Moving Beyond Binary Classification in Energy Poverty Analysis: A Probabilistic Machine Learning Framework

Noon Hussein
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
nihussei@ualberta.ca

Hanieh Taraghi Nazloo
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
htaraghi@ualberta.ca

Petr Musilek
Department of Electrical and Computer Engineering
University of Alberta
Edmonton, Canada
pmusilek@ualberta.ca

***Abstract*—The lack of reliable modern energy for basic needs affects more than 750 million people without electricity and 2.4 billion who rely on polluting fuels. Traditional metrics often fail to identify vulnerable households because of their static nature. Building on previous research, this study employs 15 years of longitudinal household data and machine-learning techniques to predict each household's probability of future energy poverty. The proposed composite index integrates both objective (e.g., energy spending) and subjective (e.g., self-reported hardship) indicators to capture hidden deprivation. Findings indicate that income stability is the most significant long-term protective factor, whereas energy-price shocks exert only short-term effects. Household size exhibits a weaker but non-linear influence, acting primarily as a contextual modifier of risk. These results support policies that prioritize income consistency and housing efficiency over reactive subsidies. The probabilistic model developed in this work enables proactive and targeted interventions beyond rigid threshold-based assessments.**

***Index Terms*—Energy poverty, machine learning, probabilistic forecasting, early detection, policy intervention**

I. Introduction

Energy poverty refers to the lack of access to modern energy services required to meet basic needs. In developing regions, it often manifests itself as the absence of electricity or clean cooking options, while in wealthier countries, it is commonly described as *fuel poverty*, a condition where energy is available but not affordable for adequate use. Globally, approximately 750 million people lack electricity and more than 2 billion rely on polluting fuels for cooking [1]. This persistent deprivation undermines health, education, and general well-being, representing a pressing global concern [2].

A. Measuring Energy Poverty

Different metrics capture different aspects of energy poverty:

- **Spending thresholds:** Households that spend more than a certain share of their income on energy (e.g., $\geq 10\%$) are deemed energy-poor. However, this can be misleading if families under-consume energy to keep bills low [1].
- **Self-reported hardship:** Survey indicators (e.g., inability to heat the home or utility bill arrears) reflect real struggles but are subjective.
- **Composite indices:** Combine multiple factors (income, energy cost, housing efficiency, etc.) into a single score.

There is often little overlap between households marked by different criteria; therefore, combining indicators provides a more complete picture.

B. Dynamic Risks and Prevention

Energy poverty status is fluid: households can fall into or escape it due to shocks (e.g., unemployment or fuel price spikes), and static indicators often miss these shifts. Forward-looking methods aim to identify at-risk households early to enable proactive support (e.g., insulation upgrades or bill assistance) before hardship occurs.

C. Contributions

This work builds on Budría et al. (2025) [2], extending their methodology and addressing key limitations:

- **Temporal modeling:** 15 years of historical panel data are used to capture lagged effects and mitigate reverse causality, addressing the partial reliance on contemporaneous data.
- **Probabilistic forecasting:** In contrast to the binary classification used in prior work, household-level risk probabilities are estimated for more nuanced and preventive interventions.
- **Enhanced index design:** The multidimensional energy poverty index is refined through prevalence-based weighting to reveal hidden deprivation, extending the fixed-threshold MEPI framework.
- **Policy insight refinement:** Findings support a shift from reactive price-based subsidies toward proactive policies centered on income stabilization, housing efficiency improvements, and risk-based targeting using probabilistic forecasts.

979-8-3315-8138-1

II. Literature Review

ML analyses highlight how combinations of moderate income, high energy needs (due to climate or home inefficiency), and energy prices together drive vulnerability. Budría et al. [3] showed that income volatility outweighs short-term price shocks in shaping vulnerability, while Al Kez et al. [4] and López-Vargas et al. [5] demonstrated that spatial and behavioral data enhance targeting accuracy. Kalfountzou et al. [6] further confirmed that ensemble ML models outperform traditional rules, underscoring the shift toward explainable, data-driven forecasting of energy poverty. Despite these advances, most ML-based energy poverty studies primarily focus on contemporaneous classification or single-horizon prediction. In contrast, this paper re-frames energy poverty as a multi-horizon probabilistic forecasting task and introduces a composite distribution-based index that fuses expenditure and self-assessed hardship measures, thus enabling new insights into how drivers shift across time scales (e.g., short-run shocks vs. long-run structural vulnerability), supports policy design using risk thresholds, early-warning screening, and prioritization under limited resources.

