



LOCATEE

LOCAL AUTHORITIES TACKLING ENERGY POVERTY
IN PRIVATE MULTI-APARTMENT BUILDINGS

Deliverable 4.2 – The microsimulation procedure behind the building-level energy poverty indicators

Sep 30, 2026

Leader Organization: Institute for Structural Research – IBS

Author: Jakub Sokołowski (IBS)



Co-funded by the European Union under n°101167621. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Disclaimer

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the European Climate, Infrastructure and Environment Executive Agency (CINEA) nor the European Commission is responsible for any use that may be made of the information contained therein.

Copyright Message

This report, if not confidential, is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0); a copy is available here: <https://creativecommons.org/licenses/by/4.0/>. You are free to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) under the following terms: (i) attribution (you must give appropriate credit, provide a link to the license, and indicate if changes were made; you may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use); (ii) no additional restrictions (you may not apply legal terms or technological measures that legally restrict others from doing anything the license permits).

Project Acronym:	LIFE23-CET-LOCATEE
Project full name:	LOCal Authorities to Tackle Energy povErty in private multi-apartment buildings
Grant Agreement No.:	101167621
Project URL:	https://ibs.org.pl/en/research/locatee-local-authorities-tackling-energy-poverty-in-private-multi-apartment-buildings/
EU Project Officer:	Vinicius Valente
Project Coordinator:	Institute for Structural Research (IBS)
Deliverable:	D4.2 – Database with energy poverty indicators
Work Package:	WP4 – Design and development of the LOCATEE toolkit
Date of Submission:	Project month 24 (30/09/2026)
Dissemination Level:	Public
Lead Beneficiary:	Institute for Structural Research (IBS)
Responsible Author:	Jakub Sokołowski (Institute for Structural Research)

About LOCATEE

LOCATEE aims to support local municipalities in addressing energy poverty through the renovation of private multi-apartment buildings for vulnerable residents. LOCATEE will achieve this goal by providing a toolkit for identifying energy-vulnerable households, matching tailored interventions to their needs, and integrating energy poverty alleviation activities into long-term strategies of municipalities such as Sustainable Energy and Climate Action Plans. LOCATEE will use administrative data to create household and building typologies to identify priority intervention locations. This process will help authorities and social partners address local energy poverty through coordinated solutions, including contact points and focus groups with housing entities, to facilitate knowledge exchange on renovation programs and targeted solutions.

The evidence based and collaborative approach will be implemented in three pilot municipalities in Central, Southern and Southeastern Europe: Piraeus (Greece), Rumia (Poland), and Torres Vedras (Portugal) and, while ensuring scaling up the LOCATEE framework to more municipalities and regions across Europe.

WHO WE ARE

	Participant full name	Short name	Country	Logo
1	Institute for Structural Research	IBS	PL	
2	Technoeconomics of Energy Systems laboratory – University of Piraeus Research Center	TEESlab - UPRC	EL	
3	Universidade NOVA de Lisboa	NOVA	PT	
4	Institute for European Energy and Climate Policy	IEECP	NL	
5	Gmina Miejska Rumia	COR	PL	
6	Municipality of Piraeus	MOP	EL	
7	Município De Torres Vedras	COTV	PT	
8	Austrian Energy Agency	AEA	AT	
9	Nadacia Habitat for Humanity International	NHFHI	SK	
10	EURACTIV Media	EURACTIV	BE	

Executive Summary

The Deliverable 4.2 is a set of spreadsheets containing building-level and aggregated outputs of the LIHC (Low Income, High Costs) energy poverty indicator for the three pilot sites in the LOCATEE project: Rumia, Torres Vedras, and Piraeus. The indicator is estimated using a microsimulation procedure which generates synthetic households based on available building, demographic and national household survey data and estimates the share that meets the LIHC energy poverty criteria. The resulting value represents the estimated probability of energy poverty for households that could plausibly occupy a given building and is intended to support the identification and prioritisation of buildings and areas for energy poverty alleviation and renovation measures in the Energy Poverty Monitoring System (EPMS), developed by the consortium. The database and methodology provide reproducible and comparable outputs while ensuring that individual households are neither identified nor assessed.

