IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17809

Enduring Inequalities: Analyzing Energy Poverty Inertia Across K-Means Clusters

Santiago Budría
Leslie Bravo Chew

MARCH 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17809

Enduring Inequalities: Analyzing Energy Poverty Inertia Across K-Means Clusters

Santiago Budría
Universidad Antonio de Nebrija and IZA

Leslie Bravo Chew
Universidad Antonio de Nebrija

MARCH 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Enduring Inequalities: Analyzing Energy Poverty Inertia Across K-Means Clusters*

Evidence on how energy poverty persistence and vulnerability to key factors are distributed across different population groups remains scarce. This paper seeks to bridge this gap by analyzing the dynamics and determinants of energy poverty within population clusters. The significance of the paper is highlighted in the integration of a two-stage Generalized Method of Moments (GMM) estimation procedure with K-means cluster analysis. K-means clustering is a fundamental tool within AI to understand and find patterns and structure in data without labeled outputs. Two key findings emerge. First, the degree of energy poverty state dependence varies substantially across clusters, with some segments of the population deeply entrenched and facing significant barriers to escape. Second, variables critical for identifying at-risk groups, such as income and energy prices, exhibit different impacts across clusters. These findings highlight the need for targeted policy interventions tailored to the specific vulnerabilities of distinct population segments.

JEL Classification: Q40, I32, C38, C33

Keywords: energy poverty, state dependence, K-means clustering, Generalized Method of Moments

Corresponding author:
Santiago Budría
Department of Business Administration
Universidad Nebrija
C/ de Sta. Cruz de Marcenado, 27
28015 Madrid
Spain

E-mail: sbudria@nebrija.es

* The author gratefully acknowledges the financial support provided by the 2022 R&D&I National Projects and 2021 Strategic Projects Oriented to the Ecological and Digital Transition by the Spanish Ministry of Sciences and Innovation (Refs: PID2022-143254OB-I00 and TED2021-132824B-I00).

1. Introduction

The societal repercussions of energy poverty are profound. Energy poverty encompasses lower developmental outcomes, reduced educational attainment, heightened income inequality, and poorer health metrics on a global scale. Extensive evidence based on international macroeconomic data has highlighted these impacts (Banerjee et al., 2021; Lee and Yuan, 2024; Bao and Liao, 2024), while microeconomic studies have revealed the detrimental effects of energy poverty on subjective well-being, health, and human capital accumulation (Zhang et al., 2021; Prakash at al, 2022). Addressing energy poverty has thus become a top priority for governments and institutions, as emphasized by the International Energy Agency (IEA, 2024), leading to significant initiatives aimed at identifying its socioeconomic drivers. Mapping these gradients ensures that anti-poverty measures prioritize vulnerable populations, fostering both equity and impact.

This paper provides new insights into the determinants of energy poverty and the extent of its persistence over time. Using panel data from HILDA, a micro-survey representative of the Australian population, and applying a Multidimensional Energy Poverty Index (MEPI), the study makes three key contributions to the literature. First, the paper introduces a methodological innovation by building on recent developments in the field of unsupervised machine learning (ML) algorithms used to group data points into a specified number of clusters based on their features. The significance of the study is highlighted in the clustering of individuals into different categories (ranging from intense to mild energy poverty) using K-means cluster analysis. K-means is considered an important method within the AI toolkit (Çılgın and Gökçen, 2024) and has recently proven useful in various areas for unveiling underlying patterns in data in studies studies on market segmentation, consumer behavior, environmental impact, real estate assessments, and electricity use (Ofetotse et al., 2021; Tabianan et al., 2022; Che and Tian, 2024; Mardianto et al., 2024).

As a second contribution, the study complements traditional analyses of energy poverty, which typically focus on contemporaneous effects and assume that control variables fully encapsulate the factors driving observed outcomes. While these approaches have provided valuable insights, they often adopt a static perspective that may overlook the potential for energy poverty to be self-perpetuating—shaped not only by current conditions but also by its own history. Furthermore, they may fail to capture the dynamic transitions individuals undergo as they move into or out of poverty over time. A growing body of literature has sought to address this gap by estimating the state dependence of energy poverty using dynamic panel models, with the lagged dependent variable serving as an explanatory factor (Alem and Demeke, 2020; Drescher and Janzen, 2021;

Halkos and Kostakis, 2023). In this paper we estimate energy poverty dynamics through a two-stage Generalized Method of Moments (GMM) estimation procedure (Arellano and Bond, 1991; Arellano and Bover, 1995). This methodological framework, enhanced by several refinements and sensitivity checks, effectively addresses endogeneity issues arising from reverse causality and unobserved heterogeneity (Leszczensky and Wolbring, 2022). It also captures the significant impact of prior energy poverty conditions on future outcomes. By integrating GMM with an AI-driven clustering approach, this study moves beyond characterizing state dependence with an "average effect." Instead, it explores whether focusing solely on averages might obscure critical variations within subpopulations.

As a third contribution, this study examines whether the relationship between energy poverty and key variables—such as income and energy prices, which are commonly used to identify at-risk groups—differs across distinct population segments. Because most research on energy poverty focuses on socio-demographic determinants at the aggregate level (Dalla Longa et al., 2021; Fry et al., 2022; Awan and Bilgili, 2022), there is limited evidence on how these factors influence specific population clusters. This study addresses this gap by documenting how key drivers uniquely impact different segments, offering valuable insights into the varying levels of vulnerability across groups.

Australia offers a valuable case study for this analysis for two key reasons. First, existing Australian programs primarily focus on price compensation and social welfare payments to assist with energy bills, and have a strong targeting of low-income groups through means-testing (Willand et al., 2023). However, fluctuations in gas and electricity prices—and consequently in compensation rates—may have only minimal effects within certain groups. Similarly, there is no guarantee that income consistently serves as a fundamental risk factor across all clusters. As a result, current policies may fail to adequately address the needs of at-risk individuals who fall outside traditional low-income classifications. Identifying differences across clusters regarding energy poverty state dependence and their sensitivity to changes in income and energy prices is therefore critical to developing more effective energy poverty reduction strategies. Second, over the past decade, electricity prices in the country have nearly tripled, raising significant concerns about energy affordability (Proctor, 2022). By 2021, the average residential electricity price in Australia ($0.22 per kWh) was approximately 60% higher than the global average ($0.14 per kWh). This sharp increase in energy costs has placed considerable strain on household budgets, intensifying challenges related to energy access and affordability (OECD, 2022).

This study is organized as follows. Section 2 reviews the literature and highlights the main determinants of energy poverty discussed in prior research. Section 3 describes the dataset and the main variables used in the analysis, including the central energy poverty measure, the MEPI. Section 4 outlines the methodological framework, detailing the K-means clustering approach and the GMM econometric procedure. Section 5 presents the results and emphasizes the existing differences across clusters. Section 6 introduces several sensitivity checks to test the validity of the model and asses the robustness of the results. Section 7 presents the concluding remarks. The paper includes and Appendix with additional results.

2. Literature review

Energy poverty is a multidimensional phenomenon characterized by deprivation and vulnerability arising from inadequate or unaffordable energy supply. While numerous conceptual definitions have been proposed, translating these into actionable measures remains a significant challenge. Policymakers and scholars employ diverse data and definitions that vary across countries and over time, reflecting ongoing debates in the academic literature (Siksnelyte-Butkiene, 2021). The literature highlights three primary approaches to measuring energy poverty: expenditure, consensual, and direct measurement (Chandrashekeran et al., 2022). The expenditure approach compares energy costs to absolute or relative thresholds as a proxy for energy deprivation. Consensual metrics rely on self-reported assessments on the ability to meet basic energy needs. Direct measurement evaluates energy use, such as cooling and heating, against established benchmarks. These approaches have led to the development of both composite and single metrics, framing energy poverty as a broader issue encompassing vulnerability, insecurity, and energy justice (Guevara et al., 2023).

Cross-national studies underscore that energy poverty is influenced by a range of macroeconomic and institutional factors, including education, governance quality, technological progress, health expenditures, and the level of economic development. The relationship between these factors and household energy deprivation often varies by GDP levels, as shown by Boţa-Avram et al. (2024). Income inequality, internal conflict and climatic conditions also contribute, though to varying degrees, depending on regional contexts (Igawa and Managi, 2022; Khalid et al., 2024). Moreover, the energy mix of a country—reflecting the sources used for electricity production—plays a pivotal role in shaping energy outcomes (Kocak et al., 2023).

At the regional and country level, energy poverty often stems from a complex interplay of factors such as energy prices, availability, and inefficiencies in building infrastructure (Cheikh et al., 2023). Characteristics like insulation, heating systems, and floor area significantly affect vulnerability to energy deprivation (Karpinska and Śmiech, 2023). Meanwhile, household-specific drivers such as low income and high energy costs exacerbate risks, leading to unpaid bills, energy debts, or supply disconnections (Best and Burke, 2019). Individual attributes also matter: educational attainment is negatively associated with energy poverty due to its role in fostering energy-efficient behaviors and economic resilience (Crentsil et al., 2019). Household size, marital status, and location—urban or rural—further interact with energy needs and costs, while age influences energy deprivation through life cycle patterns and risk preferences (Abbas et al., 2020; Drescher and Janzen, 2021). Moreover, poor health can shift household spending priorities, limiting access to energy services, while income and employment status are consistent determinants, especially in developing countries, where energy poverty is more pervasive (Awan and Bilgili, 2022; Abbas et al., 2022).