III. Methodology

A. Dataset Description

The HILDA Survey [7], an Austrailian longitudinal study examining economic, social and demographic household dynamics, is used for the analysis. Table I summarizes its data.

TABLE I: Dataset Description

Type	Value
Time Period (Years)	15
Baseline (Years)	8
Number of Households	7,977 (106,475 total observations)
Household Income ($)	58,922-133,946
Household Size (Normalized)	1.0-7.4
Age (Years)	20-65
Education (Years)	8-19
Marital Status (Binary)	Married, divorced, widowed
Parenthood	Children, no children
Health Status	Disabled, not disabled
Energy Price (¢/kWh) [8]	57.26-151.85
Growth in Gross State Product (GSP) per Capita ($)	-1.4-12.4
Employment Rate (by State)	% full-time, % part-time, % unemployed
Employment Status (Binary)	Employed, unemployed
Household Region (Discrete)	2-8

Although incorporating contemporaneous data could potentially improve model accuracy, the analysis focuses on historical information to avoid masking of lagged effects, particularly in the presence of autocorrelation. More importantly, contemporaneous variables are excluded to reduce the risk of reverse casuality between energy poverty and socio-demographic factors.

B. Energy Poverty

Five metrics were used to capture energy poverty, including: (1) expenditure-based metrics (2M, 10%, LIHC); and (2) self-assessed metrics (ability to heat home due to financial constraints, ability to pay bills on time).

Let $J = 5$ denote the set of poverty metrics, I the set of individuals, and T the set of time windows. By definition, EP_{ijt} represents the poverty status of individual i for the j^{th} metric during period t. Following the material deprivation literature [9], the resulting score, S, is:

$$S_{it} = \sum_{j \in J}(w_j \cdot EP_{ijt}), \forall i \in I, t \in T_i; T_i \subseteq T. \tag{1}$$

For each metric, the assigned weight is proportional to the percentage of population not classified as poor under its definition within a particular state:

$$w_j = \frac{1 - n_j}{\sum_{j \in J}(1 - n_j)}, \tag{2}$$

where n_j represents the proportion of energy poor individuals under metric j. Weighting is selected based on evidence showcasing that inability to access common items is a stronger indicator of deprivation than less common items [10]. Therefore, S follows a frequency-based weighting approach that improves the representation of key features.

C. Forecasting Model

Prior to training, data was standardized by removing the median and scaling to interquartile ranges (IQRs) to manage outliers, improve generalization, and prevent variables of larger magnitudes from disproportionally impacting the learning process. Specific identifiers and temporal markers were removed to prevent the model from overfitting to user-specific patterns. Given that standard ML models tend to favour the majority class, random undersampling boost (RUSBoost) was used to rectify the potential of poorly identifying the minority class (energy-poor households) by assigning higher weights to misclassified instances, thus maintaining sensitivity while ensuring robust performance.

Presented in Table II, the selected hyperparameters were optimized using a grid search, while 5-fold cross validation ensured their robustness across multiple data splits. The final configuration was chosen based on the set producing the highest Receiver Operating Characteristic Area Under the Curve (ROC-AUC) score [2]–a critical metric for this dataset, as it evaluates the model's ability to discriminate between the two classes.

TABLE II: RUSBoost Hyperparameter Description

Type	Value
Number of Estimators	50, 100, 200
Sampling Strategy	Auto, 0.5, 1.0
Window Size	2, 14
Replacement	True, False
Learning Rate	0.01, 0.1, 1.0
Algorithm	SAMME, SAMME.R
Contemporaneous Data	False

