-Foster method to uncover substantial housing and social participation deprivations in Vienna. In Spain, Arcarons et al. (2025) found that children of migrant origin experience multidimensional poverty even when income is accounted for. UNICEF (2024a) highlighted that Roma children across Central and Eastern Europe face simultaneous deprivations in housing, education, and access to services. These studies demonstrate consistent patterns of child deprivation that transcend household income measures (Bessell, 2021; UNICEF, 2024b).

Similarly, Vuorenlinna et al. (2023) found that multidimensional deprivation was associated with economic hardship, stigma, and reduced well-being among children during the COVID-19 pandemic in Finland. Although these findings rely on microdata, they highlight the deprivations that can be monitored through harmonised national-level indicators. They also strengthen the rationale for including overlapping deprivations in domains such as education, housing, and material well-being within national-level frameworks (UNICEF, 2024b; Vuorenlinna et al., 2023).

This thesis does not replicate these approaches through microdata analysis or deprivation indices but instead informs a typology-building exercise. Using harmonised Eurostat indicators, the study constructs country profiles based on multiple child-specific disadvantages. While child-level deprivation cannot be directly measured with macro data, a composite picture of child poverty across EU countries can be constructed by examining overlapping disadvantages in available indicators. Rather than generating an aggregate index or score, the study applies multivariate clustering to these indicators to group EU countries into distinct deprivation profiles. National-level indicators from Eurostat are therefore used to operationalise a conceptually grounded, multidimensional perspective on child poverty (Chzhen et al., 2018; D'Agostino et al., 2018).

2.4 METHODOLOGICAL ALTERNATIVES

Multidimensional poverty indices have historically been critiqued for relying on rigid and arbitrary weighting schemes. Both AROPE and the global Multidimensional Poverty Index (MPI) often apply equal or normative weights, which may distort deprivation profiles and reduce cross-country comparability (Alkire et al., 2022; Fabbris, 2024). Researchers have proposed empirically driven alternatives that incorporate uncertainty and improve index sensitivity. For instance, Drago (2021) developed an interval-based composite index that adjusts for variability in dimension weights. Bienkunska and Piasecki (2022) show how subjective poverty lines derived from EU-SILC data change the size and profile of poor populations, demonstrating how flexible formulations improve interpretability. Although these studies depend on microdata, the principles they promote, flexibility, transparency, and dimensional coherence, are also relevant to national-level models. They justify the move toward methods that reflect the data structure and deprivation patterns empirically, especially in child poverty measurement (Bienkunska & Piasecki, 2022; Drago, 2021).

Rather than constructing a deprivation index, this thesis adopts a typological approach based on multivariate clustering techniques. By grouping countries with similar profiles of child-specific deprivation, the analysis offers an alternative to additive or weighted-score models, reducing reliance on normative assumptions and making better use of the underlying data structure. This strategy is particularly appropriate for assessing poverty at the national level using harmonised Eurostat indicators, where direct child-level microdata is unavailable.

Cluster analysis is central to this approach. Unlike binary cut-offs or aggregate scores, clustering identifies natural groupings of countries based on patterns across multiple indicators. Frączek (2022) applied cluster analysis to classify EU countries by poverty risk among people with disabilities, while Kalinowski and Kiełbasa (2017) used it to map territorial inequalities. Antošová et al. (2021) applied k-means clustering, a Non-Hierarchical Cluster Analysis (NHCA), to group agricultural households by vulnerability using EU-SILC data. This thesis similarly applies Ward's hierarchical clustering and k-means algorithms to a standardised set of indicators, identifying typologies of multidimensional child deprivation across EU countries. This enables a more nuanced, policy-relevant classification of deprivation patterns, complementing and extending existing approaches beyond the limitations of AROPE (Antošová et al., 2021; Frączek, 2022).

In parallel with clustering, dimensionality reduction techniques are widely used to explore and visualise complex socioeconomic datasets. Classical approaches such as Principal Component Analysis (PCA) reduce dimensionality by maximising variance but often fail to capture the non-linear structures present in deprivation data (Jolliffe, 2002). More recent manifold learning techniques, including t-distributed Stochastic Neighbour Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP), capture non-linear relationships more effectively but tend to prioritise either local or global structure, making them sensitive to parameter choices and potentially unstable for comparative cross-country analysis (McInnes et al., 2018; van der Maaten & Hinton, 2008). Pairwise Controlled Manifold Approximation Projection (PaCMAP) offers an improvement over these methods by balancing local and global geometry through the structured use of near, mid-near, and further point pairs (Wang et al., 2021). In this thesis, PaCMAP is therefore used as an exploratory visualisation tool that complements the clustering procedures by providing clearer spatial representations of cross-country deprivation structures.

Advanced methods, such as the Alkire-Foster (AF) approach, fuzzy logic, and probabilistic models, have enriched child poverty measurement. Alkire and Apablaza (2016) and D'Agostino et al. (2018) applied AF methods to identify both incidence and intensity of deprivation, while Liberati et al. (2023) developed probabilistic models to assess uncertainty in deprivation status. Although these require microdata and exceed the empirical scope of this thesis, they provide a solid conceptual foundation for future applications (Alkire & Apablaza, 2016; Liberati et al., 2023).

In summary, the methodological strategy of this thesis avoids index construction and instead uses multivariate clustering to derive typologies of child deprivation. This approach leverages the strengths of available Eurostat indicators while responding to long-standing critiques of EU-level poverty measurement frameworks. It offers a transparent, comparative alternative to AROPE, aligned with both statistical best practices and conceptual insights from the literature (Cling et al., 2020; Kumar et al., 2023).

2.5 Policy Implications

How poverty is measured has profound implications for addressing it through public policy. Expanding on the critiques outlined earlier, the persistence of AROPE in EU monitoring frameworks raises concerns about its ability to identify children most in need of support (Karagiannaki & Burchardt, 2024; UNICEF, 2024b). Researchers and civil society organisations have argued that poverty measures must go beyond static benchmarks to inform policies responsive to real, multidimensional vulnerabilities (Duffy, 2020; EAPN, 2025). This section examines how multidimensional, child-focused indicators can strengthen EU monitoring frameworks and improve the effectiveness of social policy interventions.

Although the EPSR and SDG 1.2 commit member states to reducing poverty "in all its forms," current monitoring tools such as AROPE and the European Semester remain poorly equipped to address structural and intersectional child poverty (Copeland, 2023; Eurochild, 2023). Several authors have stressed the importance of disaggregating data by age, migrant background, and ethnicity to expose hidden vulnerabilities and inform equity-focused policy design (Arcarons et al., 2025; UNICEF, 2024a). For instance, UNICEF's application of multidimensional poverty analysis in Europe and Central Asia demonstrates how Roma children and those in informal housing face overlapping deprivations invisible to income-based measures (UNICEF, 2024a). Without such tools, policy frameworks risk underestimating the scope of child poverty and allocating resources in ways that do not reflect actual patterns of disadvantage (Eurochild, 2023; UNICEF, 2024a).

In addition to improving monitoring frameworks, empirical applications of child-specific and multidimensional poverty tools have shown clear benefits for policy design and intervention targeting. For example, Premrov et al. (2022) applied a city-level multidimensional child poverty index in Vienna, enabling municipal policymakers to identify domain-specific deprivations and adjust services accordingly. Similarly, Tuñón et al. (2022) highlight how cross-country comparisons of child MPIs in Spain and Argentina provided insights into effective social policy responses. The use of cluster analysis and regionally sensitive indices, as demonstrated by Antošová et al. (2021) and Vasilescu et al. (2024), further supports the development of tailored, place-specific interventions, particularly in rural or marginalised communities. While many of these applications rely on micro or local-level data, their policy insights remain relevant for constructing aggregated national-level typologies using harmonised Eurostat data, as implemented in this thesis. Civil society networks, including EAPN and Eurochild, have also played a critical role in advocating for multidimensional, rights-based monitoring frameworks that move beyond AROPE and better align with the lived realities of children across Europe (Duffy, 2020; Eurochild, 2023).

The literature therefore highlights the need for methodological innovation through child-specific, multidimensional indicators. In response, this thesis applies hierarchical and non-hierarchical cluster analysis to construct a typology of EU countries based on overlapping child deprivations. Rather than producing a composite index, this approach classifies countries into interpretable deprivation profiles that reveal cross-country patterns not captured by AROPE. By leveraging harmonised national indicators, the typology offers a statistically grounded and policy-relevant alternative to AROPE, one that better reflects the multidimensional nature of child poverty and supports more targeted, equitable, and effective EU social policy.

3. METHODOLOGY

3.1 Research Design and Objectives

This thesis adopts a quantitative, cross-sectional research design to construct a multidimensional, child-specific typology of poverty and social exclusion across European Union (EU) Member States. The primary objective is to classify countries according to overlapping child deprivation indicators, offering a statistically grounded and policy-relevant alternative to the At Risk of Poverty or Social Exclusion (AROPE) indicator, which has been widely criticised for its limited capacity to capture non-monetary and child-specific dimensions of deprivation.

Grounded in the capability approach and rights-based frameworks, namely the Convention on the Rights of the Child (CRC) and the European Pillar of Social Rights (EPSR), the proposed methodology reflects children's real opportunities to develop and thrive. It incorporates harmonised national-level indicators from Eurostat that capture deprivations in housing conditions, economic insecurity, and early education access, all of which are closely linked to the fulfilment of fundamental rights and social inclusion.

The research question guiding the empirical analysis is: How can a multidimensional, child-specific country classification based on national-level indicators better reflect structural child poverty and social exclusion patterns in the EU than existing measures such as AROPE?

The unit of analysis is the country, and the study includes only EU Member States for which comparable child-disaggregated indicators are available. The design is observational and comparative, aiming to derive deprivation profiles and identify structural and regional patterns across Europe.

Cluster analysis is applied directly to standardised deprivation indicators, complemented by dimensionality reduction techniques (PaCMAP) used for exploratory visualisation. By avoiding composite index construction, this approach sidesteps the challenges associated with weighting and aggregation and instead groups countries based on empirically observed deprivation patterns. The same analytical strategy is subsequently applied to the components of the AROPE indicator, enabling a transparent comparison between the child-specific and conventional EU poverty measurement frameworks.

3.2 Data Source and Country Coverage

The empirical foundation of this study is built on harmonised, aggregated data from Eurostat, the statistical office of the European Union. Eurostat provides publicly available, cross-nationally comparable indicators across various social, economic, and demographic domains. Indicators were selected for their institutional credibility, alignment with EU monitoring frameworks (e.g., SDGs, EPSR), and availability in child-specific national formats. This makes them appropriate for constructing a multidimensional child poverty framework at the country level.

The selected reference year is 2020, which offers the most complete indicator coverage across the EU-27 countries while avoiding systematic missingness. Although more recent data (e.g., 2023) exist, several key indicators are missing in over half the countries, rendering them unsuitable for robust cross-national analysis. In contrast, 2020 provides consistent data for all nine selected indicators across most countries.

The unit of analysis is the EU member state, and all indicators refer explicitly to children, typically defined as individuals under 18 years of age.

Lithuania was excluded from the analysis due to missing values in three of nine indicators, with no valid data for those variables across any available years. Additionally, Germany, Greece, and Poland each lacked data for one indicator in 2020. These cases will be retained using conservative regional imputation, as detailed in Section 3.3.3.

The final analytical sample comprises 26 EU countries with complete or minimally imputed information. While national-level data do not allow for intra-country or household-level analysis, they permit meaningful and policy-relevant cross-country comparisons of structural child deprivation patterns.

All countries are referenced using their ISO Alpha-2 codes to ensure clarity and consistency, as listed in Table 3.1. This facilitates data presentation in tables and visualisations throughout the empirical analysis.

Table 3.1– EU Member States and ISO Alpha-2 Codes

Belgium	(BE)	France	(FR)	Netherlands	(NL)
Bulgaria	(BG)	Croatia	(HR)	Austria	(AT)
Czechia	(CZ)	Italy	(IT)	Poland	(PL)
Denmark	(DK)	Cyprus	(CY)	Portugal	(PT)
Germany	(DE)	Latvia	(LV)	Romania	(RO)
Estonia	(EE)	Lithuania[1]	(LT)	Slovenia	(SI)
Ireland	(IE)	Luxembourg	(LU)	Slovakia	(SK)
Greece	(EL)	Hungary	(HU)	Finland	(FI)
Spain	(ES)	Malta	(MT)	Sweden	(SE)

[1] Lithuania (LT) is excluded from the final analysis due to extensive missing data across multiple indicators.

3.3 INDICATOR SELECTION AND DIMENSION JUSTIFICATION

3.3.1 INDICATOR SELECTION CRITERIA

The indicators for this multidimensional child deprivation framework were selected by four main criteria: child-specificity, data availability, cross-country comparability, and policy relevance. These principles are widely recognised in the multidimensional poverty measurement literature, particularly when developing frameworks to inform EU-level decision-making (Alkire et al., 2022; Chzhen et al., 2015; Dirksen & Alkire, 2021).

First, all indicators were required to be child-specific, meaning they either directly referred to children as the unit of measurement or were disaggregated to reflect the conditions experienced by children within their households. This addresses a key limitation of the AROPE indicator and many traditional poverty metrics, which measure poverty at the household level and assume equal intra-household distribution of resources (Bessell, 2021; Karagiannaki & Burchardt, 2024). Indicators such as overcrowding, sanitation, or joblessness reflect circumstances that can disproportionately affect children, even when household-level poverty is not detected. Their inclusion ensures that child-level vulnerabilities are directly captured (Chzhen et al., 2018; Tuñón et al., 2022).