2.1 AI applications in energy poverty research

The application of AI-based methods to predict energy poverty has gained attention recently, reflecting its potential to enhance both understanding and mitigation strategies. Although still in its infancy, this approach provides new insights into the drivers of energy poverty. For instance, studies using advanced ML frameworks such as XGBoost have identified critical predictors in diverse contexts. In the Netherlands, income, house value, and homeownership emerged as the most significant determinants of energy poverty (Dalla Longa et al., 2021). Complementarily, research utilizing Random Forest classifiers highlights the importance of both household-level factors, such as dwelling conditions and energy efficiency, and country-specific elements, including gas supplier switching rates (Spandagos et al., 2023). While the previous studies are based on a single energy poverty indicator, other studies define a multidimensional energy poverty index similar to ours. These studies have examined Asian and African contexts showing that wealth, marital status, and place of residence play pivotal roles (Abbas et al., 2020). Methodological advancements have also emerged, with ensemble models like XGBoost, Random Forest, and Artificial Neural Networks uncovering novel predictors, such as education and food security, which significantly influence energy poverty risks (Gawusu et al., 2024).

Despite these advancements in the literature, our understanding of how current circumstances influence the dynamics of energy poverty remains limited. While existing studies provide valuable insights into static relationships between predictors and energy poverty, they often fail to capture the temporal dynamics that can reveal pathways into and out of deprivation. Investigating these dynamics is essential, as it holds significant potential to inform targeted interventions.

3. Data and measures

We use the 2007-2021 waves of the Household, Income, and Labor Dynamics in Australia (HILDA) Survey, a comprehensive, nationally representative longitudinal study that examines the economic, social, and demographic dynamics of Australian households. Initiated in 2001 and conducted annually, it tracks individuals and households over time. The survey combines objective data, like income and employment, with subjective measures, such as well-being and financial stress, offering a rich dataset for understanding Australia's social fabric. Its design allows for the analysis of long-term trends and causal relationships, making it an essential tool for researchers and policymakers. The original 2001 sample included approximately 7,600 households and 13,000 individuals, with periodic updates to account for attrition.

We employ the dataset to regress energy poverty on a number of covariates, including energy prices and socio-economic factors. After dropping observations with item non-response, the estimation sample includes 172,582 observations from 23,251 individuals across 15 years. We describe the main variables below.

3.1 Energy poverty

We use five items to construct a multidimensional energy poverty index. The literature typically distinguishes between objective (expenditure-based) and subjective (self-assessed) approaches. Expenditure-based measures label a household as energy poor when the income that households spend on energy is above a specific threshold. For instance, a household may be classified as energy poor if i) its share of income spent on energy is greater than twice the national median (the *2M* indicator), ii) its share of income spent on energy exceeds 10% (the Ten Percent Rule, *TPR*), or iii) its actual energy expenditures are above the national median and, at the same time, their income net of energy costs is below the official national income poverty line (the Low Income High Costs indicator, *LIHC*). Hence, our first three items are *2M,* the *TPR*, and the *LIHC* indicators. All energy expenditures and income variables used in the paper are transformed using the OECD equivalence scale and normalized into real terms using the yearly consumer price index. These measures have been used frequently in the literature (Fry et al., 2022; Awan et al., 2022).

However, while expenditure-based measures are objective and transparent, they may overlook intentional reduction in energy consumption by low-income households. If vulnerable households limit their energy consumption to prioritise other services and goods, measures based on the actual energy costs may underestimate the true prevalence of energy poverty (Price et al., 2012).

Moreover, low-income families can resort to energy credits and repayments to smooth their monthly energy costs over time. In this case, a low monthly energy budget may hide a chronic energy deprivation status. To overcome these limitations, applied research has relied on individuals' self-evaluations of their ability to afford and access specific energy services (Prakash et al., 2022). Hence, apart from the expenditure-based measures, we also consider two self-assessed indicators based on the household's inability to pay to heat their home because of a shortage of money (*Heat*) and pay electricity, gas, or telephone bills on time (*Arrears*).

Noting that energy poverty is multifaceted, the MEPI used in the paper is based on the aforementioned items. Let J be a set of poverty indicators with element j, $j \in J, m = card(J)$. Let I be a set of individuals, with element i, $i \in I$, and EP_{ij} denote the status of the ith individual in the jth indicator. If an individual i is poor under indicator j, then EP_{ij} takes the value of one, and zero otherwise. Following the family of indexes typically described in the literature on material deprivation (Dhongde et al., 2019), individual i's weighted poverty score is given by

$$MEPI_i = \left(\sum_{j \in J} w_j EP_{ij} \right) \qquad \forall i \in I \tag{1}$$

where w_j denotes the weight assigned to the poverty indicator j, with $\sum_{j \in J} w_j = 1$. Hence, MEPI ranges from 0 to 1 and captures the percentage of dimensions in which the individual is deprived. While it is common to assign equal weights to the indicators, we emphasise the indicators where deprivation is less common, the so-called frequency-based weighting approach (Decancq & Lugo, 2013). The weight given to an indicator is proportional to the percentage of individuals *not* classified as poor under that specific indicator within a particular state. In other words, $w_j = \frac{(1-n_j)}{\sum_{j \in J}(1-n_j)}$ where n_j is the proportion of poor individuals in dimension j. This choice is motivated by the idea that not having access to common items should be a more relevant determinant of deprivation than less common items. Additionally, the weights are based on the distribution of achievements in society without considering any value judgement about what the trade-offs between items should be. For greater granularity and accuracy, the weights are calculated separately for each wave. There are two advantages to using this approach. Firstly, it allows the poverty of a given individual to increase if their conditions do not change and the conditions of all others improve. Secondly, it adapts automatically over time, considering economic conditions and social and cultural preferences when accessing items.

3.2 Energy prices and other control variables

We use annual electricity and gas prices at the state level drawn from the Australian Bureau of Statistics (ABS, 2024). The average price of gas and electricity over the sample period was \$0.012 and \$0.266 per kWh, respectively. To avoid variable proliferation, in the regression stage we introduce just one control for energy prices, defined as a weighted average between the price of gas and electricity.[1] Since energy prices are not available monthly, we construct a 12-month rolling average, $\bar{p}_{it}^{12} = k_{it} \times p_t + (1 - k_{it})p_{t-1}$, where $k_i t$ is the proportion of months elapsed from January 1 to individual's i date of the interview in year *t* and p_t is the energy price in year *t*. Thus, we do not only exploit variation in energy prices across states, but also over time and across individuals.

As additional control variables, we consider socioeconomic factors standard when accounting for individual economic outcomes. These include income, schooling, age, marital status, labor status, health and parenthood. The relevance of these variables has been emphasized in prior research and described in Section 2. We also include controls for remoteness, region of residence (the six states and two territories of Australia, reference: New South Wales), time fixed effects and variables to control for macroeconomic conditions at the regional level. The economic cycle affects the chance to find and keep jobs, and it also impacts the likelihood of having a stable income source. We include controls for the regional unemployment rate, per capita GDP, and GDP growth. We also include the regional participation rate to capture competition effects in the labour market and the labour force share of part-time workers to control for the fact that areas with a larger proportion of temporary and/or part-time workers generally have more flexibility to adapt to labour market disequilibria.

In Table 1, we report summary statistics. We classify respondents depending on whether they have some form of energy deprivation (MEPI > 0) or not (MEPI = 0). We also report summary statistics for the sample as a whole. Households experiencing some level of energy poverty (MEPI > 0) have significantly lower incomes (\$65,655.2 vs. \$116,404.2), aligning with the expectation that limited financial resources are a key driver of energy poverty. They also have fewer years of education on average (12.1 vs. 13.0), potentially restricting access to better job opportunities and income. Individuals in these households tend to be slightly older (46.4 vs. 44.7 years), possibly

[1] An average Australian household spends about twice as much on electricity as on gas (ABS, 2024). Accordingly, we assign a weight of 0.66 to electricity prices and 0.33 to gas prices. Alternative linear combinations yielded comparable results, and including separate controls for electricity and gas prices did not significantly improve the models' goodness of fit.

reflecting higher energy poverty among older populations with fixed or reduced incomes. There is also a lower proportion of married individuals (61.8% vs. 73.1%) and higher proportions of divorced (14.1% vs. 7.4%) and widowed individuals (4.1% vs. 1.8%), suggesting that the economic stability often associated with marriage reduces the risk of energy poverty. Individuals with MEPI > 0 show lower employment rates (57.3% vs. 75.4%) and have more children at home on average (1.9 vs. 1.6), which may increase energy needs and financial strain. Despite this, they tend to live in smaller households, potentially indicating a higher prevalence of single-parent families among them. They also report higher rates of ill-health (30.3% vs. 17.6%), consistent with evidence that health issues often exacerbate energy vulnerability.

-Insert Table 1 here-

4. Methodology

4.1 K-means clustering

We use the K-means clustering algorithm to group individuals based on their energy poverty level, measured by the MEPI. This approach enables us to analyze the inertia of energy poverty within each cluster and uncover patterns in the characteristics unique to each group. Given that each individual can belong to only one cluster and that energy poverty is subject to temporal fluctuations, the clustering is performed using the time-averaged MEPI for each individual over the 15 years covered in the panel. This choice aligns with standard practices for applying K-means clustering to panel data (Gök and Sodhi, 2021).