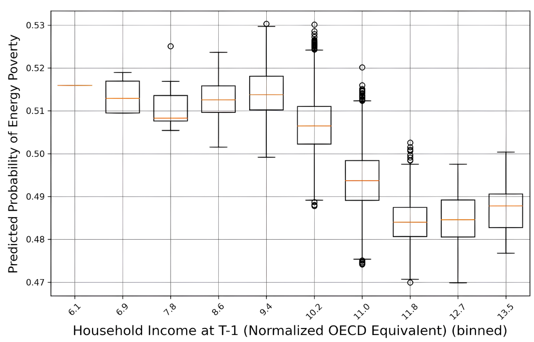
979-8-3315-8138-1

IV. Results & Discussion

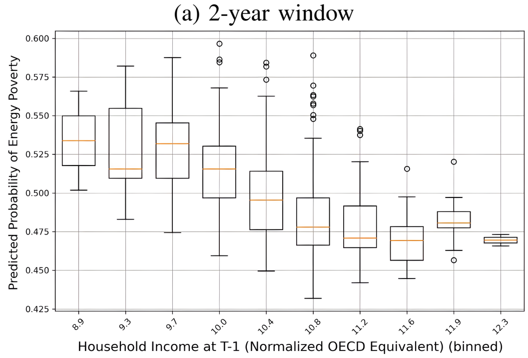
As identified in [2], 3 of the top 5 predictors for energy poverty were examined at $T-1$.

A. Household Income

Based on Fig. 1, income is the strongest single predictor in both horizons. Predicted poverty risk declines sharply across income deciles in the 2-year model, consistent with heightened short-run sensitivity to income shocks. The 14-year model preserves the negative gradient but with tighter dispersion at higher incomes, indicating greater structural resilience. In both windows, marginal risk reductions taper beyond ∼AUD 100,000, suggesting diminishing returns from additional income at the upper tail.



(a) 2-year window



(b) 14-year window

Fig. 1: Forecasted probability of energy poverty versus normalized household income at $T-1$ (OECD equivalized). OECD-normalized income = household income / equivalence factor, where the equivalence factor is defined as 1.0 for the first adult in the household, 0.5 for each additional adult (aged 14 or older), and 0.3 for each child (under 14).

In the 2-year model, median predicted probability exhibited a 13% relative reduction over a narrow income range (normalized OECD equivalent ≈6.5 to 7.8), a sharp gradient underscoring the acute vulnerability of the lowest-income households to short-term shocks.

As for the 14-year model, the gradient persists but moderates, whereas the marginal protective effect of income diminishes substantially beyond normalized income ∼10.5 (approximately the 7^{th} decile). From the 7^{th} to 10^{th} decile, median probabilities decline by only 0.01–0.02 and approach an asymptotic floor near 0.46. The resulting plateau suggests that, while higher income consistently reduces risk, structural factors independent of income (e.g., regional climate) impose a baseline vulnerability that income alone cannot eliminate.

Upper-fence outliers reveal households at elevated risk despite high income. Households in the highest income decile exhibit predicted probabilities substantially above their decile median, likely corresponding to recent income shocks (e.g., job loss, medical expenses) not yet captured in lagged income measures, or those facing high energy burdens due to inefficient housing or large family size. Such patterns reinforce the necessity of composite indicators that integrate expenditure, self-reported hardship, and dwelling characteristics to identify vulnerability that income thresholds alone would miss.

On the other hand, lower-fence outliers in the bottom income decile represent households that, despite low income, remain outside energy poverty. Such cases may benefit from subsidized housing, multi-generational cost-sharing, or receipt of energy assistance not captured in income variables. Identifying the protective mechanisms among low-income non-poor households could inform scalable intervention designs.

B. Household Size

Fig. 2 illustrates the relationship between household size at $T-1$ and forecasted energy poverty risk across short-and long-term horizons. Unlike income, household size exhibits a weakly non-linear association with risk, reflecting competing effects of resource sharing economies versus rising total energy demand as occupancy increases.

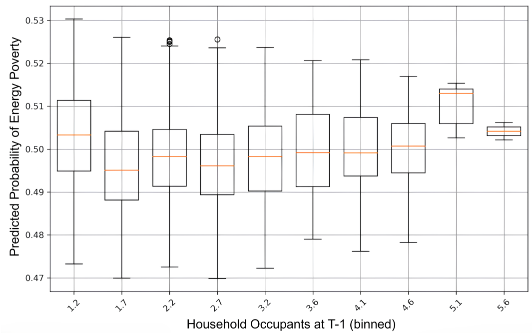
In the 2-year model, median predicted probabilities cluster tightly around 0.49–0.51 across most occupancy bins, indicating that short-run energy poverty risk is only modestly sensitive to household size. Small households (1–2 occupants) display slightly elevated medians and wider dispersion, suggesting greater vulnerability to short-term shocks such as temporary income loss or price spikes. From approximately 3-5 occupants, medians rise marginally, consistent with increasing absolute energy needs (space heating, hot water, appliance usage) that can offset economies of scale. However, the overall gradient remains shallow, implying that household size alone is not a dominant short-run driver.

The 14-year model reveals a clearer structural pattern. Median probabilities decline gradually from small households toward mid-sized households (around 3–4 occupants), indicating that, over longer horizons, multi-person households may benefit from more stable income pooling, shared fixed costs, and greater adaptive capacity. Beyond this range, risk increases again for larger households (5+ occupants), producing a U-shaped relationship. This pattern suggests that while moderate household size can be protective in the long run, very large households face persistent structural pressure from higher baseline energy demand and potential overcrowding.

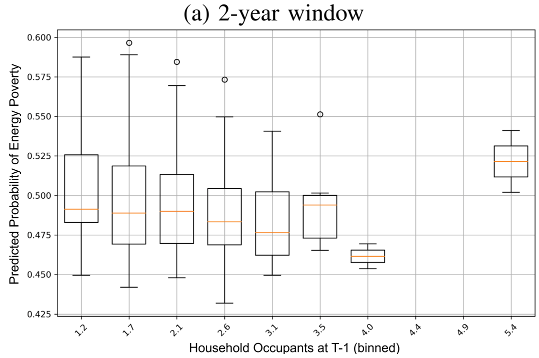
Dispersion is substantially larger among small households in both horizons, highlighting heterogeneous outcomes within

979-8-3315-8138-1

this group. Upper-fence outliers among 1–2 person households indicate individuals who remain at high risk despite small household size. Conversely, lower-fence outliers among larger households suggest cases where extended-family living arrangements, multigenerational income pooling, or access to subsidized housing mitigate risk despite high occupancy.



(a) 2-year window



(b) 14-year window

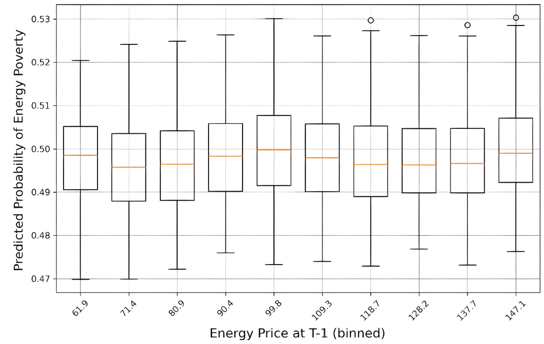
Fig. 2: Forecasted probability of energy poverty versus household occupants at $T-1$ for short and long windows.

C. Energy Price

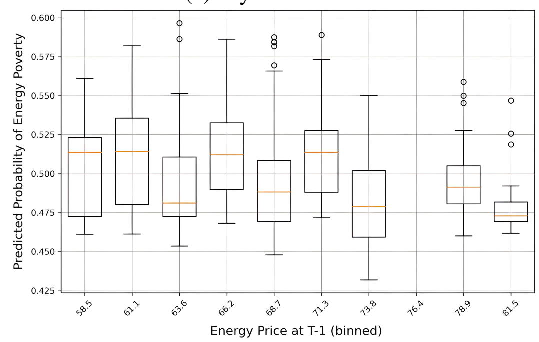
Based on Fig. 3, regional energy prices exhibit a *weak and inconsistent* association with future energy poverty. In the 2-year window, medians are nearly flat across price bins, indicating that short-run price swings are largely absorbed or offset. In the 14-year window, medians decline slightly with higher prices, suggesting either adaptation (e.g., efficiency upgrades or behavioral responses) or composition effects that mute long-run price exposure. Overall, price levels are not a primary predictor relative to household income.

