Thesis for the Degree of Doctor of Philosophy

Machine Learning-based Causal Inference in Buildings: Understanding key drivers for energy usage and occupant thermal comfort

Patrick Nzivugira Duhirwe

Department of Architectural Engineering
Graduate School
Kyung Hee University
Yongin, Korea

August 2023

Machine Learning-based Causal Inference in Buildings: Understanding key drivers for energy usage and occupant thermal comfort

Patrick Nzivugira Duhirwe

Department of Architectural Engineering
Graduate School
Kyung Hee University
Yongin, Korea

August 2023

Machine Learning-based Causal Inference in Buildings: Understanding key drivers for energy usage and occupant thermal comfort

by

Patrick Nzivugira Duhirwe

Advised by

Prof. Geun Young Yun, Ph.D.

Submitted to the Department of Architectural Engineering
and the Faculty of the Graduate School of
Kyung Hee University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Dissertation Committee:

Chairman Gon Kim

Jun Bok Lee

Sun Sook Kim

Doosun Kang

Geun Young Yun

Acknowledgements

Reflecting on my Ph.D. journey, I am reminded of the African proverb, “It takes a village to raise a child”. Indeed, this dissertation, a child of my intellectual pursuit, has been nurtured not only by my solitary efforts, but also by a community of dedicated individuals who provided unwavering support and guidance throughout this endeavor. From enriching mentorship to insightful feedback, and shared resilience in the face of challenges, each individual has left a lasting impact on my academic journey. My sincere gratitude goes out to all those who were part of this remarkable journey.

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Geun Young Yun. His mentorship, extending far beyond mere academic guidance, provided both financial support and a firm belief in my potential. His relentless quest for knowledge and fervor for pushing intellectual boundaries have catalyzed my development as an independent researcher as well as inspired me to strive for excellence. Professor Geun Young Yun’s guidance has been instrumental in the successful completion of this dissertation. For his invaluable mentorship, patience, and consistent encouragement, I am deeply indebted.

I would also like to extend my gratitude to Professor Gon Kim, Professor Jun Bok Lee, Professor Sun Sook Kim, and Professor Doosun Kang for agreeing to serve on my dissertation committee despite their busy schedules. Their insightful comments greatly enhanced the quality of this dissertation, and for that, I am truly thankful.

I wish to express my gratitude to my good friend Dr. Jack Ngarambe for providing different perspectives, creating a conducive environment for brainstorming and discussion, and consistently offering encouragement. I am also appreciative of Todd McGowan for sparing his valuable time to proofread this dissertation. Additionally, my thanks go to my colleagues from the Human Behaviour and Energy Laboratory (HuBEL), particularly Kanghee Lee and Lee Hyeji, for their assistance with data collection and the provision of pertinent information related to the South Korean building sector.

Last but not least, I would like to acknowledge the special role my family and God have played. I offer an eminently special thank you to my father and mother, who have taken care of me since my birth, loved me, believed in me, emphasized the importance of education, and taught me to be an ethical and self-reliant individual. Also, a big thank you to my best friend and eldest brother, Eric Didier, for your invaluable mentorship; and to my younger brothers, Ernest, Serge, and Arsène for their eternal brotherhood. Your presence in my life has been a source of joy, encouragement, and laughter. Finally, I thank the Almighty God for granting me a healthy body and mind throughout this journey, and for blessing me with such a loving, supportive family.

To my father and mother, my eldest brother Eric Didier, and my younger brothers Ernest, Serge, and Arsène.

Contents

Abstract xii

1 Introduction 1

1.1 Motivation 1

1.2 Problem statement 2

1.3 Objectives and research questions 3

1.4 Scope of work and dissertation structure 5

2 Causal inference analysis 6

2.1 Introduction 6

2.1.1 The distinction between association and causation 6

2.1.2 Potential outcomes and counterfactuals 7

2.1.3 Causal effects formalism 8

2.1.4 Causal assumptions 9

2.2 The need for causal inference in the built environment 9

2.2.1 Limitations of statistical and machine learning-based predictive models 9

2.2.2 The implications of applying causal inference during the design and operation stages of a building 10

3 Existing research and the novelty of the proposed machine-leaning base causal inference 1

2

3.1 Existing research 12

3.2 The proposed machine learning (ML)-based causal inference approach 12

3.2.1 Model causal mechanisms with Directed acyclic graphs (DAGs) 20

3.2.2 Estimating causal effects with Double Machine Learning (DML) 22

3.2.3 Interpreting the estimated effect in terms of changes in the outcome 23

4 Methodology 25

4.1 Data description 25

4.1.1 The Residential Energy Consumption Survey (RECS) 2015 dataset 25

4.1.2 The Household Energy Panel Survey (HEPS) dataset 31

4.1.3 The ASHRAE Global Thermal Comfort Database II 37

4.2 Causal inference analysis framework 43

4.2.1 The analysis framework 43

4.2.2 Validating the estimated causal effects 45

5 Evaluating energy policies and occupant behavior impacts on cooling energy consumption 46

5.1 The effectiveness of energy policies on EUI 46

5.1.1 Energy saving policies vs EUI considering all buildings 46

5.1.2 Energy saving policies vs EUI considering buildings that underwent energy audit 50

5.1.3 Building insulation level reflecting building thermal performance over the years 51

5.1.4 Economic analysis for implementing Energy Star qualified windows 52

5.1.5 Increasing or negligible? The effect of smart meters on EUI and potential reasons behind 53

5.2 The influence of occupant behavior on EUI 53

5.2.1 Nighttime indoor air temperature settings vs EUI 53

5.2.2 Air conditioning (AC) system usage behavior vs EUI 56

5.3 The causal effects of other variables in the DAG 59

5.4 Summary 65

5.4.1 Are there causal relationships between energy-saving policies and occupant behavior and their impact on EUI in buildings?... 66

5.4.2 If such relationships exist, can these causal relationships be quantified?.......... 66

5.4.3 What is the effectiveness of existing energy-saving policies on building EUI, what new policies could be proposed for further EUI reduction, and what role does occupant behavior play? 67

6 The role of socio-economic factors and heating equipment selection in energy consumption 69

6.1 The role of occupant socio-economic factors 69

6.1.1 Age, education level, average income, and number of household members vs EUI 69

6.1.2 Household main source of income vs EUI 71

6.1.3 The need to integrate socio-economic factors in energy saving policies........... 72

6.2 The impact of heating equipment selection on EUI 73

6.2.1 Chosen heating equipment vs EUI... 73

6.2.2 The effect of auxiliary heating on EUI 76

6.3 The causal effects of other variables in the DAG 77

6.4 Summary 84

6.4.1 Do occupant socio-economic characteristics and the choice of heating equipment serve as causal factors in determining the EUI in buildings?.. 85

6.4.2 If these are indeed causal factors, can their effect be quantified?...................... 85

6.4.3 What insights can be derived from causal inference analysis regarding the role of socio-economic factors and the adoption heating equipment towards EUI reduction?... 86

7 The effect of occupant personal and behavioral factors on thermal sensation 87

7.1 Occupant personal factors and thermal sensation 87

7.1.1 Sex, age, and BMI vs thermal sensation.. 87

7.1.2 Sex vs thermal sensation: an in-depth causal inference analysis 88

7.1.3 Age vs thermal sensation: an in-depth causal inference analysis 89

7.2 Occupant behavior and thermal sensation 91

7.2.1 The operation of blinds/curtains, fan, window, door, and heater vs thermal sensation 91

7.2.2 The operation of blinds/curtains, fan, window, door, and heater in relation to thermal sensation across sex categories 94

7.3 The causal effects of other variables in the DAG 95

7.4 Summary 98

7.4.1 Are there causal relationships between occupant personal and behavioral factors and thermal sensation? 99

7.4.2 If such relationships exist, can these identified personal and behavioral factors be quantified? 99

7.4.3 How can these identified causal factors be utilized in the development of adaptive comfort models? 100

8 Conclusions 101

8.1 Evaluating energy policies and occupant behavior impacts on cooling energy consumption 101

8.2 The role of socio-economic factors and heating equipment selection in energy consumption 102

8.3 The role of socio-economic factors and heating equipment selection in energy consumption 102

8.4 Deductions from the three analyses 103

8.5 The implications of this dissertation findings for the development of reliable energy and thermal comfort prediction models and controls 104

References 105

Table Contents

Table 3. 1. A summary of existing studies and their limitations 14

Table 4. 1. Selected variables related to cooling energy consumption ·························· 28

Table 4. 2. Selected variables related to heating energy consumption ·························· 32

Table 4. 3. Selected variables related to thermal sensation ······································ 38

Table 5. 1. Causal effects of energy saving policies on EUI ····································· 47

Table 5. 2. Causal effects of electricity payment overseer on EUI ····························· 48

Table 5. 3. Causal effects of energy saving policies on energy consumption ·················· 49

Table 5. 4. Causal effects of electricity payment overseer on energy consumption ··········· 49

Table 5. 5. Pairwise test in electricity payment overseer··· 49

Table 5. 6. Causal effects of energy saving policies on EUI in audited and non-audited buildings ·· 50

Table 5. 7. Causal effects of the house completion year on EUI ································ 52

Table 5. 8. Causal effect of nighttime temperature settings on EUI ···························· 55

Table 5. 9. Pairwise comparison for different nighttime temperature setting scenarios······· 56

Table 5. 10. Mean day and night indoor temperature·· 56

Table 5. 11. Causal effect of air conditioner (AC) equipment usage behavior on EUI········ 58

Table 5. 12. Mean indoor temperature for different AC usage behavior························ 58

Table 5. 13. Pairwise comparison for different AC usage behavior ···························· 59

Table 5. 14. Causal effects of International Energy Conservation Code on EUI ·············· 60

Table 5. 15. Causal effects of housing type on EUI ·· 61

Table 5. 16. Causal effects of AC type on EUI··· 62

Table 5. 17. Causal effects of major wall materials on EUI ···································· 63

Table 5. 18. Thermal resistance of major wall materials··· 64

Table 5. 19. Causal effects of other variables on EUI ·· 65

Table 6. 1. Causal effects of age, education level, average income, and number of household members on EUI ··· 70

Table 6. 2. Causal effects of Age, education level, average income, and number of household members on energy consumption and floor area·· 70

Table 6. 3. Causal effects of household main source of income on EUI ······················ 71

Table 6. 4. Detailed Causal effects of household main source of income on EUI ············ 72

Table 6. 5. Causal effects of house completion year on EUI ···································· 73

Table 6. 6. Causal effects of the main heating equipment on EUI ······························ 75

Table 6. 7. Pairwise statistical tests ·· 76

Table 6. 8. Causal effects of using auxiliary heating ·· 77

Table 6. 9. Causal effects of provincial climate on EUI·· 79

Table 6. 10. Causal effects of housing type on EUI ·· 80

Table 6. 11. Causal effects the ratio of double-glazed windows to exterior windows on EUI 81

Table 6. 12. Causal effects the residential floor number on EUI································ 82

Table 6. 13. Causal effects the housing orientation on EUI ········ 83

Table 6. 14. Causal effects other variables on EUI ········ 84

Table 7. 1. Causal effects of sex and age on thermal sensation ········ 87

Table 7. 2. Causal effect sex on thermal sensation in different age groups ········ 90

Table 7. 3. Casual effect of age on thermal sensation ········ 91

Table 7. 4. Causal effects of behavioral factors on thermal sensation ········ 93

Table 7. 5. Occupant behavior and associated thermal acceptability, indoor and outdoor temperatures ········ 94

Table 7. 6. Occupant behavior in relation to thermal sensation across sex categories ········ 95

Table 7. 7. Causal effects of building types on thermal sensation ········ 96

Table 7. 8. Causal effects of cooling types on thermal sensation ········ 97

Table 7. 9. Causal effects of environmental and subjective thermal factors on thermal sensation ········ 98

Figure Contents

Figure 1. 1. (a) Final energy consumption by sector (IEA, 2021) and (b) average time spent indoors by a person (Klepeis et al., 2001) 1

Figure 2. 1. A hypothetical world in casual inference analysis 8

Figure 3. 1. The novelty of the proposed machine learning-based causal inference approach 13

Figure 3. 2. An example of a directed acyclic graph 20

Figure 4. 1. Categories for the selected variables 26

Figure 4. 2. Energy use intensity (EUI) distribution in energy audit, Energy Star qualified windows usage, smart meter usage, interval data access, and energy assistance 29

Figure 4. 3. EUI distribution in electricity payment overseer 29

Figure 4. 4. EUI distribution in nighttime vs daytime indoor temperature settings 30

Figure 4. 5. EUI distribution in air condoning (AC) equipment usage behavior 30

Figure 4. 6. Categories for the selected variables 32

Figure 4. 7. EUI distribution in number of household members and head of household age · 34

Figure 4. 8. EUI distribution in head of household education level and household average monthly income 35

Figure 4. 9. EUI distribution in household source of income 36

Figure 4. 10. EUI distribution in main heating equipment used 36

Figure 4. 11. Categories for the selected variables 38

Figure 4. 12. Thermal sensation distribution in occupants' BMI and sex 39

Figure 4. 13. Thermal sensation distribution in occupants' age 40

Figure 4. 14. Thermal sensation distribution in state of blinds and fan mode 41

Figure 4. 15. Thermal sensation distribution in state of windows and doors 42

Figure 4. 16. Thermal sensation distribution in heater mode 43

Figure 4. 17. Causal inference analysis framework 44

Figure 4. 18. Joint hyperparameters optimization and estimate refutation methods 44

Figure 5. 1. Causal effects of energy saving policies on EUI 47

Figure 5. 2. Causal effect of building insulation level on EUI 51

Figure 5. 3. Causal effect of nighttime temperature settings on EUI 55

Figure 5. 4. Casual effect of air conditioning (AC) usage behavior on EUI 57

Figure 6. 1. Thermal transmittance (U-value) over the years 73

Figure 7. 1. Causal effects of BMI on thermal sensation 88

Figure 7. 2. Causal effects of behavioral factors on thermal sensation 92

Abstract

Machine Learning-based Causal Inference in Buildings:
Understanding key drivers for energy usage and occupant thermal comfort

By Patrick Nzivugira Duhirwe
Doctor of Philosophy in Architectural Engineering
Graduate School of Kyung Hee University
Advised by Prof. Geun Young Yun, Ph.D.

The increasing energy consumption in buildings, along with the necessity for comfortable indoor environments, has naturally prompted architectural engineers and policymakers to ask “what-if” questions in the pursuit of developing sustainable designs and energy-saving policies. The existing simulation, statistical, and machine learning methods fall short in answering these “what-if” questions as they primarily rely on correlation or association while calculating probabilities under static conditions. This shortfall highlights the necessity for causal inference methodologies. These methods are adept at calculating probabilities or beliefs not only under static conditions but also under varying circumstances, making them reliable for answering “what-if” questions.

Previous studies have utilized conventional causal inference methodologies such as path analysis, structural equation modeling, and propensity score matching to understand how changes in building dynamics affect energy consumption and thermal comfort. These conventional methods excel in experimental conditions where data follow certain assumptions such as linearity and the absence of measurement errors. However, the rise of the Internet of Things has resulted in an abundance of observational data, reflecting actual scenarios of building energy consumption and thermal comfort. This type of data, being nonlinear and high dimensional, constrains the application of conventional causal inference methods as they encounter selection bias problems. Furthermore, these methods do not offer a way to incorporate domain knowledge into the analysis. To address these issues, this dissertation proposes a novel machine learning-based approach that

utilizes double machine learning and integrates domain knowledge through directed acyclic graphs. The robustness of this proposed approach is demonstrated by its application to three different datasets, aiming to uncover various causal factors related to energy consumption and thermal comfort.

This dissertation first applies the proposed novel causal inference approach to the 2015 United States Residential Energy Consumption Survey data, assessing the impact of energy policies and occupant behavior on cooling energy consumption. The results highlight the effectiveness of energy audit programs, revealing no increase in Energy Use Intensity (EUI) in audited buildings, which is in contrast to a 2.543 kWh/m^2 increase in non-audited buildings due to smart meter usage. In audited buildings, interval data access led to a notable EUI reduction of 7.035 kWh/m^2, while the use of Energy Star qualified windows also proved beneficial, reducing EUI by 2.260 kWh/m^2. Well-insulated buildings also demonstrated lower EUIs. As for occupant behavior, it was found that optimal air conditioner (AC) usage and temperature settings adjustments, particularly setting the nighttime temperature higher than that of the daytime, effectively reduced EUI. Contrarily, maintaining a constant AC setpoint temperature most of the time resulted in an increased EUI. In light of these findings, three policy recommendations are proposed: first, a mandate on using Energy Star qualified windows in all buildings, with government assistance; second, government support for landlords to incorporate energy-efficient features without the need to increase rent; and third, public education to promote awareness of optimal AC usage and the benefits of varying temperature settings throughout the day. These measures would significantly enhance energy-saving strategies and promote sustainability.

The second application of the proposed approach analyzes the Korean Household Energy Panel Survey, identifying the impact of socio-economic factors and heating equipment choices on energy consumption. Transitioning from a kerosene to a gas boiler was found to decrease EUI by 6.161 kWh/m^2. However, utilizing individual heating with a briquette boiler or an electric blanket as the primary heating equipment generally led to increased EUI. Socio-economic factors like age (per 10-year increments), education level (ranging from below middle school to graduate school or higher), and average monthly income (in 2 million increments) were found to impact EUI by

10.208 kWh/m^2, -18.012 kWh/m^2, and -18.865 kWh/m^2 respectively, indicating the significance of these aspects in managing building energy consumption. Households with primary income sources not derived from occupation generally had a higher EUI, which emphasizes the importance of socio-economic considerations in energy-saving policies. The findings highlight a need for more inclusive policy formulation, integrating socio-economic considerations into energy-saving strategies, and targeted awareness campaigns, particularly for households with non-occupational primary income sources, promoting the choice of energy-efficient equipment.

The third application of the novel approach evaluates the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Global Thermal Comfort Database II, examining the impact of personal and behavioral factors on thermal sensation. Findings indicate no significant influence of age on thermal sensation, but a distinct sex-based variation is observed with females experiencing a warmer sensation by a factor of 0.031. This difference is particularly marked in the 26-35 and 56-65 age brackets, with females reporting a warmer thermal sensation by 0.066 and 0.253, respectively. In terms of Body Mass Index (BMI), there was a notable divergence of 0.205 in thermal sensation between the overweight and obese categories. Occupant behavior, such as opening windows and doors, showed changes in thermal sensation by 0.078 and -0.187 respectively. The study also quantified that heater usage is triggered by a decrease in thermal sensation of 0.479. These insights highlight the complex relationship between personal characteristics, behavioral factors, and thermal sensation, which suggest the potential for individualized thermal comfort strategies.

The findings of this dissertation demonstrated how causal inference has significant implications when selecting variables for building energy and thermal comfort prediction and optimization. By utilizing the proposed causal inference method, researchers can identify and quantify the causal relationships between various factors and energy consumption or thermal comfort outcomes. This enables a more accurate understanding of which variables truly influence energy use and thermal comfort, eliminating spurious correlations or confounding factors. Consequently, the findings from causal inference analysis can guide the selection of key variables to be included in predictive models and optimization strategies, ensuring that the focus is on the factors that have

a genuine causal impact. This approach enhances the precision and reliability of building energy and thermal comfort predictions, leading to more effective optimization techniques and ultimately facilitating the development of sustainable and comfortable built environments. By incorporating these insights into policy, design, and technology, stakeholders can collaboratively create buildings that effectively balance energy efficiency, comfort, and functionality.

Keywords: Causal inference, double machine learning, directed acyclic graphs, building energy consumption, thermal sensation, energy saving policies, occupant behavior, occupant socio-economic factors

1 Introduction

1.1 Motivation

The building sector contributes significantly to global energy demand, accounting for nearly one-third of global final energy consumption according to the International Energy Agency (IEA, 2021), as shown in Figure 1.1.a. This proportion is projected to rise in the coming years due to swift population growth, urbanization, and a mounting demand for energy-intensive appliances like air conditioning systems, particularly in developing nations. Moreover, research suggests that individuals, on average, spend around 90% of their time indoors (Klepeis et al., 2001) (see Figure 1.1.b). This underscores the importance of indoor environments on our daily lives. Consequently, the role of architectural engineers has become increasingly important in designing buildings that are not just energy-efficient and conducive to good health, but also provide a thermally comfortable environment. Gaining deeper insight from energy-saving policies, along with the dynamic occupant socio-economic and behavioral causal factors that shape building energy use and thermal comfort, offers a promising approach for deriving sustainable and efficient strategies and practices. By investing in these insights and integrating them into building design and operation stages, it is possible to create buildings that are more sustainable, more comfortable, and ultimately, more beneficial for both their occupants and the environment as a whole.

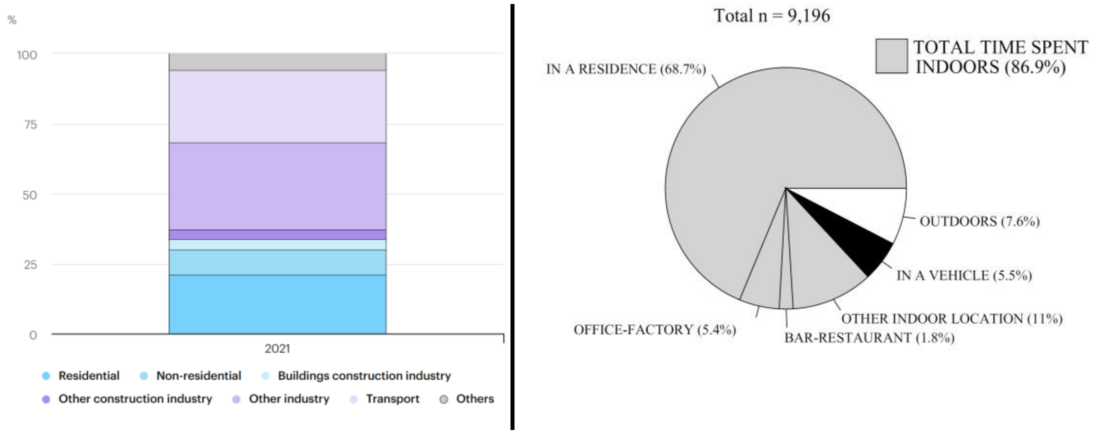


Figure 1. 1. (a) Final energy consumption by sector (IEA, 2021) and (b) average time spent indoors by a person (Klepeis et al., 2001)

1.2 Problem statement

The modern focus on sustainable and efficient building practices has led to a particular emphasis on the importance of reducing energy consumption for both cooling and heating. Yet, it is essential that this does not occur at the expense of the occupant's thermal comfort. Comfortable living and working environments are essential for individual productivity and overall well-being. To balance these objectives, it is crucial to comprehensively quantify the variables affecting energy consumption, both at the design and operation stages of the building. Moreover, beyond the physical and design characteristics of the building, the role of occupant behavior and socio-economic characteristics can significantly affect energy usage and thermal comfort. In many cases, these factors can have as much or more of an impact than the building design itself. Achieving occupant-centric and healthy environments requires a thorough understanding and quantification of these variables. However, it must be emphasized that the unjustified choice of variables and inaccurate estimations can have substantial implications. These might affect not just the occupant's well-being but also have an overall impact on the building's energy use intensity (EUI). This underlines the importance of correctly identifying and comprehensively evaluating causal parameters for both energy consumption and thermal comfort.

In practice, the choice of variables for energy and thermal comfort estimation is often based on beliefs or mere associations such as correlation. Such methods can be helpful for initial estimates but are inherently limited and can lead to errors or inaccuracies when the dynamics of the data change. This emphasizes the necessity for a more robust and flexible approach that can adapt to different contexts and changing circumstances. In addition, while there exist robust data-driven models that can predict overall energy use and thermal comfort levels, these models often fall short in providing specific and quantifiable insights. Specifically, these models are unable to quantify the contribution of each independent variable, thereby limiting their applicability for detailed analysis and improvement efforts. Such models fail to answer key questions like “How much will the thermal sensation increase by changing a specific parameter?” or "Is the observed energy reduction due to the building's insulation level or the cooling equipment used?". This reveals a significant gap in our current methods and tools for understanding and optimizing energy usage and thermal comfort in buildings.

1.3 Objectives and research questions

To address the existing problems highlighted in the previous section, the objectives of this dissertation are threefold. They aim to apply a causal inference analysis approach to efficiently identify and quantify causal factors that affect building energy consumption and thermal comfort, with the goal of deriving sustainable optimization solutions.

The first objective focuses on evaluating the impacts of energy policies and occupant behavior on cooling energy consumption. To meet this goal, the dissertation utilized causal inference to ascertain if the implementation of energy-saving policies and occupant behavior, particularly towards the use of cooling equipment, have a causal impact on EUI in residential buildings. The analysis will further quantify the causal relationships between energy-saving policies and occupant behavior, and their effect on the EUI in buildings. In doing so, it will assess the extent to which specific policies and usage habits of cooling equipment contribute to a reduction in EUI. The effectiveness of current energy-saving policies on building EUI will be evaluated, areas for improvement will be identified, and new policy recommendations and best practices for further reductions in EUI will be developed. In pursuit of this objective, the dissertation will attempt to answer the following research questions:

- Are there causal relationships between energy-saving policies and occupant behavior and their impact on EUI in buildings?
- If such relationships exist, can these causal relationships be quantified?
- What is the effectiveness of existing energy-saving policies on building EUI, what new policies could be proposed for further EUI reduction, and what role does occupant behavior play?

The second objective is to identify and quantify the role of socio-economic factors and heating equipment selection in energy consumption. To accomplish this, the dissertation employed causal inference to investigate whether socio-economic factors and the choice of specific heating equipment exert a causal impact on the EUI of buildings. The study will further quantify the causal effect of the identified socio-economic factors and the selected heating equipment on building EUI. This will enable an evaluation of the extent to which specific occupant characteristics and heating systems influence heating energy consumption. From the causal inference analysis, the dissertation will glean insights on occupant socio-economic characteristics and factors that prompt the adoption and utilization of certain heating systems. Furthermore, the analysis will explore potential strategies for promoting building energy reduction. To meet this objective, the dissertation will attempt to answer the following research questions:

- Do occupant socio-economic characteristics and the choice of heating equipment serve as causal factors in determining the Energy Use Intensity (EUI) in buildings?
- If these are indeed causal factors, can their effect be quantified?
- What insights can be derived from causal inference analysis regarding the role of socio-economic factors and the adoption of heating equipment towards EUI reduction?