Second, only indicators with harmonised national-level data available through Eurostat were considered. The study prioritised the most recent year (2020) with maximum country coverage. It relied on variables from EU Statistics on Income and Living Conditions (EU-SILC) and EU Labour Force Survey (EU-LFS) modules regularly monitored across Member States. This ensures the analysis remains replicable and policy-relevant, aligning with recommendations from EU policy reports and academic literature emphasising the use of official data sources for robust cross-country comparisons (Copeland, 2023; Liberati et al., 2023).

Third, the indicators had to be comparable across EU countries, with consistent definitions, methodologies, and disaggregation structures. This requirement excluded indicators with significant measurement gaps, methodological inconsistencies, or age groups that could not be isolated (De La Rasilla et al., 2024; UNICEF, 2024b). Specifically, child-level indicators in domains such as health, nutrition, or subjective well-being were excluded due to the absence of harmonised, child-specific data structures at EU level, which prevents consistent cross-country comparison. Ensuring methodological consistency is particularly important in clustering analysis, where spurious variation may distort typologies if data are inconsistently structured.

Finally, the selected indicators were evaluated based on their policy relevance and interpretability, especially concerning European social priorities and the Sustainable Development Goals. All indicators align with dimensions identified in prior child poverty frameworks (e.g., EU-MODA) or with key commitments under European anti-poverty strategies (EAPN, 2025). They were also selected to reflect material deprivation and institutional access (e.g., to education), supporting a broader conceptualisation of child well-being. Their inclusion thus ensures both conceptual validity and direct applicability for EU policy monitoring and reform.

3.3.2 Dimensions and Indicator List

The indicators selected for this study are grouped into three core dimensions of child-specific deprivation: housing deprivation, economic insecurity, and early education access. This dimensional structure draws on capability-based and rights-based frameworks and widely adopted approaches to multidimensional child poverty assessment, such as EU-MODA and similar regional applications.

The housing deprivation dimension captures physical inadequacies in children's living environments. Indicators reflect poor infrastructure, lack of basic amenities, insufficient lighting, and overcrowding, factors widely used in European deprivation studies and linked to adverse educational, health, and developmental outcomes.

The economic insecurity dimension includes indicators related to parental joblessness, energy poverty, and housing cost overburden. These aspects represent critical resource-based risks tied to cumulative and intergenerational disadvantage and are recognised in EU-level monitoring and poverty prevention frameworks.

The early education access dimension is represented by a single but conceptually significant indicator: participation in early childhood education. This factor is widely acknowledged as a protective determinant against later poverty and exclusion. Because this indicator originally reflects positive access (i.e., higher values imply lower deprivation), it was reverse coded prior to analysis to ensure consistency with the other indicators, where higher values reflect greater deprivation:

$$Non-participation = 100 - Participation\,rate$$

The complete list of selected indicators, along with their associated dimensions, Eurostat codes, descriptions, and variable names, is presented in Table 3.2. These variable names will be used consistently throughout the methodology and analysis for clarity and traceability.

Table 3.2 – Indicators by Dimension

DIMENSION	INDICATOR (Eurostat code)	DESCRIPTION	VARIABLE NAME
HOUSING DEPRIVATION	ilc_mdho01c	Children living in a dwelling with a leaking roof, damp walls, floors or foundation, or rot in window frames or floor	leaking_roof
	ilc_mdho02c	Children having neither a bath, nor a shower in their dwelling	no_bath
	ilc_mdho03c	Children not having indoor flushing toilet for the sole use of their household	no_toilet
	ilc_mdho04c	Children living in households considering their dwelling as too dark	too_dark
	ilc_chg05	Children living in overcrowded households	overcrowded
ECONOMIC INSECURITY	lfsi_jhh_a	Children living in jobless households	jobless
	ilc_chg04	Children living in households facing housing cost overburden	cost_overburden
	ilc_chg06	Children living in households unable to keep home adequately warm	not_warm
EARLY EDUCATION ACCESS	sdg_04_31[2]	Children (aged 3 and over) not participating in early childhood education	no_early_educ

While other domains, such as health, nutrition, or subjective well-being, are frequently included in multidimensional child poverty frameworks, they were not incorporated here due to the lack of harmonised, child-specific indicators in Eurostat. Numerous studies emphasise the conceptual importance of these additional dimensions, but without consistent EU-wide coverage, their inclusion would compromise the reliability and comparability of the analysis.

[2] This indicator is derived from Eurostat code sdg_04_31, but reverse-coded to reflect a deprivation outcome (i.e., non-participation)

This limitation is revisited in Section 6, which outlines directions for future data development and research.

3.3.3 HANDLING MISSING DATA

Although 2020 offers the most complete cross-country coverage for the selected indicators, some gaps remain. Before the exclusion of Lithuania, five of the nine indicators included missing values: four with a single missing value and one with two missing values. As explained in Section 3.3.1, Lithuania was excluded due to the absence of data for three key indicators. After this exclusion, the extent of missing data was significantly reduced. In the final analytical dataset, only two indicators remain incomplete: one with two missing values and another with one.

A targeted imputation strategy will be applied to address these residual gaps. Specifically, Germany and Poland lack data for the indicator too_dark, while Greece is missing data for no_early_educ. In the case of Greece, the most recent available value for this indicator is from 2019, when the country recorded the lowest participation rate in the EU. This may reflect a discontinuation in national data reporting rather than a temporary omission.

To ensure comparability while minimising bias, missing values will be imputed using regional averages. The regional groupings follow the EuroVoc geographical classification developed by the Publications Office of the European Union. Countries are grouped according to shared geographic and socio-political characteristics, as shown in Table 3.3.

Table 3.3 – Regional Grouping of EU Countries Based on EuroVoc

REGION	COUNTRIES
CENTRAL AND EASTERN EUROPE	BG, CZ, HR, HU, PL, RO, SI, SK
NORTHERN EUROPE	DK, EE, LV, LT[3], FI, SE
SOUTHERN EUROPE	EL, ES, IT, CY, MT, PT
WESTERN EUROPE	BE, DE, IE, FR, LU, NL, AT

Imputation will be conducted only when at least five countries in the same region have valid data for the indicator. This threshold ensures that regional averages are not overly sensitive to small group sizes, improving the robustness of the imputed values. Following best practices in multidimensional poverty studies, imputation will be carried out before data standardisation so as not to distort the scale of the analysis during subsequent clustering procedures. Additionally, all imputed values will be explicitly flagged in the dataset, and

[3] Lithuania (LT) is excluded from the final analysis due to extensive missing data across multiple indicators.

robustness checks will be conducted to assess the sensitivity of the cluster structure to these imputations.

No imputation was required for the AROPE dataset, as all seven disjoint AROPE components were fully available for the selected reference year. Therefore, the imputation procedure described in this section applies exclusively to the child-specific indicators.

3.4 Statistical Methods

3.4.1 Data Pre-processing

To ensure comparability across indicators and to avoid scale-driven distortions in subsequent distance-based analyses, all selected variables are standardised using z-scores. This transformation adjusts each indicator to have a mean of 0 and a standard deviation of 1, ensuring that no single variable disproportionately influences the grouping of countries due to differences in scale or variance. Standardisation is particularly important in this context because the subsequent analyses rely on Euclidean distance, and although all variables are expressed as percentages of children, their empirical distributions vary substantially across EU Member States.

Z-score standardisation is defined as:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

where:

- x_{ij} is the value of indicator j for country i,
- $\bar{x}_j$ is the mean of indicator j across all countries,
- s_j is the standard deviation of indicator j.

This transformation ensures that each indicator contributes equally in expectation to the Euclidean geometry of the feature space used in clustering.

3.4.2 Dimensionality Reduction with PaCMAP

Dimensionality reduction is used in this study as an exploratory tool to visualise structural patterns of similarity and dissimilarity across European Union Member States. Although clustering is performed in the full multidimensional indicator space, a two-dimensional embedding facilitates interpretation by revealing broad structural configurations that may not be immediately evident from numerical outputs alone.

For this purpose, the Pairwise Controlled Manifold Approximation Projection (PaCMAP) algorithm is applied to the z-score–standardised indicator matrix (Wang et al., 2021). Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^p$ denote the high-dimensional dataset, where x_i represents the vector of standardised indicator values for country i, p is the number of indicators, and N is the total number of observations (countries). PaCMAP seeks a low-dimensional embedding $Y = \{y_1, \ldots, y_N\} \subset \mathbb{R}^2$ that preserves both local neighbourhood relationships and global geometric structure.

PaCMAP constructs the embedding by defining three distinct types of point pairs in the high-dimensional space:

- **Near pairs**: each observation i is paired with a predefined number of nearest neighbours, denoted n_{NB}, selected using the scaled distance

$$d_{ij}^{2,\text{select}} = \frac{\| x_i - x_j \|^2}{\sigma_i \sigma_j},$$

where $\|\cdot\|$ denotes the Euclidean norm. The scaling factor σ_i is defined as the average Euclidean distance between observation i and its fourth, fifth, and sixth nearest neighbours in the original high-dimensional space, and σ_j is defined analogously for observation j. This local scaling corrects for heterogeneity in neighbourhood density across the feature space. To leverage efficient k-nearest-neighbour algorithms, the procedure first selects the $\min\,(n_{NB} + 50, N)$ closest observations to i based on Euclidean distance and then retains the n_{NB} observations with the smallest scaled distances $d_{ij}^{2,\text{select}}$. The scaled distance is used solely for neighbour selection and does not enter the optimisation stage.

- **Mid-near pairs**: for each observation i, six additional observations are randomly sampled from the dataset, and the second closest among them (according to Euclidean distance in the original space) is paired with i. The total number of mid-near pairs is defined as

$$n_{MN} = \lfloor n_{NB} \times \text{MN\_ratio} \rfloor,$$

where MN_ratio is a tuning parameter controlling the relative number of mid-near pairs, set to 0.5 by default.

- **Further pairs**: further pairs are constructed by randomly sampling non-neighbouring observations for each i. The total number of such pairs is given by

$$n_{FP} = \lfloor n_{NB} \times \text{FP\_ratio} \rfloor,$$

where FP_ratio is a turning parameter set to 2 by default.

Using these three sets of pairs, PaCMAP minimizes the following function:

$$\text{Loss}_{\text{PaCMAP}} = w_{NB} \sum_{(i,j)\in NB} \frac{\tilde{d}_{ij}}{10 + \tilde{d}_{ij}} + w_{MN} \sum_{(i,k)\in MN} \frac{\tilde{d}_{ik}}{10000 + \tilde{d}_{ik}} + w_{FP} \sum_{(i,l)\in FP} \frac{1}{1 + \tilde{d}_{il}},$$

where NB, MN, and FP denote the sets of near, mid-near, and further pairs, respectively. Indices i, j, k, l refer to observations in the dataset. The quantity

$$\tilde{d}_{ab} = \| y_a - y_b \|^2 + 1.$$

is the transformed squared Euclidean distance between embedded points y_a and y_b in the low-dimensional space.

Each loss term applies exclusively to its corresponding pair type. The near-pair term encourages local attraction between similar observations, the mid-near term guides medium-scale and global structure, and the further-pair term introduces repulsive forces that prevent collapse and improve separation between distant regions of the embedding. The weights w_{NB}, w_{MN}, and w_{FP} are updated dynamically during optimisation according to a three-phase schedule that initially emphasises mid-near pairs to establish global structure, subsequently balances global and local relationships, and finally prioritises local neighbourhood refinement, as described in Wang et al. (2021).

Optimisation is performed using the Adaptive Moment Estimation (Adam) stochastic gradient descent algorithm. Principal Component Analysis (PCA) is used to initialise the embedding to improve convergence speed, although PaCMAP is relatively insensitive to the choice of initial configuration.

In this study, PaCMAP is used exclusively for visualisation. All hierarchical and k-means clustering procedures are conducted in the original standardised indicator space. The resulting cluster memberships are subsequently projected onto the two-dimensional PaCMAP embedding to support interpretation and to visually assess the correspondence between numerical clusters and the underlying geometric structure of the data.

3.4.3 CLUSTER ANALYSIS

Cluster analysis is employed in this study to identify groups of European Union Member States with similar multidimensional child deprivation profiles. As an unsupervised learning method, clustering seeks to uncover latent structure directly from the data without imposing predefined group labels. To ensure robustness, interpretability, and consistency with the empirical analysis, a two-stage clustering strategy is adopted, combining Hierarchical Cluster Analysis and Non-Hierarchical Cluster Analysis.

All clustering procedures are conducted on the z-score-standardised indicator matrix. Standardisation ensures that each indicator contributes equally to the geometry of the feature space and prevents scale-driven distortions in distance computations. Pairwise dissimilarities between countries are quantified using Euclidean distance. For any two countries i and j, the Euclidean distance is defined as

$$d_{ij} = \| x_i - x_j \|_2 = \sqrt{\sum_{m=1}^{p} (x_{im} - x_{jm})^2},$$

where x_i and x_j denote the vectors of standardised indicator values for countries i and j, and $p$is the number of indicators. The resulting Euclidean distance matrix constitutes the fundamental input for the hierarchical clustering procedure and underpins all subsequent distance-based analyses. In the standardised space, Euclidean distance corresponds to the Mahalanobis distance with identity covariance structure, making it appropriate for indicators with comparable dispersion and measurement scales.