Beyond central tendencies, the variance structure reveals substantial heterogeneity in predicted risk within bins. The 2-year energy price model exhibits wide IQRs across most price deciles, an indicator that price alone is insufficient to stratify households. Even within the highest price bin (>130 ¢/kWh), predicted probabilities span from approximately 0.07 to 0.53, suggesting that household-specific factors (income, efficiency, behavior) dominate price exposure.

In the 14-year model, such dispersion slightly contracts, where IQRs narrow to approximately 0.06-0.09, particularly at higher price levels. At longer horizons, tighter distributions



(a) 2-year window



(b) 14-year window

Fig. 3: Forecasted probability of energy poverty versus regional energy price at $T-1$ for short and long windows.

imply that prolonged exposure to high energy costs either force adaptation or selects households with sufficient resilience to absorb costs. Persistent outliers above the 75^{th} percentile, however, identify a vulnerable subgroup that remains at elevated risk despite prevailing price levels.

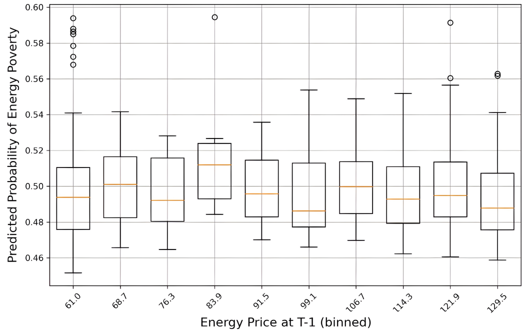
D. Baseline (T=8) Comparison

As shown in Fig. 4, the intermediate (8-year) horizon exhibits transitional behavior between short- and long-term dynamics. Predicted energy poverty risk remains weakly associated with regional energy prices (Fig. 4(a)), with median probabilities largely flat across price bins, consistent with the short-run results. Income (Fig. 4(b)) preserves a monotonic negative relationship with risk, though gradients are smoother and dispersion is reduced relative to the 2-year model, reflecting partial averaging of transitory income shocks. Household size (Fig. 4(c)) continues to act as a contextual modifier rather than a primary driver, with modestly lower risk for mid-sized households and slightly elevated risk for very small and very large households. Overall, baseline results reinforce the stability of the dominant patterns observed across horizons without introducing additional structural effects.

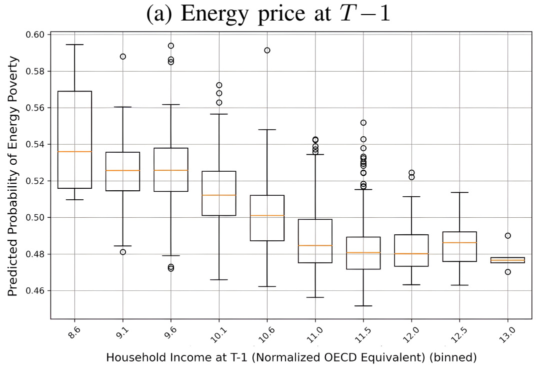
V. Interpretation and Policy Implications

Findings suggest that interventions aimed at income stabilization, energy retrofits, and targeted assistance for mid-income households are likely to yield the most durable reductions in energy poverty risk. By moving beyond a rigid yes/no

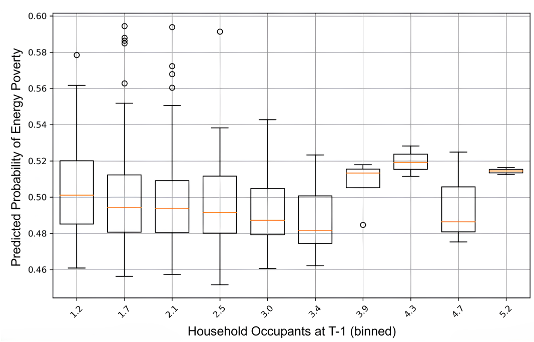
979-8-3315-8138-1



(a) Energy price at $T-1$



(b) Normalized household income at $T-1$ (OECD equivalized).



(c) Household size at $T-1$ (binned).