The algorithm works by finding the centroids (the mean of the points of each cluster) and assigning each data point to the nearest centroid, each group of points assigned to a centroid outcome in a cluster. In our case, each centroid is initialized as a random value within the range of energy poverty levels. The algorithm is randomly solved several times through an iterative process that minimizes the sum of squared distances between the points and their respective cluster centroids – inertia –. Specifically, K-means partitions the data into K clusters, with each cluster containing elements that are similar by a distance metric (such as Euclidean distance measurement method). Suppose $D = \{ x_1, \ldots, x_n\}$ corresponds to the individuals to be clustered. Following Wu (2012), K-means can be expressed by:

$$\min_{\{m_k\},1\leq k\leq K} \sum_{k=1}^{K} \sum_{x\in C_k} \pi_x \ \mathrm{dist}(x, m_k) \tag{2}$$

where π_x is the weight of x, m_k is the centroid of cluster C_k, $m_k = \sum_{x \in C_k} \frac{\pi_x x}{n_k}$, where n_k is the number of data objects assigned to C_k, K is the number of clusters and 'dist' computes the distance between objects x and centroid m_k, $1 \leq k \leq K$. The selection of the distance function is the squared Euclidean distance: $||x - m||^2$.

To determine the optimal number of clusters (*K*), we follow two key criteria: clusters must contain at least a minimum number of individuals (set at 1000), and the optimal number of clusters must be identified through The Elbow Method. The method involves iterating from *k*=1 to *k*=n (here n=10, is the hyperparameter selected), and plotting the inertia against the number of clusters. The optimal number of clusters is typically identified at the "elbow point" in the plot, where the rate of decrease in inertia slows down, indicating diminishing returns from adding more clusters (Cui, 2020).

Figure 1 shows the inertia for the analysed data, and the "elbow point" is where the inertia stops decreasing significantly. This point can be interpreted as a good candidate for the optimal number of clusters. Fulfilling the two criteria outlined above, *K* = 4 is the number of clusters best selected in the algorithm. To assess whether the clusters remain distinct and internally consistent, we compute the Silhouette Coefficient of each cluster. The silhouette score measures the similarity of each point in a cluster to points in neighbouring clusters, with values ranging from -1 to 1. A score close to -1 suggests that a point may have been incorrectly assigned to a cluster, while a value close to 0 indicates that the point is on the boundary between clusters, so its belonging is unclear, or clusters are overlapping. Scores close to +1 mean that the point is well separated from neighbouring clusters, indicating a strong and well-defined clustering. In this case, the inertia reductions for *K* = 3 and 4 are strong, as shown in Figure 1, and it is at *K* = 4 that the trend is broken. The Silhouette Coefficient of these clusters is 0.689 and 0.676, respectively. Although the score of 4 clusters is slightly lower than the 3-cluster score, it still scores relatively well, suggesting that the clusters are still well separated and coherent. For this paper, it is a reasonable choice because the aim is to capture more nuances in the data.

-Insert Figure 1 here-

Table 2 shows the distribution of individuals by cluster. The 23,251 individuals in the sample were grouped into 4 clusters based on their time-averaged energy poverty score. The cluster with the highest density of individuals (56.2%) is cluster *k*=1, with the lowest average MEPI (0.143). This is followed by cluster *k* =3 with 23.6% of respondents and a mean MEPI of 0.093 and cluster *k* =0, with 15.6% of the sample and MEPI = 0.215. The remaining 4.5% of the sample individuals

(k =2) are the poorest in energy, with an average MEPI of 0.418. Figure 2 shows each cluster with its respective centroid and the individuals density corresponding to each cluster.

-Insert Table 2 here-

-Insert Figure 2 here-

4.2 The GMM estimator

Energy poverty at time t is specified

$$MEPI_{it} = \rho MEPI_{it-1} + X'_{it}\beta + c_i + u_{it} \tag{3}$$

$(i = 1, \dots., N); (t = 2, \dots, T)$, where X_{it} is the set of covariates; c_i denotes individual-specific unobserved heterogeneity; and u_{it} is assumed to be a normally distributed error term $N(0, \sigma_u^2)$. This specification does not deal with the initial condition problem, i.e., the possibility that energy poverty at the start of the observation period is endogenously determined by the individual's past history. In such case, the unobserved heterogeneity and the initial value of the dependent variable are expected to be correlated, $E(MEPI_{i1}\, c_i) \neq 0$. Similarly, the specification does not rule out the potential correlation between the unobserved heterogeneity and the regressors, i.e., $E(X_{it}\, c_i) \neq 0$. Moreover, a consistent estimation of Eq. (1) crucially relies on the correct specification of c_i and the assumption that $E(MEPI_{it-1}u_{it}) = 0$. To address these concerns, we resort to a GMM estimation procedure that purges the individual-specific effect from the model,

$$\Delta MEPI_{it} = \rho \Delta MEPI_{it-1} + \Delta X'_{it}\beta + \Delta u_{it} \tag{4}$$

Noting that in the resulting model there is still correlation between the differenced lagged variable and the disturbance process (the former contains $MEPI_{it-1}$ and the latter contains u_{it}), we instrument $\Delta MEPI_{it-1}$ with all lags of $MEPI_{it-j}$, for $j \geq 2$. We employ a two-step GMM approach, incorporating a second-order transformation known as 'forward orthogonal deviations'. This technique involves subtracting the average of all future available observations from the current value of a variable, as opposed to subtracting the previous observations. By utilizing the two-step GMM model, we mitigate unnecessary data loss. The two-step GMM estimator is unbiased and consistent under the assumption of no second-order serial correlation in the error term, $E(MEPI_{it-j}\Delta u_{it}) = 0 \;\; \forall j \geq 2,$ a moment condition that can be tested. Strictly exogenous regressors are assumed to be uncorrelated with u_{it} and are used as instruments for themselves. It is possible that some regressors are weakly exogenous, that is, they are correlated with past errors,

$E(x_{it+s}u_{it}) \neq 0$ for $s > 0$, but uncorrelated with contemporaneous and future errors, $E(x_{it+s}u_{it}) = 0$ for $s \leq 0; t + s > 0$. If a regressor is presumed to be correlated with the contemporaneous error, it is considered endogenous. In such instances, only lagged values can be employed as valid instruments.

To select the optimal combination of exogenous and endogenous regressors, we conducted an extensive series of sensitivity checks, using tests for serial correlation in first-differenced errors and the Hansen's J-test of overidentifying restrictions as our primary screening criteria. The most robust combination assumed that all individual variables were strictly exogenous. Nonetheless, in the Appendix, we provide additional sensitivity analyses under varying assumptions for the regressors, and also include a second lag of the dependent variable. We employ standard instruments for strictly exogenous regressors, collapsed instruments for the remaining regressors, and Windmeijer-corrected robust standard errors. To mitigate the efficiency loss associated with instrument proliferation—an issue that can result in overfitting endogenous variables without adequately addressing their endogenous components— we collapse the instruments and restrict the maximum lags to three periods, as detailed by Kripfganz & Schwarz (2019).

5. Results

Before presenting the regression results separately by group, it is convenient to examine potential divergences in the socio-economic profile of the different clusters. Figure 3 presents summary statistics for the categorical variables and continuous variables used in the paper. The results suggest the presence of four hypothetical Types: the "Unemployed and Income Poor" (Cluster 0), the "Young professionals and affluent" (Cluster 1) the "Sick and Vulnerable" (Cluster 2) and the "Stable workers with moderate income" (Cluster 3).

Cluster 0 (average MEPI = 0.215) represents individuals facing significant economic constraints, characterized by low income, limited education, and a low employment rate. Cluster 1 (MEPI = 0.014) includes relatively young individuals with high employment rates, good health, and strong educational backgrounds. This group benefits from stable family structures, being more likely to be married and less likely to be divorced, and has fewer children compared to other clusters. Cluster 2 (MEPI = 0.418) is a segment disadvantaged in health, income and energy poverty. These individuals often face compounded challenges such as economic inactivity, low education levels, and ill-health, which limit employment opportunities. This cluster also tends to include older, unmarried individuals, with a significant presence of women. Finally, Cluster 3 (MEPI = 0.093) includes mostly employed individuals with relatively stable economic conditions, though less affluent than Cluster 1. While most are employed, some are economically inactive. This cluster exhibits stable family structures, with a high proportion of married individuals and relatively few divorcees, reflecting a moderate level of economic security but greater vulnerability compared to the more affluent clusters. An important takeaway from Figure 3 is that energy prices remain consistent across clusters, indicating that the differences in the underlying likelihood of experiencing energy poverty stem from socioeconomic and demographic factors rather than varying energy price rates among different groups.