Step 1: Hierarchical Cluster Analysis (HCA)

In the first stage, Hierarchical Cluster Analysis is applied using Ward's Minimum Variance method, an agglomerative algorithm in which each country initially forms its own singleton cluster. At each iteration, the pair of clusters A and B that minimises the increase in the total within-cluster sum of squares is merged. The associated merger cost is given by

$$\Delta(A, B) = \frac{|A||B|}{|A| + |B|} \| \mu_A - \mu_B \|_2^2,$$

where $|A|$ and $|B|$ denote the number of countries in clusters A and B, and μ_A and μ_B are their respective centroids in the standardised indicator space. This formulation relies on Euclidean geometry, as the increase in within-cluster variance can be expressed as a function of the squared Euclidean distance between cluster centroids; consequently, Ward's method is only valid when Euclidean distance is used.

Ward's criterion favours compact, approximately spherical clusters and avoids the chaining effects commonly associated with alternative linkage rules. The hierarchical merging process is visualised using a dendrogram, where the height of each fusion reflects the marginal loss of within-cluster homogeneity associated with that merge. Large increases in fusion height indicate substantial structural separation between clusters and provide qualitative guidance regarding plausible cluster solutions. In this study, Hierarchical Cluster Analysis is used exploratorily to reveal the global structure of deprivation patterns and to inform the range of candidate numbers of clusters. Although Ward's method minimises increases in within-cluster variance at each merge, it does not globally optimise the clustering objective for a fixed number of clusters. For this reason, Hierarchical Cluster Analysis does not determine the final cluster solution.

Step 2: Non-Hierarchical Cluster Analysis (NHCA)

In the second stage, Non-Hierarchical Cluster Analysis is conducted using the k-means algorithm, which partitions the data into a fixed number k of clusters by directly minimising the total within-cluster sum of squared Euclidean distances,

$$\min_{C_1,\dots,C_K} \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|_2^2,$$

where C_k denotes the set of countries assigned to cluster k, μ_k is the corresponding cluster centroid, and x_i is the vector of indicator values for country i. Because this optimisation problem is non-convex, the final solution depends on the initial placement of centroids. Multiple random initialisations are therefore employed, and the configuration yielding the lowest within-cluster sum of squares is retained to ensure stability.

The number of clusters is selected using the silhouette coefficient, which provides a data-driven measure of cluster cohesion and separation. For each country i, the silhouette value is defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where $a(i)$ is the average distance between country i and all other countries in its own cluster, and $b(i)$ is the minimum average distance between country i and countries in any other cluster. The value of K that maximises the average silhouette score across all observations is selected as the final clustering solution.

Once the final partition is obtained, cluster centroids are used to characterise the average deprivation profile within each group. In addition, Euclidean distances between cluster centroids are computed in the standardised indicator space to assess the degree of separation between clusters,

$$d(\mu_k, \mu_{k'}) = \| \mu_k - \mu_{k'} \|_2.$$

Larger centroid distances indicate more pronounced structural differences between deprivation profiles. Numerical clustering results are complemented by graphical tools, including boxplots, heatmaps, and projections of cluster memberships onto the two-dimensional Pairwise Controlled Manifold Approximation Projection embedding, which serve as visual diagnostics but do not influence the clustering process.

To ensure methodological symmetry, this two-stage clustering procedure is applied independently to the nine child-specific deprivation indicators and to the seven disjoint components of the At Risk of Poverty or Social Exclusion framework, enabling a statistically consistent comparison between child-focused deprivation typologies and those derived from the official European Union poverty classification.

3.5 SOFTWARE AND IMPLEMENTATION

All analyses were conducted in Python using a Google Colab environment, ensuring reproducibility and a controlled computational setting. Data manipulation and preprocessing were performed using the pandas and numpy libraries. Standardisation of the indicators was implemented using the StandardScaler function from the scikit-learn library, producing z-score–standardised variables consistent with the assumptions underlying Euclidean distance–based clustering.

Hierarchical Cluster Analysis was implemented using the AgglomerativeClustering algorithm from the sklearn.cluster module with Ward's Minimum Variance linkage and Euclidean distance. This implementation corresponds directly to the hierarchical procedure described in Section 3.4, in which clustering is based on the Euclidean distance matrix of the standardised indicators and merges are selected to minimise increases in within-cluster sum of squares. Non-Hierarchical Cluster Analysis was conducted using the KMeans algorithm from the same module, with multiple random initialisations and a fixed random seed to ensure stability and replicability of the final partition.

Silhouette coefficients were computed using the silhouette_score function from sklearn.metrics to support the selection of the number of clusters. Cluster centroids and Euclidean distances between centroids were computed directly from the standardised data to characterise cluster profiles and assess between-cluster separation, in line with the analytical framework defined in Section 3.4.

Dimensionality reduction for visualisation was performed using the Pairwise Controlled Manifold Approximation Projection algorithm, implemented via the pacmap library. Default optimisation parameters were used to balance the preservation of local neighbourhoods and global geometric structure. The resulting two-dimensional embeddings were used exclusively for visual diagnostics and interpretation and did not influence the clustering procedures.

All analytical steps were executed within a single Python notebook to ensure transparency and full replicability. Visualisations, including dendrograms, PaCMAP embeddings, boxplots, and heatmaps, were produced using matplotlib and seaborn. The same implementation pipeline was applied consistently to both the nine child-specific deprivation indicators and the seven disjoint components of the At Risk of Poverty or Social Exclusion framework.

4. EMPIRICAL STUDY

This section applies the methodological approach outlined in section 3 to identify country-level patterns of child-specific deprivation across the EU. All analysis steps are conducted using the final dataset available in Appendix A (Table A.1), including the imputed values described in section 3.3.3, ensuring full country coverage across selected indicators.

4.1 DESCRIPTIVE STATISTICS OF CHILD DEPRIVATION INDICATORS

The descriptive statistics for the nine child-specific deprivation indicators in 2020 are presented in Table 4.1. These values summarise the level and variability of deprivation across EU Member States prior to standardisation. The indicator overcrowded exhibits the highest average level (mean = 25.83%) and the widest dispersion (SD = 18.94), confirming that insufficient living space is a common and highly heterogeneous form of deprivation. In contrast, no_bath and no_toilet show very low mean values (2.08% and 2.30%, respectively), with median values close to zero, indicating that severe sanitation deprivation is rare but highly concentrated in a small number of Member States.

Indicators related to household economic insecurity - jobless, cost_overburden, and not_warm - present moderate mean levels (between 6% and 8%) and substantial variability. The indicator no_early_educ, representing reverse-coded participation in early childhood education, also displays considerable variation (mean = 8.49%), with some countries reporting values above 20%. These descriptive patterns confirm the high degree of cross-country heterogeneity that the clustering procedures aim to summarise.

Table 4.1 - Descriptive Statistics

VARIABLE	MEAN	STD DEV	MAX	MIN	LOWER QUANTILE	MEDIAN	UPPER QUANTILE
leaking_roof	15,22	7,67	36,70	4,40	8,38	15,85	19,85
no_bath	2,08	5,30	25,60	0,00	0,10	0,25	1,05
no_toilet	2,30	5,87	26,20	0,00	0,12	0,45	1,03
too_dark	5,62	2,35	11,20	2,30	3,80	5,60	6,42
overcrowded	25,83	18,94	67,40	3,50	12,05	18,85	39,03
jobless	7,72	2,83	14,70	2,50	5,52	7,60	9,25
cost_overburden	6,55	7,81	42,40	1,10	3,15	4,70	6,22
not_warm	7,00	6,43	26,10	1,80	2,95	4,65	8,15
no_early_educ	8,49	5,76	21,90	0,00	5,55	7,75	10,45

Figure 4.1 presents the boxplots for all indicators. Several clear outliers emerge across the domains. Severe sanitation deprivation is concentrated in Bulgaria, Latvia, Hungary, and Romania, which appear as outliers in both no_bath and no_toilet. Cyprus is the only outlier

for leaking_roof, suggesting a comparatively higher prevalence of structural housing defects. For too_dark, Spain stands out with a distinctly higher proportion of children living in dwellings with insufficient natural light. In the economic strain dimension, Bulgaria and Greece are outliers for cost_overburden, and both countries, together with Cyprus, also appear as outliers in not_warm, indicating substantial heating and affordability challenges. Finally, no_early_educ reveals Croatia, Romania, and Slovakia as outliers, reflecting exceptionally high non-participation in early childhood education. These outlier cases likely reflect substantive cross-country differences rather than data recording errors and were retained in the analysis to preserve overall variation. The univariate distributions corresponding to each indicator are provided in Appendix A (Figure A.1).

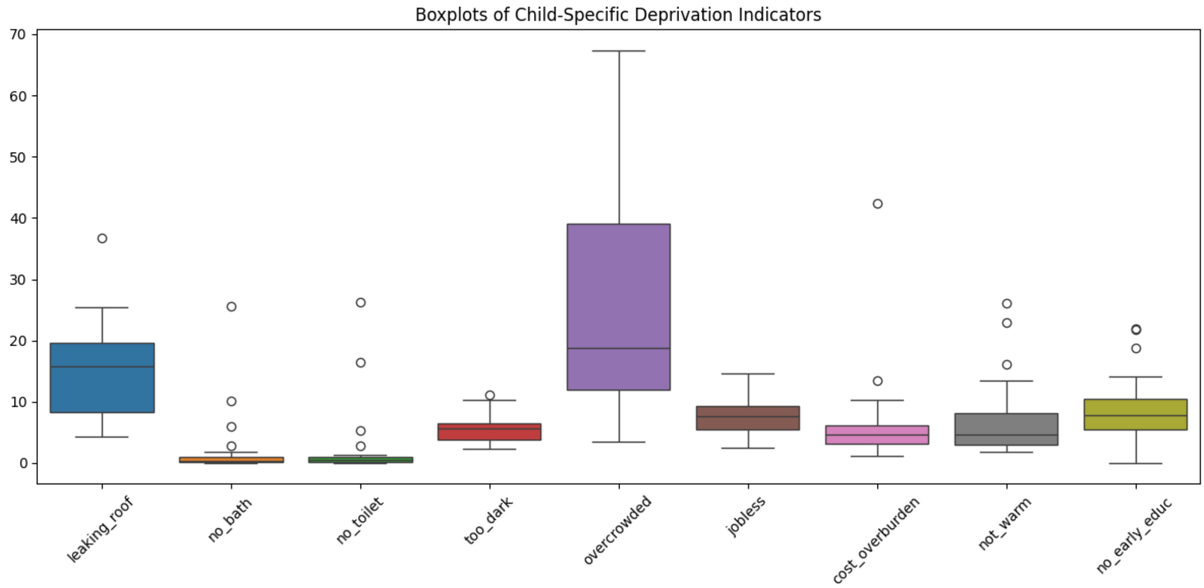

Figure 4.1 - Boxplots of the Child-Specific Deprivation Indicators

The correlation structure across indicators is shown in Figure 4.2**.** Most indicators exhibit weak to moderate correlations, confirming that they capture distinct aspects of child deprivation rather than redundant information. The strongest correlation appears between no_bath and no_toilet (r = 0.98), highlighting their shared underlying dimension of severe housing inadequacy. Moderate positive correlations are also observed between overcrowded and sanitation indicators, no_bath (r = 0.65) and no_toilet (r = 0.68), suggesting that overcrowding often co-occurs with other housing deficits. No_early_educ correlates positively with overcrowded (r = 0.52) but negatively with leaking_roof (r = –0.46) and jobless (r = –0.47), indicating that early education exclusion does not align consistently with other deprivation dimensions.

Although the very high correlation between *no_bath* and *no_toilet* suggests near-redundancy, sensitivity checks excluding one of these indicators show that the overall correlation structure and clustering results remain largely unchanged. Taken together, the descriptive statistics, outlier patterns, and correlation structure indicate substantial heterogeneity in child deprivation across EU Member States. On this basis, the full set of standardised indicators is retained for the subsequent clustering analysis, which aims to identify groups of countries with similar multidimensional deprivation profiles.