The third objective focuses on the impact of occupant personal and behavioral factors on thermal sensation. To estimate this effect, the dissertation utilized causal inference to scrutinize whether occupant personal and behavioral factors hold a causal influence on their perceived thermal sensation within a built environment. This research will also quantify the impact of personal and behavioral factors on thermal sensation, fostering a more nuanced understanding of the extent to which these factors influence occupant comfort. The dissertation will leverage the identified causal factors relating to personal and behavioral aspects to suggest enhancements to existing thermal comfort models. The goal is ultimately to enable the development of more effective and sustainable thermal comfort solutions. To fulfill this objective, the dissertation will seek to answer the following research questions:

- Are there causal relationships between occupant personal and behavioral factors and thermal sensation?
- If such relationships exist, can the effects of these identified personal and behavioral factors be quantified?
- How can these identified causal factors be utilized in the development of adaptive comfort models?

1.4 Scope of work and dissertation structure

This dissertation primarily concentrates on utilizing a machine learning-based causal inference approach to uncover and quantify the effect of energy-saving policies, as well as the influence of occupant socio-economic and behavioral factors on building heating and cooling energy consumption, and occupant thermal sensation. The dissertation is structured into eight chapters, with this subsection constituting the first chapter.

The second chapter offers an introduction to causal inference analysis and underscores its significance in the built environment. The third chapter provides a review of existing studies on causal inference in building energy consumption and thermal comfort estimations, discusses their limitations, and outlines the novelty of the proposed machine learning-based causal inference framework.

Chapter four delineates the methodology adopted for this study and provides information on the datasets employed. In chapter five, the impacts of energy policies and occupant behavior on cooling energy consumption are evaluated. The focus of chapter six is to investigate the role of socio-economic factors and the selection of heating equipment in energy consumption. Chapter seven explores the effect of occupant personal and behavioral factors on thermal sensation.

The concluding chapter, chapter eight, summarizes the findings of this dissertation and discusses potential future studies in the areas of building energy optimization and thermal comfort design.

2 Causal inference analysis

2.1 Introduction

The questions that primarily drive most studies in Architectural sciences seek to establish causal relationships rather than just associations. For instance, what is the efficacity of a proposed energy saving policy? Whether the gathered data can highlight occupants' behavior in terms or HVAC system usage? What factors influence cooling energy consumption? Do highly efficient HVAC systems contribute to building energy optimization? These types of questions are causal because they demand some knowledge that cannot be calculated neither form the data alone, nor from its distribution.

2.1.1 The distinction between association and causation

The objective of standard statistical and machine learning-based analysis, categorized by regression, classification, and hypothesis testing methods, is to evaluate parameters of a distribution with sample taken out from that distribution. Using such parameters, association among variables can be inferred, probabilities or beliefs of past and future scenarios can be estimated and updated as well with the presence of new data. As long as the data gathering conditions remain unchanged, standard statistical and machine learning methods handles these tasks efficiently. Causal analysis moves one step further by inferring probabilities or beliefs not only under static conditions but also the dynamics of these probabilities under varying conditions, for instance changes induced by external interventions. This distinction entails that association and causation concepts do not blend. For example, there is nothing in the joint distribution of type of HVAC systems an EUI that informs us that intervening on the former would or would not reduce the latter. Thus, causal assumptions help to provide this information by discovering relationships that remain unchanged when external conditions vary.

The differentiation between associational and causal analysis can be explicated as: An association is any type of relationship that can be explained by a joint distribution in the data while a causal relationship cannot be explained from a joint distribution alone. Instances of associations are correlation, regression, likelihood, condition independence, and son forth. Instances of causal concepts are confounding, effect, influence, randomization, spurious correlation, and so forth.

These differences are important in causal analysis for they assist researchers to follow assumptions needed for evaluation various types of research claims.

2.1.2 Potential outcomes and counterfactuals

Potential outcomes and counterfactuals are crucial and fundamental when defining causal effects. Suppose one is interested in the causal effect of exposure A on some outcome Y. In the Architectural engineering field, A can be the presence of a thermostat and Y EUI. Then $A = 1$ if the building has a thermostat; $A = 0$ otherwise. The potential outcomes Y^a are the outcomes that would be observed under each possible exposure option a. That is, Y^1 is the EUI that the building would have if it had a thermostat. Alternatively, Y^0 is the EUI if that building had no thermostat.

Counterfactual outcomes are those outcomes that would have been observed, had the exposure been different. If the exposure was $A = 1$, then the counterfactual outcome is Y^0. Conversely, if the exposure was $A = 0$, the counterfactual outcome is Y^1. Taking the query "Does the presence of a thermostat reduces EUI?" as an instance:

(i) What actually happened:

- The building has a thermostat.
- The actual exposure was $A = 1$.
- The observed outcome was $Y = Y^1$.

(ii) What would have happened (contrary to):

- The building has no thermostat.
- The actual exposure was $A = 0$.
- The observed outcome was $Y = Y^0$.

Potential outcomes and counterfactuals can be linked as follows. Before the exposure is applied, any outcome is a potential outcome Y^0 and Y^1. After the exposure is applied, there is an observed outcome, $Y = Y^A$ and counterfactual outcomes Y^{1-A}.

2.1.3 Causal effects formalism

Consider hypothetical world of buildings of interest and the thermostat A as the exposure, as shown in Figure 2.1. World 1 gets exposure $A = 0$ and world 2 exposure $A = 1$. Hypothetically, World 1 and World 2 have the exact same buildings. If we were able to observe both worlds simultaneously, we could obtain the outcome data and the difference between their respective means, which would be the average causal effect. This is statistically noted as

$$E = Y^1 - Y^0 \tag{1}$$

where E is the expected value, Y^1 and Y^0 are the average values if the buildings received exposures $A = 1$ and $A = 0$, respectively.

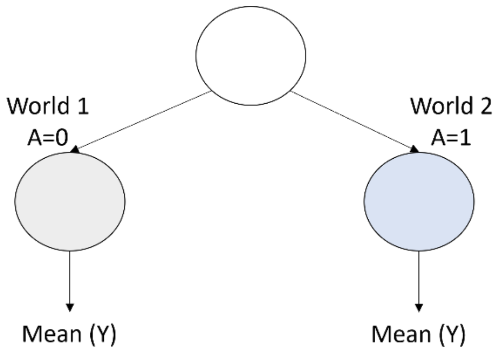


Figure 2. 1. A hypothetical world in casual inference analysis

In general, $E = Y^1 - Y^0 \neq E(Y|A = 1) - E(Y|A = 0)$ as $E(Y|A = 1)$ and $E(Y|A = 0)$ are restricting to subgroups that received the exposures $A = 1$ and $A = 0$, respectively. Hence $E(Y|A = 1) - E(Y|A = 0)$ is not a causal effect because it is comparing two different groups whilst $E = Y^1 - Y^0$ is a causal effect because it is comparing what would happen if the same group received exposure $A = 1$ versus if the same group received exposure $A = 0$. The challenge is only one exposure and one outcome can be observed for each candidate in the group. This is the fundamental problem of causal inference in observed data and some assumptions will be required to link the observed outcome to potential outcomes.

2.1.4 Causal assumptions

The identifiability of causal effects requires making some untestable assumptions. These are generally called causal assumptions. Statistically identifiability aims at identifying some parameters from the actual data and a parameter is considered identifiable if it can be estimated from the data. The most common assumptions in causal inference are:

(i) Stable unit value exposure assumption: there is no interference among units, the treatment assigned to one unit does not affect the outcomes of another unit, and there is only one version of the treatment.

(ii) The consistency assumption: the potential outcome under exposure $A = a$, Y^a is equal to the observed outcome if the actual exposure applied is $A = a$.

(iii) The ‘no unmeasured confounders’ assumption: given pre-exposure covariates X, exposure assignment is independent from the potential outcomes $(Y^1, Y^0 \coprod A|X)$. This implies that among candidates with the same values of X, it can be thought that exposure A as being randomly assigned.

(iv) The positivity assumption: it states that for every set of values for X, the exposure assignment was deterministic. That is, $P(A = a|X = x) > 0$ for all a and x. Had the exposure not been deterministic, then we would have no observed values Y for one of the exposure groups for those values of X.

2.2 The need for causal inference in the built environment

2.2.1 Limitations of statistical and machine learning-based predictive models

Statistical methods and machine learning (ML) have undoubtedly revolutionized the way we uncover crucial patterns, dependencies, and predictions in data related to energy consumption and thermal comfort in buildings. By identifying correlations and predicting future outcomes based on current and past data, these techniques have paved the way for significant advancements in the field. However, their utility has boundaries, particularly when it comes to causal inference.

These techniques, by their very nature, tend to focus predominantly on prediction rather than explanation. As such, they often encounter difficulties in inferring causal relationships from data. A fundamental limitation lies in their inability to distinguish correlation from causation. For instance, a machine learning model might discern that buildings with better insulation tend to use less energy for heating. While this may appear to be a straightforward correlation, the causation might be more complex. It's entirely possible that buildings with superior insulation are also newer constructions equipped with more efficient heating systems. In such a case, it might be the modern heating system, rather than the insulation, that's primarily responsible for reduced energy usage.

In contrast, causal inference is explicitly designed to unearth causal relationships. It combines assumptions about the data-generating process with statistical and ML methods, allowing us to estimate the effect of interventions, such as adding insulation or modifying a heating system, even when we only have observational data. This is of paramount importance in fields like energy consumption and thermal comfort in buildings, where understanding causal relationships is vital for effective decision-making.

Therefore, while statistical and ML models offer valuable insights, it is crucial to recognize their limitations in terms of causal inference. In this context, the role of causal inference becomes pivotal, complementing these predictive models and providing a more holistic and accurate understanding of the relationships within the data.

2.2.2 The implications of applying causal inference during the design and operation stages of a building

The application of causal inference during the design and operation stages of a building carries substantial implications across several key areas.

In terms of energy efficiency and thermal comfort, discerning which factors bear a causal impact on energy consumption aids in pinpointing what modifications can be implemented to enhance a building's energy efficiency. Concurrently, understanding the causal factors influencing thermal comfort can instigate design modifications that make buildings more comfortable, while also diminishing the necessity for energy-intensive heating and cooling systems.

For policy and regulation, causal inferences offer a vital tool for policymakers to create effective regulations and incentives targeted at reducing energy consumption in buildings. They can shape policies that are precisely attuned to the most influential causal factors, thereby optimizing their impact on energy conservation.

In the realm of building design, architectural engineers can harness causal inferences to design buildings that harmonize energy efficiency and comfort. A deep understanding of the causal relationships between various design choices and their effects on energy use and thermal comfort can guide the design process towards the optimal balance between comfort, efficiency, and sustainability.

From a sustainability perspective, diminishing energy consumption in residential buildings is a critical component of efforts to combat climate change. Understanding the causal factors contributing to energy use can help direct these efforts more effectively, enabling targeted interventions that have the most substantial impact on reducing greenhouse gas emissions.

Lastly, occupant behavior is a crucial determinant of a building's energy consumption and thermal comfort. Through causal inference, we can comprehend how specific behaviors—such as adjusting thermostats or opening windows—affect energy use and comfort. This knowledge can inform better educational initiatives or automated systems designed to nudge occupants towards more energy-efficient behaviors. Overall, the application of causal inference in building design and operation underscores the nuanced interplay of factors that shape a building's energy profile and occupant comfort, offering pathways towards more sustainable and comfortable built environments.

3 Existing research and the novelty of the proposed machine-leaning base causal inference

3.1 Existing research

Existing studies have employed several methods, such as propensity score matching, difference-in-differences, path analysis, structural causal models, and other statistical methods, to estimate the effects of variables on energy consumption and thermal comfort. However, these methods have several limitations primarily based on their underlying assumptions about the data distribution, such as linearity and measurement errors. Additionally, they often suffer from the selection bias problem (Bareinboim & Pearl, 2016; Flanders & Ye, 2019; Infante-Rivard & Cusson, 2018) due to their architectures. Table 1 summarizes these studies and highlight their limitations.

3.2 The proposed machine learning (ML)-based causal inference approach

This dissertation proposes a ML-based causal inference analysis approach to address the highlighted limitations in existing studies. The novelty of the proposed approach, as illustrated in Figure 3.1, lies in three aspects: 1) the method employed for modeling the causal relationships, 2) the algorithms utilized for estimating the causal effects, and 3) the manner in which the estimated effects are presented. These three aspects, detailed in subsequent sections, contribute to making the estimated effects more reliable and interpretable compared to those in previous studies.

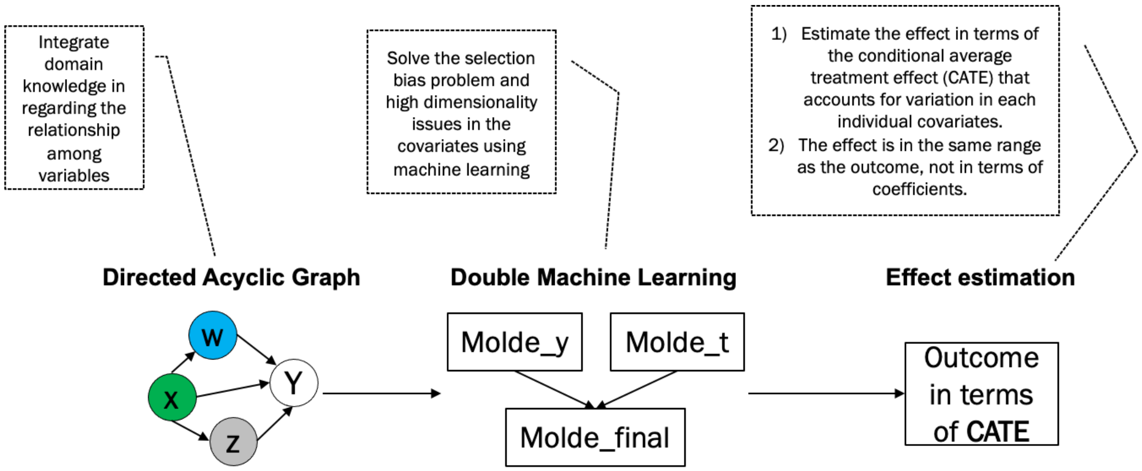


Figure 3. 1. The novelty of the proposed machine learning-based causal inference approach

Table 3. 1. A summary of existing studies and their limitations

	Study	Year	Approach	Dataset	Estimation method	Potential limitations
1	(Im et al., 2017)	2017	Propensity score matching to identify the relative impact of energy efficiency features of buildings residential buildings in the USA on the rent charged	Online data from Craiglist	Propensity score matching	- Assumes a linear relationship between the treatment and the outcome. - No treatment effect heterogeneity.
2	(Liang et al., 2022)	2022	Statistical matching, fixed effects regression, and difference-in-differences to identify the effect of switching from gas boilers to electric heat pumps n Arizona	Metered and surveys data	Propensity score matching and difference-in-difference	- Linear assumption by propensity score matching. - Selection bias and spillover effects by the difference-in-difference method

3	(Ming et al., 2023)	2023	Back propagation-artificial neural network and path analysis to identify the effect of environmental, psychological, and physiological factors on the adaptive thermal comfort in China	Field measurements and subjective questionnaire surveys	Path analysis	- Linearity and measurement error assumptions. - Causal interpretations as the results are interpreted with correlation of variable coefficients, not the actual outcome.
4	(Rentala et al., 2021)	2021	Structural causal models and investigated occupants' thermal state considering personal and behavioral factors in a study conducted in both immersed virtual and in-situ conditions	Experimental data from immersive virtual environments	Structural causal models	- Measurement error and functional form assumptions. - Complexity increases with number of variables
5	(Zhao et al., 2023)	2023	Granger causality test and the distributed lag regression model to investigate the relationship between building energy consumption and economic development, in China.	China Statistical Yearbook	Granger causality test and the distributed lag regression	The estimation methods are only applicable for timeseries data

6	(Gao & Zhang, 2021)	2021	Simultaneous equations model to evaluate the US federal government policy on investing in smart for residential buildings	EIA dataset	Simultaneous equations model	Identification problem and linear relationships assumption associated with simultaneous equations models.
7	(Yun & Steemers, 2011)	2011	Path analysis to investigate the significance of behavioural, physical and socio-economic parameters on cooling energy in order to improve energy efficiency in residential buildings.	RECS 2001	Path analysis	- Linearity and measurement error assumptions - Causal interpretations as the results are interpreted with correlation of variable coefficients, not the actual outcome.
8	(Lin & Liu, 2015)	2015	Granger causality test to analyze how the urbanization process affect building energy consumption and forecasting future effects of macroeconomic variables on energy consumption.	China Statistical Yearbook	Granger causality test	Granger causality test method is only applicable for timeseries data

9	(Reina & Kontokosta, 2017)	2017	Employed multivariate regression models to examine the factors that influence energy consumption in multi-family buildings	LL84 dataset	Multivariate regression	- No causal inference method. - Linearity, multicollinearity, and homoscedastic among other assumptions associated with multivariate linear regression that are hard to maintain in the presence of high dimensional data.
10	(Lund et al., 2010)	2010	EnergyPlan to investigate the impact of transitioning from individual heating systems to district heating in Danish residential buildings	Simulated	EnergyPlan	- No causal inference was employed. - EnergyPlan is a simulation tool, hence its estimations are based on correlation.

11	(Rupp et al., 2018)	2018	Logistic regression on associations of occupant demographics, thermal history, and obesity variables with their thermal comfort in air-conditioned and mixed-mode ventilation office buildings, with field data in Brazil	Field data	Logistic regression	- No causal inference was employed. - Linearity and multicollinearity assumptions among other assumptions associated with logistic regression that are hard to maintain in the presence of high dimensional data.
12	(Kontokosta et al., 2020)	2020	Used Bayesian regression and evaluated impact of energy audit programs on multifamily houses	LL84 dataset	Two-way ANOVA and Bayesian regression	- No causal inference method. - Reproducibility issues associated with Bayesian regression.
13	(Hope & Booth, 2014)	2014	Evaluated the effectiveness of UK government policy on providing incentives towards the attitude of landlords to improve the efficiency features of their buildings	Survey data	Statistical data	- No causal inference was employed. - The obtained estimates do not reflect the true causal effects

14	(Kim & de Dear, 2018)	2018	Statistical test to evaluate the role of operable windows and personal control on the overall thermal comfort	Field data	Statistical tests	- No causal inference was employed. - The obtained estimates do not reflect the true causal effects

3.2.1 Model causal mechanisms with Directed acyclic graphs (DAGs)

Directed acyclic graphs (DAGs) are considered useful for causal inference analysis. They provided a visual representation of causal assumptions and are helpful in identifying which variables to control for. Moreover, DAGs serve as a connection point for merging domain expertise with data-driven approaches. Developed primarily by Judea Pearl (Pearl, 1995), they share several features with structured equation models (SEMs) (Hoyle, 1995). Ideally, a DAG consists of nodes representing variables and arrows (directed edges) between these nodes to represent the direction of the causal relationship. DAGs can be thought of nonparametric SEMs (Elwert, 2013), which is the core difference between them. While SEMs assume that the relationship between variables is linear and additive (unless indicated otherwise) which is shown through arrows, DAGs can represent any type of functional relationship between variables with arrows such as polynomial, exponential, step, and sinusoidal functions. Moreover, as opposed to SEMs, DAGs only allow single-headed arrows.

Consider a simple DAG in Figure 3.2. From this DAG, several paths can be identified and categorized into three basic causal structures: chains, forks, and inverted forks (Elwert, 2013). A chain consists of a series of variables, where the first variable has a causal effect on the second variable, which then has a causal effect on the third variable, and so on. A correlation between the first and last variable in a chain suggests a genuine causal effect. An example of a chain is annual income → AC usage behavior → energy use intensity, where annual income causally influences energy use intensity via AC usage behavior.

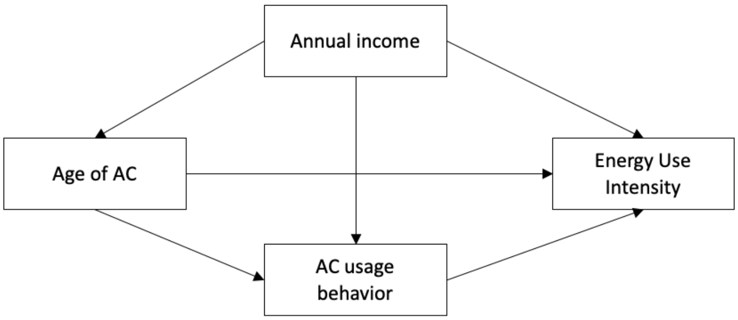


Figure 3. 2. An example of a directed acyclic graph

A fork, on the other hand, has a structure where two variables have a causal effect on a third variable. An example of a fork is AC usage behavior ← annual income → energy use intensity, where AC usage behavior and energy use intensity may be correlated because they share a common cause, annual income. Forks are relevant to the phenomenon of confounding (Dinga et al., 2020), where a variable influences both the dependent and independent variables, leading to a non-causal correlation.

An inverted fork has a structure where a single variable has a causal effect on two variables that are not directly related. An example of an inverted fork is AC usage behavior → energy use intensity ← annual income, where the correlation between AC usage behavior and annual income does not imply a correlation between AC usage behavior and energy use intensity. Inverted forks are relevant to the problem of collider bias (Ananth & Schisterman, 2017).

A path that consists only of chains can transmit a causal association, and variables that are directly or indirectly causally affected by a certain variable are called its descendants, while variables that directly or indirectly affect a certain variable are considered its ancestors. A path that also contains forks can still transmit an association, but it is no longer a causal association due to the confounding variable (in this case, annual income). A path that contains an inverted fork is blocked, and no association is transmitted.

The primary challenge when dealing with observational data is the issue of confounding, which occurs when a common cause exists that is behind both the independent variable (often referred to as the treatment or exposure) and the dependent variable (outcome) of interest These common causes are found on the back-door paths, which are paths that start with an arrow pointing to the independent variable and end with an arrow pointing to the dependent variable. The influence of these common causes can lead to what is called a spurious correlation. To eliminate the undesirable noncausal association, it is necessary to block the all back-door paths (Pearl, 2000).

3.2.2 Estimating causal effects with Double Machine Learning (DML)

Double machine learning (Chernozhukov et al., 2018) is a method developed to estimate treatment effects in the presence of high dimensional covariates. It is particularly effective in addressing a problem called "model selection bias" which occurs when the same data are used to choose the model and estimate the treatment effects.

DML relies on three main components: model_y, model_t, and model_final. The first, model_y, is a machine learning model that estimates the expected value of the outcome variable based on observed features and the exposure variable. Examples of such models include linear regression, decision trees, and neural networks. The second component, model_t, is another machine learning model that estimates the expected value of the exposure variable given the observed features. This model can utilize a variety of algorithms, such as logistic regression, support vector machines, and decision trees. Lastly, model_final is the ultimate model that estimates the causal effect of the exposure variable on the outcome variable, taking into account the predictions of both model_y and model_t. Common examples of model_final include linear regression, instrumental variable regression, and other linear models.

The DML method proceeds in several steps:

(i) Divide the dataset into K folds for cross-fitting purposes.

(ii) For each fold, execute the following:

a) Train model_y on the remaining K-1 folds using observed features (X) and exposure variable (T) to predict the outcome variable (Y). Then, predict Y for the held-out fold with the trained model_y.

b) Train model_t on the remaining K-1 folds using X to predict T. Then, predict T for the held-out fold with the trained model_t.

(iii) Calculate residuals for each observation in the dataset:

a) Compute the residual of the Y by subtracting the predicted value of Y from the observed value of Y for each observation.

b) Compute the residual of T by subtracting the predicted value of T from the observed value of T for each observation.

(iv) Train the model_final on the entire dataset, using the residuals of Y and T as dependent and independent variables, respectively. The estimated coefficients from model_final represent the causal effects of T on the Y, adjusted for X.

DML offers several advantages, such as reducing bias introduced by unobserved confounding variables, providing robust causal effect estimates, and being compatible with a wide range of ML techniques. However, it assumes that both the treatment assignment mechanism and the outcome process are correctly specified and that the models can estimate the true conditional expectations. As a result, researchers must carefully select and validate their models to ensure accurate causal effect estimation.

In this dissertation, the extreme gradient boosting trees (Xgboost) (T. Chen & Guestrin, 2016) algorithm was used for both model_y and model_t, while the lasso regression with cross validation (Obuchi & Kabashima, 2016) model was employed for model_final. Xgboost, is a highly regarded and efficient ML algorithm known for its effectiveness in managing high-dimensional data. It utilizes a gradient boosting framework to create an ensemble of decision trees, successively enhancing model predictions by minimizing a specified objective function. Xgboost excels with high-dimensional data due to its inherent regularization, which regulates model complexity and averts overfitting. Moreover, its column block and sparsity-conscious design facilitate rapid processing of extensive datasets, while its compatibility with parallel and distributed computing expedites training time.

3.2.3 Interpreting the estimated effect in terms of changes in the outcome

Based on the exposure in question, we may be interested in different causal relationships and effects for specific conditions, subsets of data, or individual instances of buildings. The Average Treatment Effect (ATE) is a widely used measure that aids in determining whether an exposure should be implemented. Evaluation of ATE learning is well-suited for regression error metrics; however, in our case, the ATE may be less relevant because each building represents a unique case. In such situations, the Conditional Average Treatment Effect (CATE) is employed to learn the causal effect encompassing heterogeneous groups, relying on graph information (causal relationships) demonstrated by DAGs and graph-based rules, such as do-calculus (Pearl, 2000). Fundamentally, CATE enables data-driven interventions and ensures accurate effect learning within each homogeneous group, preventing misleading or contradictory conclusions (Hernán et al., 2011). Therefore, in this dissertation, the causal effects are expressed in terms of CATE.

4 Methodology

4.1 Data description

To validate the robustness of the proposed ML-based causal inference approach, it was applied to three different datasets: the Residential Energy Consumption Survey (RECS) 2015, Korean Household Energy Panel Survey (KHEPS) 2018-2019, and the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Global Thermal Comfort Database II.