-Insert Figure 3 here-

5.1 The determinants of energy poverty

In Table 3 we present results from the GMM estimator. The lagged dependent variable, representing state dependence, indicates a significant energy poverty inertia among individuals in Australia, even after accounting for a broad range of socio-economic determinants. Across all clusters, individuals with a high MEPI at time $t-1$ tend to exhibit relatively high MEPIs at time t. However, the magnitude of this effect varies across clusters, reflecting differences in the persistence of energy poverty. Cluster 0 exhibits the highest inertia (0.108), pointing to deeply entrenched energy poverty in this group, while Cluster 1 shows the lowest state dependence (0.052). To contextualize these findings, we reference three earlier studies, though methodological differences—such as reliance on a binary definition of energy poverty (yes/no) and the use of Wooldridge's conditional maximum likelihood estimator—make direct comparisons challenging. Drescher and Janzen (2021), using data from the German Socio-Economic Panel, report state dependence effects of 3.8 to 7.5 percentage points (pp) in energy poverty likelihood. This aligns with the low persistence observed in Clusters 1 and 2 for Australia, reflecting conditions typical of developed nations with strong social safety nets. In contrast, Alem and Demeke (2020), employing data from the Ethiopian Urban Socio-economic Survey, estimate lagged poverty effects between 9.8 and 16.4 pp, highlighting greater persistence in lower-income settings. Similarly, Halkos and Kostakis (2023) report an approximate 12 pp effect for Greece, a middle-income country. Hence, the high state dependence observed in Cluster 0 aligns more closely with estimates from low- and middle- income contexts, thus suggesting significant socioeconomic disparities within Australia.

-Insert Table 3 here-

The effect of income also differs significantly across clusters. Cluster 2 ("Sick and vulnerable") shows the strongest negative effect (-0.249), indicating that improvements in income are particularly critical for reducing energy poverty in this group, given their low income levels and high vulnerability. Cluster 0 ("Unemployed and income poor") also exhibits a strong negative effect (-0.185), while Cluster 3 ("Stable workers with moderate income") shows a moderate negative effect (-0.137), reflecting that income improvements are still relevant but less critical compared to the more vulnerable clusters. While the negative relationship between income and energy poverty is well-documented (Churchill and Smyth, 2021; Dalla Longa et al., 2021; van Hove et al., 2022) and forms the foundation for income-based policies to address energy poverty (Simshauser and Miller, 2023), most studies in the literature depict this relationship in a linear, average fashion. Our findings suggest significant heterogeneity in the income-energy poverty relationship across individual profiles, supporting the need for targeted, cluster-specific policy interventions.

Energy prices are a key tool for policy and the monitoring of at-risk groups. The results in Table 3 highlight the vulnerability of different clusters to energy cost fluctuations. Cluster 2 shows an acute sensitivity to energy price increases likely due to their constrained resources and higher energy needs. The estimate indicates that a 1% increase in the energy price leads to 0.0034 points increase in the MEPI, which represents a 0.34 pp variation. Cluster 0 also experiences a significant rise (0.21 pp), emphasizing their vulnerability to energy costs, while the effect in Cluster 3 is smaller but still positive (0.063 pp). In sharp contrast, Cluster 1 presents a small negative effect (-0.0274 pp), suggesting that this group is least affected by energy price increases and may even benefit slightly, perhaps due to efficient energy use and price resilience.

As for the remaining socio-economic variables, we also detect relevant differences. In Cluster 0, there is a significant dependency on unemployment and ill-health. Clusters 1 and 3 show greater sensitivity to aging, widowhood, and divorce, while having children at home raises the energy poverty score in Cluster 3. All clusters except Cluster 2 respond positively to favorable aggregate conditions such as labor market flexibility (part-time employment rate) and GDP growth to a lesser extent. Additionally, Cluster 3 is negatively impacted by labor market competition, as reflected by the participation rate, consistent with the fact that this group primarily consists of employed individuals. To illustrate these differences, Figure 4 presents a summary of the relative sizes of the coefficients by cluster. For a given covariate (e.g., energy prices), the coefficients are summed to total 100%, with the colors indicating the relative contribution of each cluster within that 100%. This visualization highlights the varying importance of each covariate across clusters, emphasizing the heterogeneity in the underlying relationships.

-Insert Figure 4 here-

5.2 Model validation and sensitivity checks

In the lower section of Table 3, we report the diagnostic tests for the GMM model, including the test for serial correlation in the first-differenced errors and Hansen's J-test of overidentifying restrictions. The results indicate no evidence of second-order serial correlation, supporting the validity of the lagged instruments used in the model, as their relevance depends on the error term not being serially correlated beyond the first order. Additionally, Hansen's J-test fails to reject the null hypothesis that the instruments are uncorrelated with the error term, confirming their appropriateness. These diagnostics suggest that the model satisfies the key assumptions underpinning GMM estimation, ensuring that the estimates are not biased by instrument invalidity or misspecified error structures.

In the Appendix, we present results from variations of the GMM estimator, incorporating alternative assumptions. Table A1 reports the outcomes for a model that includes a second lag of the dependent variable. The diagnostic results are less satisfactory, showing evidence of second-order autocorrelation in two clusters (Clusters 0 and 3) and a rejection of the Hansen's J-test of overidentifying restrictions at the 10% significance level in Cluster 1. Furthermore, the second lag of the dependent variable is not statistically significant at the 5% level in three of the clusters, lending support to the poverty dynamics assumed in the baseline specification. Despite these limitations, the results remain broadly consistent with those of the baseline model, with Clusters 0 and 2 exhibiting significant sensitivity to income and energy prices and state dependence being most pronounced in Cluster 0. In Table A2, we explore a model where income and the employment variables are assumed to be endogenous. For Cluster 0, the Hansen's J-test suggests model misspecification, as the null hypothesis is rejected. Nevertheless, the estimated coefficients remain largely consistent with those obtained in the baseline model. Finally, Table A3 presents results under the assumption of endogeneity for the full set of individual regressors. This specification does not produce any meaningful differences in the main conclusions of the paper, reinforcing the robustness of the baseline findings.

5.3 Assessing endogeneity in sample attrition

Although the sample exhibits moderate levels of attrition, with entry and exit rates averaging 8.7% and 7.2%, respectively, there remains a concern about whether these dynamics are systematically influenced by factors related to energy poverty. To investigate this, we conducted a series of analyses to determine whether sample attrition is random or exhibits a degree of endogeneity. First, we assessed the likelihood of individuals leaving the sample by regressing an indicator variable for sample exit in the following year on energy poverty status and a set of control variables. The coefficient for energy poverty was statistically insignificant (-0.004, p-value = 0.387) suggesting that the likelihood of leaving the sample is unrelated to energy poverty. This finding alleviates concerns about selective dropout based on energy vulnerability.

We then turned to the entry process, analysing whether energy poverty influences the probability of joining the sample. Here, energy poverty showed a small but statistically significant negative effect (-0.006, p-value = 0.058), indicating that individuals who are less likely to experience energy poverty are slightly overrepresented among new entrants. This non-random selection could reflect ease of recruitment, as individuals in more stable socioeconomic situations may be more accessible to participate in longitudinal studies. While this potential bias in new panelist inclusion warrants acknowledgment, the randomness of attrition post-recruitment supports the representativeness of the sample over time.

6. Conclusions

This paper examined the dynamics of energy poverty by integrating advanced machine learning methods with a dynamic Generalized Method of Moments (GMM) model. Using data from HILDA, representative of the Australian population, and a multidimensional index of energy poverty, the paper identified key determinants of energy poverty and examined its persistence across different population groups.

The study contributes to the existing literature by addressing key limitations of traditional analyses. It introduces an innovative methodology that leverages unsupervised clustering algorithms, such as K-means, to segment the population into relevant clusters. In combination with a dynamic panel model, this enables the identification of substantial differences in state dependence and the factors driving energy poverty within each group, challenging the conventional reliance on population averages and contemporaneous, static effects.

The results of this study identify four clusters —the "Unemployed and Income Poor," "Young Professionals and Affluent," "Sick and Vulnerable," and "Stable Workers with Moderate Income"—. These clusters are associated to different latent levels of energy poverty. The results emphasize the diversity of energy poverty dynamics across clusters. For example, the "Unemployed and Income Poor" group experiences the highest level of persistence, with current deprivation heavily influenced by past conditions. In contrast, the "Young Professionals and Affluent" group demonstrates resilience, with low energy poverty persistence and reduced sensitivity to income and energy price fluctuations. The "Sick and Vulnerable" group, meanwhile, is marked by acute sensitivity to income and energy prices, reflecting their precarious economic and health circumstances.

These differences are critical for policy design, as they show that one-size-fits-all solutions are inadequate. The results document the persistent nature of energy poverty in certain clusters, revealing that this condition is often more strongly tied to historical deprivation than to contemporaneous factors like income or energy prices. Hence, policies must move beyond static approaches, with particular attention to structural interventions for clusters “trapped” in energy poverty. Moreover, by distinguishing between clusters, the paper demonstrates that commonly employed policy levers, such as income support or energy subsidies, have vastly different effects across groups. For example, income support may have the strongest effect on reducing energy poverty among the "Sick and Vulnerable," while energy price interventions are most impactful for the "Unemployed and Income Poor". This highlights the need for highly targeted interventions tailored to the specific characteristics of each segment. For some clusters, policies addressing affordability may be effective, while for others, where energy poverty is driven by structural issues, such measures may prove insufficient. The pronounced persistence of energy poverty in certain groups suggests that policies aimed solely at mitigating current deprivation need to be complemented by strategies targeting its root causes. Structural improvements, such as enhancing energy infrastructure, providing access to efficient energy systems, and addressing systemic barriers to employment, may be essential to break the cycle of energy poverty in these populations.

This study has several limitations that warrant further investigation. The findings emphasize the importance of understanding why certain groups remain trapped in energy poverty, suggesting the need for future research to investigate the structural, institutional, and cultural factors perpetuating this persistence. Although these factors were not explicitly analyzed, they likely play a critical role in shaping outcomes. Additionally, while clustering enhances precision by segmenting the population, it does not address intra-cluster heterogeneity, potentially overlooking important regional or geographic variations. Furthermore, the study abstracts from the financial strategies and economic behaviors households may employ to mitigate the impact of adverse life events, despite evidence that variables such as income and employment significantly influence how individuals navigate energy challenges (Burlinson et al., 2024). Future research could bridge these gaps by incorporating spatial data and integrating both structural and idiosyncratic factors, offering a more nuanced understanding of the mechanisms driving energy poverty.