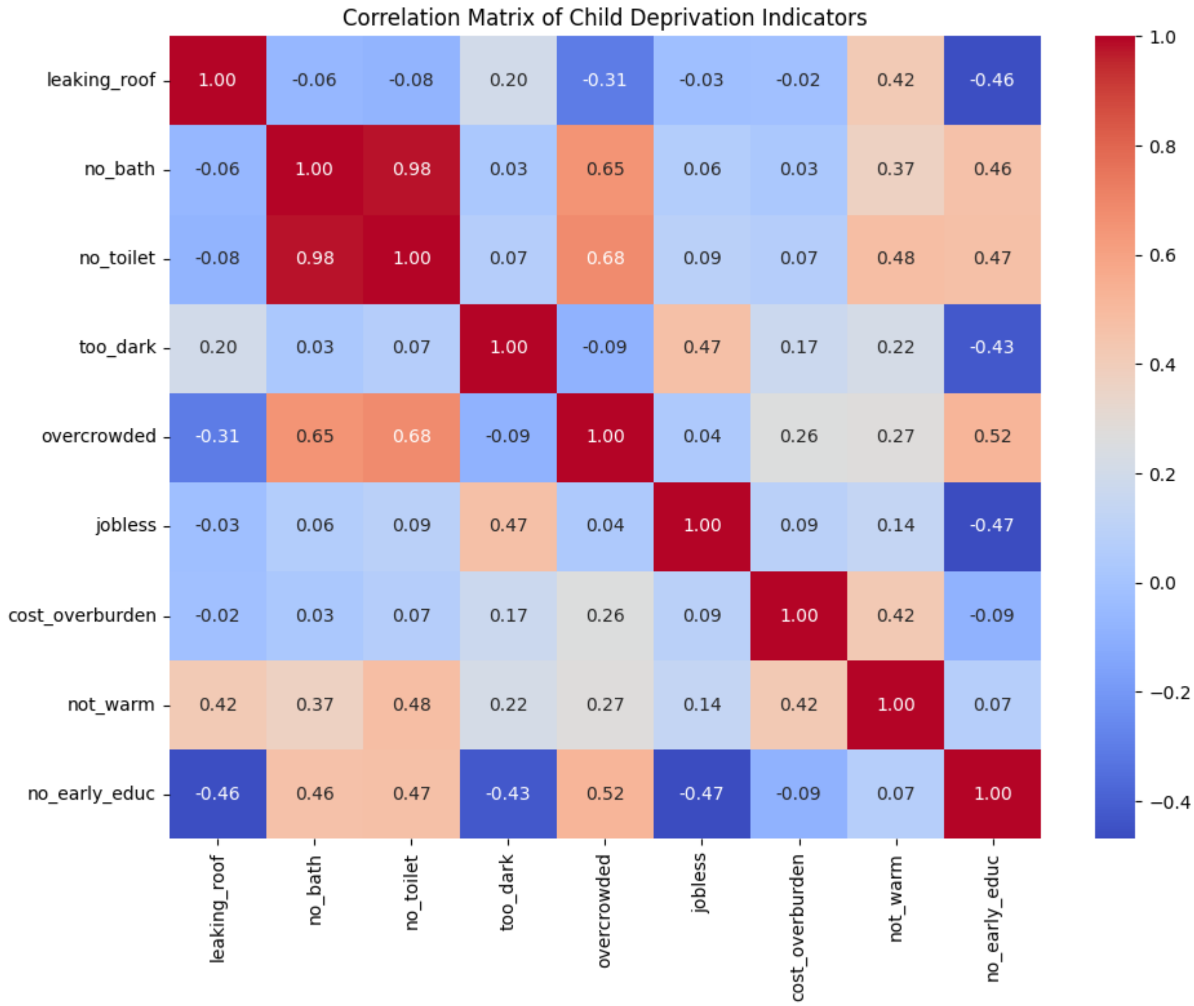

Figure 4.2 - Correlation Matrix of Child-Specific Deprivation Indicators

4.2 Indicator Standardisation

All indicators were standardised using z-scores prior to the clustering procedures, following the approach described in Section 3.4.1. Standardisation places all variables on a common scale by centring them at zero and rescaling them to unit variance, ensuring that no single indicator disproportionately influences the Euclidean distances used in subsequent analyses. The complete standardised dataset is provided in Appendix A (Table A.2). The resulting standardised dataset forms the basis for the PaCMAP embedding and clustering results presented in Sections 4.3 to 4.5.

4.3 PaCMAP Embedding

PaCMAP was applied to the standardised dataset to explore the global and local structure of multidimensional child deprivation across EU Member States. The resulting two-dimensional embedding is shown in Figure 4.3. As described in Section 3.4.2, PaCMAP constructs the low-dimensional representation by optimising the relative distances of near, mid-near, and further point pairs, allowing it to preserve both neighbourhood-level relationships and broader geometric structure. This balance makes PaCMAP particularly suitable for socio-economic datasets in which both fine-grained and macro-level differences among observations are analytically relevant.

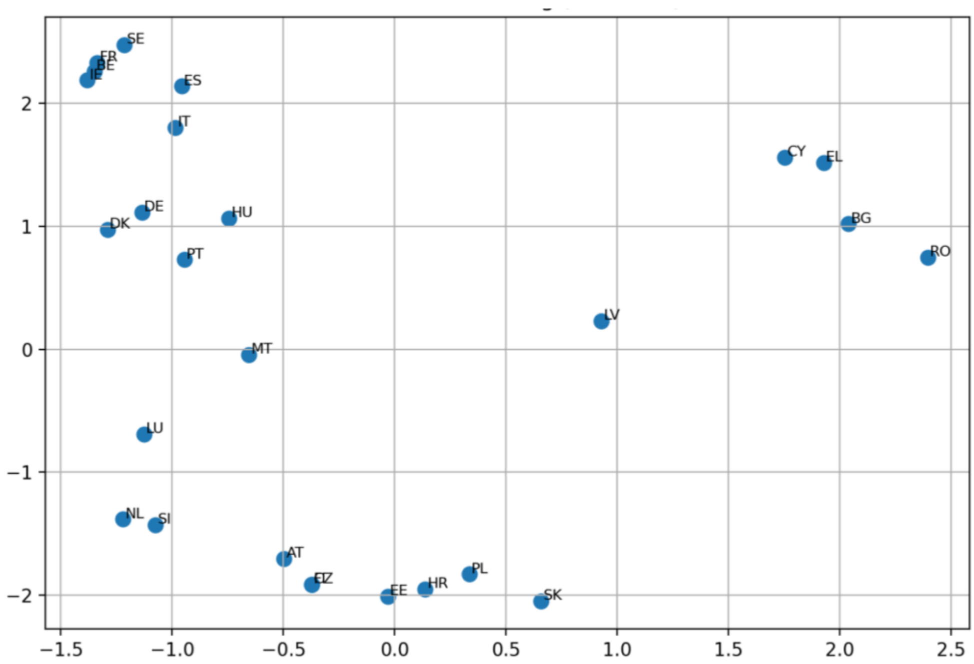

Figure 4.3 – PaCMAP embedding

The embedding reveals several substantive patterns. First, countries are widely dispersed across the two dimensions, indicating substantial heterogeneity in deprivation profiles. A small number of countries appear located at the periphery of the map. Their relative isolation is consistent with the extreme values detected in the descriptive statistics, particularly in indicators such as overcrowding, severe sanitation deprivation, inadequate heating, or non-participation in early childhood education.

Countries located in the central area of the map typically exhibit moderate values on most indicators. Their proximity suggests broadly similar deprivation structures, such as average levels of housing adequacy, economic insecurity, or early education access, although their positioning does not imply full homogeneity.

Broader regional tendencies are also visible. Several Northern and Western European countries are positioned toward the upper-left region, reflecting comparatively favourable deprivation profiles. In contrast, various Central and Eastern European countries appear toward the lower or right-hand areas of the embedding, consistent with higher levels of material deprivation or structural housing challenges. While these regional patterns are not absolute, they align with established socio-economic gradients documented in the European poverty literature.

The PaCMAP representation also highlights local structure. Countries positioned close together tend to share similar deprivation patterns across multiple dimensions simultaneously, whereas those located far apart differ on several indicators at once. This multi-scale interpretability illustrates why non-linear manifold learning methods offer advantages over examining individual indicators or bivariate plots alone.

It is important to emphasise that the PaCMAP embedding is used exclusively for exploratory purposes. Clustering is not performed in the reduced two-dimensional space. All formal clustering procedures, Hierarchical Cluster Analysis and k-means, are applied directly to the full set of standardised indicators to avoid distortions inherent to non-linear embeddings. Nonetheless, the PaCMAP projection provides valuable intuition about potential groupings, outlier structures, and global patterns, and it offers a complementary geometric perspective to the numerical clustering results presented in Sections 4.4 and 4.5.

4.4 Hierarchical Cluster Analysis (HCA)

Hierarchical Cluster Analysis (HCA) was conducted using Ward's Minimum Variance method on the standardised indicators. Ward's method was selected because it minimises the increase in total within-cluster variance at each agglomeration step, yielding compact and internally coherent clusters. This property makes it particularly suitable for multidimensional deprivation research, where clusters are expected to reflect distinct structural profiles rather than elongated or chained patterns.

The resulting dendrogram is shown in Figure 4.4. The vertical axis represents the fusion height, that is, the loss of within-cluster homogeneity incurred when two clusters merge. Large jumps in fusion height indicate substantial dissimilarity between the groups being combined and therefore signal potential boundaries between meaningful clusters.

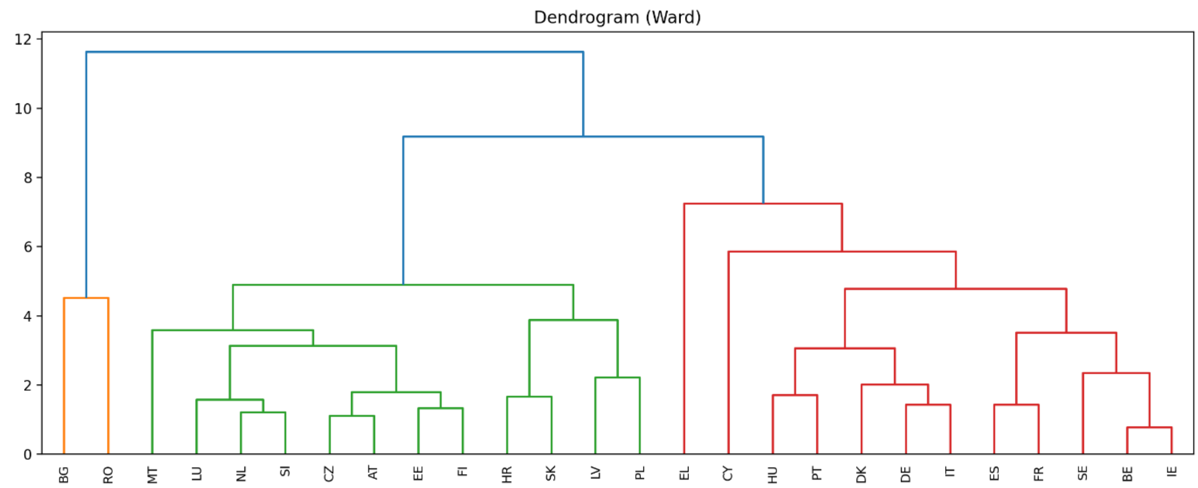

Figure 4.4 – Hierarchical clustering dendrogram (Ward's method)

Visual inspection of the dendrogram reveals one such major increase in fusion height, suggesting that a four-cluster solution provides a parsimonious and interpretable partition of the data. Cutting the dendrogram at this level yields four well-defined groups of countries:

- **Cluster 1:** BG, RO
- **Cluster 2:** AT, CZ, EE, FI, HR, LU, LV, MT, NL, PL, SI, SK
- **Cluster 3**: BE, CY, DE, DK, ES, FR, HU, IE, IT, PT, SE
- **Cluster 4:** EL

These groups correspond closely to the patterns observed in the descriptive statistics and in the PaCMAP embedding. Bulgaria and Romania consistently appear as a high-deprivation cluster across multiple indicators; Greece forms a distinct single-country outlier; several Central and Northern European countries group together with relatively favourable deprivation profiles; and a broader set of Western and Southern European countries forms an intermediate cluster.

Although HCA provides valuable insight into the global structure of the data, it is sensitive to early linkage decisions and does not directly optimise within-cluster variance for a given value of *k*. For this reason, HCA is treated purely as an exploratory step. The final cluster solution is obtained using k-means, following the silhouette-based selection of the number of clusters, as detailed in Section 4.5.

4.5 NON-HIERARCHICAL CLUSTER ANALYSIS (NHCA)

K-means clustering was applied to the standardised deprivation indicators to obtain the final classification of EU Member States. In contrast to HCA, which provides a descriptive view of the agglomeration structure, k-means directly optimises cluster membership by minimising the within-cluster sum of squared Euclidean distances. For this reason, it serves as the definitive clustering method in this study.

To determine the optimal number of clusters, silhouette coefficients were computed for solutions ranging from $k = 3$ to $k = 6$. The silhouette coefficient measures the degree to which a country is closer to the members of its own cluster than to the members of neighbouring clusters. As shown in Figure 4.5, the highest silhouette value (0.246) was obtained for $k = 4$, indicating that this solution offers the best balance between within-cluster cohesion and between-cluster separation.

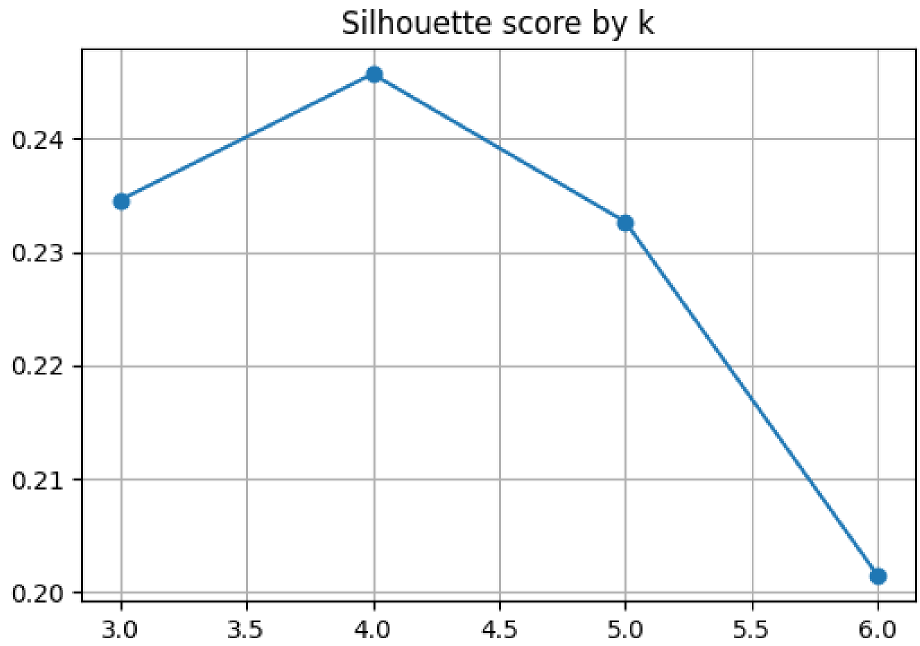

Figure 4.5 - Silhouette scores for k-means (k = 3–6)

Based on this result, k-means was estimated with four clusters. The algorithm was run with multiple random initialisations to ensure stability, and the final configuration was selected based on the lowest within-cluster sum of squares. The resulting clusters are:

- **Cluster 1:** BE, CY, DE, DK, ES, FR, HU, IE, IT, PT, SE
- **Cluster 2:** AT, CZ, EE, FI, HR, LU, LV, MT, NL, PL, SI, SK
- **Cluster 3:** EL
- **Cluster 4:** BG, RO

These clusters are identical to those derived from HCA, indicating strong structural stability across clustering methods. This consistency reinforces the robustness of the four-group typology.