4.1.1 The Residential Energy Consumption Survey (RECS) 2015 dataset

4.1.1.1. Data collection

To evaluate energy policies and occupant behavior impacts on cooling energy consumption, the RECS 2015 data was used. RECS is a national survey that collects energy-related data for housing units occupied as a primary residence and the households that live in them, in the United States of America. The data include household energy use and expenditures, structural characteristics of homes, appliances used, demographic characteristics of residents, and behaviors toward energy use.

RECS is conducted by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. The survey began in 1978 and is currently conducted every four years. Data from RECS is collected using multi-stage sampling procedures to ensure representation of the nine Census divisions and urban and rural population areas. The survey typically involves an interview and sometimes a physical inspection of the energy-related aspects of the dwelling.

4.1.1.2. Inputs selection cleaning, and grouping

The survey description, expert knowledge, and findings from existing studies were utilized to handpick variables related to cooling energy consumption. The cooling energy consumption was then normalized by the floor area to determine the EUI. Mobile home records were omitted from the analysis. Furthermore, use cases with fewer than 10 records were also excluded from the investigation. The data cleaning process resulted in a reduction in the number of records from

5686 to 4701. Six categories were defined to classify all inputs, based on their influence on the outcome. These categories, as depicted in Figure 4.1 and Table 4.1, included climate factors, building characteristics, occupant socio-economic factors, occupant behavioral factors, equipment characteristics, and energy-saving policies.

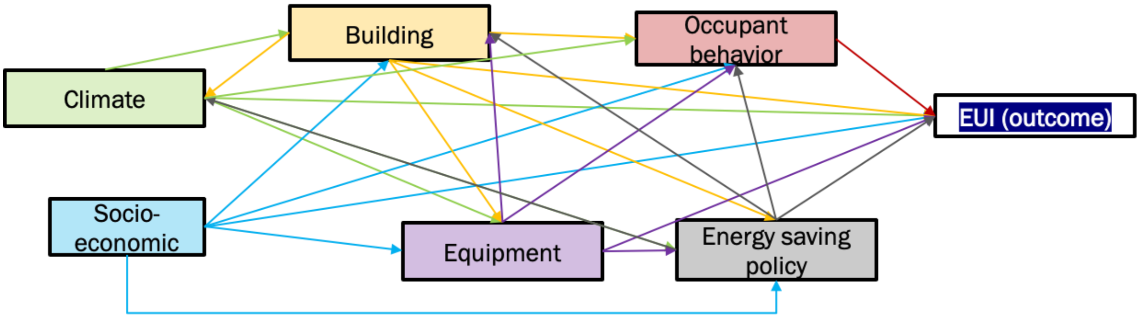


Figure 4. 1. Categories for the selected variables

4.1.1.3. EUI distribution in targeted exposures

Statistical tests were employed to evaluate the significance of hypothesized relationships and construct a DAG. Given that the EUI distribution in the targeted exposures deviated from normality, non-parametric tests were leveraged. These tests included Spearman's correlation, Kruskal-Wallis, Mann-Whitney, and Chi-square tests.

In the energy saving policies, the targeted exposures, as shown in Figures 4.2 and 4.3, are:

- Energy audit: a home energy audit refers to the process where a skilled expert investigates the energy consumption patterns throughout a residence. Following a comprehensive analysis of the house, the energy auditor will offer a series of recommendations designed to lower energy consumption and consequently, save money on energy-related expenses.
- Energy Star qualified windows: this variable determines if the windows installed in the building meet the criteria for Energy Star certification.
- Smart meter: logs electricity consumption in short time intervals and automatically sends this data to the corresponding utility service provider.
- Interval data access: the household can access data regarding their electricity consumption on an hourly or daily basis, which is recorded by their smart meter.

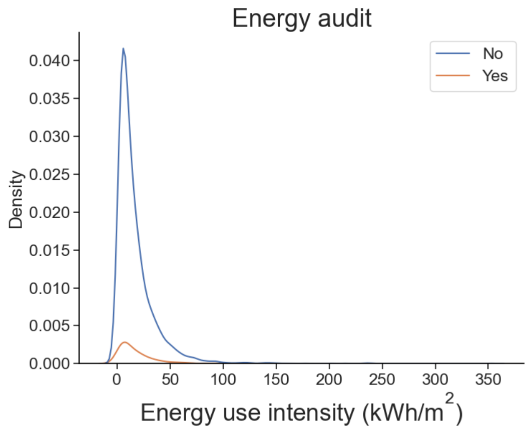
- Energy assistance: the household has engaged in a home energy support scheme that assists in covering energy expenses or repairing malfunctioning appliances.
- Electricity payment overseer: outlines the individual who bears the responsibility for covering the costs of electricity consumption within the household.

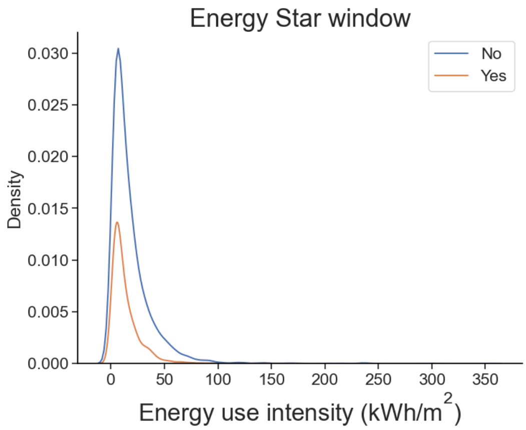
In the occupant behavior category, the targeted exposures are:

- Nighttime indoor temperature settings: describes how indoor air temperature (IAT) at night is set compared to the daytime in the cases when an individual is present at home during daytime hours (Scenario 1) and when the house is unoccupied during the day (Scenario 2). This comparison resulted in three cases, with case 1: nighttime IAT is greater than daytime IAT, 0: nighttime IAT is equal to daytime IAT, and -1: nighttime IAT is less than daytime IAT. EUI distribution for these cases is shown in Figure 4.4.
- Air conditioner (AC) usage behavior: describes how the household controls the AC equipment most of the time, with cases 1: Maintain a constant temperature setting for most of the time, 2: Adjust the temperature manually at night or when the residence is vacant, 3: Set up the thermostat to modify the temperature automatically at certain times during day and night, and 4: Operate the AC unit as required. The EUI distribution for this variable is shown in Figure 4.5.

Table 4. 1. Selected variables related to cooling energy consumption

1. CDD65	10. Type of house unit	19. Energy Star qualified windows
2. IECC climate zone	11. Roof material	20. Insulation level
3. Weather and shielding factor	12. Wall material	21. Home energy audit
4. Number of people	13. Number of windows	22. Electricity payment overseer
5. Annual income	14. Year built range	23. Smart meter
6. Cooling type	15. Number of rooms	24. Meter data access
7. Thermostat present	16. Total area	25. Energy assistance
8. Night vs Day indoor temperature	17. Total cooled area	26. Solar energy used
9. AC usage behavior	18. Latent heat infiltration	





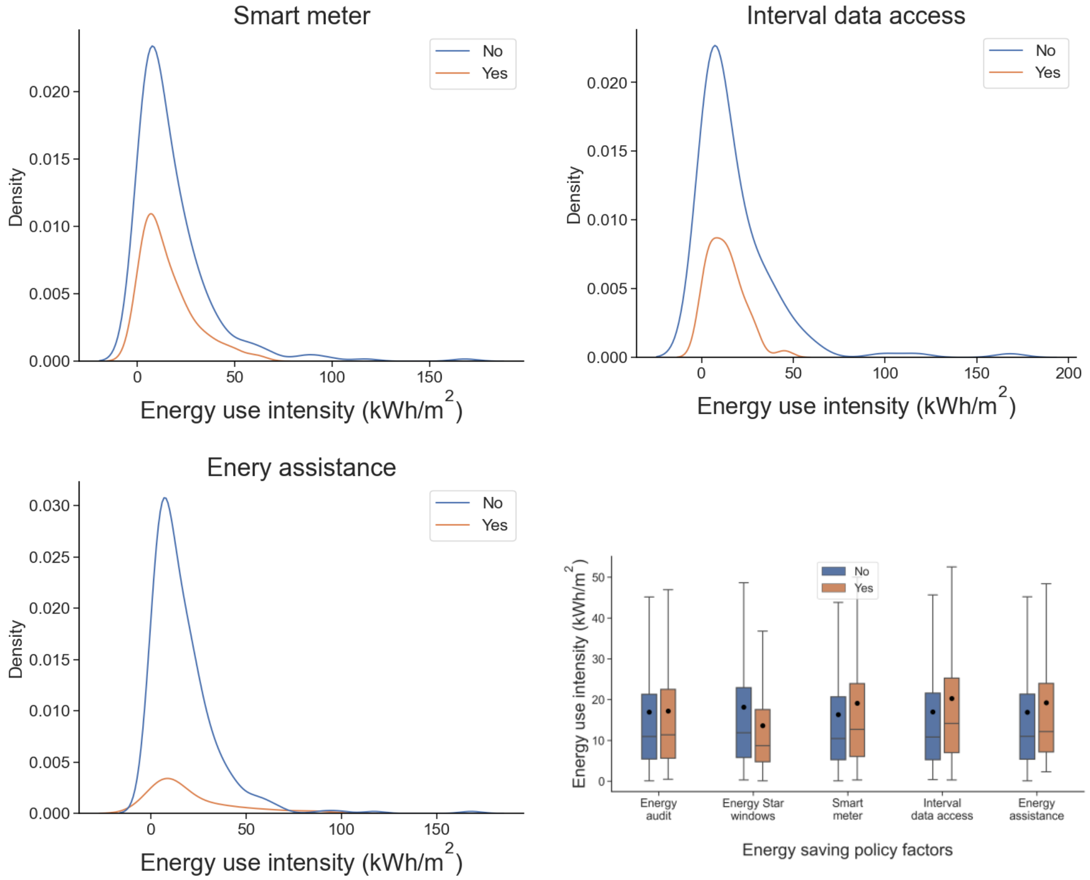


Figure 4. 2. Energy use intensity (EUI) distribution in energy audit, Energy Star qualified windows usage, smart meter usage, interval data access, and energy assistance

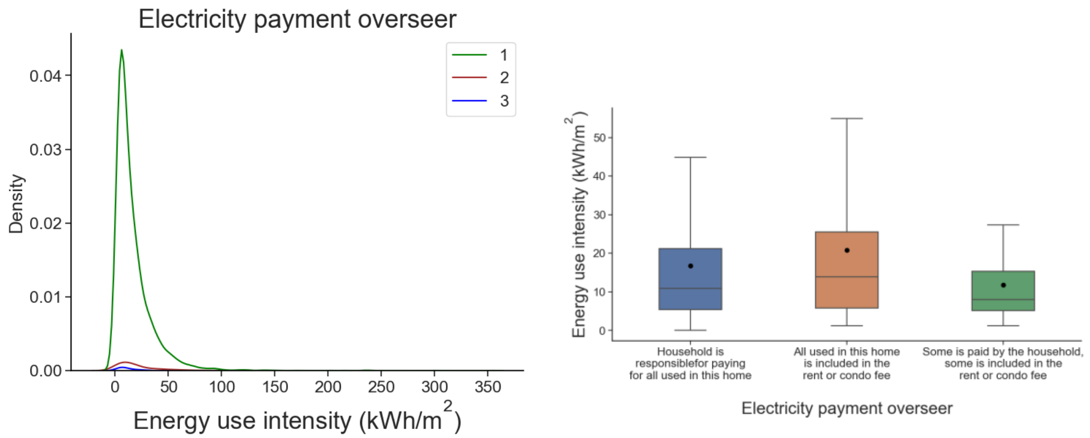


Figure 4. 3. EUI distribution in electricity payment overseer

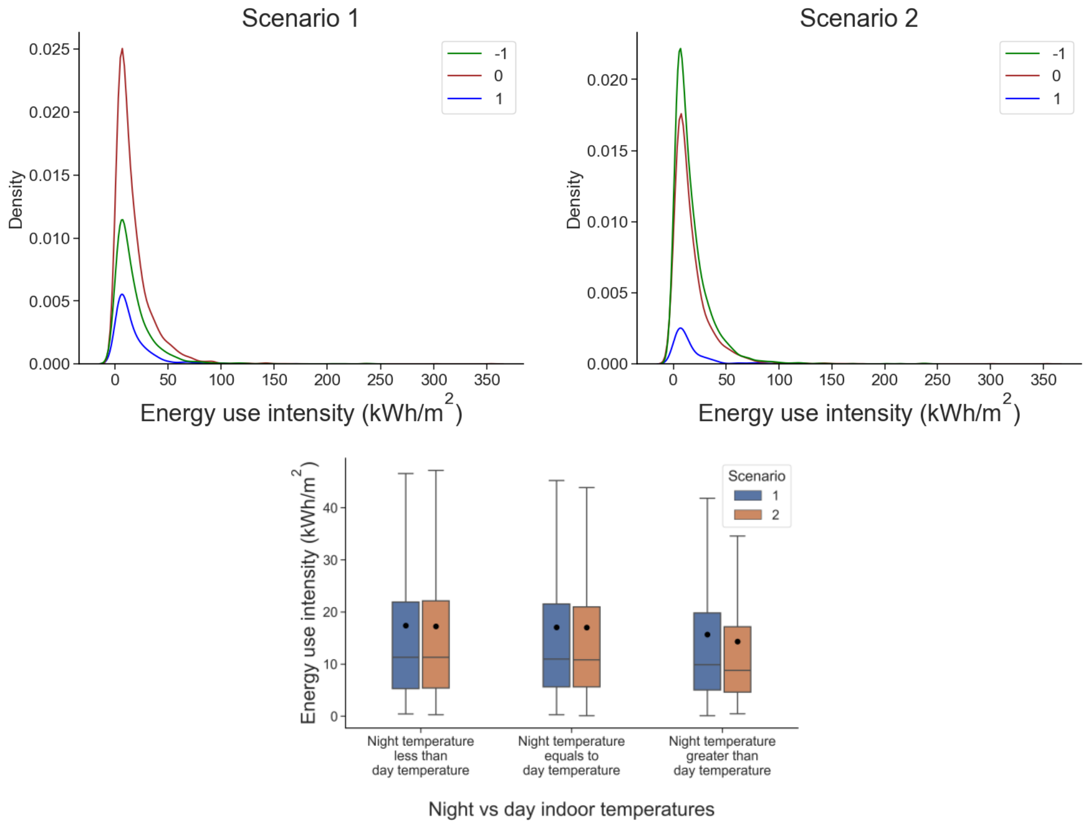


Figure 4. 4. EUI distribution in nighttime vs daytime indoor temperature settings

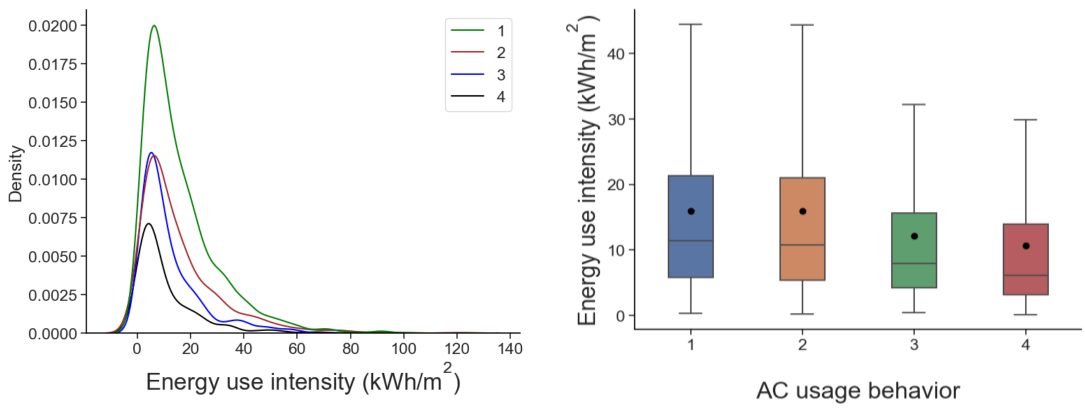


Figure 4. 5. EUI distribution in air condoning (AC) equipment usage behavior

4.1.2 The Household Energy Panel Survey (HEPS) dataset

4.1.2.1. Data collection

To identify role of socio-economic factors and heating equipment selection in energy consumption, the HEPS 2018-2019 data was used. HEPS is conducted by Korea Energy Economics Institute to understand the energy consumption and consumption behavior of households in order to use it as a basis for national energy policy. The data include household energy use and expenditures, structural characteristics of homes, appliances used, socio-economic characteristics of residents.

The survey began in 2011 and is currently conducted every year. Data collection methods involve surveys of a representative sample of households, with questions about their energy use, housing characteristics, household characteristics, energy-using appliances, and possibly even attitudes toward energy use and conservation.

4.1.2.2. Inputs selection cleaning, and grouping

The survey description, expert knowledge, and findings from existing studies were utilized to handpick variables related to heating energy consumption. The selected data covered the heating period starting from November 2018 to February 2019. The heating energy consumption was then normalized by the floor area to determine the EUI. Cases with less than 10 records were excluded from the investigation. The data cleaning process resulted in a reduction in the number of records from 2520 to 2118. Six categories were defined to classify all inputs, based on their influence on the outcome. These categories, as depicted in Figure 4.6 and Table 4.2, included climate factors, building characteristics, occupant socio-economic factors, occupant behavioral factors, equipment characteristics, and energy-saving policies.

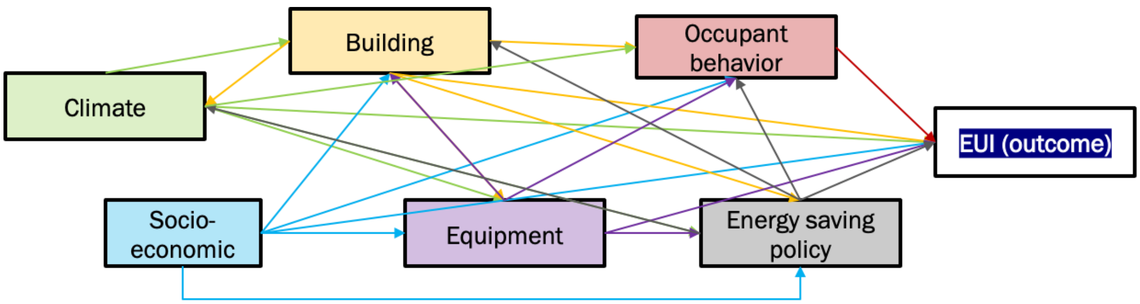


Figure 4. 6. Categories for the selected variables

Table 4. 2. Selected variables related to heating energy consumption

1. HDD18.5	10. Number of heated living rooms	19. Head of the household education level
2. Provincial climate	11. Number of heated bathrooms	20. Head of the household occupational status
3. Housing type	12. Number of external wall windows	21. Head of the household main source of income
4. Type of residential floor	13. Ratio of exterior windows to double-glazed windows	22. household member pregnant woman
5. Number of external walls	14. Number of double-glazed windows	23. Average monthly income
6. Housing orientation	15. House occupancy type	24. Main heating facility
7. Completion year (U-values)	16. Number of household members	25. Kerosene boiler heating efficiency
8. Residential floor area	17. The head of household gender	26. Gas boiler heating efficiency
9. Number of heated rooms	18. Head of the household age	

4.1.2.3. EUI distribution in the targeted exposures

Statistical tests were conducted to evaluate the significance of hypothesized relationships to construct a DAG. Given that the EUI distribution in the targeted exposures deviated from normality, non-parametric tests were leveraged. These tests included Spearman's correlation, Kruskal-Wallis, Mann-Whitney, and Chi-square tests.

In the occupant socio-economic category, the targeted exposures, as shown in Figures 4.7, 4.8, and 4.9, are the number of household members, age of the head of the household, head of the household education level, household average monthly income and income source. Income source (Figure 4.9) has cases such as 1: Your/spouse's work or occupation, 2: deposit, installment savings, 3: public pension, 4: real estate, 5: children, and 6: Assistance from the state and local governments. For the type of equipment, as shown in Figure 4.10, cases include 1: District heating, 2: Central heating with city gas, 3: Individual heating with kerosene boiler, 4: Individual heating with propane gas boiler, 5: Individual heating with city gas boiler, 6: Individual heating with electric boiler, 7: Individual heating with briquette boiler, and 8: Individual heating with electric blanket.

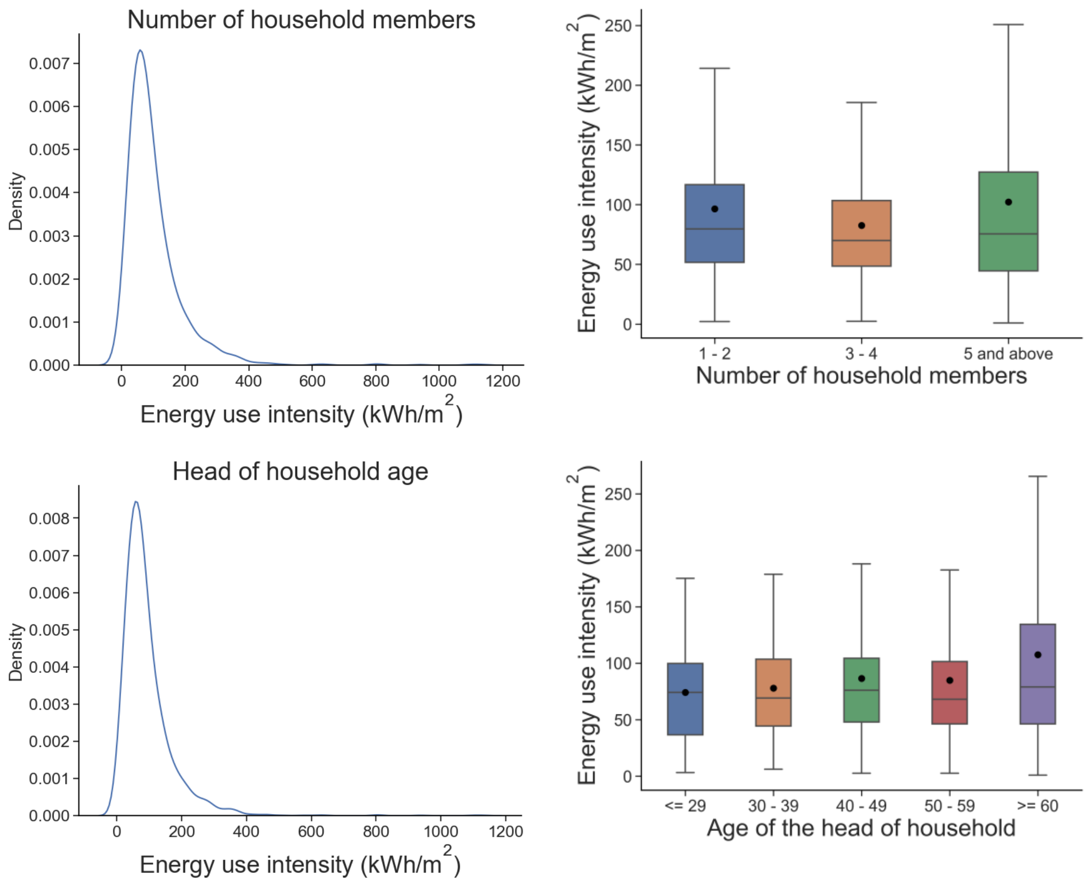


Figure 4. 7. EUI distribution in number of household members and head of household age

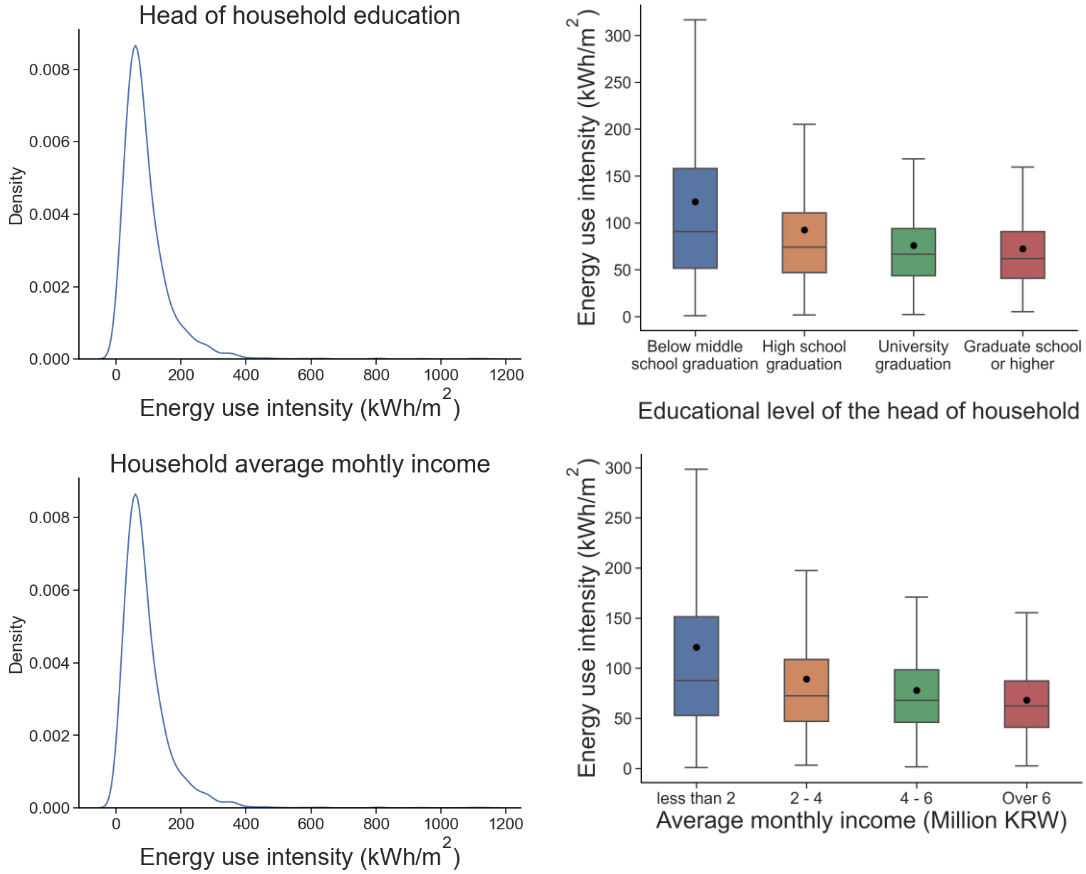


Figure 4. 8. EUI distribution in head of household education level and household average monthly income