Tables

Table 1. Summary statistics by energy poverty status

	All	MEPI > 0	MEPI = 0
MEPI	0.077	0.271	0.000
	(0.146)	(0.151)	(0.000)
Household income	102003.654	65655.224	116404.246
	(86,716.2)	(48,174.6)	(94,067.1)
Energy Price	0.242	0.241	0.242
	(0.017)	(0.017)	(0.017)
Years of education	12.720	12.076	12.975
	(2.468)	(2.282)	(2.492)
Age	45.17	46.39	44.68
	(15.369)	(15.846)	(15.148)
Married	0.699	0.618	0.731
	(0.459)	(0.486)	(0.443)
Divorced	0.093	0.141	0.074
	(0.290)	(0.348)	(0.262)
Widowed	0.024	0.041	0.018
	(0.154)	(0.198)	(0.131)
Employed	0.702	0.573	0.754
	(0.457)	(0.495)	(0.431)
Unemployed	0.032	0.052	0.024
	(0.176)	(0.221)	(0.154)
Children at home	1.667	1.898	1.575
	(1.491)	(1.564)	(1.450)
Ill-health	0.212	0.303	0.176
	(0.409)	(0.459)	(0.381)
Household size	1.812	1.739	1.841
	(0.580)	(0.594)	(0.573)

Notes: Standard errors in parentheses.

Table 2. Distribution of individuals across clusters and average MEPI by group

Cluster	*k*=0:	*k* =1:	*k* =2:	*k* =3
Nº. of individuals	3,631	13,076	1,051	5,493
% of individuals	15.6%	56.2%	4.5%	23.6%
Average MEPI	0.2153	0.0143	0.4180	0.0928

Notes: Distribution of individuals among the four clusters (k=0 to k=3), along with the average MEPI for each group.

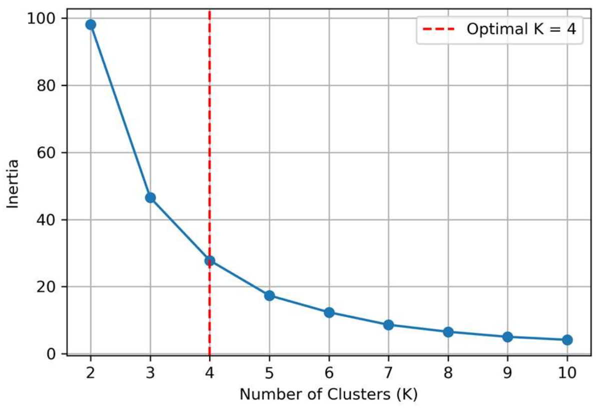
Table 3. GMM estimates for energy poverty, by cluster

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
$MEPI_{t-1}$	0.1080***	0.0524***	0.0722***	0.0950***
	(0.0163)	(0.0088)	(0.0260)	(0.0090)
Ln (Household income)	-0.185***	-0.0401***	-0.249***	-0.137***
	(0.0068)	(0.0018)	(0.0151)	(0.0039)
Ln (Energy Price)	0.210***	-0.0274**	0.340***	0.0634**
	(0.0527)	(0.0136)	(0.106)	(0.0251)
Ln (Years of education)	-0.143	-0.0110	-0.0620	-0.0388
	(0.123)	(0.0166)	(0.218)	(0.0554)
Age	-0.0005	0.0012	0.0276	0.0092**
	(0.0101)	(0.0021)	(0.0178)	(0.0036)
Age^2	-6.59e-05	2.38e-05**	-0.0003*	-5.81e-05*
	(7.63e-05)	(1.13e-05)	(0.0002)	(3.01e-05)
Married	0.0174	-0.0003	-0.0450	-0.0084
	(0.0221)	(0.0034)	(0.0592)	(0.0113)
Divorced	-0.0110	0.0093*	-0.0992	0.0032
	(0.0288)	(0.0055)	(0.0718)	(0.0152)
Widowed	-0.0023	0.0379**	-0.0803	0.0488*
	(0.0493)	(0.0174)	(0.0825)	(0.0283)
Employed	0.0091	0.0002	-0.0190	-0.0124**
	(0.0106)	(0.0022)	(0.0206)	(0.0056)
Unemployed	0.0329**	0.0017	-0.0551**	0.0091
	(0.0132)	(0.0033)	(0.0259)	(0.0080)
Children at home	-0.0175	-0.0046	0.0247	0.0134*
	(0.0267)	(0.0029)	(0.0618)	(0.0079)
Ill-health	0.0169**	0.0004	0.0282*	0.0046
	(0.0066)	(0.0013)	(0.0150)	(0.00338)
Ln (Household size)	-0.0760***	-0.0039	-0.0999***	-0.0253***
	(0.0152)	(0.0025)	(0.0296)	(0.0049)
Participation rate	0.0160	-0.0045	0.0281	0.0129*
	(0.0120)	(0.0034)	(0.0253)	(0.0066)
Part-time rate	-0.0150*	-0.0063**	-0.0156	-0.0145***
	(0.0078)	(0.0028)	(0.0168)	(0.0052)
Unemployment rate	-0.0085	-0.0002	0.0086	0.00552
	(0.0088)	(0.0022)	(0.0165)	(0.0056)
Per capita GDP	1.50e-06	1.01e-06	-1.52e-05	3.26e-07
	(9.34e-06)	(3.09e-06)	(1.71e-05)	(5.55e-06)
GDP growth	-0.0078*	-0.0013	0.0043	-0.0022
	(0.0043)	(0.0012)	(0.0080)	(0.0021)
Constant	0.7630	0.8370***	0.4420	0.5650
	(0.7410)	(0.2340)	(1.5850)	(0.4010)
Year fixed-effects	Yes	Yes	Yes	Yes
Region fixed-effects	Yes	Yes	Yes	Yes
No autocorrelation of order 1	0.0000	0.0000	0.0000	0.0000
No autocorrelation of order 2	0.1438	0.1355	0.7433	0.7826
Valid overidentifying	0.1095	0.4911	0.1275	0.4321
Observations	22,652	97,622	4,794	47,514