Figure 4.6 presents the PaCMAP embedding coloured by k-means cluster assignments. The four clusters occupy distinct regions of the two-dimensional landscape, confirming that the numerical solution corresponds closely to the underlying geometric structure of the deprivation space. This alignment between the embedding and the clustering solution supports the validity of the partition.

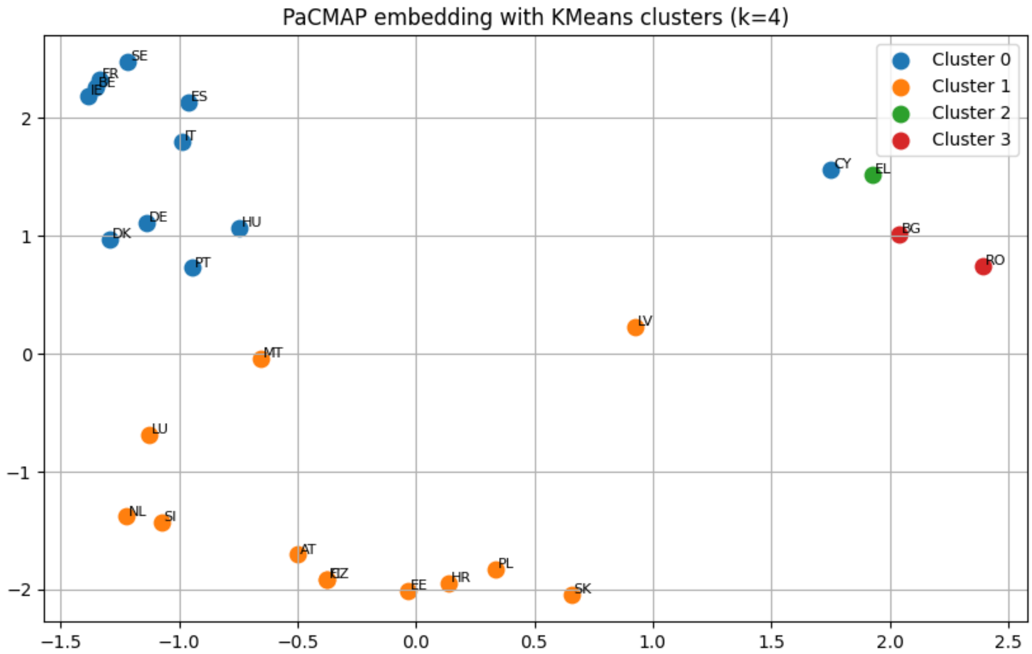

Figure 4.6 - PaCMAP embedding with k-means clusters

Table 4.2 reports the mean values of the nine indicators in their original scale for each cluster. Clear differences emerge across groups. Cluster 4 (BG, RO) represents the highest-deprivation group, with extreme values in overcrowding, sanitation deprivation, inadequate heating, and non-participation in early childhood education. Cluster 3 (EL) forms a distinct single-country outlier, characterised by very high cost_overburden and not_warm values alongside several elevated deprivation measures. Cluster 1 comprises mostly Western and Southern EU countries with moderate deprivation profiles, while Cluster 2, dominated by Central and Northern European countries, shows the most favourable deprivation levels across all dimensions.

Table 4.2 - Clusters Means

Cluster	leaking_roof	no_bath	no_toilet	too_dark	overcrowded	jobless	cost_overburden	not_warm	no_early_educ
1	20,77	0,59	0,54	6,89	16,99	9,42	5,62	8,26	4,17
2	10,77	0,96	0,91	4,32	26,10	5,98	3,89	2,96	11,23
3	12,40	0,30	0,50	5,90	43,20	7,80	42,40	16,20	6,70
4	12,75	17,90	21,30	6,35	64,20	8,80	9,70	19,80	16,75

To further assess separation between groups, Euclidean distances between cluster centroids were computed in the standardised indicator space (Table 4.3). The results show that Clusters 3 (EL) and 4 (BG–RO) are the most distant, with centroid distances exceeding 6, reflecting their notably distinct deprivation profiles. Clusters 1 and 2 are closer (distance = 2.668), consistent with their more moderate deprivation patterns. Overall, these centroid distances confirm the distinctiveness of the four groups and support the validity of the chosen solution.

Table 4.3 - Euclidean distances between k-means centroids

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0	2,668	5,348	6,358
Cluster 2	2,668	0	5,668	6,160
Cluster 3	5,348	5,668	0	6,907
Cluster 4	6,358	6,160	6,907	0

Taken together, the k-means analysis identifies four coherent and well-separated profiles of child-specific deprivation across EU Member States. These clusters constitute the final typology used in this study. Section 4.6 applies the same analytical pipeline to the AROPE components to compare these child-focused patterns with those derived from the official EU poverty measure.

Because centroid-based clustering methods are sensitive to extreme observations, an additional robustness check was conducted to assess the influence of multivariate outliers. Multivariate extremeness was evaluated using Euclidean distance in the standardised indicator space, and the three countries with the largest distances (Romania, Greece, and Bulgaria) were temporarily excluded. Re-estimating the clustering solutions without these observations did not materially alter the number of clusters, their composition, or their substantive interpretation. This suggests that the identified typology of child deprivation is not driven by a small number of extreme country profiles.

4.6 APPLICATION OF THE METHOD TO AROPE

To enable a systematic comparison between the child-specific deprivation typology and the EU's conventional poverty measure, the same analytical pipeline was applied to the AROPE framework. Instead of using the aggregate AROPE rate, the analysis relies on its seven mutually exclusive components, which jointly decompose all possible combinations of the three underlying sub-indicators: At Risk of Poverty (AROP), Severe Material and Social Deprivation (SMSD), and Very Low Work Intensity (VLWI). These seven components include three single-dimension categories, three double overlaps, and one triple overlap. Because the aggregate AROPE rate conceals differences in how deprivation dimensions overlap across countries, this disaggregated structure allows the clustering procedure to capture variation in the composition and intensity of deprivation rather than only its overall level. Lithuania was excluded due to missing disaggregated data. The complete dataset is shown in Appendix A (Table A.3). As in the child-specific analysis, all AROPE components were standardised using z-scores prior to the PaCMAP embedding and clustering procedures to ensure comparability across variables and prevent scale-driven distortions in Euclidean distance calculations.

The seven mutually exclusive AROPE components are defined as follows:

- **arop_only** — children in AROP only
- **smsd_only** — children in SMSD only
- **vlwi_only** — children in VLWI only
- **arop_smsd** — children simultaneously in AROP and SMSD
- **arop_vlwi** — children simultaneously in AROP and VLWI
- **smsd_vlwi** — children simultaneously in SMSD and VLWI
- **arop_smsd_vlwi** — children simultaneously experiencing all three deprivation dimensions

PaCMAP Embedding

PaCMAP was applied to the standardised AROPE components to examine the global and local structure of deprivation patterns. The resulting embedding (Figure 4.7) displays a clear tri-partite structure: a large cluster of countries with low component values, a smaller intermediate cluster, and a sharply isolated high-deprivation cluster (BG, EL, HU, RO). Compared with the child-specific indicators, the AROPE manifold shows stronger polarisation and clearer separation across the two-dimensional space.

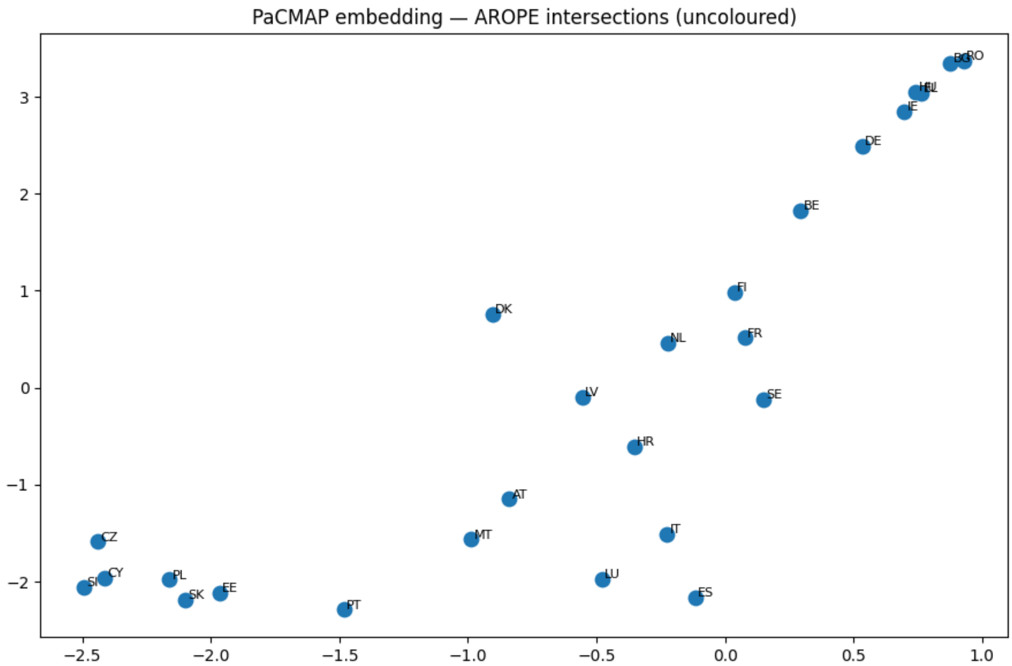

Figure 4.7 - PaCMAP Embedding of AROPE components

Hierarchical Cluster Analysis

Ward's minimum variance method was applied to the standardised AROPE components. The dendrogram (Figure 4.8) shows two prominent jumps in fusion height, indicating that a three-cluster structure provides a meaningful representation of the data.

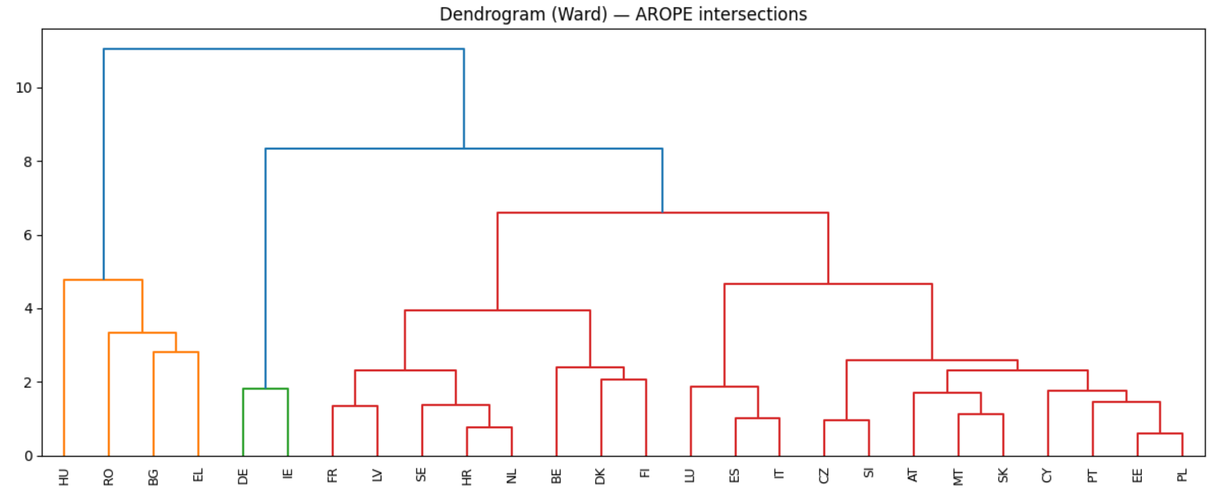

Figure 4.8 - Hierarchical clustering dendrogram of AROPE components

Cutting the dendrogram at the three-cluster level yields the following groups:

- **Cluster 1:** CZ, DK, EE, ES, FR, HR, IT, CY, LV, LU, MT, NL, AT, PL, PT, SI, SK, SE
- **Cluster 2:** BE, DE, IE, FI
- **Cluster 3:** BG, EL, HU, RO

These groups align with the patterns visible in the PaCMAP embedding: a large low-overlap cluster, an intermediate cluster, and a distinct high-deprivation cluster.

Non-Hierarchical Cluster Analysis

Silhouette coefficients were computed for $= 3$ to $k = 6$. The highest value (0.3758) occurred at $k = 3$, indicating strong internal cohesion and clear between-cluster separation. This silhouette value is substantially higher than the one obtained for the child-specific indicators (0.246), confirming a more polarised structure in the AROPE space.