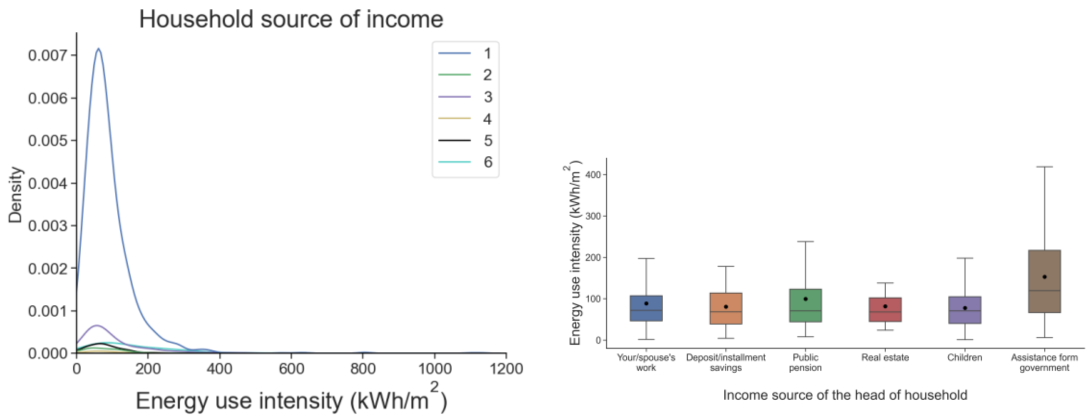


Figure 4. 9. EUI distribution in household source of income

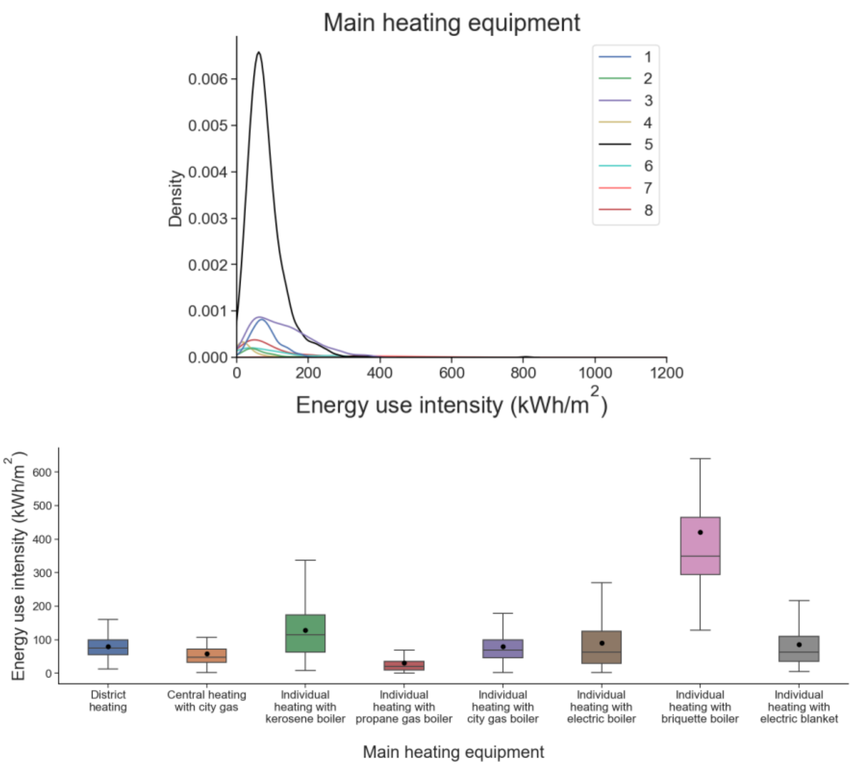


Figure 4. 10. EUI distribution in main heating equipment used

4.1.3 The ASHRAE Global Thermal Comfort Database II

4.1.3.1. Data collection

To assess the impact of occupant personal and behavioral factors on thermal sensation, the ASHRAE global thermal comfort database II (Földváry Ličina et al., 2018) was utilized. This database comprises 55 thermal comfort studies conducted in diverse climate zones across 27 countries. It offers valuable information regarding subjects' personal details, subjective thermal comfort ratings, instrumental thermal comfort measurements, calculated indices, and environmental control variables. The comprehensive data available in this database enables researchers to analyze and develop thermal comfort prediction models for various indoor environments, encompassing residential, commercial, and institutional settings.

4.1.3.2. Inputs selection cleaning, and grouping

For a focused analysis, the six variables (air velocity, air temperature, mean radiant temperature and relative humidity, clothing, and metabolism rate) used in PMV calculation were firstly selected, to which personal and behavioral factors were added. Personal factors-related variables included age, gender, and BMI (calculated from weight and height), and behavioral factors like the usage of blinds/curtains, windows, doors, fans, and heaters.

All inputs were categorized into environmental factors, building characteristics, personal, behavioral, and subjective thermal comfort-related factors, as shown in Figure 4.11 and Table 4.3. As the ASHRAE db II is a combination of many surveys, a stepwise data cleaning by removing null values was adopted. That is, based on the variable under consideration, null values were removed. Thus, the number of observations varied accordingly. The original dataset has 108,235. For instance, considering personal factor like age yields in 21,810 observations while BMI yields 30,147 observations.

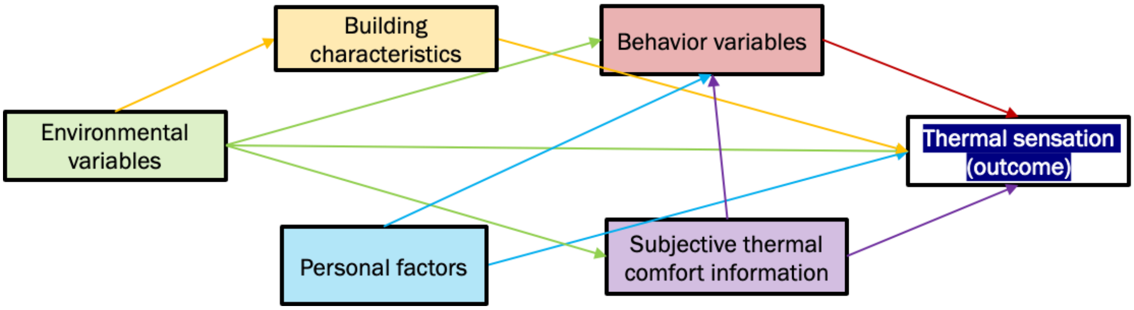


Figure 4. 11. Categories for the selected variables

Table 4. 3. Selected variables related to thermal sensation

1. Building type	7. Clo	13. Blind (curtain)
2. Cooling strategy	8. Met	14. Fan
3. Age	9. Air temperature	15. Window
4. Gender	10. Radiant temperature	16. Door
5. Subject's Weight	11. Relative humidity	17. Heater
6. Subject's Height	12. Air velocity	

4.1.3.3. Thermal sensation distribution in the targeted exposures

Statistical tests were employed to evaluate the significance of hypothesized relationships and construct a DAG. EUI distribution for personal factors (BMI: n = 21810 and sex: n = 30147) is shown in Figures 4.12 and 4.13. Occupant's age (Figure 4.13) was grouped into bins of 10 years interval (n = 21810). Occupant behavior characteristics (state of blinds: n = 1986, fan mode: n = 5833, state of the windows: n = 4984, state of doors: n = 1108, and heater mode: n = 4741) are shown in Figures 4.14, 4.15, and 4.16.

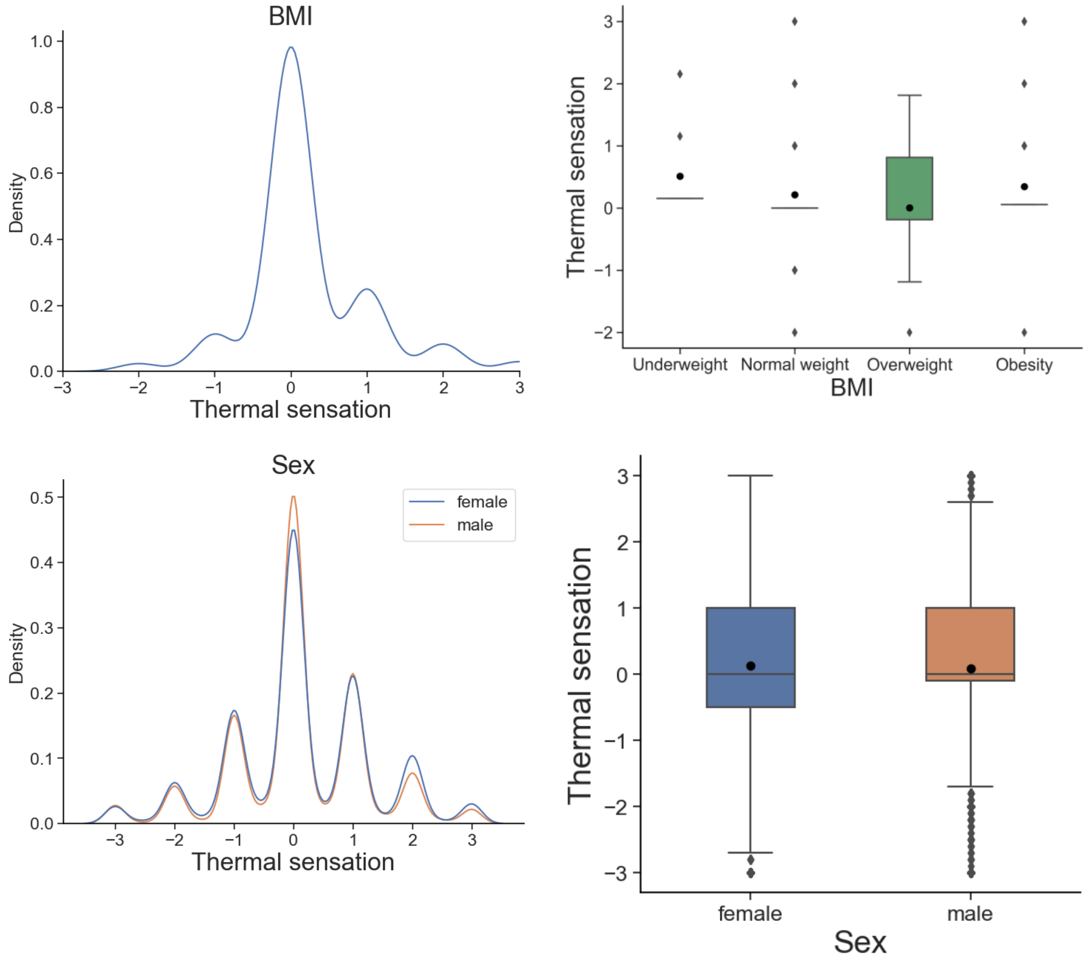


Figure 4. 12. Thermal sensation distribution in occupants' BMI and sex

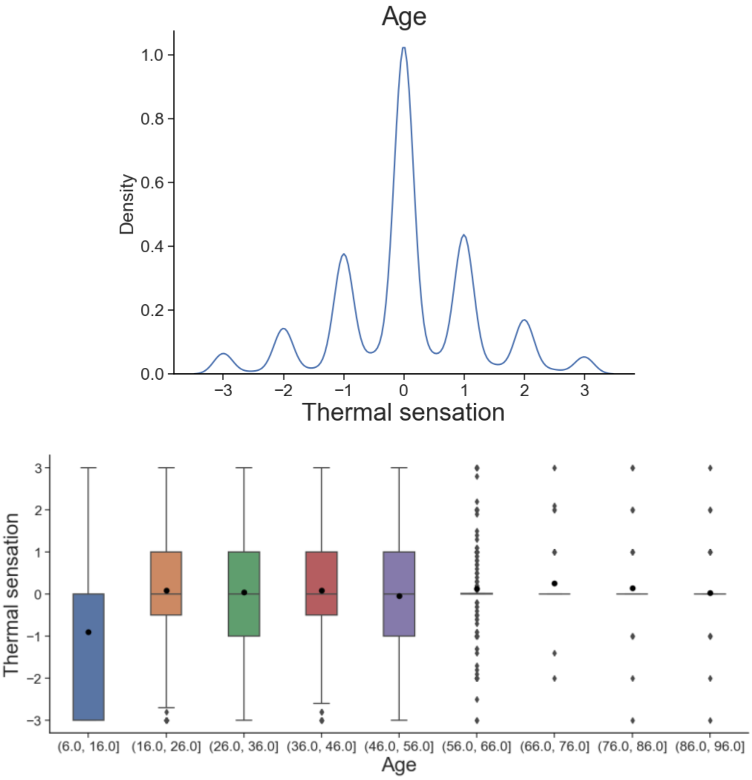


Figure 4. 13. Thermal sensation distribution in occupants' age

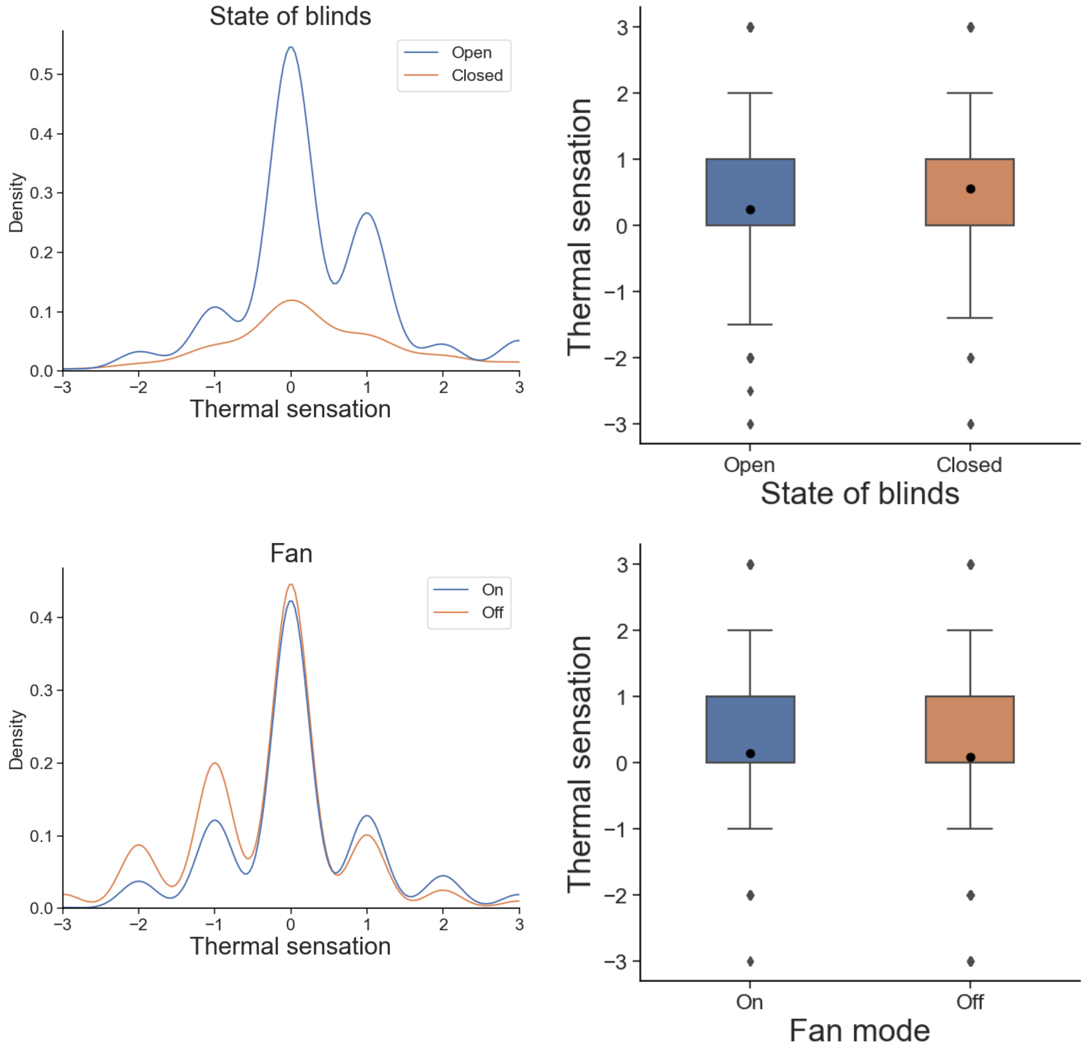


Figure 4. 14. Thermal sensation distribution in state of blinds and fan mode

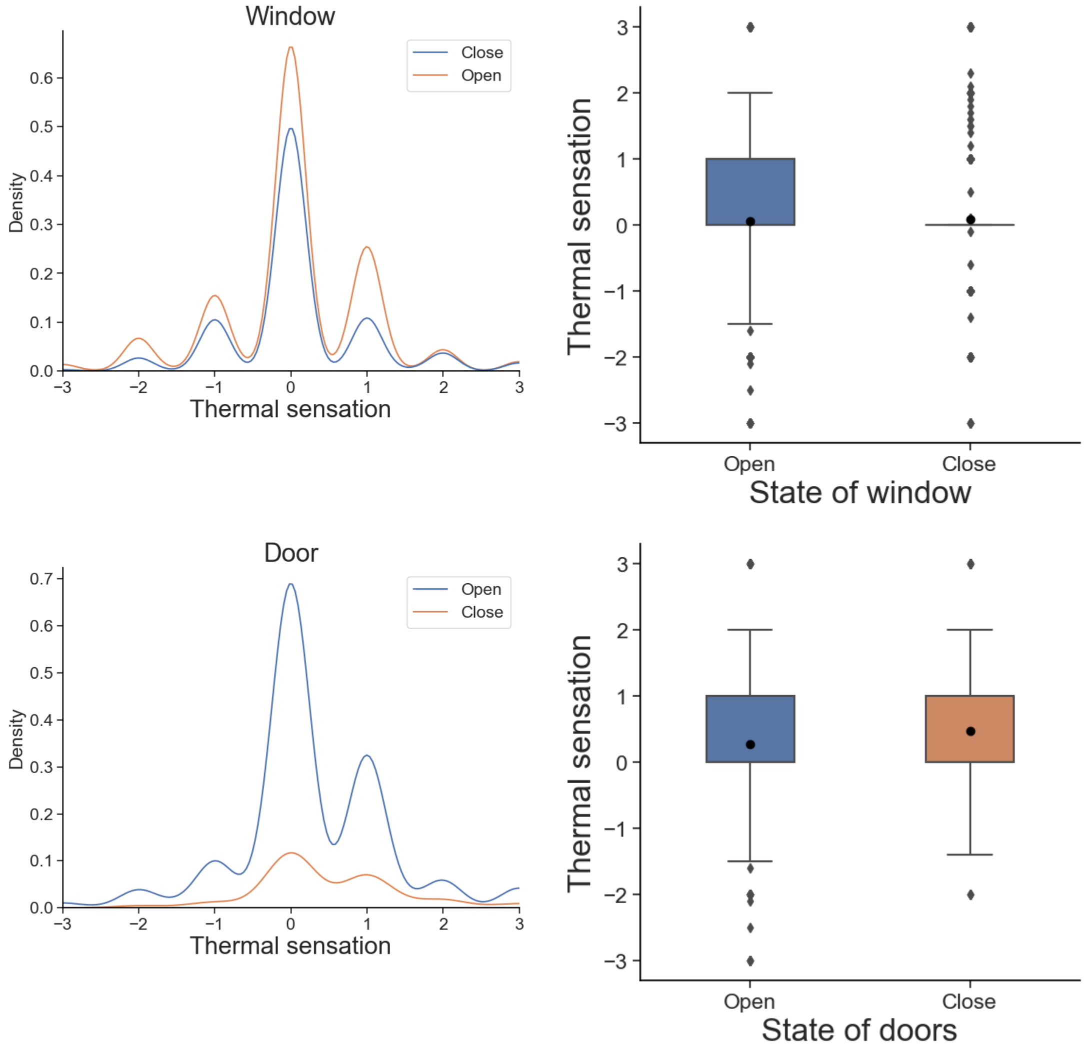


Figure 4. 15. Thermal sensation distribution in state of windows and doors

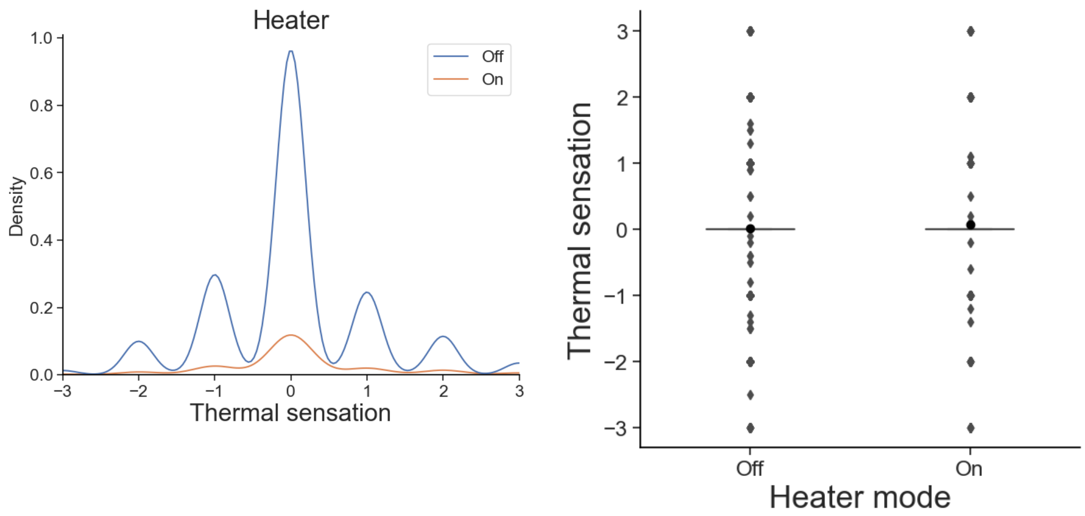


Figure 4. 16. Thermal sensation distribution in heater mode

4.2 Causal inference analysis framework

4.2.1 The analysis framework

Figure 4.17 illustrates the causal inference analysis framework employed in this dissertation to estimate the causal effects of several variables on both the cooling/heating energy and thermal comfort. The process begins with the construction of a Directed Acyclic Graph (DAG), which is based on the input dataset and domain knowledge assumptions regarding the interrelationships among variables (step 1). In step 2, the target estimand is identified through an analysis of all present confounders. Subsequently, the causal effect is estimated (step 3). To ensure the reliability of the estimated causal effect, the final step involves validating the results using various refutation methods (step 4). This comprehensive framework provides a systematic approach to understanding the causal relationships between the variables and their effects on cooling/heating energy and thermal comfort. To obtain the best hyperparameters for accurate estimates, a joint hyperparameter optimization was employed and the estimated effects were validated with refutation methods (see the next section). Figure 4.18 illustrates this process.

In terms of tools used, DoWhy (Sharma & Kiciman, 2020) and EconML (Battocchi et al., 2019) python-based libraries were utilized for causal inference analysis. DoWhy stands out among the existing causal inference libraries such as CausalML (H. Chen et al., 2020) and Causallib (Shimoni et al., 2019) as it support the last step of the analysis framework (i.e., Refute the estimate), providing a unified framework to validate the estimated causal inference. DoWhy reads the DAG and identify the estimate, EconML uses ML algorithms to estimate the effect, which is then send back to DoWhy for refutation.

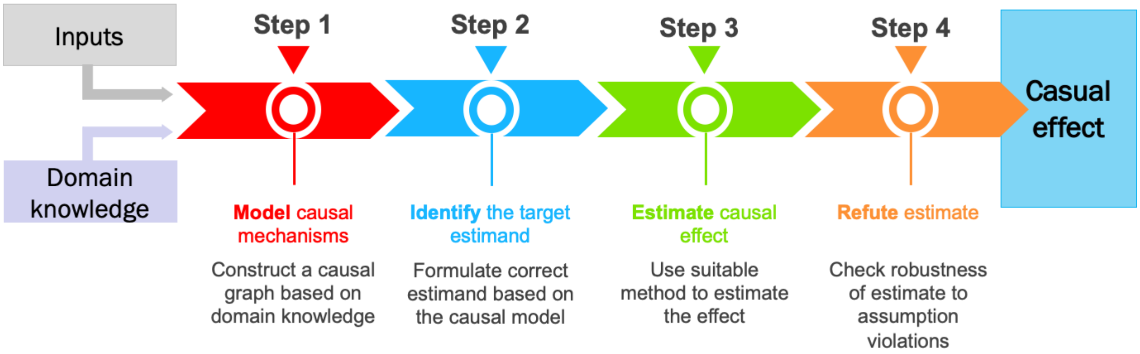


Figure 4. 17. Causal inference analysis framework

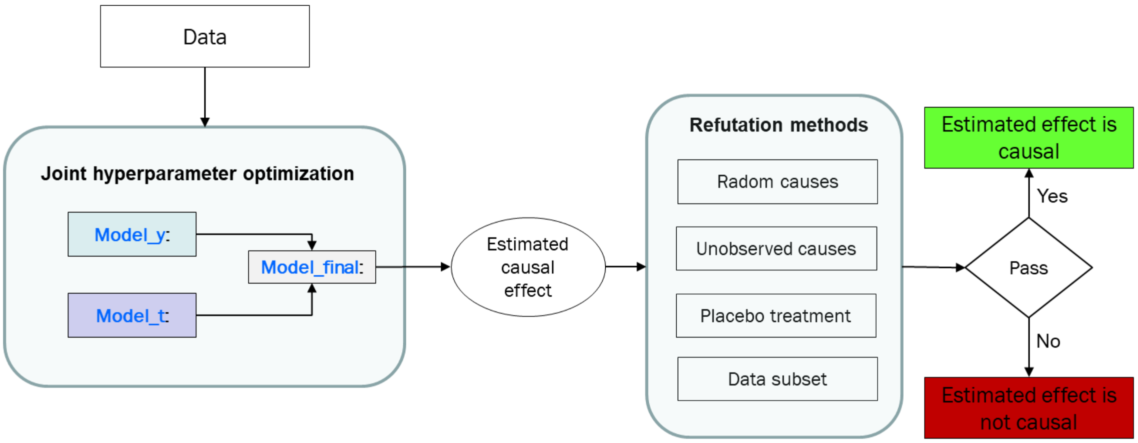


Figure 4. 18. Joint hyperparameters optimization and estimate refutation methods

4.2.2 Validating the estimated causal effects

To validate the estimated causal effect, this dissertation employed refutation methods. These include random common causes, unobserved causes, placebo treatment, and data subset analysis.