1. Introduction

The D4.2 database is a set of outputs, one spreadsheet per pilot site, with one row per building. Each row carries an estimated probability of energy poverty for that building. The estimate is produced by a microsimulation: a procedure that generates large numbers of synthetic households consistent with what is known about each building, and measures how many of them would be energy poor.

This note explains that procedure. It is organised so that the reasoning can be followed at two levels. A reader who only needs to understand what the numbers mean should read Section 2. A reader who needs to reproduce, audit or extend the estimates should read Sections 3 to 6, which give the formal specification, and the annex, which lists the parameters.

The database reports a single indicator, LIHC (Low Income, High Costs). The reasons for reporting one indicator rather than several are given in Section 3.

1.1 The estimand

For each building, the quantity being estimated is a conditional probability: given everything the administrative registers record about this building, what share of the households that could plausibly occupy it would meet the energy poverty criteria? Formally, for building b with attribute vector x_b ,

$$p_b = \Pr(\text{household is energy poor} \mid x_b)$$

This is a statement about a building and the class of households compatible with it, not about the specific people currently living there. Section 2.3 develops that distinction, which governs how the results may legitimately be used.

2. The procedure

2.1 Why simulation is necessary

Energy poverty can be defined by two facts about a household: how much it earns and how much it must spend on energy. Neither fact is recorded at the level of an individual building. Income is held by tax and statistical authorities and released only in anonymised or aggregated form; energy consumption is held by utilities and is confidential; and no municipality can survey every household in its territory. What municipalities hold is administrative information about buildings (e.g., location, age, size, heating system, energy performance) and the age profile of registered residents, and this information is nearly complete.

The gap is therefore structural: the information that exists concerns buildings, whereas the information required concerns households. Microsimulation bridges it by generating households that could plausibly occupy these particular buildings.

2.2 What the procedure does

For each building, the procedure repeats a single operation several thousand times. It constructs a synthetic household by drawing a household size, income, electricity cost, and heating cost, each from distributions taken from the national household budget survey and adjusted for what is known about that specific building. It then applies the energy poverty definition to the synthetic household's budget and records whether the household is poor. After several thousand repetitions, the share of synthetic households that emerged poor is the building's estimated probability of energy poverty.

The building attributes enter through the adjustments. A synthetic household placed in a poorly insulated pre-war block is assigned a higher heating cost than one in a certified modern building, because the survey shows that households in such buildings spend more on heat. A household in a low-income region is given a lower income than one in a wealthy region. The simulation is, in effect, a disciplined way of carrying what is known about buildings into an estimate of what is unknown about households, while preserving the uncertainty.

The microsimulation

We cannot see which households struggle to pay for energy, because income and energy bills are not recorded at the level of a single address. So instead of observing them, we simulate them. For every building, we generate thousands of realistic households, i.e. plausible sizes, plausible incomes for the region, plausible bills for a building of that age, size and heating system, and count how many would fall below the energy poverty line. That share is the building's score. A high score does not mean the people living there are in poverty; it means a building like that one, in a place like that one, is where energy poverty is most likely to occur, and therefore where support should first be directed.

2.3 Building-level probability

The interpretation of the estimate is fixed by how it is produced.

Table 1: Interpretation of the building-level probability

The probability is	The probability is not
A property of a class of buildings: among households that could plausibly occupy a building like this one, this share would meet the energy poverty criteria.	A property of the actual occupants. Their incomes are neither observed nor individually inferred.
A device for ranking buildings, so that limited renovation and outreach resources can be directed where energy poverty is most likely.	A test of individual eligibility for any benefit, allowance or subsidy.
An estimate with a quantified sampling error and an unquantified calibration uncertainty.	A measurement. No real household budget is observed at any point.
Reproducible: identical inputs and random seed reproduce identical values.	Comparable across model versions. Recalibration shifts the values; comparisons hold within a version.

Source: own elaboration.