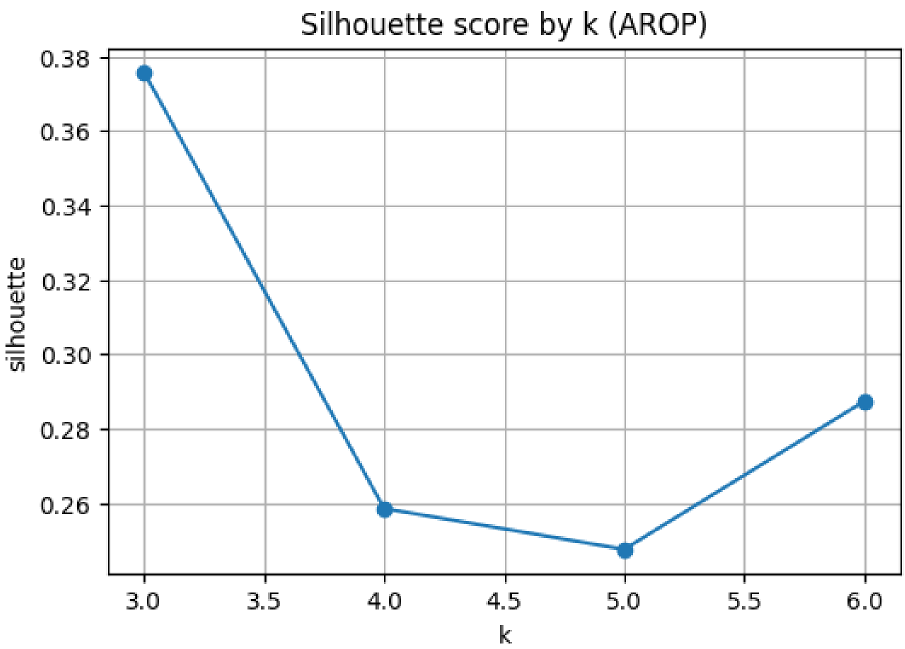

Figure 4.9 - Silhouette scores of AROPE components

K-means was then estimated with $k = 3$, using multiple random initialisations. The resulting classification replicates the HCA partitions exactly:

- **Cluster 1:** CZ, DK, EE, ES, FR, HR, IT, CY, LV, LU, MT, NL, AT, PL, PT, SI, SK, SE
- **Cluster 2:** BE, DE, IE, FI
- **Cluster 3:** BG, EL, HU, RO

The PaCMAP embedding with cluster assignments (Figure 4.10) shows three clearly separated areas of the manifold, confirming the strong geometric organisation of the AROPE components.

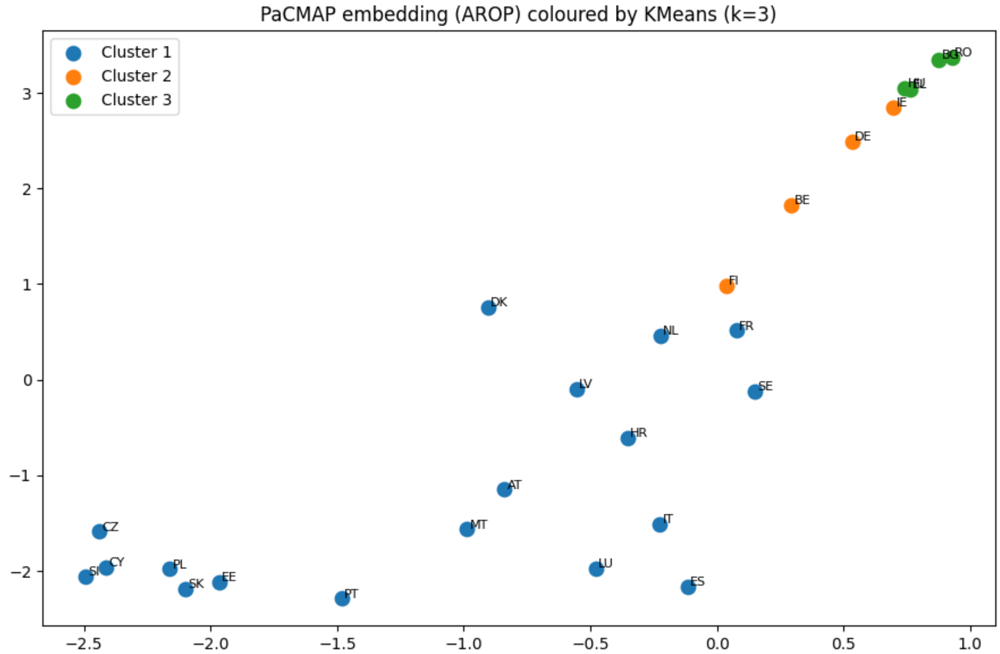

Figure 4.10 - PaCMAP Embedding with k-means clusters of AROPE components

Cluster means in the original scale (Table 4.4) reveal marked contrasts between groups:

- **Cluster 1** shows the lowest incidence across all categories, reflecting more favourable socio-economic conditions.
- **Cluster 2** exhibits moderate overlap, driven mainly by VLWI-related components.
- **Cluster 3** clearly represents the most deprived countries, with extremely high values in SMSD and triple-overlap indicators.

Table 4.4 - Cluster means of AROPE components

Cluster	arop_only	smsd_only	vlwi_only	arop_smsd	arop_vlwi	smsd_vlwi	arop_smsd_vlwi
1	11.106	1.361	1.017	1.756	3.017	0.172	1.472
2	7.425	1.825	3.025	1.025	4.250	1.100	1.625
3	7.325	8.675	0.825	8.800	1.525	0.500	4.550

Euclidean distances between cluster centroids seen in Table 4.5 confirm strong separation, cluster 3 is the most distant from both cluster 1 and cluster 2, highlighting its distinct disadvantage profile.

Table 4.5 – Euclidean distances between cluster centroids of AROPE components

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	0.000	3.445	4.259
Cluster 2	3.445	0.000	5.209
Cluster 3	4.259	5.209	0.000

It is important to acknowledge that k-means clustering assumes approximately spherical clusters with relatively homogeneous within-cluster variance in Euclidean space. Although the AROPE partitions are strongly separated, the PaCMAP embedding suggests that the underlying geometry of the data may not fully conform to these spherical assumptions, as some groups appear elongated or unevenly distributed in the two-dimensional representation. Consequently, differences between the visual manifold structure and the numerical k-means partition are not unexpected. The k-means solution should therefore be interpreted as a variance-minimising partition in Euclidean space rather than as a definitive representation of the intrinsic geometric structure of AROPE-based deprivation patterns.

5. RESULTS AND DISCUSSION

5.1 CLUSTER COMPARISON AND INTERPRETATION

The cluster analyses based on the child-specific deprivation indicators and the AROPE components reveal markedly different patterns of cross-country classification, reflecting the distinct conceptual foundations and dimensional structures of the two frameworks. Although both approaches identify broad gradients of deprivation across EU Member States, the composition, separation, and substantive meaning of the clusters diverge significantly.

Using the nine child-specific indicators, the analysis produced a four-cluster solution that captures a wide spectrum of deprivation profiles. This typology distinguishes between high-deprivation contexts, intermediate patterns characterised by mixed disadvantages, and low-deprivation countries with consistently favourable outcomes across most domains. The resulting clusters reflect meaningful differences in housing quality, sanitation, overcrowding, access to early childhood education, and energy deprivation. For example, Bulgaria and Romania form the highest-deprivation cluster, driven by severe sanitation deficits, extreme overcrowding, and low early education participation. Greece emerges as a distinctive single-country cluster, reflecting its unusually high housing cost overburden and heating deprivation. A larger group of Central and Northern European countries constitutes the lowest-deprivation profile, while several Western and Southern European Member States form a moderate-deprivation cluster with differentiated patterns across domains.

In contrast, the clustering of the seven disaggregated AROPE components yielded a three-cluster solution characterised by a more polarised structure. One large cluster contains the majority of EU Member States, reflecting low levels of income poverty, material and social deprivation, and low work intensity. A second, much smaller cluster groups Belgium, Germany, Ireland, and Finland, indicating moderate levels of risk across intersections of AROP, SMSD, and VLWI. The highest-risk cluster, comprising Bulgaria, Greece, Hungary, and Romania, is clearly separated from the rest of the distribution, driven by high rates of income poverty and overlapping forms of material deprivation and work-intensity exclusion. This more binary structure reflects the narrower dimensional scope of the AROPE framework and its strong reliance on income- and employment-based indicators.

Comparing the two typologies illustrates how the choice of indicators shapes cross-country classifications. Several Member States that appear similar under AROPE diverge markedly when child-specific deprivations are considered. For instance, Portugal, Italy, Spain, Belgium, and Germany are grouped together within the same AROPE cluster, but they fall into different clusters in the child-specific typology due to contrasts in housing quality, overcrowding rates, and early education access. These divergences imply different policy priorities. In countries where housing-related deprivation is pronounced, measures such as targeted housing renovation programmes, investment in social housing supply, and stricter enforcement of minimum housing standards could directly address structural deficits. Where overcrowding is prevalent, expanding affordable family housing and providing rental subsidies for larger households may be more effective. In contexts where non-participation in early childhood education is elevated, policies aimed at reducing enrolment costs, increasing the availability of public childcare places, and expanding outreach to disadvantaged families could help close access gaps. By contrast, where income poverty remains the dominant driver, reinforcing child benefits or targeted income-support schemes may be more appropriate. The comparison therefore highlights that relying solely on AROPE may lead to uniform policy prescriptions, whereas a child-specific approach supports more differentiated and context-sensitive interventions.

The centroid distances between clusters further highlight these differences. In the child-specific typology, moderate clusters are separated by relatively small distances, while the distances to the extreme clusters (Greece and the Bulgaria–Romania group) exceed six standardised units, indicating substantial differences in deprivation severity. In the AROPE-based typology, by contrast, the two low-risk clusters are relatively close together, while the high-risk cluster exhibits much larger distances from the rest. This reinforces the more polarised nature of the AROPE classification, whereby a small group of countries accounts for the majority of extreme poverty and social exclusion risks.

Interpreting the clusters within each framework reveals additional contrasts. The child-specific typology captures multiple dimensions of disadvantage that tend to co-occur in distinct ways across Member States, producing a richer set of profiles. Some countries experience severe housing and energy deprivations while performing moderately in joblessness; others show stronger outcomes in infrastructure but weaker performance in early education participation. The AROPE typology, however, primarily differentiates countries based on the intensity of income poverty and low work intensity, with limited capacity to identify multi-domain deprivation patterns that disproportionately affect children.

Overall, the comparison demonstrates that the child-specific indicators support a more nuanced and multidimensional understanding of deprivation across EU Member States, whereas the AROPE components produce a more compressed and polarised classification. Both methods identify a similar high-deprivation cluster, centred on Bulgaria and Romania, but diverge substantially in their classification of intermediate- and low-deprivation countries. These differences likely reflect structural contrasts in the underlying data, as both typologies were derived using identical clustering procedures applied to z-standardised variables. The geometric patterns underlying these contrasts are explored in Section 5.2.

5.2 Structural Insights from the PaCMAP Embedding

The PaCMAP embeddings provide additional insight into the structure of deprivation patterns that is not visible from the clustering results alone. Because PaCMAP preserves both local neighbourhoods and broader global relationships, the two-dimensional maps reveal how countries are positioned relative to one another in the underlying multidimensional space. These geometric patterns help explain why the clustering of child-specific indicators and the clustering of the AROPE components produce different typologies.

In the embedding based on the nine child-specific indicators, countries display wide dispersion across the two-dimensional space. This reflects the multidimensional nature of the data, where housing quality, sanitation, overcrowding, energy deprivation, joblessness, and early education access vary independently across Member States. Several countries occupy extreme positions: Bulgaria and Romania appear as a clearly separated pair due to high sanitation deprivation, severe overcrowding, and very low early education participation, while Greece lies in an isolated region of the embedding, driven by its unusually high housing cost overburden and heating difficulties. Meanwhile, a large group of Central and Northern European countries forms a compact low-deprivation neighbourhood, indicating similarly low disadvantage across most indicators. The embedding also shows intermediate substructures, countries that are not extreme outliers but nonetheless form distinct local groupings, highlighting gradations of deprivation that align with the four-cluster solution.

The PaCMAP embedding for the AROPE components reveals a markedly different geometry. Most Member States appear tightly compressed into a single dense region, reflecting the limited dimensional variation across the seven disjoint components of AROPE. Only a small number of countries - Bulgaria, Romania, Greece, and Hungary - fall clearly outside this region, forming a distinct high-deprivation group. This geometric polarisation mirrors the three-cluster AROPE solution, where the majority of EU countries cluster together while a small subset forms a sharply separated high-risk cluster. The absence of intermediate neighbourhoods or gradations in the embedding indicates that the AROPE framework captures fewer latent structural differences across countries, largely focusing on the intensity of overlapping income poverty, severe material and social deprivation, and low work intensity.

Comparing the two embeddings highlights the fundamental contrast between the child-specific and AROPE deprivation structures. The child-indicator embedding displays multiple coherent neighbourhoods, suggesting a continuous and multidimensional deprivation landscape with several meaningful intermediate profiles. In contrast, the AROPE embedding compresses most countries into a single cluster and isolates only the most deprived Member States, producing a largely binary separation between extreme and non-extreme cases. This geometric difference helps explain why the child-specific clustering supports four clusters while the AROPE-based clustering collapses into three: the richer variation in the child indicators creates more distinct structural regions in the space, whereas the AROPE components generate only one dense core and one extreme peripheral group.