The random Common cause refutation method introduces an independent random variable as a common cause to the original DAG and then re-estimates the causal effect. This method aims to test if the inclusion of an additional confounder significantly changes the original causal estimate, which could indicate the presence of unmeasured confounders. If the newly estimated causal effect remains similar to the original estimate, it suggests that the initial analysis is robust against potential hidden confounders (Rosenbaum & Rosenbaum, 2002).

The Unobserved Causes refutation method attempts to test the sensitivity of the original causal estimate to potential unobserved confounding variables. This approach involves systematically introducing biases into the estimated causal relationship, simulating the presence of an unobserved confounder, and then re-estimating the causal effect. By examining how the original causal estimate changes in response to varying degrees of unobserved confounding, this method provides insights into the robustness of the original analysis with respect to hidden biases (Rosenbaum & Rubin, 1983). Ideally, the newly estimated causal effect should not deviate far from the original estimate.

The placebo exposure refutation method substitutes the original treatment variable with a randomly generated placebo exposure that has no causal relationship with the outcome variable. It then re-estimates the causal effect using the placebo exposure. The goal is to test the sensitivity of the original causal estimate to random noise. If the newly estimated causal effect is substantially different from the original estimate, it indicates that the original analysis is robust and not driven by noise or confounding (Banerjee et al., 2009). The data subset refutation method assesses the stability of the original causal estimate by re-estimating the causal effect on different subsets of the data. This approach helps evaluate whether the causal effect is consistent across various subsamples, providing an indication of the generalizability and robustness of the original analysis (Imbens & Rubin, 2015).

5 Evaluating energy policies and occupant behavior impacts on cooling energy consumption

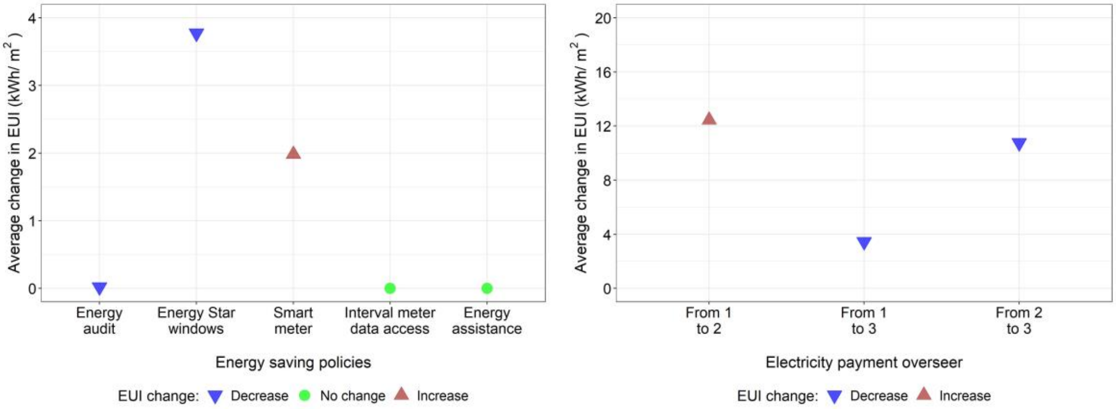
5.1 The effectiveness of energy policies on EUI

5.1.1 Energy saving policies vs EUI considering all buildings

Figure 5.1 provides empirical insights into the differential impacts of various energy-saving policies on EUI. From Figure 5.1.a, the execution of energy audits appears to have a marginal positive effect, increasing EUI by 0.021 kWh/m². In stark contrast, the adoption of Energy Star qualified windows markedly decreases EUI by -3.772 kWh/m², the magnitude of which underscores the significant potential of these technologies for substantial energy savings. The installation of smart meters somewhat increases the EUI by 1.982 kWh/m², suggesting these devices might lead to an increase in energy use, perhaps due to increased energy consciousness leading to rebound effects. The policies of interval data access and energy assistance, interestingly, have a neutral effect, contributing no change to the EUI. Table 5.1 provide more details about these results.

In Figure 5.1.b, the causal relationship between the allocation of electricity costs and EUI, shows that the transition from a household being solely responsible for all electricity costs to a model where all electricity is included in the rent or condo fee precipitates a significant increase in EUI of 12.44 kWh/m². Conversely, transitioning from a scenario where a household is responsible for all costs to a hybrid model – where some electricity costs are covered by the household and some included in the rent or condo fee – results in a decrease in EUI of -3.445 kWh/m². This outcome suggests that shared responsibility promotes more efficient energy use. Lastly, a shift from all electricity costs included in the rent or condo fee to the hybrid model generates a substantial reduction in EUI of -10.762 kWh/m². This further bolsters the view that partial cost responsibility enhances energy efficiency. Table 5.2 provide more details about these results.

(a) (b)



1: The household bears the responsibility of paying for all the utilities consumed in this home. 2: The utilities consumed in this residence are covered by the rent or condominium fee. 3: The cost of certain utilities is borne by the household, while the rest is incorporated in the rent or condominium fee

Figure 5. 1. Causal effects of energy saving policies on EUI

Table 5. 1. Causal effects of energy saving policies on EUI

	Estimated average change in EUI	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Energy audit	-0.021	-0.021	-0.020	0	-0.021
Energy Star qualified windows	-3.772	-3.772	-3.780	0	-3.772
Smart meter	1.982	1.982	1.984	0	1.982
Interval meter data access	0	0	0	0	0
Energy assistance	0	0	0	0	0

Table 5. 2. Causal effects of electricity payment overseer on EUI

Electricity payment overseer	**Estimated average change in EUI**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
1 to 2	12.442	12.442	12.442	0	12.402
1 to 3	-3.445	-3.445	-3.445	0	-3.445
2 to 3	-10.762	-10.762	-10.760	0	-10.764

1: The household bears the responsibility of paying for all the utilities consumed in this home. 2: The utilities consumed in this residence are covered by the rent or condominium fee. 3: The cost of certain utilities is borne by the household, while the rest is incorporated in the rent or condominium fee

In an in-depth analysis to confirm the causal effects on EUI, another analysis focusing on energy consumption as the outcome (see Tables 5.3, 5.4 and 5.5) revealed that the utilization of Energy Star qualified windows actually reduces overall energy consumption, whereas other energy-saving policies do not contribute to EUI reduction. The quantified benefits of Energy Star qualified windows align with the research conducted by (Nevin, 2010), who emphasized the potential advantages of Energy Star qualified windows in reducing energy consumption and enhancing building efficiency. Furthermore, when examining the factor of "who pays for the electricity" a similar reduction is observed when the responsibility for all electricity bills shifts from the tenant to a shared approach between the landlord and the tenant. This reduction could be attributed to the reduced motivation for energy efficiency in cases where tenants are solely responsible for paying the bills. This finding supports the studies conducted by (Hope & Booth, 2014; Im et al., 2017), which found that energy-efficient features increase rent, potentially leading tenants to opt for less energy-efficient buildings.

Table 5. 3. Causal effects of energy saving policies on energy consumption

Variable	**Estimated average change in EUI (kWh/m^2)**	**Average change in energy (kWh)**
Energy audit	0.021	0
Energy Star qualified windows	-3.772	-73.959
Smart meter	1.982	619.507
Interval data access	0	415.548
Energy assistance	0	-545.330

Table 5. 4. Causal effects of electricity payment overseer on energy consumption

Variable	**Estimated average change in EUI (kWh/m^2)**	**Average change in energy (kWh)**	**Mean area (m^2)**
From 1 to 2	12.442	-2305.998	164.031
From 1 to 3	-3.445	-1159.097	63.973
From 2 to 3	-10.762	0	85.779

1: The household bears the responsibility of paying for all the utilities consumed in this home. 2: The utilities consumed in this residence are covered by the rent or condominium fee. 3: The cost of certain utilities is borne by the household, while the rest is incorporated in the rent or condominium fee

Table 5. 5. Pairwise test in electricity payment overseer

Pairwise comparisons		
Code	**Test statistic**	**Significance**
1-2	-256.288	$p < 0.05$
1-3	291.144	$p < 0.05$
2-3	547.432	$p < 0.05$

1: The household bears the responsibility of paying for all the utilities consumed in this home. 2: The utilities consumed in this residence are covered by the rent or condominium fee. 3: The cost of certain utilities is borne by the household, while the rest is incorporated in the rent or condominium fee

5.1.2 Energy saving policies vs EUI considering buildings that underwent energy audit

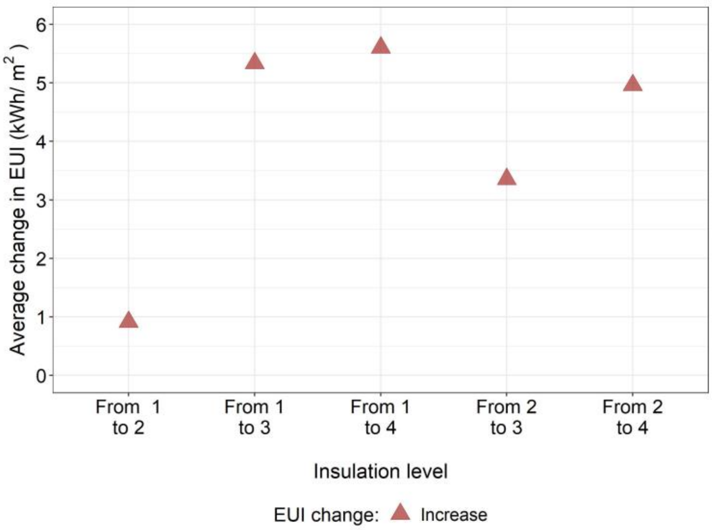
Narrowing down the analysis and focusing on buildings that implemented energy audit suggested measures (see Table 5.6), the impact of Energy Star qualified windows remains significant, although smaller compared to buildings without the audit. Another noteworthy finding is that access to interval meter data demonstrates a reduction in EUI. On the other hand, the energy assistance program does not show effectiveness in reducing EUI, partially aligning with the study conducted by (Reina & Kontokosta, 2017), which reported an increase in EUI (occupant, climate, and behavioral factors were not considered in the analysis). Surprisingly, the usage of smart meters has no contribution to EUI reduction (but raises EUI when all buildings are considered), conflicting with the findings of (Gao & Zhang, 2021), who reported reduced EUI through econometric models without considering occupant, building, and equipment factors. Our comprehensive framework, which accounts for all relevant factors and data non-linearity, enhances the reliability of our findings. It is evident that existing policies such as energy audits and assistance programs do not achieve their intended EUI reduction goals, consistent with studies such as (Kontokosta et al., 2020), which found little difference in energy consumption between audited and non-audited properties.

Table 5. 6. Causal effects of energy saving policies on EUI in audited and non-audited buildings

Variable	**Estimated average change in EUI (kWh/m^2) for building that receive energy audit**	**Estimated average change in EUI (kWh/m^2) for building that did not receive energy audit**
Energy Star qualified windows	-2.260	-4.108
Smart meter	0	2.543
Interval data access	-7.035	0
Energy assistance	0	0

5.1.3 Building insulation level reflecting building thermal performance over the years

Figure 5.2 elucidates the causal effect of building insulation level on EUI, reflecting the implementation of energy-saving policies through building codes. There is a noticeable trend indicating that as insulation quality decreases, EUI increases. A clear example of this trend is observed when transitioning from well-insulated buildings to those without insulation, resulting in a significant EUI increase of 5.603 kWh/m². This change in EUI based on the insulation level is believed to reflect government policies regarding building codes (see Table 7), which have evolved over the years. Therefore, it is evident that buildings constructed in recent decades exhibit lower EUI, validating the effectiveness of insulation-related energy policies.



1: Thoroughly insulated, 2: Satisfactorily insulated, 3: Inadequately insulated, and 4: Lacking insulation

Figure 5. 2. Causal effect of building insulation level on EUI

Table 5. 7. Causal effects of the house completion year on EUI

House completion year	
Cases	**Average change in EUI (kWh/m^2)**
From 1 to 2	1.333
From 2 to 3	0
From 3 to 4	-4.995
From 4 to 5	0
From 5 to 6	-2.484
From 6 to 7	-0.013
From 7 to 8	-3.383

1: Before 1950, 2: 1950 to 1959, 3: 1960 to 1969, 4: 1970 to 1979, 5: 1980 to 1989, 6: 1990 to 1999, 7: 2000 to 2009, and 8: 2010 to 2015

5.1.4 Economic analysis for implementing Energy Star qualified windows

The installation of energy-efficient windows with an Energy Star rating offers numerous benefits. While the average replacement cost can vary between $473 and $3,109 per window depending on factors like brand, window type, and chosen features, the long-term savings are substantial. The Environmental Protection Agency's EnergyStar program estimates that replacing single-pane windows in an average U.S. home can result in annual savings of $101 to $583, and double-pane windows can save $27 to $197 per year with EnergyStar-qualified replacements. With an average lifespan of over 30 years, Energy Star qualified windows are a wise long-term investment. Furthermore, these windows can potentially increase a home's selling price by up to $12,000. To promote the adoption of energy-efficient windows, the government could provide incentives or discounts, not only for replacement projects but also for new construction, encouraging homeowners to reap the benefits of energy savings and increased property value.

5.1.5 Increasing or negligible? The effect of smart meters on EUI and potential reasons behind

The results revealed that the utilization of smart meters was linked to an overall increase in EUI across all buildings included in the analysis. However, when examining the subset of buildings that underwent an energy audit, the impact of smart meters on EUI was found to be negligible, contrary to the anticipated energy savings predicted by (Gao & Zhang, 2021).

For buildings that received an energy audit:

- The influence of smart meters on EUI can primarily be attributed to an excessive reliance on automation. While smart meters and connected systems have the potential to optimize energy usage, occupants may excessively depend on automation, disregarding more energy-efficient manual practices.
- This phenomenon is exemplified by the analysis of air conditioning (AC) usage behavior, discussed in subsequent section, where manually turning the AC on and off proves to be more energy-efficient than relying solely on automated control. The data reveals that 31.6% of households utilize automated AC control, whereas only 6.3% employ manual control.
- Additionally, the study found that 42% of households tend to maintain a constant temperature setting for their AC most of the time, a practice that is not energy efficient.

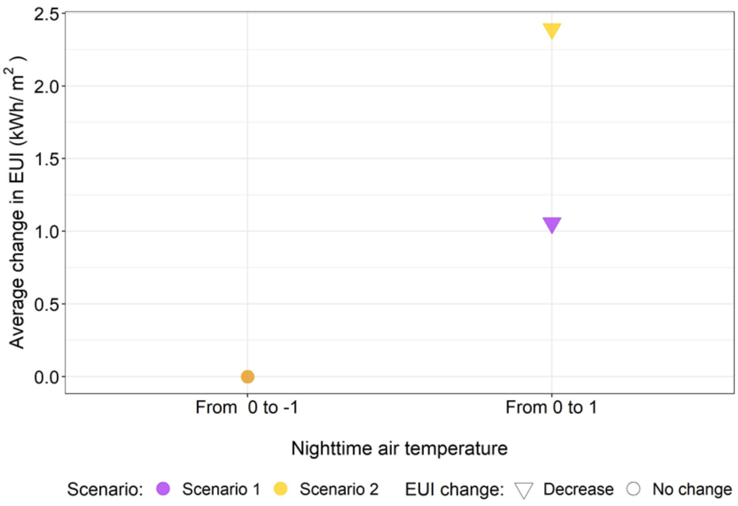
For buildings that did not undergo an energy audit: It is hypothesized that the observed increase in EUI may be attributed to a feedback effect. This dissertation proposes that providing real-time feedback on energy consumption can inadvertently lead to an upsurge in energy use. This is because individuals, upon realizing they are using less energy than their neighbors or less than their perceived "normal," may subsequently increase their energy consumption.

5.2 The influence of occupant behavior on EUI

5.2.1 Nighttime indoor air temperature settings vs EUI

Figure 5.3 presents the causal impacts of nighttime indoor air temperature (IAT) settings on Energy Use Intensity (EUI), as measured in kilowatt-hours per square meter (kWh/m²), under two distinct occupancy scenarios. In both scenarios, when nighttime IAT is set to be lower than daytime IAT, it results in no change in EUI, likely due to reduced energy consumption during the non-peak hours. However, intriguing differences emerge when nighttime IAT is set to be higher than the daytime IAT. In Scenario 1, where someone is home during the day, this leads to a reduction in EUI by -1.056 kWh/m². Even more substantial savings are observed in Scenario 2, where no one is home during the day, with a decrease in EUI by -2.393 kWh/m². Tables 5.8 and 5.9 provide detailed causal inference results.

Since this is summertime, the decrease in EUI is suspected to be caused by the lower temperature gradient (temperature difference between the indoors and outdoors) leading to less heat flows into the building (see Table 5.10). Therefore, the cooling system doesn't need to work as hard, leading to lower energy consumption as explained by (Lechner, 2014). These outcomes suggest that increasing nighttime IAT relative to daytime settings, particularly when the home is unoccupied during the day, can lead to notable energy savings. These findings underscore the importance of smart temperature management in homes, especially in the context of unoccupied daytime periods, for effective energy conservation. The differential effects in the two scenarios also point to the significance of home occupancy patterns in the effective design and implementation of energy-saving strategies.



-1: Night temperature less than day temperature, 0: Night temperature equals to day temperature and 1: Night temperature greater than day temperature.

Scenario 1: when an individual is present at home during daytime hours and Scenario 2: when the house is unoccupied during the day

Figure 5. 3. Causal effect of nighttime temperature settings on EUI

Table 5. 8. Causal effect of nighttime temperature settings on EUI

		Estimated average change in EUI	Refutation			
			Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Scenario 1	From 0 to -1	0	0	0	0	0.001
	From 0 to 1	-1.056	-1.056	-1.064	0	-1.056
Scenario 2	From 0 to -1	0	0	0	0	0.002
	From 0 to 1	-2.393	-2.393	-2.379	0	-2.39

-1: Night temperature less than day temperature, 0: Night temperature equals to day temperature and 1: Night temperature greater than day temperature.

Scenario 1: when an individual is present at home during daytime hours and Scenario 2: when the house is unoccupied during the day

Table 5. 9. Pairwise comparison for different nighttime temperature setting scenarios

	Pairwise comparisons		
	Code	Test statistic	Significance
Scenario 1	0- -1	-24.876	$p > 0.05$
	0-1	168.471	$p < 0.05$
Scenario 2	0- -1	-31.514	$p > 0.05$
	0-1	247.769	$p < 0.05$

-1: Night temperature less than day temperature, 0: Night temperature equals to day temperature and 1: Night temperature greater than day temperature.

Scenario 1: when an individual is present at home during daytime hours and Scenario 2: when the house is unoccupied during the day

Table 5. 10. Mean day and night indoor temperature

	Case	Mean day temperature	Mean night temperature
Scenario 1	-1	23.522 (SD=2.446)	21.019 (SD=2.740)
	0	22.400 (SD=2.487)	22.400 (SD=2.487)
	1	21.482 (SD=3.031)	23.571 (SD=2.609)
Scenario 2	-1	25.082 (SD=2.720)	21.824 (SD=2.714)
	0	22.407 (SD=2.599)	22.407 (SD=2.599)
	1	20.572 (SD=3.534)	23.238 (SD=2.930)

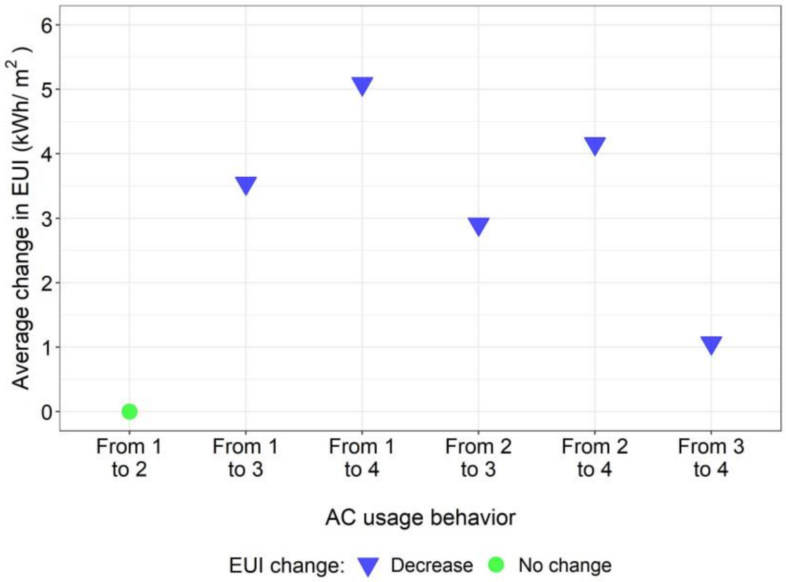
-1: Night temperature less than day temperature, 0: Night temperature equals to day temperature and 1: Night temperature greater than day temperature.

Scenario 1: when an individual is present at home during daytime hours and Scenario 2: when the house is unoccupied during the day

5.2.2 Air conditioning (AC) system usage behavior vs EUI

Figure 7 delineates the causal implications of diverse occupant behaviors in utilizing AC on EUI. Transitioning from a static temperature setting to manual temperature adjustments during the night or when the home is unoccupied results in no change in EUI. However, transitioning from a static setting to programming the thermostat for automatic adjustments at certain times results in a significant reduction in EUI by -3.542 kWh/m². Further energy savings can be realized when switching from a static setting to turning the equipment on or off as needed, resulting in a decrease in EUI by -5.086 kWh/m². When switching from manual adjustments to programmed thermostat settings or to turning the equipment on/off as needed, there are additional decreases in EUI, with respective reductions of -2.907 kWh/m² and -4.156 kWh/m². Lastly, transitioning from programmed thermostat settings to turning the equipment on/off as needed leads to a modest reduction of -1.068 kWh/m² in EUI. Detailed results regarding causal inference, the mean indoor temperature and AC usage behavior, and pairwise statistical comparisons are provided in Tables 5.11, 5.12, and 5.13, respectively.

This dissertation provides empirical evidence, confirming that active and strategic management of AC usage, particularly through the programming of thermostats and discretionary equipment usage, can result in substantial energy savings. This is a quantification of the demand response approach (Aghniaey & Lawrence, 2018; Wai et al., 2014) that those households may have followed, that consequently reduced EUI.



1: Maintain a constant temperature setting for most of the time, 2: Adjust the temperature manually at night or when the residence is vacant, 3: Set up the thermostat to modify the temperature automatically at certain times during day and night, and 4: Operate the AC unit as required

Figure 5. 4. Casual effect of air conditioning (AC) usage behavior on EUI

Table 5. 11. Causal effect of air conditioner (AC) equipment usage behavior on EUI

	Estimated average change in EUI	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	0	0	0	0	0.001
From 1 to 3	-3.542	-3.542	-3.556	0	-3.542
From 1 to 4	-5.086	-5.086	-5.065	0	-5.086
From 2 to 3	-2.907	-2.907	-2.908	0	-2.907
From 2 to 4	-4.156	-4.156	-4.128	0	-4.156
From 3 to 4	-1.068	-1.068	-1.065	0	-1.067

1: Maintain a constant temperature setting for most of the time, 2: Adjust the temperature manually at night or when the residence is vacant, 3: Set up the thermostat to modify the temperature automatically at certain times during day and night, and 4: Operate the AC unit as required

Table 5. 12. Mean indoor temperature for different AC usage behavior

Code	Description (AC usage behavior)	Mean temperature	
1	Maintain a constant temperature setting for most of the time	Day	22.532 (SD=2.492)
		Night	22.307 (SD = 2.526)
2	Adjust the temperature manually at night or when the residence is vacant	Day	22.842 (SD=2.317)
		Night	22.273 (SD=2.503)
3	Set up the thermostat to modify the temperature automatically at certain times during day	Day	23.217 (SD=2.094)
		Night	22.716 (SD=2.451)
4	Operate the AC unit as required	Day	23.005 (SD = 2.750)
		Night	22.300 (SD=2.915)

Table 5. 13. Pairwise comparison for different AC usage behavior

Pairwise comparisons		
Code	**Test statistic**	**Significance**
1-2	55.870	$p > 0.05$
1-3	342.637	$p < 0.05$
1-4	574.276	$p < 0.05$
2-3	286.767	$p < 0.05$
2-4	518.406	$p < 0.05$
3-4	231.639	$p < 0.05$

1: Maintain a constant temperature setting for most of the time, 2: Adjust the temperature manually at night or when the residence is vacant, 3: Set up the thermostat to modify the temperature automatically at certain times during day and night, and 4: Operate the AC unit as required

5.3 The causal effects of other variables in the DAG

Table 5.14 delineates the effects of transitioning among various International Energy Conservation Code (IECC) climate zones on EUI. These shifts in climate zones, which are determined by unique regional weather characteristics, largely depict a decrease in energy usage per square meter. The transition from a hot-dry climate (2B) to a warm-humid one (3A) leads to an EUI reduction of 8.865 kWh/m^2, implying energy conservation due to climatic conditions. Similarly, the transition from 3A to mixed-dry (3B) to mixed-humid (4B), and then to a marine climate (3C), results in reductions of 8.118 kWh/m^2 and 6.854 kWh/m^2 respectively, underscoring the benefits of moderate weather conditions in these zones. However, an exception is observed when transitioning from the marine climate (3C) to a mixed-humid climate (4A), which surprisingly shows an increase in EUI by 10.049 kWh/m^2, potentially due to greater humidity or temperature extremes. The trend reverts when moving to a marine west coast climate (4C), causing a reduction of 8.066 kWh/m^2. The transition to a cool, moist continental climate (5A) marginally increases EUI by 2.325 kWh/m^2, potentially due to increased heating needs. Lastly, moving from 5A to 5B-5C, both characterized as cool and dry, results in a slight reduction in EUI, amounting to -1.771 kWh/m^2, reflecting the impact of cooler, drier climates on reduced energy usage.