Fig. 4: Baseline (8-year window): Forecasted probability of energy poverty versus (a) regional energy price and (b) normalized household income (c) household size at $T-1$.

definition of energy poverty to a more flexible alternative, the probabilistic approach provides a continuous risk spectrum for each household. Such granularity is crucial: it allows to identify and support borderline cases who may fall just below a binary threshold yet still face significant risk. Moreover, it reflects the diminishing returns of protective factors (for example, additional income beyond a certain point yields smaller reductions in risk rather than an abrupt status change). In practice, this enables early and more proactive, targeted interventions for households with high or moderate risk, and ultimately ensures that no at-risk household is overlooked due to the constraints of a single threshold.

VI. Conclusions & Future Work

The long-term (14-year) model, designed to capture broader socioeconomic divergence, combined with the short-term (2-year) model reflecting near-term variability, reveals that household income consistently emerges as the dominant predictor of future energy poverty, with diminishing returns beyond approximately AUD 100,000. Household size exhibits a weaker non-linear influence, with modestly lower long-run risk for mid-sized households and elevated dispersion among very small and very large households, indicating that demographic structure is a contextual modifier rather than a primary driver.

Future work should refine this framework by: (1) *extending the assessment to all features* under varying window sizes; (2) *applying SHAP-based temporal analysis* to trace evolving feature importance; and (3) *conducting pairwise interaction analysis* to uncover compound effects among socioeconomic drivers. Collectively, these extensions aim to enhance interpretability and inform proactive policy design.

VII. Acknowledgments

This work has been supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, by the Government of Alberta under the Major Innovation Fund project resilient and Clean Energy Systems Initiative, and by the Indigenous and Black Engineering and Technology (IBET) Momentum Fellowship.

References

[1] International Energy Agency, "World energy outlook 2024: Executive summary," 2024, quote: "750 million people ... without access to electricity and more than 2 billion without clean cooking fuels.". [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2024/executive-summary

[2] S. Budria, E. Fermé, and D. Freitas, "Toward proactive policy design: Identifying "to-be" energy-poor households using shap for early intervention," *SSRN Electronic Journal*, 2025, 10.2139/ssrn.5133997.

[3] S. Budría, E. Fermé, and D. N. Freitas, "Toward proactive policy design: Identifying "To-Be" energy-poor households using SHAP for early intervention," IZA Institute of Labor Economics, Discussion Paper 17669, Feb. 2025.

[4] D. Al Kez, A. Foley, Z. K. Abdul, and D. Furszyfer Del Rio, "Energy poverty prediction in the united kingdom: A machine learning approach," *Energy Policy*, vol. 184, p. 113909, Jan. 2024, dOI:10.1016/j.enpol.2023.113909.

[5] A. López-Vargas, A. Ledezma-Espino, and A. Sanchis-de-Miguel, "Methods, data sources and applications of artificial intelligence in the energy poverty context: A review," *Energy and Buildings*, vol. 268, p. 112233, Aug. 2022, dOI:10.1016/j.enbuild.2022.112233.

[6] E. Kalfountzou, L. Papada, C. Tourkolias, S. Mirasgedis, D. Kaliampakos, and D. Damigos, "A comparative analysis of machine learning algorithms in energy poverty prediction," *Energies*, vol. 18, no. 5, p. 1133, Mar. 2025, dOI:10.3390/en18051133.

[7] D. of Social Services, M. I. of Applied Economic, and S. Research, "The Household, Income and Labour Dynamics in Australia (HILDA) Survey, RESTRICTED RELEASE 21 (Waves 1-21)," 2022, 10.26193/24EJST.

[8] Australian Bureau of Statistics, "Consumer price index, australia," March 2024, aBS. [Online]. Available: https://www.abs.gov.au/statistics/economy/price-indexes-and-inflation/consumer-price-index-australia/mar-quarter-2024

[9] D. Al Kez, A. Foley, C. Lowans, and D. F. Del Rio, "Energy poverty assessment: Indicators and implications for developing and developed countries," *Energy Conversion and Management*, vol. 307, p. 118324, May 2024, 10.1016/j.enconman.2024.118324.

[10] D. Nock, A. J. Jones, S. Bouzarovski, H. Thomson, and D. J. Bednar, "Reducing energy burden in the power sector: Metrics for assessing energy poverty," *IEEE Power and Energy Magazine*, vol. 22, no. 4, pp. 26–37, 2024, 10.1109/MPE.2024.3408770.

979-8-3315-8138-1