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

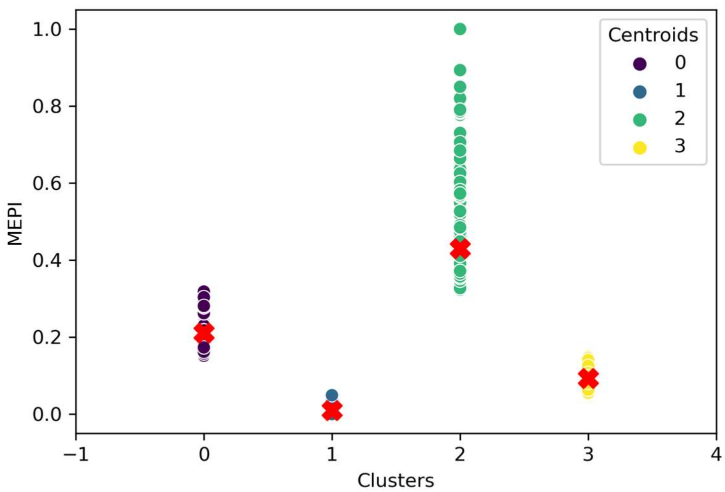
Figures

Figure 1. Inertia by number of clusters



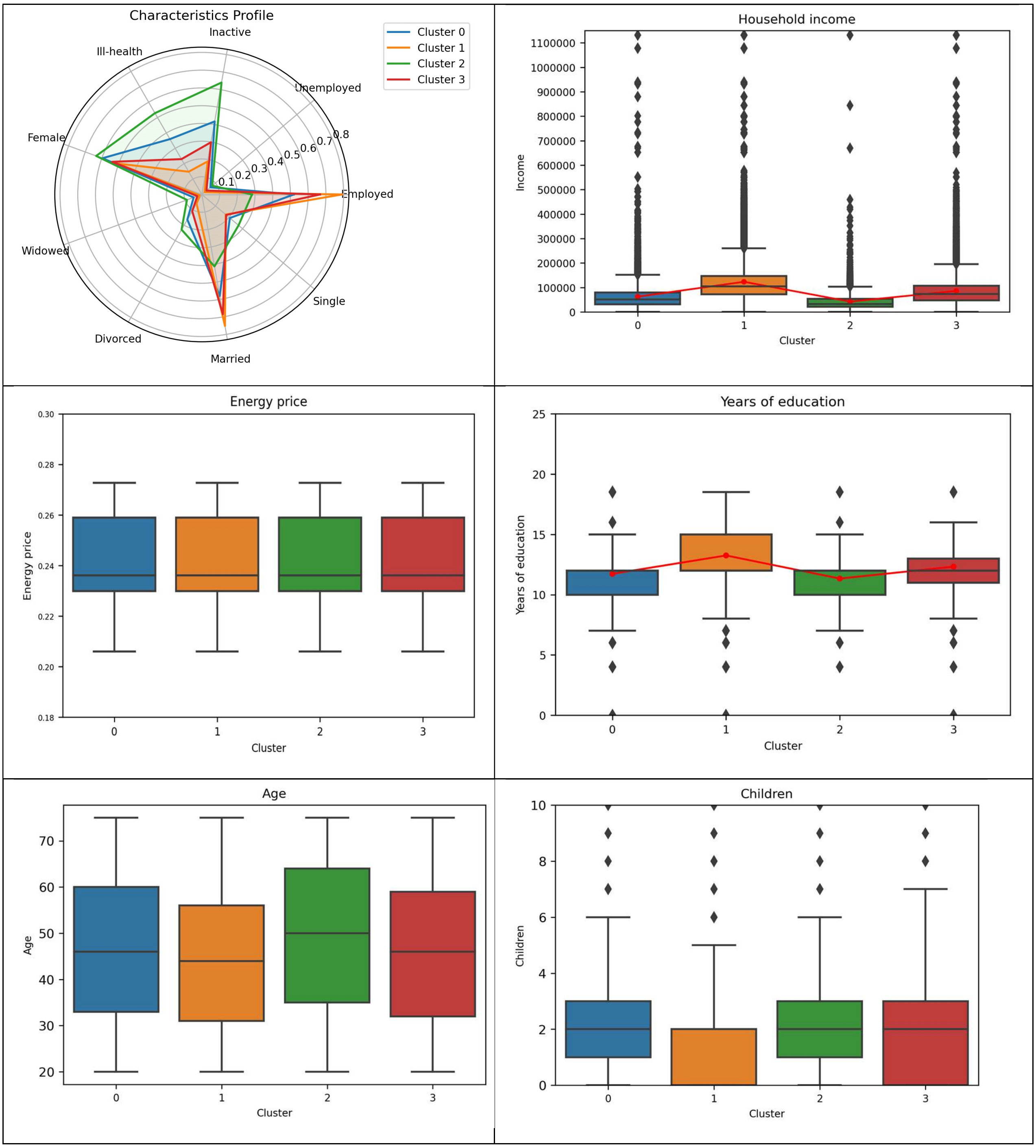
Notes: The Elbow Method determines the optimal number of clusters by iterating from k=1 to k=n (here n=10) and identifying the "elbow point", where the decrease in inertia significantly slows. The optimal number of clusters can be highlighted using a red line on the plot, K=4 is selected as the optimal number of clusters.

Figure 2. Energy poverty centroids by clusters



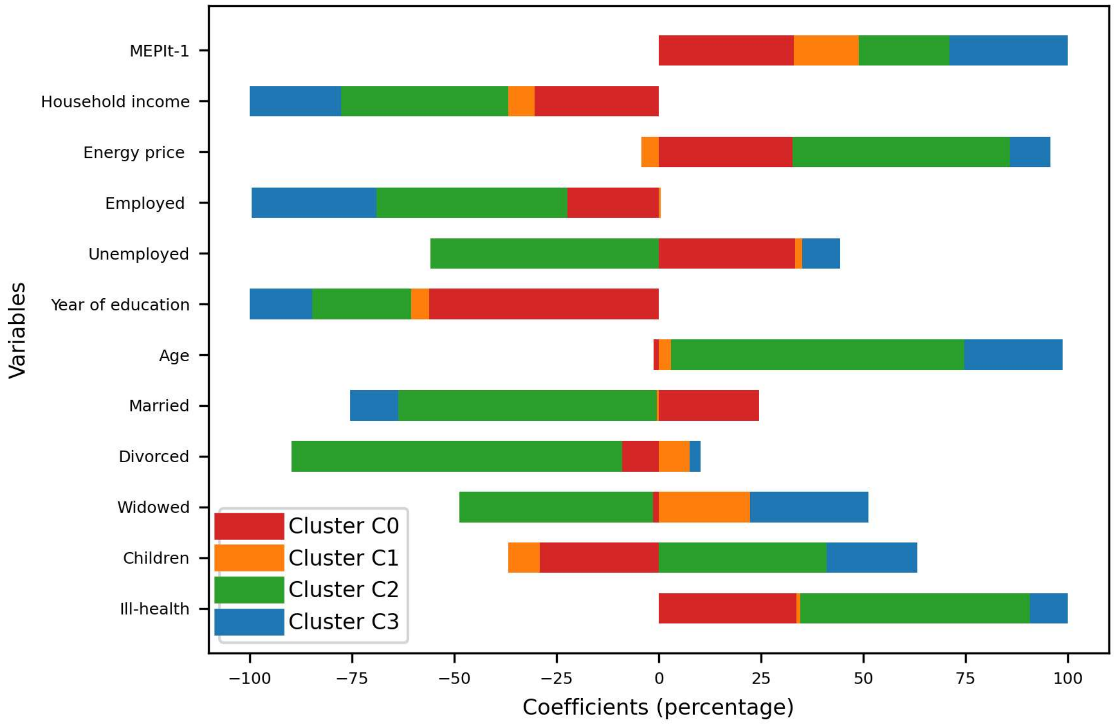
Notes: The figure illustrates the energy poverty centroids for the four clusters identified in the k-means analysis. Each centroid represents the average values of the variables within its respective cluster, summarizing the key characteristics of individuals assigned to that group. Data points are grouped based on their similarity, with each point assigned to the nearest centroid. The centroids depicted in the figure serve as the central "profiles" around which each cluster is formed.

Figure 3. Summary statistics by cluster



Notes: Summary statistics across clusters. The first panel is a radar chart depicting categorical attributes. The remaining panels are boxplots to show the distribution and variability of variables across clusters. These visualizations summarize key demographic and socioeconomic patterns for each cluster.

Figure 4. Summary of coefficients by cluster



Notes: Percentage contribution of each variable's coefficients to the four identified clusters (C0, C1, C2, C3). Variables are shown along the y-axis, while the x-axis represents the coefficient percentages, with positive and negative values indicating the direction of their effect. Each horizontal bar totals 100%, divided proportionally among the clusters. This visualization highlights how key variables contribute differently across clusters.

Appendix

Table A1. GMM estimates for energy poverty by cluster – Including two lags of the dependent variable

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
$MEPI_{t-1}$	0.111***	0.0688***	0.0163	0.0985***
	(0.0199)	(0.0105)	(0.0426)	(0.0105)
$MEPI_{t-2}$	0.0233	0.0206***	0.0221	0.0154*
	(0.0154)	(0.0070)	(0.0290)	(0.0086)
Ln (Household income)	-0.194***	-0.0428***	-0.241***	-0.137***
	(0.0079)	(0.0020)	(0.0166)	(0.0042)
Ln (Energy Price)	0.224***	-0.0188	0.290**	0.0918***
	(0.0540)	(0.0151)	(0.124)	(0.0262)
Ln (Years of education)	-0.273	-0.0204	-0.0189	-0.0027
	(0.175)	(0.0176)	(0.216)	(0.0641)
Age	-0.0018	0.0044**	0.0457**	0.0078**
	(0.0110)	(0.0021)	(0.0217)	-0.0037
Age^2	-8.24e-05	4.01e-07	-0.0006***	-2.90e-05
	-1.01E-05	(1.14e-05)	-1.03E-05	(3.28e-05)
Married	0.0433	-0.003	0.0199	-0.0078
	(0.0276)	-0.0034	(0.0852)	(0.0121)
Divorced	0.0054	0.0031	-0.0417	1.80e-05
	(0.0332)	(0.0058)	(0.0906)	(0.0160)
Widowed	0.0112	0.0434**	0.0034	0.0754**
	(0.0548)	(0.0185)	(0.0976)	(0.0332)
Employed	0.0052	-0.0025	-0.0027	-0.0110*
	(0.0116)	(0.0022)	(0.0244)	(0.0066)
Unemployed	0.0303**	0.0001	-0.0482	0.0073
	(0.0142)	(0.0035)	(0.0299)	(0.0097)
Children at home	0.0292	-0.00141	0.0273	0.0206**
	(0.0316)	(0.0028)	(0.0673)	(0.0096)
Ill-health	0.0096	0.0008	0.0124	0.005
	(0.0073)	(0.0014)	(0.0171)	(0.0039)
Ln (Household size)	-0.0741***	-0.0072***	-0.139***	-0.0258***
	(0.0154)	(0.0021)	(0.0317)	(0.0054)
Participation rate	0.00580	-0.0002	0.0435	0.0161**
	(0.0126)	(0.0033)	(0.0269)	(0.0068)
Part-time rate	-0.0253***	-0.0032	-0.0187	-0.0152***
	(0.0094)	(0.0028)	(0.0188)	(0.0055)
Unemployment rate	0.0012	0.00112	0.0246	0.0093*
	(0.0086)	(0.0021)	(0.0172)	(0.0056)
Per capita GDP	1.03e-05	-2.12e-07	-1.71e-06	-3.18e-06
	(1.08e-05)	(3.30e-06)	(2.26e-05)	(6.11e-06)
GDP growth	-0.0079*	-0.0007	0.0009	-0.0008
	(0.0047)	(0.0012)	(0.0104)	(0.0022)
Constant	1.116	0.587***	-2.229	0.423
	(0.834)	(0.226)	(1.581)	(0.436)
Year fixed-effects	Yes	Yes	Yes	Yes
Region fixed-effects	Yes	Yes	Yes	Yes
No autocorrelation of order 1	0.0000	0.0000	0.0000	0.0000
No autocorrelation of order 2	0.0936	0.2461	0.5202	0.0200
Valid overidentifying	0.1829	0.0879	0.1920	0.2335
Observations	16,779	78,395	3,331	37,988

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A2. GMM estimates for energy poverty by cluster – Endogenous income & employment

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
$MEPI_{t-1}$	0.0856***	0.0623***	0.0488***	0.103***
	(0.0121)	(0.0064)	(0.0169)	(0.0079)
Ln (Household income)	-0.253***	-0.0309***	-0.274***	-0.130***
	(0.0226)	(0.0051)	(0.0260)	(0.0154)
Ln (Energy Price)	0.157***	-0.0011	0.282***	0.0513***
	(0.0302)	(0.0039)	(0.0558)	(0.0135)
Ln (Years of education)	0.0036	-0.0136	-0.0695	0.0099
	(0.0816)	(0.0117)	(0.1450)	(0.0438)
Age	0.0193***	0.0029***	0.0200*	0.0128***
	(0.0057)	(0.0009)	(0.0108)	(0.0027)
Age^2	-0.0002***	0.0000	-0.0003***	-0.0001***
	(0.0001)	(0.0000)	(0.0001)	(0.0000)
Married	-0.0011	-0.0035*	-0.0274	-0.0063
	(0.0161)	(0.0021)	(0.0284)	(0.0077)
Divorced	-0.0063	0.0028	-0.0958**	0.0077
	(0.0217)	(0.0039)	(0.0391)	(0.0115)
Widowed	0.0265	0.0344**	-0.0899*	0.0520*
	(0.0384)	(0.0145)	(0.0474)	(0.0268)
Employed	-0.0836***	-0.0081	0.0028	-0.0570***
	(0.0277)	(0.0079)	(0.0285)	(0.0175)
Unemployed	-0.188***	0.0063	-0.0546	-0.0378
	(0.0435)	(0.0178)	(0.0398)	(0.0390)
Children at home	-0.0537**	-0.0024	-0.0028	0.0051
	(0.0231)	(0.0027)	(0.0339)	(0.0083)
Ill-health	0.0131**	0.0006	0.0320***	0.0027
	(0.0055)	(0.0011)	(0.0100)	(0.0030)
Ln (Household size)	-0.0462***	-0.0079***	-0.106***	-0.0270***
	(0.0103)	(0.0013)	(0.0173)	(0.0041)
Participation rate	0.0187***	0.0021**	0.0247**	0.0071**
	(0.0057)	(0.0008)	(0.0115)	(0.0031)
Part-time rate	-0.0116**	-0.0026***	-0.0047	-0.0053**
	(0.0046)	(0.0007)	(0.0073)	(0.0023)
Unemployment rate	-0.0012	0.0006	0.0097	0.0020
	(0.0047)	(0.0007)	(0.0091)	(0.0026)
Per capita GDP	-0.0000	-0.0000***	-0.0000	-0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GDP growth	-0.0038	0.0002	-0.0018	0.0010
	(0.0026)	(0.0003)	(0.0047)	(0.0011)
Constant	1.410***	0.399***	0.4310	1.059***
	(0.4250)	(0.0711)	(0.8550)	(0.2260)
Year fixed-effects	Yes	Yes	Yes	Yes
Region fixed-effects	Yes	Yes	Yes	Yes
No autocorrelation of order 1	0.0000	0.0000	0.0000	0.0000
No autocorrelation of order 2	0.6778	0.2862	0.1591	0.0898
Valid overidentifying	0.0035	0.1490	0.4630	0.8527
Observations	22,652	97,622	4,794	47,514

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table A3. GMM estimates for energy poverty by cluster – Endogenous regressors

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
$MEPI_{t-1}$	0.0732***	0.0595***	0.0119***	0.0993***
	(0.0083)	(0.0051)	(0.0013)	(0.0063)
Ln (Household income)	-0.204***	-0.0237***	-0.244***	-0.120***
	(0.0099)	(0.0027)	(0.0014)	(0.0076)
Ln (Energy Price)	0.1640***	-0.0032	0.2570***	0.0398***
	(0.0192)	(0.0027)	(0.0045)	(0.0104)
Ln (Years of education)	-0.1180	-0.0098	0.555***	0.2170
	(0.1980)	(0.0237)	(0.0243)	(0.1410)
Age	0.0113***	0.0005	0.0245***	0.0070**
	(0.0043)	(0.0009)	(0.0010)	(0.0028)
Age^2	-0.0001***	0.0000***	-0.0003***	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Married	-0.0377	0.0000	-0.0763***	-0.0289*
	(0.0283)	(0.0047)	(0.0060)	(0.0168)
Divorced	0.0063	-0.0038	-0.161***	-0.0331
	(0.0447)	(0.0094)	(0.0056)	(0.0302)
Widowed	-0.0828	0.0717***	-0.187***	0.0056
	(0.0540)	(0.0101)	(0.0073)	(0.0634)
Employed	-0.0487***	0.0060	0.0062***	-0.0423***
	(0.0148)	(0.0046)	(0.0019)	(0.0118)
Unemployed	-0.0298	0.0028	-0.0394***	0.0403*
	(0.0208)	(0.0106)	(0.0032)	(0.0210)
Children at home	0.0488	0.0144***	-0.0942***	0.0192
	(0.0341)	(0.0047)	(0.0056)	(0.0157)
Ill-health	0.0379**	-0.0026	-0.0215***	-0.0002
	(0.0148)	(0.0022)	(0.0019)	(0.0098)
Ln (Household size)	-0.0789***	-0.0063***	-0.0961***	-0.0170***
	(0.0103)	(0.0014)	(0.0022)	(0.0041)
Participation rate	0.0134***	0.0013**	0.0070***	0.0059**
	(0.0038)	(0.0006)	(0.0009)	(0.0023)
Part-time rate	-0.0069**	-0.0014***	0.0027***	-0.0036**
	(0.0030)	(0.0005)	(0.0006)	(0.0017)
Unemployment rate	-0.0026	-0.0003	0.0076***	0.0023
	(0.0032)	(0.0005)	(0.0008)	(0.0018)
Per capita GDP	-0.0000***	-0.0000***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
GDP growth	-0.0009	0.0002	0.0055***	0.0010
	(0.0013)	(0.0002)	(0.0004)	(0.0007)
Constant	1.707***	0.345***	0.284***	0.4940
	(0.4980)	(0.0685)	(0.0722)	(0.3660)
Year fixed-effects	Yes	Yes	Yes	Yes
Region fixed-effects	Yes	Yes	Yes	Yes
No autocorrelation of order 1	0.0000	0.0000	0.0000	0.0000
No autocorrelation of order 2	0.9332	0.1855	0.0952	0.1482
Valid overidentifying	0.0003	0.0210	1.0000	0.6206
Observations	22,652	97,622	4,794	47,514

Notes: Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

References

1. Abbas, K., Li, S., Xu, D., Baz, K., & Rakhmetova, A. (2020). Do socioeconomic factors determine household multidimensional energy poverty? Empirical evidence from South Asia. *Energy Policy, 146*, 111754. https://doi.org/10.1016/j.enpol.2020.111754
2. Australian Bureau of Statistics, 2024. Consumer Price Index, Aus843 tralia. URL: https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/consumer-price-index-australia/mar-quarter-2024#data-downloads. (Accessed on 29 March 2024).
3. Alem, Y., & Demeke, E. (2020). The persistence of energy poverty: A dynamic probit analysis. *Energy Economics*, *90*, 104789. https://doi.org/10.1016/j.eneco.2020.104789
4. Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*(1), 29-51. https://doi.org/10.1016/0304-4076(94)01642-D
5. Awan, A., Bilgili, F., & Rahut, D. B. (2022). Energy poverty trends and determinants in Pakistan: Empirical evidence from eight waves of HIES 1998–2019. *Renewable and Sustainable Energy Reviews, 158*, 112157. https://doi.org/10.1016/j.rser.2022.112157
6. Banerjee, R., Mishra, V., & Maruta, A. A. (2021). Energy poverty, health and education outcomes: evidence from the developing world. *Energy economics, 101*, 105447. https://doi.org/10.1016/j.eneco.2021.105447
7. Bao, Y., & Liao, T. (2024). Multidimensional poverty and growth: Evidence from India 1998–2021. *Economic Modelling*, *130*, 106586. https://doi.org/10.1016/j.econmod.2023.106586
8. Best, R., & Burke, P. J. (2019). Factors contributing to energy-related financial stress in Australia. *Economic Record*, *95*(311), 462-479. https://doi.org/10.1111/1475-4932.12504
9. Boardman, B. (1991). Fuel poverty is different. *Policy Studies, 12*(4), 30–41. https://doi.org/10.1080/01442879108423600.
10. Boţa-Avram, C., Apostu, S. A., Ivan, R., & Achim, M. V. (2024). Exploring the impact of macro-determinant factors on energy resource depletion: Evidence from a worldwide cross-country panel data analysis. *Energy Economics*, *130*, 107341. https://doi.org/10.1016/j.eneco.2024.107341
11. Bouzarovski, S., & Simcock, N. (2017). Spatializing energy justice. *Energy Policy*, 107, 640-648.
12. Burlinson, A., Davillas, A., Giulietti, M., & Price, C. W. (2024). Household energy price resilience in the face of gas and electricity market crises. *Energy Economics*, 132, 107414.