Table 5. 14. Causal effects of International Energy Conservation Code on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 2B to 3A	-8.865	-8.865	-8.904	0	-8.865
From 3A to 3B-4B	-8.118	-8.118	-8.128	0	-8.118
From 3B-4B to 3C	-6.854	-6.854	-6.881	0	-6.854
From 3C to 4A	10.049	10.049	10.0749	0	10.048
From 4A to 4C	-8.066	-8.066	-8.079	0	-8.066
From 4C to 5A	2.325	2.325	2.346	0	2.325
From 5A to 5B-5C	-1.771	-1.771	-1.759	0	-1.771

1A: Hot-Humid, 1B: Hot-Dry, 2A: Warm-Humid, 2B: Warm-Dry, 3A: Mixed-Humid, 3B: Warm-Dry, 3C: Mixed-Marine, 4A: Mixed-Humid, 4B: Mixed-Dry, 4C: Mixed-Marine,5A: Cool-Humid, 5B: Cool-Dry, 5C: Cool-Marine, 6A: Cold-Humid, 6B: Cold-Dry, 7: Very Cold, 8: Subarctic

Table 5.15 elucidates the impact of transitioning between different housing types on EUI. Interestingly, no change in EUI is observed when moving from a single-family detached house to a single-family attached house, indicating that the shared walls in attached housing do not significantly impact energy consumption. On the other hand, a transition from a single-family detached house to an apartment in a building with 5 or more units leads to a decrease in EUI by 4.442 kWh/m^2. A smaller reduction of EUI, specifically -0.824 kWh/m^2, is noted when transitioning from a single-family attached house to an apartment in a building with 5 or more units. Lastly, transitioning from an apartment in a building with 2-4 units to an apartment in a building with 5 or more units results in a reduction in EUI of 3.316 kWh/m^2.

Table 5. 15. Causal effects of housing type on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	0	0	0	0	-0.008
From 1 to 3	0	0	0	0	-2.014
From 1 to 4	-4.442	-4.442	-4.433	0	-4.442
From 2 to 3	0	0	0	0	-2.001
From 2 to 4	-0.824	-0.824	-0.816	0	-0.823
From 3 to 4	-3.316	-3.316	-3.316	0	-3.316

1: Single-family detached house , 2: Single-family attached house, 3: Apartment in a building with 2 to 4 units, 4: Apartment in a building with 5 or more units

Table 5.16 presents the implications of various AC system types on EUI. Transitioning from a central air conditioning system to individual window, wall, or portable units results in an increase in EUI by 11.858 kWh/m^2. This suggests that centralized AC systems tend to be more energy-efficient due to their integrated nature and potential for energy recovery, in contrast to standalone units which may incur higher energy consumption due to their localized cooling and potential inefficiency. However, a shift from a central AC system to a setup that utilizes both a central system and individual units only marginally increases the EUI by 1.470 kWh/m^2, indicating that the use of individual units in conjunction with central AC only slightly impacts energy efficiency, possibly due to the mitigating effect of the central system's efficiency. Conversely, moving from individual window, wall, or portable units to a combined system results in a decrease of EUI by 7.171 kWh/m^2, underscoring the energy-saving potential of integrating central systems into environments initially cooled by individual units.

Table 5. 16. Causal effects of AC type on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	11.858	11.858	11.861	0	11.858
From 1 to 3	1.470	1.470	1.460	0	1.470
From 2 to 3	-7.171	-7.171	-7.197	0	-7.171

1: Central air conditioning system, 2: Individual window/wall or portable units, 3: Both a central system and individual units

Table 5.17 illustrates the impacts of transitioning between different major wall materials on EUI, factoring in their respective thermal resistance values (see Table 5.18), or R-values. The data indicate a nominal increase in energy usage (0.075 kWh/m^2) when transitioning from brick (R-value 0.14089 $m^2 \cdot K/W$) to wood (R-value 0.11976 $m^2 \cdot K/W$), likely due to wood's lower thermal resistance. Conversely, a shift from brick to siding (R-value 0.31701 $m^2 \cdot K/W$), which offers higher thermal resistance, results in a substantial decrease in EUI (-3.628 kWh/m^2), highlighting the insulation benefits of siding. When transitioning from brick or wood to concrete or concrete block (R-value 0.07749 $m^2 \cdot K/W$), the EUI significantly increases, suggesting that despite its durability, concrete's lower thermal resistance contributes to higher energy consumption. A shift from wood to siding yields similar results as observed with brick. The changes in EUI for transitions from siding to other materials, including stucco (R-value 0.009 $m^2 \cdot K/W$), shingle (composition, R-value 0.07749 $m^2 \cdot K/W$), stone (R-value 0.01408 $m^2 \cdot K/W$), and concrete or concrete block, further underline the insulative benefits of siding, given that all these transitions result in increases in EUI. Lastly, transitioning from stucco to shingle (composition) or concrete block results in increased EUI, reflective of the exceptionally low thermal resistance of stucco. These results, in essence, underscore the critical role of a material's thermal resistance in determining the energy efficiency of a building.

Table 5. 17. Causal effects of major wall materials on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	0.075	0.075	0.079	0	0.075
From 1 to 3	-3.628	-3.628	-3.630	0	-3.628
From 1 to 4	0	0	0	0	-1.001
From 1 to 5	1.112	1.112	1.131	0	2.112
From 1 to 6	0	0	0	0	0.579
From 1 to 7	11.768	11.768	11.792	0	11.768
From 1 to 9	-0.681	-0.683	0	0	-0.657
From 2 to 3	-3.694	-3.694	-3.681	0	-3.693
From 2 to 4	-0.408	-0.408	-0.437	0	-0.108
From 2 to 5	0	1.340	0	0	1.781
From 2 to 6	0.411	0.416	0.280	0	0.462
From 2 to 7	11.845	11.845	11.907	0	11.845
From 2 to 9	0	0	0	0	-1.001
From 3 to 4	2.456	2.456	2.466	0	2.456
From 3 to 5	6.015	6.015	6.000	0	6.015
From 3 to 6	3.079	3.079	3.043	0	3.079
From 3 to 7	16.476	16.476	16.487	0	16.476
From 3 to 9	0	0	0	0	0
From 4 to 5	1.670	1.670	1.666	0	1.670
From 4 to 6	0	0	0	0	0.200
From 4 to 7	11.912	11.912	11.928	0	11.912
From 4 to 9	-1.433	-1.843	-1.452	0	-2.433
From 5 to 6	0	0.673	0	0	0
From 5 to 7	6.737	5.612	6.789	0	6.737
From 5 to 9	-1.975	-2.915	-1.942	0	-1.975
From 6 to 7	8.893	8.893	8.885	0	2.893
From 6 to 9	0	0	0.450	0	0
From 7 to 9	-6.079	-6.086	-6.078	0	-4.889

1: Brick, 2: Wood, 3: Siding, 4: Stucco, 5: Shingle (composition), 6: Stone, 7: Concrete or concrete block, 9: Other

Table 5. 18. Thermal resistance of major wall materials

Code	**Description (wall material)**	**R value ($m^2 \cdot K/W$)**
1	Brick	0.14089
2	Wood	0.11976
3	Siding	0.31701
4	Stucco	0.009
5	Shingle (composition)	0.07749
6	Stone	0.01408
7	Concrete or concrete block	0.07749
9	Other	-

The results in Table 5.19 depict a multifaceted interplay between various structural, environmental, and demographic variables with energy use intensity. With an increase in Cooling Degree Days (CDD65) by 500 degree days, EUI shows an upward trend, increasing by 3.854 kWh/m^2, illustrating the demand for more cooling energy with higher external temperature. Conversely, an increment of 0.1 in the weather and shielding factor results in a reduction of 8.116 kWh/m^2, indicating the effectiveness of external factors in mitigating energy consumption. The number of occupants in the dwelling has no observable impact on energy usage, while an increment of $10,000 in annual income decreases energy use intensity by 3.43 kWh/m^2, possibly due to large homes or energy-conscious practices associated with higher income brackets. The presence of a thermostat further exemplifies this, leading to a decrease of 17.620 kWh/m^2, underscoring its role in regulating and conserving energy. Incremental changes in building characteristics such as the number of windows, the year built, the number of rooms, the total area, and the cooled area all indicate reductions in EUI. Lastly, the latent heat infiltration is associated with an increase in energy usage by 1.29 kWh/m^2, indicating that heat infiltration can contribute to higher energy consumption.

Table 5. 19. Causal effects of other variables on EUI

Variable (increment)	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
CDD65 (500 DD)	3.854	3.854	3.854	0	3.854
Weather and shielding factor (0.1)	- 8.116	-8.116	-8.101	0	-8.116
Number of people (2)	0	0	0	0	0
Annual income (10k)	-3.430	-3.430	-3.442	0	-3.430
Thermostat present	-17.620	-17.620	-17.630	0	-17.620
Number of windows (3)	-0.855	-0.855	-0.852	0	-0.855
Year built range (10)	-1.378	-1.378	-1.377	0	-1.378
Number of rooms (2)	-1.730	-1.730	-1.725	0	-1.730
Total area (97)	-3.374	-3.374	-3.371	0	-3.373
Total cooled area (93)	-4.483	-4.483	-4.483	0	-4.483
Latent heat infiltration (1)	1.290	1.290	1.290	0	1.290

5.4 Summary

This chapter has showcased the implementation of the proposed ML-based causal inference framework to assess and quantify the effects of energy-saving policies and occupant behavior on cooling energy consumption. The derived causal inference results offer insights into the research questions posed in this chapter and underscore their implication within the built environment.

5.4.1 Are there causal relationships between energy-saving policies and occupant behavior and their impact on EUI in buildings?

Yes, the implemented causal inference framework, utilizing DML, has successfully ascertained causal relationships between energy saving policies and cooling EUI. The distinct advantage of this approach lies in its ability to disentangle and assess complex interactions, thereby identifying causality with precision. The determined causal associations considered a comprehensive range of determinants including climatic factors, building characteristics, occupant behavior, and socio-economic factors. Furthermore, this approach has leveraged expert domain knowledge, which was integrated via the use of DAG. Thus, the adopted DML framework provides a robust and comprehensive method for discerning the effects of energy saving policies and occupant behavior on cooling EUI, demonstrating its potential for aiding in the development of effective energy conservation strategies.

5.4.2 If such relationships exist, can these causal relationships be quantified?

Indeed, this dissertation has not only identified, but also quantified the causal relationships between energy-saving policies, occupant behavior, and EUI. In relation to energy-saving policies, the efficacy of the energy audit program is evident. Specifically, for buildings that participated in energy audits, the use of smart meters didn't contribute to an increase in EUI, in contrast to the outcome observed in buildings that didn't undergo such audits (an increase of 2.543 kWh/m^2). Moreover, in audited buildings, the provision of interval data access showed a substantial reducing effect on EUI, lowering it by -7.035 kWh/m^2. Energy Star qualified windows, too, had a significant impact, resulting in a decrease of -2.260 kWh/m^2 in EUI, albeit this reduction was less than that achieved in non-audited buildings (-4.108 kWh/m^2). The effectiveness of building insulation policy is also manifest, with well-insulated buildings exhibiting lower EUI compared to their poorly insulated counterparts.

Moving onto occupant behavior, the dissertation has quantified its impact, particularly in relation to AC usage and temperature settings. The study revealed that setting the nighttime temperature higher than the daytime temperature led to a reduction in EUI, an effect likely tied to the resulting temperature gradient. Furthermore, the practice of turning the AC on and off as needed led to the most substantial EUI savings, whereas maintaining a consistent AC setpoint temperature most of the time was associated with an increase in EUI. Collectively, these findings underscore the significant roles of both policy measures and occupant behaviors in determining energy use and efficiency.

5.4.3 What is the effectiveness of existing energy-saving policies on building EUI, what new policies could be proposed for further EUI reduction, and what role does occupant behavior play?

This dissertation establishes the effectiveness of the energy audit policy in influencing the use of smart meters, facilitating access to interval data, and promoting the implementation of Energy Star qualified windows. Additionally, the building insulation policy has demonstrated success as well-insulated buildings have exhibited a decrease in EUI. However, it's important to note that some policies, such as energy assistance, don't appear to affect EUI. It is surmised that the primary objective of such programs might not be to reduce EUI but rather to aid households in managing their electricity bills and fixing malfunctioning equipment.

Given these findings, there are several potential policy directions for furthering energy efficiency. The first approach involves mandating the use of Energy Star qualified windows in both new and existing buildings, with government assistance provided for their installation. This policy could foster the adoption of energy-efficient windows, which have been demonstrated to significantly reduce EUI, resulting in substantial energy savings.
The second policy recommendation proposes government assistance for landlords to incorporate energy-efficient features into their buildings. This support would mitigate the need for rent increases, thereby incentivizing tenants to choose and remain in these energy-efficient properties.

The third approach calls for a focus on occupant behavior, specifically regarding AC usage.

Increased public awareness about optimal AC operation is necessary. By educating households about the ineffectiveness of maintaining a constant AC setpoint temperature, they could be encouraged to adopt more energy-efficient habits, such as turning the AC on only as needed, or, alternatively, programming the thermostat to make automatic temperature adjustments at specific times. Through these proposed policy measures, we could significantly enhance our energy-saving strategies and make strides towards a more sustainable future.

6 The role of socio-economic factors and heating equipment selection in energy consumption

6.1 The role of occupant socio-economic factors

6.1.1 Age, education level, average income, and number of household members vs EUI

Table 6.1 provides a comprehensive analysis of the causal implications of various socio-economic attributes of the occupants on the heating EUI. It is observed that an increment of two in the number of household members leads to a decrease in heating EUI by 9.808 kWh/m^2. This suggests a more efficient utilization of heating energy per person in larger households, possibly due to the shared heating environment. The age of the household head, on the other hand, shows a positive relationship with heating EUI, with every decade of increase leading to a rise of 10.208 kWh/m^2. This could be indicative of older household heads' preferences for higher indoor temperatures, or the ownership of older, less energy-efficient heating appliances. The education level of the household head inversely influences heating EUI, decreasing it by 18.012 kWh/m^2, likely due to better knowledge and application of energy-saving practices among more educated individuals. Moreover, a two million increment in the average monthly income results in a reduction of 18.865 kWh/m^2 in heating EUI, which might be a reflection of the financial capacity to purchase and maintain more energy-efficient heating systems.

A detailed analysis to find whether the floor area had any implication in these results showed that indeed the age of the head of the household and education level contribute to EUI increment and decrement, respectively as their relationships with the floor area were not causal. These results are presented in Table 6.2, and the grey highlighted numbers indicate non-causal relationships.

Table 6. 1. Causal effects of age, education level, average income, and number of household members on EUI

Variable (unit increment)	**Estimated average change in EUI**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Number of household members (2)	-9.808	-9.808	-9.801	0	-9.808
Head of the household age (10)	10.208	10.208	10.207	0	10.208
Head of the household education level	-18.012	-18.023	-18.011	0	-18.011
Average monthly income (2M)	-18.865	-18.863	-18.857	0	-18.332

Table 6. 2. Causal effects of Age, education level, average income, and number of household members on energy consumption and floor area

Variable (unit increment)	**Average change in EUI (kWh/m^2)**	**Average change in energy (kWh)**	**Average change in floor area (m^2)**
Number of household members (2)	-9.808	468.772	10.671
Head of the household age (10)	10.208	52.142	-2.998
Head of the household education level	-18.012	75.512	11.567
Average monthly income (2M)	-18.865	775.425	11.777

6.1.2 Household main source of income vs EUI

Table 6.3 provides an analytical exploration of the causal effects of changes in the primary sources of household income on heating EUI. A shift in the main income source from work or occupation, or one's own or their spouse's employment, to assistance from state and local governments causes an increase in heating EUI by 48.124 kWh/m^2. Similarly, moving from deposit/installment savings to governmental assistance results in a substantial rise in heating EUI of 74.602 kWh/m^2. A change from a public pension to state and local government assistance also leads to an increase in heating EUI, specifically by 54.842 kWh/m^2. Lastly, if the primary income source changes from children's support to governmental assistance, the heating EUI sees an increase of 62.205 kWh/m^2. Furthermore, it is observed that this increase in EUI was due to the reduction in floor area, as shown except from children to assistance from the state and local governments. Table 6.4 details these results, and the highlighted gray numbers represent non-causal relationships.

These findings suggest a significant increase in heating EUI when the main source of income shifts to assistance from state and local governments, regardless of the previous income source. This could potentially be attributed to less disposable income for investing in energy-efficient heating solutions or less agency over housing conditions, implying that energy conservation strategies and financial assistance programs need to be appropriately tailored for households predominantly reliant on government assistance.

Table 6. 3. Causal effects of household main source of income on EUI

Cases	Average change in EUI (kWh/m^2)	Average change in energy (kWh)
From 1 to 6	48.124	0
From 2 to 6	74.602	52.791
From 3 to 6	54.842	-25.424
From 5 to 6	62.205	775.425

1: Your/spouse's work or occupation, 2: deposit, installment savings, 3: public pension, 4: real estate, 5: children, 6: Assistance from the state and local governments

Table 6. 4. Detailed Causal effects of household main source of income on EUI

	Estimated average change in EUI	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	-8.089	-8.020	-7.469	0	-5.414
From 1 to 3	13.815	13.815	13.824	0	13.824
From 1 to 4	0	0	0	0	0
From 1 to 5	0	0	0	0	0
From 1 to 6	48.124	48.172	47.804	0	48.134
From 2 to 3	14.158	14.175	18.082	0	16.019
From 2 to 4	16.119	16.050	15.097	0	-0.739
From 2 to 5	0	0	0	0	0
From 2 to 6	74.602	74.693	73.981	0	74.652
From 3 to 4	0	0	0	0	0
From 3 to 5	0	0	0	0	0
From 3 to 6	54.842	54.966	54.798	0	54.751
From 4 to 5	-0.449	-0.386	-0.506	0	-1.295
From 4 to 6	85.662	85.772	84.587	0	85.683
From 5 to 6	62.205	62.244	62.639	0	62.639

1: Your/spouse's work or occupation, 2: deposit, installment savings, 3: public pension, 4: real estate, 5: children, 6: Assistance from the state and local governments

6.1.3 The need to integrate socio-economic factors in energy saving policies

There is a pressing need for government entities to consider socio-economic determinants in the formulation of energy-saving policies. Currently, energy policies largely concentrate on the physical aspects of buildings, emphasizing thermal performance. Nevertheless, evidence from Table 6.5 reveals the causal effects of the construction year of a building on its EUI, a reflection of the applied U-value policy. Data illustrates that buildings constructed more recently exhibit a low EUI, demonstrating the effectiveness of contemporary thermal regulations. However, Figure 6.1 indicates a plateau in the U-value, suggesting a limit to the benefits obtainable solely through improving building structures and thermal performance. Given these findings, this dissertation proposes that energy policies should be expanded beyond the physical domain of buildings. Specifically, policymakers should account for the socio-economic characteristics of occupants, which, as the previous analyses indicate, significantly influence energy consumption patterns.

This broader, more holistic approach could further improve the effectiveness of energy-saving policies and interventions, thereby contributing more energy substantially goals.

Table 6. 5. Causal effects of house completion year on EUI

Cases	Average change in EUI (kWh/m^2)
From 1 to 2	-1.825
From 2 to 3	-37.537
From 3 to 4	-19.602
From 4 to 5	0
From 5 to 6	-7.295

1: Before 1969, 2: 1970-1979, 3: 1980-1989, 4: 1990-1999, 5: 2000-2009, and 6: After 2010

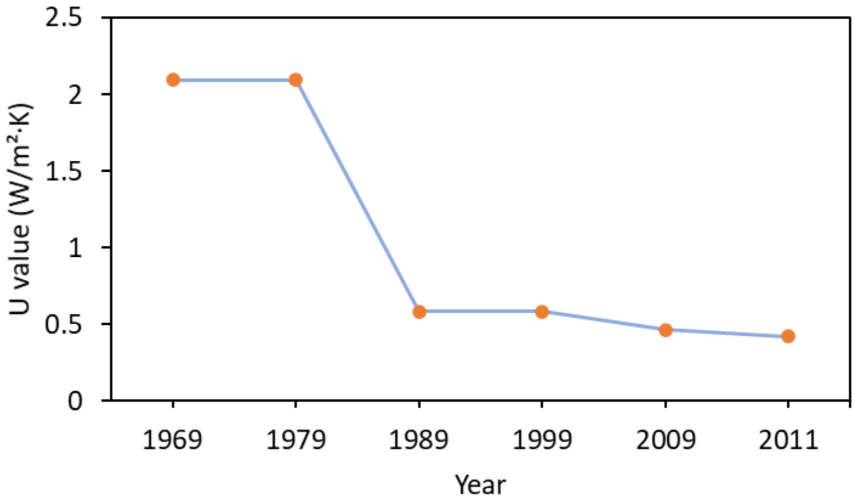


Figure 6. 1. Thermal transmittance (U-value) over the years

6.2 The impact of heating equipment selection on EUI

6.2.1 Chosen heating equipment vs EUI

Table 6.6 provides an intriguing analysis of the causal effects of shifts in primary heating equipment on the heating EUI. Transitioning from district heating to central heating with city gas leads to a decrease in heating EUI by 20 kWh/m^2, suggesting superior energy efficiency of city gas central heating systems. This does not align with the findings by (Lund et al., 2010) that find that it was beneficial to transition to district heating. However, unlike the EnergyPlan simulation model used by (Lund et al., 2010), which is based on associations/correlations, our causal inference approach provides more accurate results. Switching from central heating with city gas to individual heating systems manifests varied outcomes.

There is no change in heating EUI when shifting to a kerosene boiler, but a considerable surge of 365.742 kWh/m^2 occurs with a briquette boiler. Within individual heating systems, transitions from a kerosene boiler to other systems mostly result in no change, except for a modest decrease of 6.161 kWh/m^2 with a city gas boiler. Similarly, changes from a propane gas boiler to other individual heating systems present varied effects, with an increase of 48.551 kWh/m^2 to a city gas boiler and a considerable rise of 393.87 kWh/m^2 to a briquette boiler, while no change is seen with electric boilers or electric blankets.

Lastly, a transition from individual heating with a city gas boiler to individual heating with a briquette boiler leads to a significant surge in heating EUI by 339.859 kWh/m^2. Table 6.7 shows the results for pairwise statistical tests. It is evident that post non-causal and null relationships are present where there was no significant statistical relationship between the groups. This analysis reveals that the type of heating equipment significantly influences the heating EUI, with changes to less efficient systems like briquette boilers resulting in drastic increases in energy use, suggesting a need for focused policy interventions and consumer education around heating system choices.

Table 6. 6. Causal effects of the main heating equipment on EUI

	Estimated average change in EUI	Average change in energy (kWh)	Refutation			
			Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	-20.605	-1270.830	-20.612	-20.665	0	-20.606
From 1 to 3	0	0	0	0	0	0
From 1 to 4	-47.276	-4013.916	-15.276	-47.379	0	-32.244
From 1 to 5	0	-927.998	0	0	0	0
From 1 to 6	0	-142.034	0	0	0	-2.637
From 1 to 7	343.540	8133.400	343.565	173.331	0	247.561
From 1 to 8	-0.003	-2125.465	-0.003	-0.003	0	-0.003
From 2 to 3	0	0	0	0	0	0
From 2 to 4	-24.198	-2844.295	-24.122	-24.124	0	-36.900
From 2 to 5	22.237	0	22.238	14.455	0	20.249
From 2 to 6	0	0	0	9.021	0	8.149
From 2 to 7	365.742	9568.751	365.778	365.337	0	365.786
From 2 to 8	9.640	-1079.923	9.939	10.677	0	3.052
From 3 to 4	0	-2304.479	0	0	0	0
From 3 to 5	-6.161	0	-6.161	-6.143	0	-6.165
From 3 to 6	0	0	0	0	0	0
From 3 to 7	190.877	6053.872	190.855	184.655	0	196.290
From 3 to 8	0	0	0	0	0	0
From 4 to 5	48.551	3164.935	48.551	48.533	0	48.547
From 4 to 6	0	554.265	0	0	0	0
From 4 to 7	393.870	11266.683	393.891	393.350	0	393.905
From 4 to 8	0	1766.403	0	0	0	0
From 5 to 6	0	0	0	5.904	0	7.851
From 5 to 7	339.859	9182.202	339.851	338.852	0	339.994
From 5 to 8	0	-391.018	0	0	0	0
From 6 to 7	49.975	1857.824	50.025	50.533	0	63.112
From 6 to 8	0	0	0	0	2.980	0
From 7 to 8	-2.018	-106.145	-2.023	-2.066	0	-84.432

1: District heating, 2: Central heating with city gas, 3: Individual heating with kerosene boiler, 4: Individual heating with propane gas boiler, 5: Individual heating with city gas boiler, 6: Individual heating with electric boiler, 7: Individual heating with briquette boiler, 8: Individual heating with electric blanket

Table 6. 7. Pairwise statistical tests

	Test statistic	**Significance**
1 vs 2	7.054	$p < 0.05$
1 vs 3	36.463	$p > 0.05$
1 vs 4	31.117	$p < 0.05$
1 vs 5	2.380	$p > 0.05$
1 vs 6	1.285	$p > 0.05$
1 vs 7	43.522	$p > 0.05$
1 vs 8	2.280	$p > 0.05$
2 vs 3	25.337	$p > 0.05$
2 vs 4	10.431	$p > 0.05$
2 vs 5	5.915	$p > 0.05$
2 vs 6	2.747	$p > 0.05$
2 vs 7	50.495	$p < 0.05$
2 vs 8	3.867	$p > 0.05$
3 vs 4	38.578	$p < 0.05$
3 vs 5	60.211	$p < 0.05$
3 vs 6	9.458	$p > 0.05$
3 vs 7	38.765	$p > 0.05$
3 vs 8	25.743	$p < 0.05$
4 vs 5	25.167	$p < 0.05$
4 vs 6	14.929	$p < 0.05$
4 vs 7	65.855	$p < 0.05$
4 vs 8	22.354	$p < 0.05$
5 vs 6	0.514	$p > 0.05$
5 vs 7	35.994	$p < 0.05$
5 vs 8	1.002	$p > 0.05$
6 vs 7	52.508	$p > 0.05$
6 vs 8	0	$p > 0.05$
7 vs 8	46.995	$p < 0.05$

1: District heating, 2: Central heating with city gas, 3: Individual heating with kerosene boiler, 4: Individual heating with propane gas boiler, 5: Individual heating with city gas boiler, 6: Individual heating with electric boiler, 7: Individual heating with briquette boiler, 8: Individual heating with electric blanket

6.2.2 The effect of auxiliary heating on EUI

Table 6.8 provides a critical examination of the causal impacts of the usage of auxiliary heating equipment alongside main heating equipment on the EUI and the highlighted gray numbers represent non-causal effects. The utilization of auxiliary equipment with district heating leads to an increase in heating EUI of 4.244 kWh/m^2, indicating that additional energy consumption occurs when supplementary heating devices are used in tandem with district heating systems. In contrast, employing auxiliary heating equipment with an individual heating system that utilizes a propane gas boiler result in a reduction in heating EUI by 7.366 kWh/m^2, suggesting potential energy-saving synergies. However, a similar combination with an individual heating system using a city gas boiler leads to an increase in heating EUI by 7.389 kWh/m^2. Lastly, the employment of auxiliary equipment in conjunction with an individual heating system using an electric boiler causes a modest increase of 3.784 kWh/m^2 in heating EUI. These observations imply that the impact of auxiliary heating equipment on heating EUI is contingent on the type of primary heating system, with potential implications for energy efficiency strategies and the design of energy policy interventions.

Table 6. 8. Causal effects of using auxiliary heating

Main heating equipment	**Estimated average change in EUI (kWh/m^2) due to auxiliary heating**	**Estimated average change in energy consumption (kWh) due to auxiliary heating**
1	4.244	-15.584
2	-1.220	-49.278
3	-2.602	-105.057
4	-7.366	-376.611
5	7.389	0
6	3.784	334.502
7	0	0
8	-51.117	0

1: District heating, 2: Central heating with city gas, 3: Individual heating with kerosene boiler, 4: Individual heating with propane gas boiler, 5: Individual heating with city gas boiler, 6: Individual heating with electric boiler, 7: Individual heating with briquette boiler, 8: Individual heating with electric blanket

6.3 The causal effects of other variables in the DAG

Table 6.9 elucidates the impacts of transitioning between different provincial climates on EUI. Moving from the climate of Seoul to the climate of Gwangju leads to a significant decrease in

EUI, specifically by -26.668 kWh/m^2. The transition from Seoul to Daejeon results in no change in EUI, indicating a comparable energy profile among these locations, possibly due to similar temperature ranges or building practices. Conversely, the transition from Seoul to Gangwon results in a substantial increase in EUI by 81.77 kWh/m^2. Transitions from Seoul to Jeonbuk and Jeonnam, both situated in the southwestern part of the country, lead to relatively small increases in EUI, reflecting the slightly different climatic conditions of these regions compared to Seoul. Overall, these findings underscore the influence of regional climates on energy use, reflecting the variable demands for heating across different geographical locations.

Table 6.10 delineates the effect of transitioning between different housing types on EUI. Transitioning from general detached houses, which are individual structures with potentially more exposed walls to the outside environment, to row houses leads to a significant reduction in EUI by -48.498 kWh/m^2. This suggests that the shared walls in row houses may provide better insulation and hence result in lower energy usage. A further decrease in EUI, by -55.003 kWh/m^2, is observed when transitioning from general detached houses to apartments. The most substantial reduction in EUI is observed when transitioning from general detached houses to officetels (a form of studio apartment popular in South Korea), with a decrease of -102.237 kWh/m^2. Lastly, moving from multi-household housing to apartments results in a decrease in EUI by -24.745 kWh/m^2. These findings underline the role of building type and configuration in determining energy use profiles.

Table 6. 9. Causal effects of provincial climate on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
Seoul vs Busan	0	2.320	1.209	0	-0.231
Seoul vs Daegu	0.001	0.001	0.001	0	2.101
Seoul vs Incheon	0	0	0	0	0
Seoul vs Gwangju	-26.668	-26.665	-26.677	0	-26.667
Seoul vs Daejeon	0	0	0	0	0
Seoul vs Ulsan	0	0	0	0	-2.022
Seoul vs Gyeonggi	0	1.023	0	0	1.111
Seoul vs Gangwon	81.770	81.764	81.774	0	81.774
Seoul vs Chungbuk	0.3	0.3	0.3	0	0.3
Seoul vs Chungnam	0	0	0	0	1.201
Seoul vs Jeonbuk	8.996	8.996	0.192	0	2.992
Seoul vs Jeonnam	8.740	8.742	8.755	0	8.749
Seoul vs Gyeongbuk	-0.137	-0.138	-0.137	0	-2.137
Seoul vs Gyeongnam	0.051	0.051	0.053	0	1050
Seoul vs Jeju	0	0	0	0	0

Table 6. 10. Causal effects of housing type on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	-39.750	-39.752	-39.758	0	-39.758
From 1 to 3	-41.424	-41.425	-41.735	0	-24.598
From 1 to 4	-48.498	-48.521	-48.930	0	-50.013
From 1 to 5	-30.250	-30.249	-29.993	0	-30.248
From 1 to 6	-55.003	-55.009	-55.016	0	-54.916
From 1 to 7	0	0	0	0	0
From 1 to 8	-102.237	-102.282	-105.842	0	-106.159
From 2 to 3	0	0	0	0	0
From 2 to 4	0	0	0	0	0
From 2 to 5	8.138	8.125	7.900	0	2.902
From 2 to 6	-23.106	-23.128	-23.289	0	-24.066
From 2 to 7	33.797	33.837	31.439	0	26.874
From 2 to 8	-48.673	-48.514	-51.698	0	-45.501
From 3 to 4	0	0	0	0	0
From 3 to 5	0	0	0	0	0
From 3 to 6	0	0	0	0	0
From 3 to 7	0	0	0	0	0
From 3 to 8	-45.669	-45.569	-45.483	0	-19.799
From 4 to 5	16.952	16.946	16.986	0	16.976
From 4 to 6	0	0	0	0	0
From 4 to 7	27.063	21.667	0	0	15.004
From 4 to 8	-39.439	-39.578	-38.881	0	-23.949
From 5 to 6	-24.745	-24.740	-24.700	0	-24.376
From 5 to 7	0	0	0	0	0
From 5 to 8	-81.654	-81.571	-77.350	0	-50.266
From 6 to 7	34.810	34.792	34.792	0	35.309
From 6 to 8	-41.488	-41.618	-38.997	0	-38.587
From 7 to 8	-63.008	-62.764	-62.605	0	-42.208

1: General detached house, 2: Multi-family detached house, 3: Detached house for commercial use, 4: Row house, 5: Multi-family housing, 6: Apartment, 7: Housing in non-residential buildings, 8: Officetels

Table 6.11 expounds on the impacts of varying the ratio of double-glazed windows to exterior windows on EUI. Transitioning from having no glazing to having less than 25% of windows double-glazed results in a decrease in EUI by -2.819 kWh/m^2. This finding highlights the benefits of double glazing in terms of improved thermal insulation and resultant energy savings, even with a relatively small proportion of double-glazed windows. Interestingly, however, an increase in the double-glazed window ratio beyond this point, whether from no glazing to 25-50%, no glazing to above 50%, less than 25% to 25-50%, less than 25% to above 50%, or 25-50% to above 50%, does not appear to further impact the EUI, as these transitions all result in a 0 kWh/m2 change. This could imply that while the initial introduction of double-glazing can enhance energy efficiency, the benefits may plateau beyond a certain threshold, perhaps due to other factors in the building's construction or use influencing overall energy consumption. These results underline the complex interplay between building features and energy use, and the potential for energy-saving interventions to have diminishing returns in certain contexts.

Table 6. 11. Causal effects the ratio of double-glazed windows to exterior windows on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	-2.819	-2.820	-2.821	0	-2.814
From 1 to 3	0	0	0	0	0
From 1 to 4	0	0	0	0	0
From 2 to 3	0	0	0	0	0
From 2 to 4	0	0	0	0	0
From 3 to 4	0	0	0	0	0

1: Doesn’t exist, 2: less than 25%, 3: 25-50%, 4: Above 50%

Table 6.12 illuminates the relationship between residential floor level and EUI. The transition from lower to higher floors consistently results in a decrease in EUI. Specifically, moving from the first floor to higher floors leads to substantial reductions in EUI, ranging from -34.323 kWh/m^2 for the second floor to -60.263 kWh/m^2 for the eleventh floor and above. Similarly, transitions from the second floor to higher levels result in continued decreases in EUI, albeit at a lesser magnitude, suggesting that the energy efficiency benefits of higher floors continue but at a diminishing rate. The smallest reductions are observed when transitioning between the higher floor ranges (3rd-5th, 6th-10th, 11th and above), indicating the more marginal energy efficiency gains at these levels.

Table 6. 12. Causal effects the residential floor number on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1st floor to 2nd	-34.323	-34.327	-34.295	0	-34.271
From 1st floor to 3rd~5th	-47.622	-47.626	-47.495	0	-46.903
From 1st floor to 6th~10th	-54.583	-54.578	-54.269	0	-54.848
From 1st floor to 11th and above	-60.263	-60.263	-60.025	0	-60.222
From 2nd to 3rd~5th	-15.583	-15.586	-15.348	0	-15.565
From 2nd to 6th~10th	-20.505	-20.527	-20.416	0	-20.574
From 2nd to 11th and above	-28.079	-28.059	-27.992	0	-28.043
From 3rd~5th to 6th~10th	-4.435	-4.432	-4.479	0	-4.324
From 3rd~5th to 11th and above	-11.797	-11.798	-11.689	0	-11.471
From 6th~10th to 11th and above	-7.476	-7.468	-7.460	0	-6.654

Table 6.13 unravels the impacts of changing housing orientation on heating EUI. It is evident that the direction a home faces can significantly affect its heating energy consumption. Transitioning from a West-facing orientation to a South-facing orientation results in a substantial decrease in EUI, specifically by -10.674 kWh/m^2. This result highlights the benefits of southern exposure, which can utilize passive solar gain to help heat a home, reducing the need for additional heating. Similarly, moving from a North-facing orientation to a South West-facing orientation results in a significant decrease in EUI by -11.872 kWh/m^2, underscoring again the energy-saving potential of southern exposure. In contrast, the change in EUI for transitions involving North East, South East, or North West orientations are minimal, indicating these orientations have less of an effect on heating EUI. Interestingly, transitioning from a South East to North East orientation results in a slight increase in EUI by 0.127 kWh/m^2, suggesting a slightly higher heating demand for North East-facing homes. These findings reinforce the importance of building orientation in energy-

efficient design and the potential for strategic orientation to reduce heating energy demands.

Table 6. 13. Causal effects the housing orientation on EUI

Cases	Estimated average change in EUI (kWh/m^2)	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
From 1 to 2	0	0	0	0	0
From 1 to 3	0	0	0	0	0
From 1 to 4	7.591	7.521	7.510	0	0.257
From 1 to 5	0	0	0	0	0
From 1 to 6	0	0	0	0	0
From 1 to 7	0	0	0	0	0
From 1 to 8	-1.518	-1.518	-1.536	0	-1.536
From 2 to 3	-10.674	-10.674	-10.600	0	-10.674
From 2 to 4	-0.099	-0.095	-0.096	0	-0.478
From 2 to 5	-21.361	-21.369	-21.186	0	-16.318
From 2 to 6	2.192	2.194	2.252	0	-4.804
From 2 to 7	-0.173	-0.173	-0.173	0	-0.173
From 2 to 8	0	0	0	0	0
From 3 to 4	0	0	0	0	0
From 3 to 5	-5.444	-5.450	-5.447	0	-5.446
From 3 to 6	0	0	0	0	0
From 3 to 7	10.886	10.856	0	0	-0.041
From 3 to 8	-1.045	-1.044	-1.027	0	-1.041
From 4 to 5	-0.225	-0.224	-0.232	0	-0.229
From 4 to 6	-11.872	-11.864	-11.894	0	-11.869
From 4 to 7	0	0	0	0	0
From 4 to 8	0	0	0	0	0
From 5 to 6	0	0	0	0	0
From 5 to 7	0.127	0.120	0.115	0	0.116
From 5 to 8	0	0	0	0	0
From 6 to 7	6.039	6.007	5.710	0	-1.116
From 6 to 8	-4.587	-4.587	-4.701	0	-2.296
From 7 to 8	0	0	0	0	0

1: East, 2: West, 3: South, 4: North, 5: South East, 6: South West, 7: North East, 8: North West

Table 6.14 elucidates the impacts of various factors on heating EUI. First, an increase in the Heating Degree Days (HDD18.5) by 10-degree days causes an increase in EUI by 8.466 kWh/m^2, underscoring the role of ambient temperature in dictating heating demand. Similarly, the addition of an external wall increases the EUI by 11.75 kWh/m^2, possibly due to the added exposure to outdoor weather conditions, increasing heat loss and consequently heating demand. However, an increase in residential floor area by 139 m^2 (The determination of this range was aimed at obtaining the same number of bins as those used for the average monthly income) is associated with a substantial decrease in EUI by -41.539 kWh/m^2. Furthermore, an increase in the number of heated rooms and bathrooms both correspond to decreases in EUI by -0.204 kWh/m^2 and -3.828 kWh/m^2 respectively. Lastly, each additional external wall window causes a significant decrease in EUI by -11.944 kWh/m^2.

Table 6. 14. Causal effects other variables on EUI

Cases	**Estimated average change in EUI (kWh/m^2)**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
HDD18.5 (10 HDD)	8.466	8.466	8.466	0	8.466
Number of external walls (1)	11.75	11.754	11.745	0	11.787
Residential floor area (139)	-41.539	-41.539	-41.539	0	-41.539
Number of heated rooms (1)	-0.204	-0.204	-0.202	0	-0.202
Number of heated bathrooms (1)	-3.828	-3.828	-3.850	0	-3.826
Number of external wall windows (1)	-11.944	-11.945	-11.949	0	-11.949

6.4 Summary

This chapter has demonstrated the implementation of the proposed ML-based causal inference framework to assess and quantify the role of socio-economic factors and heating equipment selection in energy consumption. The derived causal inference results offer insights into the research questions posed in this chapter and accentuate their significance within the built environment.

6.4.1 Do occupant socio-economic characteristics and the choice of heating equipment serve as causal factors in determining the EUI in buildings?

Occupant socio-economic characteristics and choices in heating equipment indeed serve as causal factors in determining the EUI in buildings. This conclusion has been substantiated by a rigorous causal inference analysis mechanism using DML. The identified causal effects emerge from an intricate interplay of a multitude of variables, including climatic factors, building characteristics, occupant behavior, and socio-economic attributes. These factors have been meticulously analyzed and integrated using directed acyclic graphs. By harnessing expert domain knowledge, the analysis comprehensively illustrates how socio-economic factors of occupants and their choices of heating equipment can influence energy usage patterns in buildings.

6.4.2 If these are indeed causal factors, can their effect be quantified?

Certainly, the causal factors of occupant socio-economic characteristics and the choice of heating equipment on the EUI in buildings can be quantified, providing practical and actionable insights for policy-making and energy efficiency strategies. An in-depth analysis of the primary heating facilities utilized in the dataset demonstrates that a transition from a kerosene boiler to a gas boiler could potentially result in a decrease of EUI by 6.161 kWh/m^2. When taking all equipment into account, using individual heating with either a briquette boiler or an electric blanket as the main heating equipment generally contributes to an increase in EUI. Additionally, the employment of auxiliary heating typically leads to an increase in EUI, with the notable exception of scenarios involving individual heating with a propane gas boiler, where auxiliary heating use results in a decrease in EUI.

The effects of occupant socio-economic characteristics have also been quantified. For instance, the age, education level, and average monthly income of the household occupants have been found to impact EUI by 10.208 kWh/m^2, -18.012 kWh/m^2, and -18.865 kWh/m^2, respectively.

Lastly, households with primary income sources not derived from occupation are typically associated with higher EUI, underscoring the importance of socio-economic considerations in understanding and managing building energy consumption.

6.4.3 What insights can be derived from causal inference analysis regarding the role of socio-economic factors and the adoption heating equipment towards EUI reduction?

Regarding heating equipment, the results strongly suggest that efforts should be made to decrease the usage of individual heating methods such as briquette boilers and electric blankets, which have been found to consume the highest EUI. It would be beneficial for the government to raise awareness about the energy efficiency of individual heating equipment like gas, electric, and propane boilers. Furthermore, central heating systems have been shown to be more advantageous in terms of energy use compared to district heating equipment.

These findings underscore the imperative to consider socio-economic factors in energy-saving policies. The results from the causal inference analysis illuminate the necessity for the government to integrate considerations of occupant socio-economic characteristics into such policies. The pressing need to incorporate these socio-economic factors is evident, as existing energy policies that primarily focus on building characteristics, such as thermal performance, seem to have reached a plateau in their efficacy. In contrast, socio-economic factors like age, education level, and average monthly income have demonstrated promising potential for reducing EUI. This further amplifies the argument for a more inclusive approach in policy formulation, integrating socio-economic considerations into energy-saving strategies.

Furthermore, an important insight derived from the causal results is the necessity for targeted awareness campaigns, especially among households whose main source of income is not derived from the head's or spouse's occupation. These households tend to have a higher EUI consumption, indicating the need for tailored interventions. For instance, the government could focus on raising awareness about the choice of energy-efficient equipment among these households, contributing to more sustainable energy consumption patterns.

7 The effect of occupant personal and behavioral factors on thermal sensation

7.1 Occupant personal factors and thermal sensation

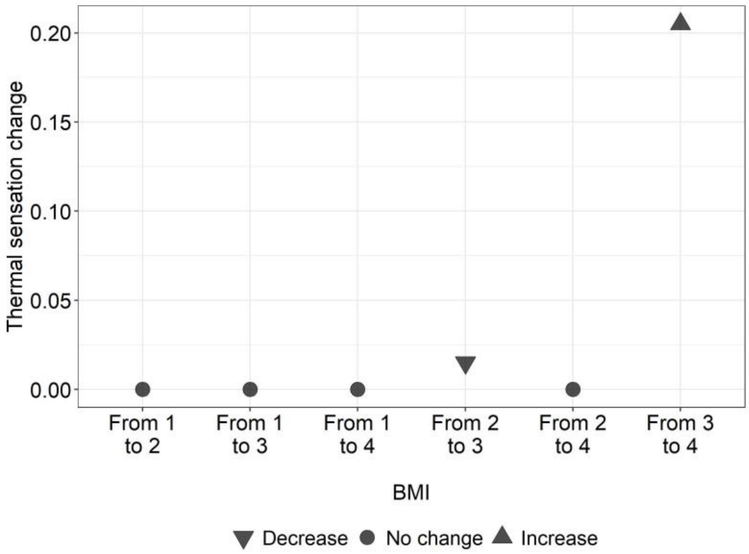
7.1.1 Sex, age, and BMI vs thermal sensation

The analysis of the effects of sex, age, and BMI on thermal sensation reveals interesting patterns. Regarding sex, it appears that males on average possess a thermal sensation that is 0.031 units higher than that of females, suggesting a mild sex difference in thermal perception. Conversely, age does not appear to significantly impact an individual's thermal sensation. This observation holds true even when males and females are considered separately, suggesting a relative consistency in thermal sensation across different age groups. Table 7.1 provide more details.

The relationship between BMI and thermal sensation, as shown in Figure 7.1, indicates that occupants classified as obese have a thermal sensation that is 0.205 units higher than those classified as overweight, echoing the findings of (Ming et al., 2023) who observed similar results among subjects with higher BMIs. Conversely, occupants in the overweight category were found to have a slightly lower thermal sensation, by 0.015 units, compared to those of normal weight. These findings illustrate the multifaceted influences of personal characteristics on thermal sensation, and the need for more nuanced understanding and considerations in designing thermally comfortable environments.

Table 7. 1. Causal effects of sex and age on thermal sensation

Variable (unit increment)	**Thermal sensation change**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Gender (from female to male)	0.031	0.031	0.030	0.031	0.031
Age (10)	0	0	0	0	0
Age (10) in females	0	0	0	0	0
Age (10) in males	0	0	0	0	0



1: Underweight, 2: Normal weight 3: Overweight, 4: Obesity

Figure 7. 1. Causal effects of BMI on thermal sensation

7.1.2 Sex vs thermal sensation: an in-depth causal inference analysis

The causal inference analysis, as represented in Table 3, reveals a nuanced impact of sex on thermal sensation across different age groups, underlining its importance in designing adaptive comfort for indoor spaces such as schools and offices where occupants often belong to the same age bracket. Intriguingly, in the 16-25 years age group, a modest negative effect of -0.058 is observed, signifying that females in this age group may perceive temperatures slightly colder than their male peers, a finding that is consistent with the observations made by (Ming et al., 2023). Conversely, for those in the 26-35 years age group, a positive effect of 0.066 is recorded, suggesting that females in this group might perceive temperatures as slightly warmer compared to males. This reversal in trend illustrates the complex interplay between age, sex, and thermal sensation.

Notably, there are no observed effects in the 36-45 and 46-55 age groups, indicating comparable thermal perceptions between males and females within these brackets. However, the most substantial divergence is witnessed among the 56-65 years age group, where a significant positive effect of 0.253 is observed, suggesting that females in this age group perceive temperatures as considerably warmer than their male counterparts. These findings underscore the need for a more personalized approach to thermal comfort design in indoor spaces, taking into consideration age and sex-specific thermal perceptions.

7.1.3 Age vs thermal sensation: an in-depth causal inference analysis

Table 7.3 illustrates the distinctive causal impact of age, segmented into decadal age groups, on thermal sensation, while differentiating between males and females. The findings for the group aged below 16 years reveal no perceivable impact on thermal sensation for females. In the 16-25 age bracket, a slight decrease in thermal sensation (-0.015) for females contrasts with a moderate increase (0.046) for males. For the 26-35 age demographic, females exhibit no changes, while males exhibit a modest increase (0.026) in thermal sensation. The 36-45 age group findings demonstrate a mild increase for females (0.078), with no effect on males. In both the 46-55 and 56-65 age groups, no discernible changes in thermal sensation for either sex are observed. However, in the 56-65 age bracket, males present a notable decrease (-0.313) in thermal sensation. Lastly, in the above 65 age group, a marginal decrease is seen in thermal sensation for both sexes, with females and males showing -0.015 and -0.025, respectively. These observations resonate with and quantify the ambiguous causal effect of sex on thermal sensation documented in prior research studies (Karjalainen, 2012; Lan et al., 2008; Liu et al., 2018; Wang et al., 2018). This dissertation, therefore, advances our understanding of the intricate interplay between age, sex, and thermal sensation, reinforcing the need for age and sex-specific considerations in creating comfortable thermal environments.

Table 7. 2. Causal effect sex on thermal sensation in different age groups

Age (From male to female)	Thermal sensation change	Refutation			
		Random cause	Unobserved Common Causes	Placebo Treatment	Data subset
All age range	0.031	0.031	0.033	0	0.031
Below 16 (only females)	-	-	-	-	-
Age range: 16~25	-0.058	-0.058	-0.052	0	-0.058
Age range: 26~35	0.066	0.066	0.064	0	0.066
Age range: 36~45	0	0	0	0	0
Age range: 46~55	0	0	0	0	0
Age range: 56~65	0.253	0.253	0.255	0	0.253
Age range: above 65	-0.013	-0.013	-0.011	0	-0.013

Table 7. 3. Casual effect of age on thermal sensation

Age range (2 years increment)	Sex	Thermal sensation change
Below 16	Female	0
Age range: 16~25	Female	-0.015
	Male	0.046
Age range: 26~35	Female	0
	Male	0.026
Age range: 36~45	Female	0.078
	Male	0
Age range: 46~55	Female	0
	Male	0
Age range: 56~65	Female	0
	Male	-0.313
Age range: above 65	Female	-0.015
	Male	-0.025

7.2 Occupant behavior and thermal sensation

7.2.1 The operation of blinds/curtains, fan, window, door, and heater vs thermal sensation

The findings unveiled in Figure 7.2 provide critical insights into the behavioral tendencies of occupants with regard to their thermal comfort. It appears that occupants are inclined to utilize fans when their thermal sensation surpasses their comfort zone by 0.384, exemplifying an adaptive behavior to achieve thermal comfort. A similar pattern is discernible in the act of window opening, wherein occupants resort to this action when their thermal sensation rises 0.078 above their normal comfort level. Significantly, the data shows that the activation of heaters is associated with a thermal sensation that is 0.479 lower than when heaters are not in use, illustrating their essential function in creating a warmer ambiance. In contrast, the opening of doors seems to correspond to a lower thermal sensation by 0.187 compared to when doors are kept closed, implying the potential of door opening to fresh ventilation. These results are detailed in Table 7.4. These findings have profound implications for understanding and predicting occupant behavior relative to thermal sensation management, which may be instrumental in developing efficient and user-friendly thermal regulation strategies.

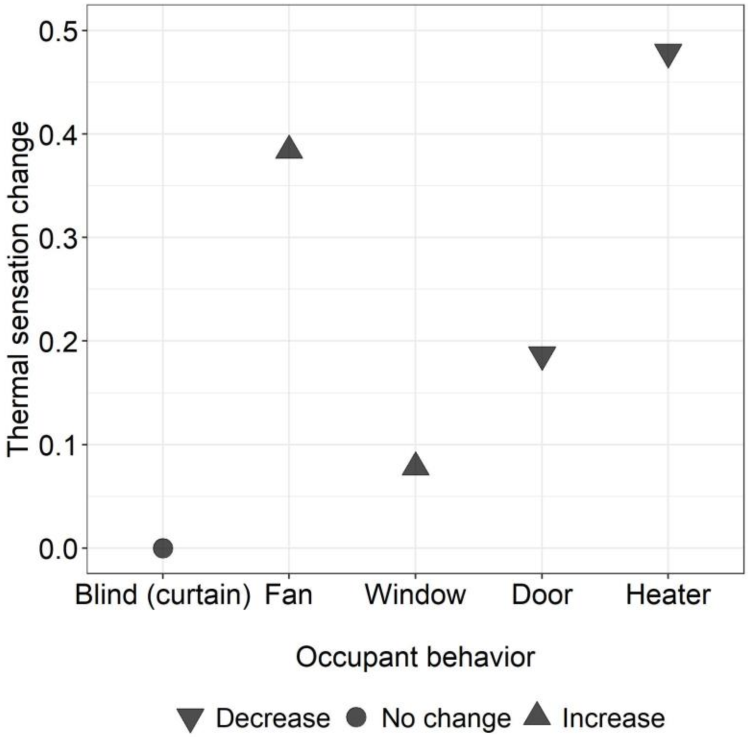


Figure 7. 2. Causal effects of behavioral factors on thermal sensation

Upon a detailed examination of the indoor/outdoor temperature and thermal acceptability data delineated in Table 7.5, this dissertation hypothesizes that occupants opening windows experienced a somewhat warmer thermal environment than those who kept their windows shut. This behavior implies the use of windows as a strategy to introduce cooler external air, potentially in a bid to enhance their thermal comfort. Intriguingly, the case with doors is different. It is observed that occupants who opted to open doors reported lower thermal sensations, suggesting that the primary intent behind this behavior might be to improve indoor air quality rather than temperature control. These interpretations resonate well with the findings from the seminal study by (Kim & de Dear, 2018), wherein respondents were reported to use windows and doors as conduits for facilitating fresh air inflow or creating a cooler indoor environment. These trends reinforce the complex and multi-faceted nature of occupant behavior in thermal comfort management, highlighting the critical role of ventilation through windows and doors.