Overall, the PaCMAP results reinforce the findings of the cluster comparison. The child-specific indicators exhibit a multidimensional structure with substantial internal differentiation, enabling a more nuanced typology of Member States. In contrast, the AROPE components display limited global dispersion and a highly concentrated low-risk region, resulting in a simpler and more polarised clustering. These embeddings therefore clarify the structural origins of the differences between the two typologies and demonstrate how the broader conceptual scope of the child indicators translates into a more complex and informative deprivation landscape.

6. CONCLUSIONS AND FUTURE RESEARCH

This thesis set out to develop a multidimensional, child-specific typology of deprivation across EU Member States using harmonised national-level indicators from Eurostat. By applying a combination of hierarchical clustering, k-means analysis, and PaCMAP embedding techniques, the study produced a detailed classification of cross-country deprivation profiles and compared these with the conventional AROPE-based typology. The results demonstrate that the choice of indicators has a significant effect on how countries are grouped, how deprivation is conceptualised, and how policy-relevant inequalities are identified.

The child-specific indicators generated a four-cluster solution that captured a broad spectrum of deprivation patterns across the EU. These clusters reveal that child poverty is multidimensional and that several Member States share similar levels of disadvantage despite differing income profiles. The AROPE components, by contrast, produced a more polarised three-cluster solution, with most countries concentrated in a single low-risk cluster and a small number forming a clearly separated high-risk group. This difference reflects the narrower conceptual scope of AROPE and its reliance on income poverty, material deprivation, and low work intensity as its primary dimensions.

The PaCMAP embeddings further clarified the structural differences between the two typologies. While the child-specific indicators formed a dispersed, multidimensional landscape with several meaningful substructures, the AROPE components produced a compressed configuration with limited internal variation. These findings highlight that child deprivation, when measured through housing conditions, energy poverty, early education access, and related indicators, reveals patterns that remain invisible when relying solely on income-based measures. The analysis therefore demonstrates the added value of multidimensional, child-focused approaches for identifying structural disadvantages across EU Member States.

From a policy perspective, the study underscores the importance of incorporating child-specific deprivation measures into EU-wide monitoring frameworks. While AROPE plays a central role in the European Semester and SDG monitoring, its limited dimensional scope risks underestimating important aspects of child well-being. A multidimensional approach, particularly one grounded in domains such as housing, energy access, and early education, can contribute to more targeted national and EU-level policies and strengthen the alignment with the European Pillar of Social Rights and the EU's ambition to "leave no child behind."

Despite its contributions, the study faces several limitations that offer avenues for future research. First, the analysis relies on national-level aggregates, which limits the ability to capture intra-country disparities and the lived experiences of children within households. Access to EU-SILC microdata would allow for a more granular, child-level analysis using methods such as the Alkire-Foster approach or fuzzy-set multidimensional indices. Second, some relevant dimensions, such as child health, subjective well-being, and access to services, could not be included due to gaps in harmonised child-disaggregated data. Developing consistent EU-wide indicators in these areas is essential for more comprehensive monitoring. Third, the study applied clustering techniques using Euclidean distance and z-score standardisation; future work could explore alternative distance metrics, probabilistic clustering, or model-based approaches that account for uncertainty and indicator interactions. Finally, extending the analysis longitudinally would allow researchers to study the evolution of child deprivation over time and assess the impact of major socioeconomic events, such as inflationary shocks or the COVID-19 pandemic.

In summary, this thesis provides a robust, data-driven typology of child-specific deprivation across the European Union and demonstrates the analytical value of multidimensional, non-income-based indicators. The findings highlight clear differences between the child-focused typology and the AROPE framework, offering new insights for research, monitoring, and policy design. A more consistent integration of child-specific, multidimensional indicators into EU policy frameworks would support more equitable and evidence-based strategies to improve child well-being across Member States.

7. BIBLIOGRAPHICAL REFERENCES

Aguayo, I. H., Díaz Herráiz, E., Montenegro Marques, E., Machado, I., & Almeida, S. (2016). Child at Risk of Poverty or Social Exclusion: Comparative View Between Spain and Portugal in the European Context. *Social Indicators Research*, *129*(3), 961–978. https://doi.org/10.1007/s11205-015-1119-y

Alkire, S., & Apablaza, M. (2016). *Multidimensional Poverty in Europe 2006–2012: Illustrating a Methodology*.

Alkire, S., Kanagaratnam, U., Nogales, R., & Suppa, N. (2022). Revising the Global Multidimensional Poverty Index: Empirical Insights and Robustness. *Review of Income and Wealth*, *68*(S2). https://doi.org/10.1111/roiw.12573

Antošová, I., Hazuchová, N., & Stávková, J. (2021). Income situation of agricultural households of EU countries. *Agricultural Economics (Zemědělská Ekonomika)*, *67*(4), 121–128. https://doi.org/10.17221/474/2020-AGRICECON

Arcarons, A. F., Lanau, A., Marí-Klose, P., & Muñoz-Comet, J. (2025). Migrant Origin Children and Child Poverty in Spain: A Decomposition Analysis. *Child Indicators Research*. https://doi.org/10.1007/s12187-025-10228-8

Bárcena-Martín, E., Pérez-Moreno, S., & Rodríguez-Díaz, B. (2020). Rethinking multidimensional poverty through a multi-criteria analysis. *Economic Modelling*, *91*, 313–325. https://doi.org/10.1016/j.econmod.2020.06.007

Bessell, S. (2021). Rethinking Child Poverty. *Journal of Human Development and Capabilities*, *23*(4), 539–561. https://doi.org/10.1080/19452829.2021.1911969

Bienkunska, A., & Piasecki, T. (2022). *Measuring Objective and Subjective Aspects of Poverty. Discussion of Different Approaches and Measurement Methods Based on the Polish EU-SILC Survey*.

Biggeri, M., & Cuesta, J. A. (2020). An Integrated Framework for Child Poverty and Well-Being Measurement: Reconciling Theories. *Child Indicators Research*, *14*(2), 821–846. https://doi.org/10.1007/s12187-020-09774-0

Chzhen, Y., & Bruckauf, Z. (2017). *Sustainable Development Goal 1.2: Multidimensional child poverty in the European Union*.

Chzhen, Y., Bruckauf, Z., & Toczydlowska, E. (2018). Monitoring progress towards sustainable development: Multidimensional child poverty in the European Union. *Journal of Poverty and Social Justice*, *26*(2), 129–150. https://doi.org/10.1332/175982718X15154249173514

Chzhen, Y., De Neubourg, C., Plavgo, I., & De Milliano, M. (2015). Child Poverty in the European Union: The Multiple Overlapping Deprivation Analysis Approach (EU-MODA). *Child Indicators Research*, *9*(2), 335–356. https://doi.org/10.1007/s12187-015-9321-7

Clerici, E. (2022). *Measuring social exclusion: Strengths and limits of the European indicator AROPE*.

Cling, J.-P., Eghbal-Teherani, S., Orzoni, M., & Plateau, C. (2020). The interlinkages between the SDG indicators and the differentiation between EU countries: It is (mainly) the economy! *Statistical Journal of the IAOS*, *36*(2), 455–470. https://doi.org/10.3233/SJI-190507

Copeland, P. (2023). Poverty and social exclusion in the EU: Third-order priorities, hybrid governance and the future potential of the field. *Transfer: European Review of Labour and Research*, *29*(2), 219–233. https://doi.org/10.1177/10242589231171091

D'Agostino, A., Giusti, C., & Potsi, A. (2018). Gender and Children's Wellbeing: Four Mediterranean Countries in Perspective. *Child Indicators Research*, *11*(5), 1649–1676. https://doi.org/10.1007/s12187-018-9526-7

Darvas, Z. (2018). Why is it So Hard to Reach the EU's Poverty Target? *Social Indicators Research*, *141*(3), 1081–1105. https://doi.org/10.1007/s11205-018-1872-9

De La Rasilla, P., Stamos, I., Proietti, P., & Siragusa, A. (2024). Closing the Loop: Enhancing Local Monitoring of Child Poverty to Leave No Child Behind. *Children*, *11*(1), 67. https://doi.org/10.3390/children11010067

Dirksen, J., & Alkire, S. (2021). Children and Multidimensional Poverty: Four Measurement Strategies. *Sustainability*, *13*(16), 9108. https://doi.org/10.3390/su13169108

Domínguez-Serrano, M., & Del Moral-Espín, L. (2022). The Capability Approach and Child Well-Being: A Systematic Literature Review. *Child Indicators Research*, *15*(6), 2043–2063. https://doi.org/10.1007/s12187-022-09953-1

Drago, C. (2021). The Analysis and the Measurement of Poverty: An Interval-Based Composite Indicator Approach. *Economies*, *9*(4), 145. https://doi.org/10.3390/economies9040145

Duffy, K. (2020). What is poverty and how to combat it? *EAPN UK*. https://www.eapn.eu/wp-content/uploads/2020/04/EAPN-Poverty-Explainer-Web-1-4331.pdf

EAPN. (2025). *EU 2024 Poverty Watch*. https://www.eapn.eu/wp-content/uploads/2025/04/eapn-EU-2024-PovertyWatchReport-6131.pdf

Eurochild. (2023). *2023 Spring Package of the European Semester* [Policy Paper]. Eurochild.

Eurostat. (n.d.). *Glossary: At risk of poverty or social exclusion (AROPE)*. Eurostat - Statistics Explained. Retrieved 30 January 2025, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At_risk_of_poverty_or_social_exclusion_(AROPE)

Fabbris, L. (2024). COUNTING THE POOR IN ITALY AND EU. *Statistica Applicata – Italian Journal of Applied Statistics*, *35*(3), 1. https://doi.org/10.26398/IJAS.0035-013

Frączek, B. (2022). The Use of Cluster Analysis to Assess the Threats of Poverty or Social Exclusion in EU Countries: The Case of People with Disabilities Compared to People without Disabilities. *Sustainability*, *14*(21), 14223. https://doi.org/10.3390/su142114223

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd edn). Springer.

Kalinowski, S., & Kiełbasa, B. (2017, February 15). RISK OF POVERTY AND SOCIAL EXCLUSION IN THE EUROPEAN UNION IN THE CONTEXT OF SUSTAINABLE DEVELOPMENT. *Proceedings of the 8th International Scientific Conference Rural Development 2017*. RURAL DEVELOPMENT, Aleksandras Stulginskis University, Lithuania. https://doi.org/10.15544/RD.2017.044

Karagiannaki, E., & Burchardt, T. (2024). Living Arrangements, Intra-Household Inequality and Children's Deprivation: Evidence from EU-SILC. *Child Indicators Research*, *17*(5), 2319–2359. https://doi.org/10.1007/s12187-024-10149-y

Kumar, S., Chakraverty, S., & Sethi, N. (2023). Multidimensional Poverty: CMPI Development, Spatial Analysis and Clustering. *Social Indicators Research*, *169*(1–2), 647–670. https://doi.org/10.1007/s11205-023-03181-y

Liberati, P., Resce, G., & Tosi, F. (2023). The probability of multidimensional poverty: A new approach and an empirical application to EU-SILC data. *Review of Income and Wealth*, *69*(3), 668–700. https://doi.org/10.1111/roiw.12598

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. https://arxiv.org/abs/1802.03426

Premrov, T., Varlamova, M., Fuchs, M., & Sandu, V. (2022). *An index to measure child poverty in European cities: City of Vienna case study*.

Robeyns, I., & Byskov, M. F. (2025). *The Capability Approach*. https://plato.stanford.edu/entries/capability-approach/

Tuñón, I., Lamarmora, G., & Sánchez, M. E. (2022). Pobreza multidimensional infantil en España y Argentina. Un ejercicio de construcción, compatibilización y análisis de robustez. *Empiria. Revista de metodología de ciencias sociales*, *55*, 57–96. https://doi.org/10.5944/empiria.55.2022.34182

UNICEF. (2024a). *In Focus: Ending Child Poverty*. https://www.unicef.org/eca/media/37981/file/In%20Focus%20Ending%20child%20poverty.pdf

UNICEF. (2024b). *Policy Brief 1: Child Poverty*. https://www.unicef.org/eu/media/2571/file/Child%20poverty%20brief.pdf

van der Maaten, L., & Hinton, G. (2008). Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Vasilescu, M. D., Stănilă, L., Popescu, M. E., Militaru, E., & Marin, E. (2024). Using panel data clustering regression analysis to revisit income inequalities in the European Union. *Applied Economics Letters*, 1–6. https://doi.org/10.1080/13504851.2024.2358186

Vuorenlinna, E., Välimäki, S., Lindberg, M., Sarkia, A., Hakovirta, M., & Nygård, M. (2023). Poverty as a Driver of Stigma among Finnish Children during the Covid-19 Pandemic–Evidence from the 2021 Children's Voice Survey. *Child Indicators Research*, *16*(6), 2631–2652. https://doi.org/10.1007/s12187-023-10069-3

Wang, S., Sun, C., Wan, C., & Shao, B. (2021). PaCMAP: Enhancing Structure Preservation in Manifold Learning. *Proceedings of The Web Conference 2021*, 638–647. https://doi.org/10.1145/3442381.3449838

Zieleńska, M., & Wnuk, M. (2024). Benchmarking and the Technicization of Academic Discourse: The Case of the EU at-Risk of Poverty or Social Exclusion Composite Indicator. *Minerva*, *63*(1), 23–43. https://doi.org/10.1007/s11024-024-09537-z