13. Che, C., & Tian, J. (2024). Analyzing patterns in Airbnb listing prices and their classification in London through geospatial distribution analysis. *Advances in Engineering Innovation, 12*, 53-59. https://doi.org/10.54254/2977-3903/12/2024132
14. Chandrashekeran, S., Noka, V., & Bouzarovski, S. (2022). Energy poverty: Measurement and governance in Europe and lessons for Australia. *Australian Economic Review*, 55(4), 491-502.
15. Cheikh, N. B., Zaied, Y. B., & Nguyen, D. K. (2023). Understanding energy poverty drivers in Europe. *Energy Policy*, *183*, 113818.
16. Churchill, A. S., & Smyth, R. (2021). Energy poverty and health: Panel data evidence from Australia. *Energy Economics, 97*, 105219. https://doi.org/10.1016/j.eneco.2021.105219
17. Çılgın, C., & Gökçen, H. (2024). A Hybrid Machine Learning Model Architecture with Clustering Analysis and Stacking Ensemble for Real Estate Price Prediction. *Computational Economics,* 1-52. https://doi.org/10.1007/s10614-024-10703-4
18. Crentsil, A. O., Asuman, D., & Fenny, A. P. (2019). Assessing the determinants and drivers of multidimensional energy poverty in Ghana. *Energy Policy*, *133*, 110884. https://doi.org/10.1016/j.enpol.2019.110884
19. Cui, M. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance, 1*(1), 5-8. https://doi.org/10.23977/accaf.2020.010102
20. Dalla Longa, F., Sweerts, B., & van der Zwaan, B. (2021). Exploring the complex origins of energy poverty in The Netherlands with machine learning. *Energy Policy*, *156*, 112373.
21. Decancq, K., & Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An Overview. *Econometric Reviews, 32*(1), 7-34. https://doi.org/10.1080/07474938.2012.690641
22. Dhongde, S., Pattanaik, P. K., & Xu, Y. (2019). Well-Being, Deprivation, and the Great Recession in the U.S.: A Study in A Multidimensional Framework. *Review of Income and Wealth, 65*(S1), S281-S306. https://doi.org/10.1111/roiw.12411
23. Drescher, K., & Janzen, B. (2021). Determinants, persistence, and dynamics of energy poverty: An empirical assessment using German household survey data. *Energy Economics, 102*, 105433. https://doi.org/10.1016/j.eneco.2021.105433
24. Fry, J. M., Farrell, L., & Temple, J. B. (2022). Energy poverty and retirement income sources in Australia. *Energy Economics, 106*, 105793. https://doi.org/10.1016/j.eneco.2021.105793
25. Gawusu, S., & Ahmed, A. (2024). Analyzing variability in urban energy poverty: A stochastic modeling and Monte Carlo simulation approach. *Energy*, 304, 132194.
26. Gök, A., & Sodhi, N. (2021). The environmental impact of governance: a system-generalized method of moments analysis. *Environmental Science and Pollution Research, 28*(25), 32995-33008. https://doi.org/10.1007/s11356-021-12903-z

27. Guevara, Z., Mendoza-Tinoco, D., & Silva, D. (2023). The theoretical peculiarities of energy poverty research: A systematic literature review. *Energy Research & Social Science*, 105, 103274.
28. Halkos, G., & Kostakis, I. (2023). Exploring the persistence and transience of energy poverty: Evidence from a Greek household survey. *Energy Efficiency, 16*(6), 50. https://doi.org/10.1007/s12053-023-10137-1
29. Huang, Y., Jiao, W., Wang, K., Li, E., Yan, Y., Chen, J., & Guo, X. (2022). Examining the multidimensional energy poverty trap and its determinants: An empirical analysis at household and community levels in six provinces of China. *Energy Policy*, *169*, 113193.
30. International Energy Agency, 2024. World Energy Outlook 2024. URL:https://www.iea.org/reports/world-energy-outlook-2024. (Accessed on 2 September 2024).
31. Igawa, M., & Managi, S. (2022). Energy poverty and income inequality: An economic analysis of 37 countries. *Applied Energy, 306*, 118076. https://doi.org/10.1016/j.apenergy.2021.118076
32. Karpinska, L., & Śmiech, S. (2020). Invisible energy poverty? Analysing housing costs in Central and Eastern Europe. *Energy Research & Social Science*, *70*, 101670. https://doi.org/10.1016/j.erss.2020.101670
33. Khalid, U., Shafiullah, M., & Chaudhry, S. M. (2024). Does conflict aggravate energy poverty?. *Energy Policy*, *194*, 114317.
34. Kocak, E., Ulug, E. E., & Oralhan, B. (2023). The impact of electricity from renewable and non-renewable sources on energy poverty and greenhouse gas emissions (GHGs): Empirical evidence and policy implications. *Energy, 272*, 127125. https://doi.org/10.1016/j.energy.2023.127125
35. Kripfganz, S., & Schwarz, C. (2019). Estimation of linear dynamic panel data models with time-invariant regressors. *Journal of Applied Econometrics, 34*(4), 526-546. https://doi.org/10.1002/jae.2681
36. Lee, C. C., & Yuan, Z. (2024). Impact of energy poverty on public health: A non-linear study from an international perspective. *World Development*, *174*, 106444. https://doi.org/10.1016/j.worlddev.2023.106444
37. Leszczensky, L., & Wolbring, T. (2022). How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study. *Sociological Methods & Research, 51*(2), 837-865. https://doi.org/10.1177/0049124119882473
38. Lin, B., & Okyere, M. A. (2021). Does energy poverty affect the well-being of people: Evidence from Ghana. *Sustainable Production and Consumption, 28,* 675-685. https://doi.org/10.1016/j.spc.2021.06.031

39. Mardianto, M. F. F., Siregar, N. R. A. A., Soewignjo, S., Putri, F. F. R., Prayogi, H., Imama, C., ... & Dewi, D. A. (2024). Time Series Clustering Analysis for Increases Food Commodity Prices in Indonesia Based on K-Means Method. *Journal of Human, Earth, and Future, 5*(3), 319-329. https://doi.org/10.28991/HEF-2024-05-03-02
40. Nussbaumer, P., Bazilian, M., & Modi, V. (2012). Measuring energy poverty: Focusing on what matters. *Renewable and Sustainable Energy Reviews, 16*(1), 231-243. https://doi.org/10.1016/j.rser.2011.07.150
41. OECD. (2022). OECD Economic Outlook, Volume 2022 Issue 2: Preliminary version. OECD. https://doi.org/10.1787/f6da2159-en
42. Ofetotse, E. L., Essah, E. A., & Yao, R. (2021). Evaluating the determinants of household electricity consumption using cluster analysis. *Journal of Building Engineering, 43*, 102487. https://doi.org/10.1016/j.jobe.2021.102487
43. Prakash, K., Churchill, S. A., & Smyth, R. (2022). Are you puffing your Children's future away? Energy poverty and childhood exposure to passive smoking. *Economic modelling*, *114*, 105937.
44. Price, W.-C., Brazier, K., & Wang, W. (2012). Objective and subjective measures of fuel poverty. *Energy Policy, 49*, 33-39. https://doi.org/10.1016/j.enpol.2011.11.095
45. Proctol, D. (2022, marzo 8). Coal Use Rises, Prices Soar as War Impacts Energy Markets. https://www.powermag.com/coal-use-rises-prices-soar-as-war-impacts-energy-markets/
46. Phoumin, H., & Kimura, F. (2019). Cambodia's energy poverty and its effects on social wellbeing: Empirical evidence and policy implications. *Energy Policy*, *132*, 283-289. https://doi.org/10.1016/j.enpol.2019.05.032
47. Pondie, T. M., Engwali, F. D., Nkoa, B. E. O., & Domguia, E. N. (2024). Energy poverty and respiratory health in Sub-Saharan Africa: Effects and transmission channels. *Energy*, *297*, 131158. https://doi.org/10.1016/j.energy.2024.131158.
48. Siksnelyte-Butkiene, I., Streimikiene, D., Lekavicius, V., & Balezentis, T. (2021). Energy poverty indicators: A systematic literature review and comprehensive analysis of integrity. *Sustainable Cities and Society*, 67, 102756.
49. Spandagos, C., Reaños, M. A. T., & Lynch, M. Á. (2023). Energy poverty prediction and effective targeting for just transitions with machine learning. *Energy Economics*, 128, 107131.
50. Stevens, M. M., Ambrosio-Albala, P., van Ooij, C., Searle, B. A., Garcia, N., Rocher, E., ... & van Grieken, A. (2024). Exploring the relationship between energy poverty and mental wellbeing: a qualitative study. *European Journal of Public Health*, *34*(Supplement_3), ckae144-028. https://doi.org/10.1093/eurpub/ckae144.028

51. Sy, S. A., & Mokaddem, L. (2022). Energy poverty in developing countries: A review of the concept and its measurements. *Energy Research & Social Science, 89*, 102562. https://doi.org/10.1016/j.erss.2022.102562
52. Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability, 14*(12), 7243. https://doi.org/10.3390/su14127243
53. Willand, N., Torabi, N., & Horne, R. (2023). Recognition justice in Australia: hidden energy vulnerability through the experiences of intermediaries. *Energy Research & Social Science*, *98*, 103013.
54. Wu, J. (2012). *Advances in K-means clustering: a data mining thinking*. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-29807-3
55. Zhang, Z., Shu, H., Yi, H., & Wang, X. (2021). Household multidimensional energy poverty and its impacts on physical and mental health. *Energy Policy*, *156*, 112381.