Table 7. 4. Causal effects of behavioral factors on thermal sensation

Age (From male to female)	**Thermal sensation change**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Blind/curtain	0	0	0	0	0
Fan	0.374	0.374	0.372	0	0.373
Window	0.078	0.078	0.071	0	0.076
Door	-0.187	-0.187	-0.185	0	-0.189
Heater	-0.479	-0.479	-0.480	0	-0.475

Table 7. 5. Occupant behavior and associated thermal acceptability, indoor and outdoor temperatures

	Thermal sensation change	State	Indoor temperature	Outdoor temperature	Thermal acceptability
Blind /curtain	0	Close	24.528 (SD=4.888)	22.039 (SD=4.809)	Acceptable: 70.4 % Unacceptable: 29.6 %
		Open	23.647 (SD=3.091)	22.039 (SD=5.153)	Acceptable: 76.5 % Unacceptable: 23.5%
Fan	0.374	Off	22.090 (SD=5.319)	19.210 (SD=7.449)	Acceptable: 87.9 % Unacceptable: 12.1%
		On	29.897 (SD=2.089)	28.395 (SD=2.624)	Acceptable: 83.8 % Unacceptable: 16.2 %
Window	0.078	Close	27.249 (SD=4.210)	25.376 (SD=5.024)	Acceptable: 85.3 % Unacceptable: 14.7 %
		Open	23.980 (SD=5.071)	22.219 (SD=7.398)	Acceptable: 80.9 % Unacceptable: 19.1 %
Door	-0.187	Close	23.969 (SD=1.945)	21.857 (SD=5.406)	Acceptable: 69.4 % Unacceptable: 30.6 %
		Open	23.719 (SD=1.946)	22.436 (SD=5.068)	Acceptable: 77.4 % Unacceptable: 22.6 %
Heater	-0.479	Off	26.105 (SD=5.277)	23.816 (SD=6.824)	Acceptable: 84.9 % Unacceptable: 15.1 %
		On	19.388 (SD=3.565)	14.789 (SD=4.890)	Acceptable: 72.6 % Unacceptable: 27.4 %

7.2.2 The operation of blinds/curtains, fan, window, door, and heater in relation to thermal sensation across sex categories

In a deeper exploration of the topic, this dissertation segregates and scrutinizes the causal effect of occupant behavior on thermal sensation by gender. From Table 7.6, it is observed that the manipulation of blinds/curtains elicits a notable causal effect on thermal sensation. Both female and male occupants who opted to open their blinds/curtains experienced a significantly lower thermal sensation, by 0.736 and 1.162 respectively, compared to those who left them closed.

Interestingly, gender differences emerge in the reasons for window opening, with females primarily seeking cooler air, whereas males aimed for the introduction of fresh air. These observations reinforce and quantify the insights drawn from the study by (Kim & de Dear, 2018) asserting that occupants with a high degree of perceived control over their environment tend to maintain a comfort temperature that aligns closely with the temperature they experience. This dissertation, therefore, underscores the importance of understanding and integrating occupant behavior in the quest to optimize thermal comfort in residential and office settings.

Table 7. 6. Occupant behavior in relation to thermal sensation across sex categories

	Sex	Thermal sensation change
Blind/curtain	Female	-0.736
	Male	-1.162
Fan	Female	0
	Male	0.958
Window	Female	0.213
	Male	-0.353
Door	Female	-0.223
	Male	0
Heater	Female	-0.459
	Male	-0.799

7.3 The causal effects of other variables in the DAG

Table 7.7 delineates the causal impacts of building type on thermal sensation. There appears to be no perceptible change in thermal sensation when transitioning from a classroom setting to a multifamily housing, office, or senior center environment, indicating that the thermal sensation experiences across these building types might be highly similar. Similarly, no substantial difference is observed when moving from a multifamily housing environment to an office. However, an intriguing shift is noticed when transitioning from multifamily housing to 'Others', reflected in a 0.126 increase in thermal sensation. This suggests that the 'Others' category of buildings might embody unique thermal characteristics or occupant behaviors influencing thermal comfort. Moreover, transitioning from an office setting to a senior center leads to a marginal increase in thermal sensation by 0.069, possibly reflecting differences in environmental conditions or user demographics between these building types. However, further transitions between these settings do not seem to affect thermal sensation, denoting some consistency in thermal comfort experiences across various building types. These results underline the complexity of thermal comfort, influenced not only by physical parameters but also building type-specific factors.

Table 7. 7. Causal effects of building types on thermal sensation

Cases	**Average change in thermal sensation**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
From Classroom to Multifamily housing	0	0	0	0	0
From Classroom to Office	0	0	0	0	0
From Classroom to Senior center	0	0	0	0	0
From Classroom to Others	0	0	0	0	0
From Multifamily housing to Office	0	0	0	0	0
From Multifamily housing to Others	0.126	0.126	0.126	0	0.126
From Multifamily housing to Others	0	0	0	0	0
From Office to Senior center	0.069	0.069	0.069	0	0.069
From Office to Senior center	0	0	0	0	0
From Office to Senior center	0	0	0	0	0

Table 7.8 presents the causal effects of different cooling types on thermal sensation. Transitioning from a naturally ventilated environment to an air-conditioned one results in a decrease in thermal

sensation by -0.385, indicating that air conditioning generally provides a cooler perceived environment than natural ventilation. Similarly, shifting from a naturally ventilated setting to a mixed-mode one – involving both natural ventilation and mechanical cooling – corresponds to a decrease in thermal sensation by -0.205, albeit less pronounced than the shift to solely air-conditioned spaces. This suggests that mixed-mode cooling, while providing cooling benefits, might not achieve the same level of perceived coolness as full air conditioning. On the other hand, transitioning from an air-conditioned environment to a mixed-mode environment is associated with an increase in thermal sensation by 0.145. This underscores that although mixed-mode cooling can provide some level of comfort, it might not match the perceived cooling effect of dedicated air conditioning. These results underline the significant role of cooling type in shaping thermal sensation and, by extension, thermal comfort.

Table 7. 8. Causal effects of cooling types on thermal sensation

Cases	**Average change in thermal sensation**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
From Naturally ventilated to Air conditioned	-0.385	-0.385	-0.385	0	-0.385
From Naturally ventilated to mixed mode	-0.205	-0.205	-0.205	0	-0.205
From Air conditioned to mixed mode	0.145	0.145	0.145	0	0.145

Table 7.9 elucidates the causal effects of various environmental and individual factors on thermal sensation. A 2K increment in air temperature yields a positive effect of 0.344, suggesting that as the air temperature increases, the perceived thermal sensation becomes warmer. A similar trend is evident with radiant temperature; a 2K increment prompts an increase in thermal sensation by 0.317, reinforcing the impact of environmental temperatures on thermal comfort. In contrast, a 5% increment in relative humidity appears to have no effect on thermal sensation, indicating that

humidity levels may not directly influence individuals' perception of warmth or coolness. An interesting result is observed with air velocity; a 0.5 m/s increase causes a notable rise in thermal sensation by 0.370. As for subjective thermal factors, a 0.5 clo increment in clothing insulation level corresponds to a decrease in thermal sensation by -0.456, confirming that occupants wear clothes when their thermal sensation has reduced. Lastly, a 0.5 met increment in metabolic rate results in an increase in thermal sensation by 0.25, which is likely a consequence of the additional internal heat produced at higher metabolic rates. These results illustrate the multifaceted interplay of both environmental and individual factors in determining thermal sensation.

Table 7. 9. Causal effects of environmental and subjective thermal factors on thermal sensation

Variable (unit increment)	**Thermal sensation change**	**Refutation**			
		Random cause	**Unobserved Common Causes**	**Placebo Treatment**	**Data subset**
Air temperature (2K)	0.344	0.344	0.344	0	0.344
Radiant temperature (2K)	0.317	0.317	0.317	0	0.317
Relative humidity (5%)	0	0	0	0	0
Air velocity (0.5 m/s)	0.370	0.370	0.370	0	0.370
Clothing insulation level (0.5 clo)	-0.456	-0.456	-0.456	0	-0.456
Metabolic rate (0.5 met)	0.25	0.250	0.250	0	0.250

7.4 Summary

This chapter has validated the application of the proposed ML-based causal inference framework to assess and quantify the effect of occupant personal and behavioral factors on thermal sensation. The derived causal inference results provide insights into the research questions posed in this chapter and highlight their significance in the pursuit of creating comfortable indoor environments.

7.4.1 Are there causal relationships between occupant personal and behavioral factors and thermal sensation?

The findings affirmatively indicate the existence of such relationships. The proposed causal inference analysis mechanism employed in this dissertation has been effective in discerning and quantifying the causal effects of various occupant personal and behavioral factors on thermal sensation. Notably, this analytical mechanism has been crafted to incorporate a broad spectrum of potentially influential factors, offering a holistic perspective. The identified causal effects not only consider personal and behavioral aspects of occupants but also adequately take into account the overarching environmental variables and building characteristics. This comprehensive approach aids in a more nuanced understanding of the intricate web of factors influencing thermal sensation, thereby enhancing the scope and depth of the insights drawn from this research.

7.4.2 If such relationships exist, can these identified personal and behavioral factors be quantified?

In the ambit of this dissertation, quantifiable causal effects of personal and behavioral factors on thermal sensation have been discerned. Personal attributes such as age appeared to exert no appreciable influence on thermal sensation. However, a subtle but distinct variation in thermal sensation between males and females was observed, with females experiencing a warmer sensation by a factor of 0.031. This sex-specific variation in thermal sensation was particularly pronounced in the age brackets of 26-35 and 56-65, where females reported a warmer thermal sensation by 0.066 and 0.253, respectively. Moreover, when considering BMI groups, an interesting divergence was revealed between the overweight and obese categories, with a differential of 0.205 in thermal sensation.

Shifting focus to occupant behavior, the study quantified the impact of manipulating windows and doors on thermal sensation. The act of opening windows and doors corresponded to a change in thermal sensation by 0.078 and -0.187, respectively. Moreover, the thermal sensation that motivates occupants to use heaters has been quantified, revealing that the usage of heaters is influenced by a decrease in thermal sensation of 0.479. These nuanced insights underline the intricate interplay between personal characteristics, behavioral factors, and thermal sensation, illuminating avenues for tailoring thermal comfort strategies to individual needs and preferences.

7.4.3 How can these identified causal factors be utilized in the development of adaptive comfort models?

As gleaned from this study, integrating environmental control variables such as doors, blinds, windows, fans, and heaters that notably shape occupant behavior into building design and management systems could prove beneficial. Such incorporation would likely augment the degree of perceived control over the environmental conditions, consequently enhancing thermal sensation and overall occupant comfort.

Furthermore, in the conceptualization and implementation of adaptive comfort models, it is prudent to consider occupant personal factors like age and sex. In settings like offices, senior homes, and schools where the occupant demographic predominantly pertains to specific age groups, age is shown to be a salient factor influencing thermal sensation, as evidenced by the causal inference findings. Therefore, age-focused thermal comfort strategies could significantly enhance the lived experiences of occupants in these environments.

For a more nuanced and personalized approach to comfort modeling in spaces particularly designed for a specific gender, the deployment of blinds/curtains could play a pivotal role. The causal inference outcomes highlight distinct gender-based discrepancies in the way thermal sensation is influenced by the use of blinds or curtains. Hence, cognizance of these gender-specific thermal comfort dynamics could guide the design of more inclusive and effective thermal comfort strategies.

8 Conclusions

This dissertation introduces a novel approach to applying causal inference analysis in the context of the built environment for sustainable energy reduction and thermal comfort solutions. The proposed approach emphasizes the utilization of the Double Machine Learning algorithm to effectively identify and quantify the causal effect of variables amidst observed and high-dimensional data, thus countering the selection bias faced by conventional models. Moreover, this approach employs directed acyclic graphs to integrate domain knowledge into the causal inference analysis. The robustness of the proposed approach is validated by its application to three distinct datasets (RECS 2018, KHEPS 2018 and 2019, and ASHRAE database II).

8.1 Evaluating energy policies and occupant behavior impacts on cooling energy consumption

The proposed causal inference methodology was first applied to the RECS 2018 data, revealing the causal effects of energy-saving policies and occupant behavior on cooling EUI. The results demonstrated that the energy audit policy was somewhat effective; buildings that underwent the audit did not experience an increase in EUI due to smart meter usage. In contrast, an increase was observed in buildings that did not undergo the audit. Additionally, audited buildings that had access to interval meter data experienced a reduction in EUI, whereas non-audited buildings showed no such effect. The research also found that building insulation codes and Energy Star qualified windows were effective in reducing EUI. With regard to occupant behavior, the most energy-efficient practice was turning the AC on and off according to need, whereas maintaining a constant temperature most of the time resulted in an increased EUI. Based on these findings, this dissertation proposed three policy recommendations that could contribute to significant energy savings and promote sustainability. The first approach involves mandating the use of Energy Star qualified windows with government assistance, reducing EUI in both new and existing buildings. The second recommendation suggests government support for landlords to incorporate energy-efficient features, providing tenants with affordable options. The third approach emphasizes educating occupants on optimal AC usage and encouraging energy-efficient habits such as turning the AC on as needed or using programmable thermostats. These policy measures aim to enhance energy-saving strategies and advance sustainability efforts.

8.2 The role of socio-economic factors and heating equipment selection in energy consumption

The proposed causal inference methodology was secondly applied to the KHEPS 2018-2019 data to assess the impact of socio-economic factors and heating equipment selection on heating EUI. The analysis revealed causal relationships between occupant socio-economic characteristics and heating EUI. Specifically, characteristics such as higher education levels and higher average monthly incomes were associated with reduced EUI, while occupant age showed an increase in EUI. Additionally, households with primary income sources not derived from occupation demonstrated higher EUI, emphasizing the significance of socio-economic considerations in understanding and managing building energy consumption. Regarding heating equipment selection, using individual heating with briquette boilers or electric blankets generally led to increased EUI, while gas or kerosene boilers had the potential to reduce EUI. These findings emphasize the need for the government to integrate occupant socio-economic characteristics into energy-saving policies. While existing policies primarily focus on building characteristics, incorporating socio-economic factors such as age, education level, and average monthly income shows promise in reducing EUI. Additionally, targeted awareness campaigns are crucial, particularly for households with alternative income sources, to promote energy-efficient equipment choices and foster sustainable energy consumption patterns.

8.3 The role of socio-economic factors and heating equipment selection in energy consumption

Finally, the proposed causal inference analysis approach was applied to the ASHRAE thermal comfort database II to investigate the impact of occupants' personal and behavioral factors on thermal sensation. Personal attributes, such as age, did not appear to significantly influence thermal sensation. However, a subtle yet distinct variation in thermal sensation was observed between males and females, with females experiencing a warmer sensation. This sex-specific difference in thermal sensation was particularly notable within the age brackets of 26-35 and 56-

65. Additionally, when considering BMI groups, occupants in the obese category reported feeling warmer compared to those in the overweight category. In terms of occupant behavior, opening windows was associated with a cooler thermal sensation, while opening doors was linked to a warmer sensation, highlighting the need for fresh air and a cooler environment, respectively. The thermal sensation related to the use of heaters was also quantified. This dissertation revealed that integrating environmental control variables such as doors, blinds, windows, fans, and heaters into building design and management systems can enhance occupant comfort by increasing perceived control over environmental conditions. Consideration of occupant personal factors, particularly age and sex, is crucial in the conceptualization and implementation of adaptive comfort models. Age-focused thermal comfort strategies can improve the experiences of occupants in settings like offices, senior homes, and schools with specific age demographics. Sex-specific thermal comfort dynamics, influenced by the use of blinds/curtains, should be acknowledged for a more nuanced and personalized approach to comfort modeling in gender-specific spaces.

8.4 Deductions from the three analyses

Occupant personal and behavioral factors played key roles in cooling, heating, and thermal sensation across all three analyses. In particular, equipment usage behavior, such as turning equipment on and off as needed, was found to reduce cooling EUI in the RECS data and has the potential to reduce EUI in South Korea. Regarding building characteristics, analysis of KHEPS data showed that an increased number of windows reduced EUI in contrast to an increased number of external walls, indicating the beneficial thermal resistance of windows. Therefore, adopting Energy Star qualified windows, as in the RECS analysis, could further help reduce EUI. In terms of thermal comfort, windows still appeared to be a key parameter in achieving the desired thermal sensation while at the same time increasing the degree of perceived control. It can be deduced that using Energy Star qualified windows, managing cooling or heating equipment effectively, and maximizing natural ventilation can serve as a sustainable approach to reducing EUI while maintaining thermal comfort.

8.5 The implications of this dissertation findings for the development of reliable energy and thermal comfort prediction models and controls

Overall, the proposed causal inference approach has large implications in the assessment and selection of variables for building energy and thermal comfort prediction and designs. By utilizing the proposed causal inference method, researchers can discern and measure the causal relationships between various factors and energy consumption or thermal comfort outcomes. This provides a more detailed comprehension of which variables truly have an impact on energy usage and thermal comfort, thereby avoiding misleading correlations or extraneous variables. Hence, the insights derived from causal inference analysis can steer the choice of crucial variables to be incorporated in predictive models and optimization plans, ensuring that attention is paid to factors with genuine causal effects. This method heightens the accuracy and dependability of predictions related to energy usage and thermal comfort in buildings, leading to improved optimization methods, and ultimately supporting the creation of sustainable and comfortable built environments. Incorporating these findings into policymaking, design choices, and technological developments allows stakeholders to collaboratively establish buildings that proficiently harmonize energy efficiency, comfort, and functionality.

References

Aghniaey, S., & Lawrence, T. M. (2018). The impact of increased cooling setpoint temperature during demand response events on occupant thermal comfort in commercial buildings: A review. *Energy and Buildings*, *173*, 19–27.

Ananth, C. V., & Schisterman, E. F. (2017). Confounding, causality, and confusion: The role of intermediate variables in interpreting observational studies in obstetrics. *American Journal of Obstetrics and Gynecology*, *217*(2), 167–175.

Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., & Chaudhury, S. (2009). Hypothesis testing, type I and type II errors. *Industrial Psychiatry Journal*, *18*(2), 127.

Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, *113*(27), 7345–7352.

Battocchi, K., Dillon, E., Hei, M., Lewis, G., Oka, P., Oprescu, M., & Syrgkanis, V. (2019). EconML: a Python package for ML-based heterogeneous treatment effects estimation. *GitHub*.

Chen, H., Harinen, T., Lee, J.-Y., Yung, M., & Zhao, Z. (2020). Causalml: Python package for causal machine learning. *ArXiv Preprint ArXiv:2002.11631*.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). *Double/debiased machine learning for treatment and structural parameters*.

Dinga, R., Schmaal, L., Penninx, B. W., Veltman, D. J., & Marquand, A. F. (2020). Controlling for effects of confounding variables on machine learning predictions. *BioRxiv*, 2020–08.

Elwert, F. (2013). Graphical causal models. *Handbook of Causal Analysis for Social Research*, 245–273.

Flanders, W. D., & Ye, D. (2019). Limits for the magnitude of M-bias and certain other types of structural selection bias. *Epidemiology*, *30*(4), 501–508.

Földváry Ličina, V., Cheung, T., Zhang, H., de Dear, R., Parkinson, T., Arens, E., Chun, C., Schiavon, S., Luo, M., Brager, G., Li, P., Kaam, S., Adebamowo, M. A., Andamon, M. M., Babich, F., Bouden, C., Bukovianska, H., Candido, C., Cao, B., … Zhou, X. (2018). Development of the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, *142*, 502–512. https://doi.org/10.1016/j.buildenv.2018.06.022

Gao, Y., & Zhang, J. (2021). *Studying the Impacts of Federal Funding on Residential Smart Meter Adoption and Utilities' Performance in the US: A Simultaneous Equation Approach*. 416–423.

Hernán, M. A., Clayton, D., & Keiding, N. (2011). The Simpson's paradox unraveled. *International Journal of Epidemiology*, *40*(3), 780–785.

Hope, A. J., & Booth, A. (2014). Attitudes and behaviours of private sector landlords towards the energy efficiency of tenanted homes. *Energy Policy*, *75*, 369–378.

Hoyle, R. H. (1995). *The structural equation modeling approach: Basic concepts and fundamental issues.*

IEA. (2021). *Final energy consumption in the buildings sector, 2021*.

Im, J., Seo, Y., Cetin, K. S., & Singh, J. (2017). Energy efficiency in US residential rental housing: Adoption rates and impact on rent. *Applied Energy*, *205*, 1021–1033.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Infante-Rivard, C., & Cusson, A. (2018). Reflection on modern methods: Selection bias—A review of recent developments. *International Journal of Epidemiology*, *47*(5), 1714–1722.

Karjalainen, S. (2012). Thermal comfort and gender: A literature review. *Indoor Air*, *22*(2), 96–109.

Kim, J., & de Dear, R. (2018). Thermal comfort expectations and adaptive behavioural characteristics of primary and secondary school students. *Building and Environment*, *127*, 13–22.

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., Behar, J. V., Hern, S. C., & Engelmann, W. H. (2001). The National Human Activity Pattern Survey (NHAPS): A resource for assessing exposure to environmental pollutants. *Journal of Exposure Science & Environmental Epidemiology*, *11*(3), 231–252.

Kontokosta, C. E., Spiegel-Feld, D., & Papadopoulos, S. (2020). The impact of mandatory energy audits on building energy use. *Nature Energy*, *5*(4), 309–316.

Lan, L., Lian, Z., Liu, W., & Liu, Y. (2008). Investigation of gender difference in thermal comfort for Chinese people. *European Journal of Applied Physiology*, *102*, 471–480.

Lechner, N. (2014). *Heating, cooling, lighting: Sustainable design methods for architects*. John wiley & sons.

Liang, J., Qiu, Y. (Lucy), & Xing, B. (2022). Impacts of electric-driven heat pumps on residential electricity consumption: An empirical analysis from Arizona, USA. *Cleaner and Responsible Consumption*, *4*, 100045. https://doi.org/10.1016/j.clrc.2021.100045

Lin, B., & Liu, H. (2015). China's building energy efficiency and urbanization. *Energy and Buildings*, *86*, 356–365. https://doi.org/10.1016/j.enbuild.2014.09.069

Liu, W., Yang, D., Shen, X., & Yang, P. (2018). Indoor clothing insulation and thermal history: A clothing model based on logistic function and running mean outdoor temperature. *Building and Environment*, *135*, 142–152. https://doi.org/10.1016/j.buildenv.2018.03.015

Lund, H., Möller, B., Mathiesen, B. V., & Dyrelund, A. (2010). The role of district heating in future renewable energy systems. *Energy*, *35*(3), 1381–1390.

Ming, R., Li, B., Du, C., Yu, W., Liu, H., Kosonen, R., & Yao, R. (2023). A comprehensive understanding of adaptive thermal comfort in dynamic environments–An interaction matrix-based path analysis modeling framework. *Energy and Buildings*, *284*, 112834.

Nevin, R. (2010). Energy-efficient housing stimulus that pays for itself. *Energy Policy*, *38*(1), 4–

11.
Obuchi, T., & Kabashima, Y. (2016). Cross validation in LASSO and its acceleration. *Journal of Statistical Mechanics: Theory and Experiment*, *2016*(5), 053304.
Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688.
Pearl, J. (2000). Models, reasoning and inference. *Cambridge, UK: CambridgeUniversityPress*, *19*(2).
Reina, V. J., & Kontokosta, C. (2017). Low hanging fruit? Regulations and energy efficiency in subsidized multifamily housing. *Energy Policy*, *106*, 505–513.
Rentala, G., Zhu, Y., & Mukhopadhyay, S. (2021). Application of Causal Inference to the Analysis of Occupant Thermal State and Energy Behavioral Intentions in Immersive Virtual Environments. *Frontiers in Sustainable Cities*, *3*, 730474.
Rosenbaum, P. R., & Rosenbaum, P. R. (2002). *Overt bias in observational studies*. Springer.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.
Rupp, R. F., Kim, J., de Dear, R., & Ghisi, E. (2018). Associations of occupant demographics, thermal history and obesity variables with their thermal comfort in air-conditioned and mixed-mode ventilation office buildings. *Building and Environment*, *135*, 1–9. https://doi.org/10.1016/j.buildenv.2018.02.049
Sharma, A., & Kiciman, E. (2020). DoWhy: An end-to-end library for causal inference. *ArXiv Preprint ArXiv:2011.04216*.
Shimoni, Y., Karavani, E., Ravid, S., Bak, P., Ng, T. H., Alford, S. H., Meade, D., & Goldschmidt, Y. (2019). An evaluation toolkit to guide model selection and cohort definition in causal inference. *ArXiv Preprint ArXiv:1906.00442*.
Wai, C. H., Beaudin, M., Zareipour, H., Schellenberg, A., & Lu, N. (2014). Cooling devices in demand response: A comparison of control methods. *IEEE Transactions on Smart Grid*, *6*(1), 249–260.
Wang, Z., de Dear, R., Luo, M., Lin, B., He, Y., Ghahramani, A., & Zhu, Y. (2018). Individual difference in thermal comfort: A literature review. *Building and Environment*, *138*, 181–193. https://doi.org/10.1016/j.buildenv.2018.04.040
Yun, G. Y., & Steemers, K. (2011). Behavioural, physical and socio-economic factors in household cooling energy consumption. *Applied Energy*, *88*(6), 2191–2200. https://doi.org/10.1016/j.apenergy.2011.01.010
Zhao, G., Wang, X., Zheng, D., & Yang, C. (2023). Analysis of the Sustainable Driving Effect of Building Energy Consumption on Economic Development Based on the Sustainable Driving Force Model. *Buildings*, *13*(5), 1180.