Two consequences follow. First, the estimates support decisions about buildings and areas, e.g. which envelopes to insulate, which heating systems to replace, where to concentrate outreach, but not decisions about individuals. Second, the procedure identifies energy poverty that is driven by the building. A low-income household in an efficient modern building is not detected, because nothing observable distinguishes that building from its neighbours. The estimates are a targeting instrument for building-related energy poverty, complementary to, and not a substitute for, the case knowledge held by social services.

3. The energy poverty definition

3.1 Notation

A synthetic household h has monthly disposable income Y , monthly electricity cost E_{el} , monthly space-heating cost E_{ht} , total energy cost $E = E_{el} + E_{ht}$, and size n persons. Income and energy are equivalised — adjusted for household size — using a scale $s(n)$; equivalised values are written $\frac{Y}{s(n)}$ and $\frac{E}{s(n)}$. The equivalence scale used at each site is given in the annex.

3.2 The LIHC indicator

A household is classified as energy poor under LIHC (Low Income, High Costs) when it is simultaneously below the relative income poverty line and above the median energy expenditure:

$$LIHC(h) = 1 \iff \frac{Y}{s(n)} < 0.6 \cdot \text{median} \left[\frac{Y}{s(n)} \right] \wedge \frac{E}{s(n)} > \text{median} \left[\frac{E}{s(n)} \right].$$

The first condition is the standard European at-risk-of-poverty line: an income below 60% of the national median equivalised income. The second condition restricts attention to households with energy costs above the median, which distinguishes energy poverty from income poverty in general. A household is energy-poor under LIHC only if both conditions hold.

LIHC compares each household to the distribution on two dimensions at once. If the level of simulated energy costs is uniformly too high or too low, the most likely form of calibration error is that the simulated observations and the simulated median move together, and the classification is largely preserved. An indicator compared against a fixed national threshold has no such protection: a uniform error in energy costs moves every household across the threshold in the same direction. In validation against the Polish and Greek surveys, LIHC closely reproduced the observed national prevalence.

3.3 Thresholds

The thresholds in the LIHC definition are national reference values, supplied to the procedure from an external source. This external anchoring is what makes the estimates falsifiable: because the national prevalence is defined outside the model, the simulated prevalence can be compared against the observed national rate, and a discrepancy is informative. For each site, the income line is the published at-risk-of-poverty threshold, and the energy line is the median equivalised household energy expenditure from the national household budget survey. The values used are listed in the annex.

4. Inputs

The procedure draws on two kinds of input, kept deliberately separate. The building layer describes the physical stock of a municipality and is derived from administrative registers. The calibration layer describes how households of different types earn and spend, and is drawn from the national household budget survey. The first tells the procedure what exists; the second tells it how to populate it. This separation is what makes the procedure transferable: applying it to a new country requires a new calibration layer but no change to the procedure itself.

4.1 The building layer

One record per building, from national and municipal registers. The attributes actually used to condition the simulation are as follows: not every site provides every attribute, and the procedure degrades when one is absent.

Table 2: Building attributes and their role.

Attribute	Role in the simulation
Floor area and, where available, number of dwellings	Determines the dwelling area, which scales the heating cost
Construction period	Adjusts income and heating cost, the single strongest predictor of risk
Heating system	Selects the electricity and heating cost distributions
Energy performance certificate, or building condition where no certificate exists	Scales the heating cost to the specific building rather than to its type
Region	Selects the regional income distribution
Registered residents by age band	Adjusts income and energy for the age profile of occupants
Location	Not used in the estimate; retained for mapping

Source: own elaboration.

The registers are administrative by-products rather than research instruments, and their imperfections propagate into the estimates. The most consequential are that registered residency is not the same as actual occupancy; that energy performance certificates cover only part of the stock and are biased toward newer, larger, professionally managed buildings; and that floor area is often recorded for the whole building rather than the dwelling. The procedure records, for every building, how it resolved the dwelling area (see section 5.3), so that a reader can see how this last imperfection affects a given estimate.

4.2 The calibration layer

From the national household budget survey, the calibration layer supplies the household size distribution; the regional income distributions; the distribution of electricity cost per person by building type and heating system; the distribution of heating cost per square metre by construction period and heating system; multiplicative adjustments to income and energy for construction period, building type and occupant age; the energy performance adjustment; and the national thresholds and validation target. The calibration values for each site are listed in the annex.

5. The simulation procedure

The procedure is a static, cross-sectional Monte Carlo microsimulation. It synthesises household budgets conditional on building attributes and evaluates the LHC rule on each. It contains no behavioural dynamics and no time dimension: it is a snapshot conditioned on a dated register extract.

5.1. Household size

Each synthetic household is assigned a size drawn from the national distribution of household sizes:

$$n \sim \text{Categorical}(q_1, \dots, q_5).$$

5.2. Income

Income is drawn at the household level from a regional lognormal distribution, shifted by the building's construction period, type and occupant age profile, then adjusted for household size:

$$\log Y \sim \text{Normal}(\mu_r + \log(f_\kappa \cdot f_\tau \cdot f_a), \sigma_r^2) \cdot g(n), \quad Y \leftarrow \max(Y, Y_{\min}).$$

Here μ_r and σ_r are the regional log-income parameters, f_κ , f_τ and f_a are the period, type and age adjustments, and $g(n)$ grosses a per-household draw up by size. The lower bound, $Y_{\min}$ prevents implausible draws in the extreme tail. The building adjustments encode residential sorting, i.e. the observed regularity that households in older stock have systematically lower incomes, rather than any causal effect of buildings on income.

5.3. Dwelling area

Heating costs are charged per square metre, so the area to which they apply must be the area of a dwelling, not the whole building. Supplying a building's total floor area where a dwelling is required would significantly inflate heating costs. The procedure, therefore, derives dwelling area by an explicit order of preference and records, for each building, which route it used:

1. Total floor area divided by the number of dwellings, where the dwelling count is available.
2. Total floor area divided by the number of registered residents, bounded to a plausible range and multiplied by an assumed household size.
3. Total floor area used directly, where it already falls within plausible dwelling bounds.
4. A national default dwelling size.

The result is bound to a plausible dwelling range. Because the route is recorded per building, a reader can identify estimates that rest on a weaker area assumption, and the delivered data include this indicator (the *area_{source}* column).

5.4. Energy cost

Energy cost has two components with different scaling logic. Electricity is drawn per person and multiplied by household size; heating is drawn per square metre and multiplied by dwelling area. Both are drawn from distributions whose mean and dispersion are taken from the calibration:

$$E_{el} \sim \text{LogNormal}(\text{mean} = m_{el(\tau, \sigma)}, sd = s_{el(\tau, \sigma)}) \cdot n,$$

$$E_{ht} \sim \text{LogNormal}(\text{mean} = m_{ht(\kappa, \sigma)} \cdot f_\kappa \cdot f_{cond} \cdot f_a, sd = s_{ht(\kappa, \sigma)}) \cdot A.$$

The heating mean is adjusted by construction period (f_k), by building condition where recorded (f_{cond}), and by occupant age (f_a). The lognormal form keeps energy costs positive and preserves the right-hand tail of the distribution, where energy poverty is concentrated.

5.5. Energy performance adjustment

Where a building carries an energy performance certificate, the drawn heating cost is scaled by the building's certified performance relative to the national reference, and the electricity cost by a damped version of the same factor. Where no certificate exists, as in the Greek site, the same role is played by the construction-period and building-condition adjustments in section 5.4.

$$\varphi_b = \text{clip}\left(\frac{\text{performance}_b}{\text{reference}}, 0.3, 2.5\right), \quad E_{ht} \leftarrow \varphi_b \cdot E_{ht}, \quad E_{el} \leftarrow (1 + 0.15(\varphi_b - 1)) \cdot E_{el}.$$

The bounds on φ_b prevent extreme certificate values, which in practice are frequently data-entry errors rather than genuine outliers, from dominating the estimate. The damping on electricity reflects that only a minority of electricity use depends on the building envelope.

5.6. Behavioural adjustment

Households in the lower part of the income distribution tend to ration heating rather than accumulate arrears, a well-documented response that causes expenditure-based measures to understate hidden energy poverty. The procedure represents this as a smooth reduction in heating costs that increases as income falls below a low percentile of the distribution, rather than as an abrupt cut at a fixed income, so that the adjustment does not create a discontinuity or jump when prices or incomes are reindexed. This adjustment is a documented assumption rather than an estimated relationship, and its parameters are listed in the annex.

5.7. Classification and aggregation

Each synthetic household's budget is equalised and tested against the LIHC rule. The building's estimated probability is the share of its synthetic households classified as energy poor:

$$p_b = \left(\frac{1}{N}\right) \sum_i \text{LIHC}(h_{\{b,i\}}).$$

Alongside the probability, the procedure retains the mean simulated income, energy cost, and energy share for each building, as well as the two components of the LIHC condition separately (the share below the income line and the share above the cost line). These summaries are delivered with probabilities and allow an estimate to be traced to its cause.

5.8. Reproducibility

All draws come from a seeded pseudo-random generator, and the seed for each building is derived from the building's identifier and a global seed. A building's estimate, therefore, depends only on that building and not on its position in the file, so adding or removing buildings does not disturb any other estimate, and successive annual runs remain comparable. Every run records the input and configuration checksums, the seed, the thresholds used and the validation outcome, so that any delivered figure can be reproduced exactly and audited.

6. Validation

The mean simulated prevalence across a site's buildings is compared against the observed national or regional prevalence from the household budget survey. Because the thresholds are anchored externally (section 3.3), this comparison is meaningful: the simulated rate is free to differ from the observed rate, and its closeness is evidence about the calibration.

Three testable sites reproduce the benchmark within two percentage points. Rumia lands within a tenth of a point of the Polish national rate; Piraeus lands 1.6 points above the Attiki regional rate, in the expected direction, since Piraeus is an older and denser municipality than the regional average. Torres Vedras cannot be tested because Portugal publishes no national LIHC rate; its estimates are internally consistent and validly ranked, but their overall level is not externally confirmed.

Annex — Parameters and reference values

A.1 Thresholds and validation targets

Site	Income line (monthly equiv.)	Energy line (monthly, equiv.)	Benchmark LIHC	Source
Rumia / Poland	2,100 PLN	250 PLN	10.7%	BBGD 2023
Torres Vedras	€591	€48	none	INE 2023 (2022 incomes); IDEF 2022
Piraeus	€702	€73.6	5.6%	HBS 2020, Attiki region

Source: own elaboration.

Table A.1 — National reference values by site. Polish values are placeholders pending final calibration; the Greek income and energy dispersion are provisional pending confirmation against HBS microdata

A.2 Equivalence scales

Rumia and the Polish reference use the modified OECD scale, $s(n) = 1 + 0.5(n - 1)$. Torres Vedras and Piraeus use the national discrete scale, $s = 1.0, 1.5, 1.8, 2.1, 2.4$ for one to five persons. One scale is applied consistently to both income and energy at each site.

A.3 Fixed procedure constants

Constant	Value
Energy performance factor bounds	0.3 to 2.5
Electricity damping on the performance factor	0.15
Behavioural (under-heating) trigger / maximum reduction	below the 30th income percentile / 20%
Draws per building	5,000 (1,000 for the Polish reference sample)
Random seed	42, combined with each building identifier

Source: own elaboration.

A.4 Energy performance adjustments

Poland scales heating cost by the certified primary energy demand relative to a national median of 127.6 kWh/m²/yr. Portugal scales by certificate class, from 0.50 for A+ to 1.90 for F. Greece, which has no certificates, scales by construction period (1.35 for pre-1920 stock down to 0.90 for post-1985) and by building condition (1.50 for derelict down to 0.90 for good condition). The full per-site calibration tables are held with the pipeline configuration and are reproduced in the accompanying user documentation.



LOCATEE



EURACTIV

#LOCATEE



Co-funded by the European Union under n°101167621. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.