A. APPENDIX A

Table A.1 - Final Dataset

country	leaking_roof	no_bath	no_toilet	too_dark	overcrowded	jobless	cost_overburden	not_warm	no_early_educ
BE	18,3	0,1	0,4	6,2	9,7	12,9	4,8	5,1	1,7
BG	12,9	10,2	16,4	6,8	61,0	9,6	13,5	26,1	11,7
CZ	7,8	0,2	0,3	3,7	24,5	5,4	6,0	2,0	14,2
DK	20,4	1,8	0,2	4,2	11,9	7,6	5,8	3,3	2,4
DE	15,3	0,0	0,0	**6,1**	16,8	8,8	7,9	8,2	6,3
EE	8,2	1,9	1,3	2,5	20,9	7,9	3,1	2,1	8,1
IE	17,8	0,0	0,2	6,1	5,0	11,1	4,4	4,2	0,0
EL	12,4	0,3	0,5	5,9	43,2	7,8	42,4	16,2	**6,7**
ES	21,3	0,2	0,5	11,2	12,8	9,4	10,3	10,6	2,8
FR	21,3	0,5	0,6	10,3	15,0	11,8	3,8	8,0	0,0
HR	8,1	0,3	0,8	3,6	49,2	4,2	3,0	3,4	18,8
IT	18,1	0,7	0,7	5,5	39,2	10,4	7,7	7,6	5,4
CY	36,7	0,1	0,1	3,2	3,5	6,0	1,5	23,0	8,9
LV	18,3	6,0	5,3	4,4	58,2	8,1	3,7	4,4	6,0
LU	19,7	0,0	0,0	5,7	12,5	5,3	8,2	3,1	10,5
HU	24,9	2,9	2,9	8,5	33,7	5,9	5,7	6,0	7,2
MT	4,7	0,0	0,2	9,3	6,3	6,5	2,6	5,2	10,9
NL	16,4	0,1	0,0	3,0	6,1	4,9	3,3	2,9	8,3
AT	10,4	0,8	0,7	5,2	23,9	6,6	6,3	1,8	10,3
PL	6,1	1,1	1,1	**5,0**	48,1	7,6	2,5	2,1	9,2
PT	25,5	0,2	0,3	6,5	16,5	5,0	5,3	11,3	7,1
RO	12,6	25,6	26,2	5,9	67,4	8,0	5,9	13,5	21,8
SI	19,2	0,1	0,1	4,1	16,0	2,5	3,9	1,8	7,4
SK	4,4	0,9	1,1	2,3	38,5	7,6	3,0	4,9	21,9
FI	5,9	0,1	0,0	3,0	9,0	5,2	1,1	1,8	9,1
SE	8,9	0,0	0,0	8,0	22,8	14,7	4,6	3,5	4,1

Table A.2 - Standardised Dataset

country	leaking_roof	no_bath	no_toilet	too_dark	overcrowded	jobless	cost_overburden	not_warm	no_early_educ
BE	0.410378	-0.38124	-0.33085	0.250267	-0.86878	1.865512	-0.22841	-0.30192	-1.20356
BG	-0.30804	1.562724	2.44959	0.510545	1.893505	0.676352	0.907133	3.028306	0.568386
CZ	-0.98654	-0.362	-0.34822	-0.83422	-0.07186	-0.83712	-0.07179	-0.79352	1.011373
DK	0.689762	-0.05404	-0.3656	-0.61733	-0.75032	-0.04435	-0.09789	-0.58736	-1.07953
DE	0.011257	-0.40049	-0.40036	0.206887	-0.48648	0.388071	0.176206	0.189688	-0.38847
EE	-0.93333	-0.03479	-0.17445	-1.35478	-0.26571	0.063754	-0.4503	-0.77766	-0.06951
IE	0.343858	-0.40049	-0.3656	0.206887	-1.12185	1.216879	-0.28062	-0.44464	-1.50479
EL	-0.37456	-0.34275	-0.31347	0.120128	0.935051	0.027719	4.679238	1.458344	-0.31759
ES	0.809498	-0.362	-0.31347	2.419249	-0.70186	0.604282	0.48946	0.570285	-1.00865
FR	0.809498	-0.30425	-0.29609	2.028832	-0.5834	1.469125	-0.35894	0.157972	-1.50479
HR	-0.94663	-0.34275	-0.26133	-0.8776	1.258125	-1.26955	-0.46336	-0.5715	1.826469
IT	0.38377	-0.26576	-0.27871	-0.05339	0.719668	0.964633	0.150101	0.094539	-0.54794
CY	2.858316	-0.38124	-0.38298	-1.05112	-1.20262	-0.62091	-0.65914	2.536701	0.072241
LV	0.410378	0.754342	0.520663	-0.53057	1.742737	0.135825	-0.37199	-0.41292	-0.44162
LU	0.596634	-0.40049	-0.40036	0.033369	-0.71801	-0.87316	0.215362	-0.61908	0.355753
HU	1.288442	0.157679	0.103598	1.247999	0.423517	-0.65695	-0.11094	-0.15919	-0.22899
MT	-1.39897	-0.40049	-0.3656	1.595036	-1.05185	-0.44074	-0.51556	-0.28606	0.426631
NL	0.157601	-0.38124	-0.40036	-1.13788	-1.06262	-1.0173	-0.4242	-0.6508	-0.03408
AT	-0.64064	-0.24651	-0.27871	-0.18353	-0.10417	-0.4047	-0.03263	-0.82524	0.320314
PL	-1.21271	-0.18877	-0.2092	-0.27029	1.198895	-0.04435	-0.52862	-0.77766	0.125399
PT	1.368266	-0.362	-0.34822	0.380406	-0.50263	-0.98126	-0.16315	0.681293	-0.24671
RO	-0.34795	4.526793	4.152606	0.120128	2.238117	0.09979	-0.08484	1.030173	2.358054
SI	0.530114	-0.38124	-0.38298	-0.66071	-0.52955	-1.88214	-0.34589	-0.82524	-0.19355
SK	-1.43888	-0.22726	-0.2092	-1.44154	0.681976	-0.04435	-0.46336	-0.33363	2.375773
FI	-1.23932	-0.38124	-0.40036	-1.13788	-0.90647	-0.90919	-0.71135	-0.82524	0.10768
SE	-0.8402	-0.40049	-0.40036	1.0311	-0.1634	2.514144	-0.25452	-0.55565	-0.77829

Table A.3 - AROPE Dataset

country	arop_only	smsd_only	vlwi_only	arop_smsd	arop_vlwi	smsd_vlwi	arop_smsd_vlwi
BE	6,7	2,9	2,8	1,6	4,4	0,7	2,9
BG	7,9	6,2	0,6	10,7	2,3	0,5	7,4
CZ	7,3	0,9	0,8	0,6	2,1	0,1	1,1
DK	5,1	1,4	1,8	1,1	1,9	0,2	1,9
DE	9,6	2,2	3,2	1,0	3,6	1,4	1,2
EE	12,6	0,8	1,1	0,8	1,7	0,2	0,1
IE	7,1	2,0	3,6	1,0	4,8	2,0	1,8
EL	9,0	7,8	1,6	6,2	1,2	0,5	4,5
ES	16,8	3,0	0,8	4,7	3,5	0,3	2,5
FR	8,9	2,6	1,3	2,4	5,1	0,2	3,0
HR	10,0	0,4	1,2	1,3	4,6	0,0	1,0
IT	16,6	2,5	1,0	2,5	4,2	0,4	1,8
CY	10,3	1,6	0,7	3,2	1,5	0,6	1,0
LV	8,4	2,6	0,9	1,7	3,7	0,4	2,0
LU	16,7	0,1	1,1	2,1	3,7	0,0	0,7
HU	2,6	10,6	0,8	3,1	1,7	0,8	2,0
MT	12,4	1,1	0,9	2,8	2,5	0,0	2,8
NL	7,8	0,3	1,3	0,5	4,9	0,1	1,0
AT	11,9	1,7	1,7	1,8	2,8	0,1	1,9
PL	11,0	1,4	1,1	0,5	1,6	0,1	0,4
PT	14,4	1,5	1,0	2,1	1,1	0,2	1,5
RO	9,8	10,1	0,3	15,2	0,9	0,2	4,3
SI	8,2	1,0	0,3	0,6	1,5	0,2	0,3
SK	11,6	1,1	0,3	1,9	1,6	0,0	2,0
FI	6,3	0,2	2,5	0,5	4,2	0,3	0,6
SE	9,9	0,5	1,0	1,0	6,3	0,0	1,5

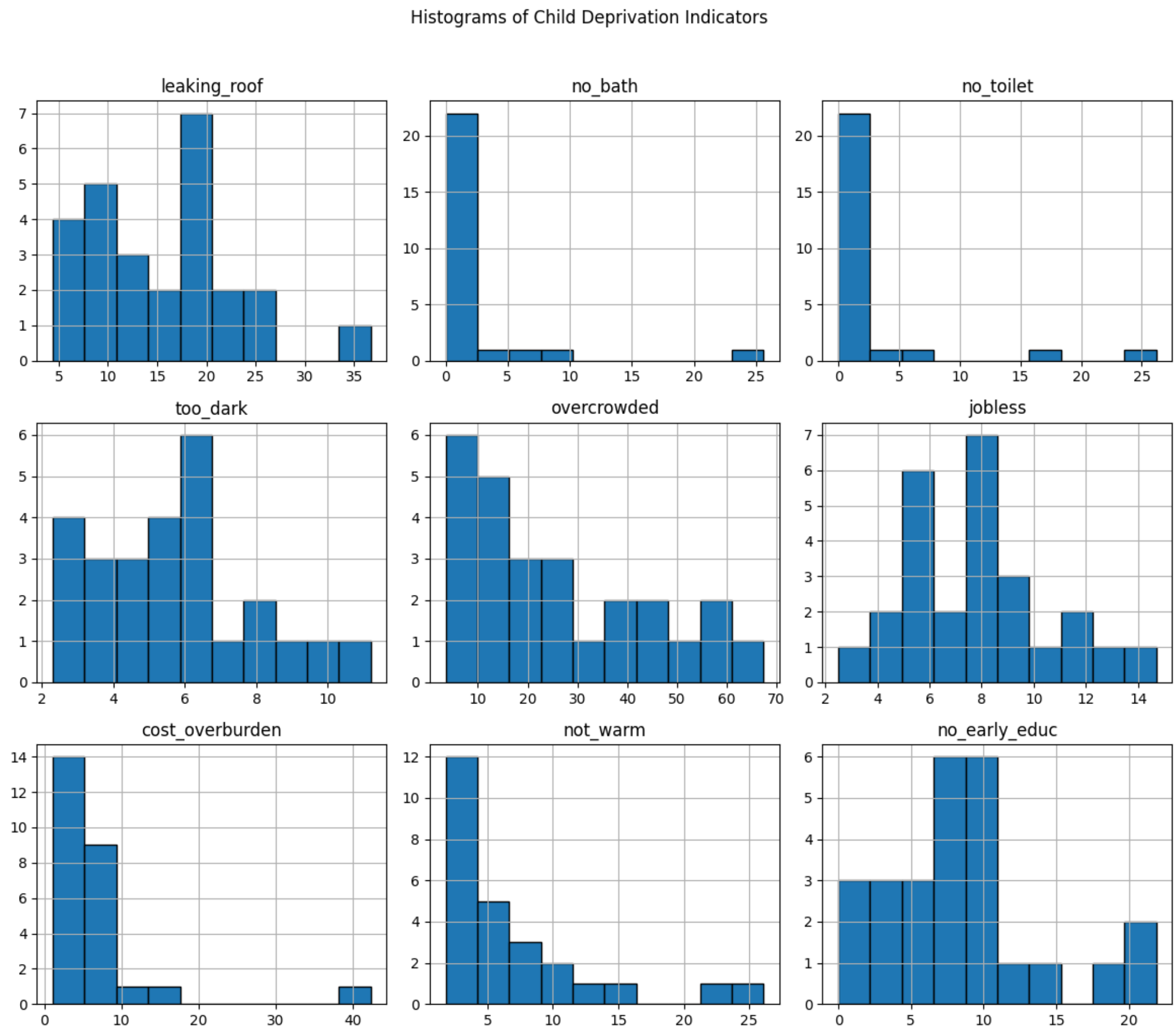

Figure A.1 – Histograms of Child Deprivation Indicators

B. APPENDIX B

Ethics Committee <ethicscommittee@novaims.unl.pt> 21 de junho de 2025 às 10:15
Para: "itsmariaserra@gmail.com" <itsmariaserra@gmail.com>, Maria Helena Miranda Flores Baptista <mhbaptista@novaims.unl.pt>
Cc: Ethics Committee <ethicscommittee@novaims.unl.pt>

[illegible]

[illegible]

[illegible]

[illegible] **STAT2025-6-167522**

[illegible] **Rethinking Child Poverty in the EU: A cluster-based alternative to AROPE**

[illegible] **Maria João Gonçalves Dutra Serra**

[illegible] **APPROVED** [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

ethicscommittee@novaims.unl.pt

This email serves as formal proof of ethical approval. If required for inclusion in a thesis, dissertation, or any other academic documentation, a PDF version of this message may be created and attached accordingly.

Cristina Oliveira

Gestora executiva do centro de investigação MagIC | *Executive manager of the Information Management Research Center (MagIC)*

Find out more about our research at https://magic.novaims.unl.pt/en/

Team member of RM Roadmap - Co-creating the future of Research Management (https://rmroadmap.eu/)

https://orcid.org/0000-0002-0887-7961

NOVA Information Management School
Universidade Nova de Lisboa

Campus de Campolide, 1070-312 Lisboa
Tel. +351 213 828 610

www.novaims.unl.pt

Data with purpose

Figure B.1 – Ethics Committee Report

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa