UNIVERSITY OF OTTAWA

DOCTORAL THESIS

# Essays in Energy and Environmental Economics

*Author:*
Kareman YASSIN

*Supervisor:*
Dr. Abel BRODEUR
*Co-Supervisor:*
Dr. Maya PAPINEAU

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

Department of Economics
Faculty of Social Sciences

November 28, 2023

# Declaration of Authorship

I, Kareman Yassin, declare that this thesis titled, "Essays in Energy and Environmental Economics " and the work presented in it are my own. Chapter one of this thesis is done by myself jointly with Dr. Anthony Heyes. My contribution is equal to his. The second chapter of this thesis is done jointly with Dr. Maya Papineau, Guy Newsham, and Sarah Brice. My contribution as well as Dr. Papineau's contribution accounts for one third each, while the combined contribution of Guy Newsham and Sarah Brice constitutes one third. The third chapter of this thesis is research done solely by myself.
I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Kareman YASSIN

Date: November 28, 2023

UNIVERSITY OF OTTAWA

# *Abstract*

Faculty of Social Sciences

Doctor of Philosophy

**Essays in Energy and Environmental Economics**

by Kareman YASSIN

This dissertation employ applied microeconomics techniques with a specific emphasis on behavioral dynamics within the realms of energy and environmental economics. In Chapter one, we investigates the impact of outdoor temperature on productivity in the service sector, using data from the India Human Development Survey. Our findings suggest a precisely estimated zero effect on interview duration, ruling out significant productivity impacts. In Chapter two, we employs a conditional demand analysis on a Canadian electricity consumer data set, highlighting the effectiveness of local heat pumps and thermostat setbacks for electricity savings. Results also reveal trends favoring newer homes in electricity consumption decline. In Chapter three, I study the causal relationship of spatial peer effects from Canada's largest home energy efficiency retrofit program on energy consumption. My results show that close neighbors to energy efficiency retrofitted homes experience a significant reduction in monthly natural gas and electricity consumption. Moreover, visible retrofits, such as windows and doors, significantly impact peer energy savings compared to less visible retrofits.

Keywords: energy economics, environmental economics, behavioral economics, residential sector, energy savings, grid decarbonization, conditional demand analysis, peer effects

# *Acknowledgements*

First and foremost, all praise is to Allah, the almighty, the entirely merciful, and especially merciful, for His countless blessings in my life.

I would like to express my deepest appreciation to my thesis supervisors, Abel Brodeur and Maya Papineau, for their rigorous academic training, continuous support, and kindness. Their timely comments and feedback on this dissertation and throughout my doctoral studies have been invaluable. Their support has been humbling, especially during challenging times when I navigated a pandemic, two pregnancies, and motherhood, all while being a full-time student.

I am deeply indebted to Abel Brodeur for consistently dispelling my negative thoughts and guiding me back on the right track during difficult moments, for inspiring me to set high goals, and for generously sharing his time and wisdom, setting an exceptional example as both a person and a researcher.

I am extremely grateful to Maya Papineau for introducing me to the field of Energy and Environment, opening doors to new research opportunities, and facilitating valuable networking experiences. Throughout my time as her student and coauthor, Maya never failed to give me guidance and to inspire me to enhance both my personal growth and research abilities.

I thank my thesis committee, composed of internal examiners Anthony Heyes, Nicholas Rivers, Myra Mohnen and Louis-Philippe Morin for their suggestions, ideas, and recommendations during my first and second thesis workshops and thesis defense, as well as during departmental seminars. I also thank Juan Moreno-Cruz for agreeing to evaluate my thesis as an external examiner and for his comments and suggestions. I extend my appreciation to all faculty members and my colleagues at the University of Ottawa, Carleton University, and the participants in various seminars, including CREEA, CEA, and departmental seminars.

Lastly, I could not have undertaken this journey without the support of my mother, Wafaa, who ingrained in me the values of perseverance, hard work, and unwavering faith; my husband, Amro, who believed in me and taught me to dream big; my daughters, Amina and Salima, whose presence has given profound meaning to this entire journey; my sisters, Sherihan and Eiman, who are a safety net in every aspect of my life; my brother-in-law, Rami, who is a constant source of support; my nephews, Ahmed and Yahia, whose joyful spirits have brought immense happiness to my life; and my dear friends, Nanis, Yasmine, Riham, Lobna, Nini and Noran, who always share with me life's joys and challenges.

*Dedicated to the memory of my beloved father, Ahmed Fouad*

# Contents

# List of Figures

# List of Tables

# General Introduction

This dissertation consists of three essays in applied microeconomics within the field of energy and environmental economics. The first and third chapters incorporate a behavioral dimension to better understand individual behaviors in the context of environment and energy.

In the first chapter, we attempt to understand how economies function at higher outdoor temperature. Evidence show that labor productivity in the manufacturing sector declines on hot days. We extend this research agenda to common work tasks in the service sector. We test the effect of outdoor temperature on the time taken to complete household questionnaire. Data is drawn from India Human Development Survey (IHDS) 29,890 interviews executed by 204 field-workers. In our preferred specification which includes spatial, temporal, and interviewer fixed effects, and controls for various potential confounders, the coefficient on temperature is a precisely-estimated zero in climate-protected and unprotected homes. We rule out effects on interview duration bigger than an increase of 2 percent in response to 10°C increase in maximum temperature on the day of the interview with 95% assurance.

In the second chapter, we implement a conditional demand analysis (CDA) using a large dataset of electricity consumers in a Canadian province with a high market share of electric heating technologies. In doing so we also provide a unifying review of the breadth of interdisciplinary applications of CDA, beginning from the earliest studies up to the present, and test for evidence of unobservable variable bias from random effects panel data estimators. We find that local (i.e. minisplit) heat pumps and thermostat setbacks show the largest electricity savings. Central heat pumps generally do not save heating electricity compared to electric baseboards, and exhibit higher cooling season consumption compared to local heat pumps. We also observe a consistent decline in electricity consumption for newer homes. Our results can inform research to identify promising technologies that support a shift towards large-scale electrification and decarbonization of energy end-uses.

In the third chapter, I study the spatial peer effects generated by the largest home energy efficiency retrofit program in Canada, and its impact on energy consumption within a city in the Prairie provinces. I construct a comprehensive panel database for all single-family homes in the city, comprising ten years of monthly energy consumption data, tax assessment variables, program participation and house audit data. I find that close neighbors to energy efficiency retrofitted homes reduce their monthly natural gas and electricity consumption by an average of 2.4% and 1% respectively. Estimates using three different comparison groups yield similar results. Examining the mechanisms, I find that visible retrofits like windows, doors, and exterior wall insulation have a significantly stronger impact on peer energy savings (three and a half times) compared to less visible retrofits like natural gas furnace upgrades. Additionally, the effect diminishes as the distance to neighbors increases.

# Chapter 1

# Outdoor Temperature and Indoor Productivity in an Archetypal Service Sector Task: Evidence from the Filling in of 29,890 Forms

## 1.1 Introduction

A central unknown in projecting how economic performance will be impacted by the higher temperatures projected for most places by mainstream climate models is how labor productivity will be affected in common work tasks.

An insightful recent paper by Somanathan et al., 2021 reports evidence of daily temperature variations influencing worker-level productivity in a number of Indian manufacturing settings. In particular they observed cloth weaving, garment sewing, diamond cutting and steel rolling activities across various locations, finding that on hot days worker productivity declines by 2 to 4 percent per degree Celsius, and in addition that the whole of the negative effect of hot weather on plant-level output can be accounted for by variations in the productivity of labor. This is a big effect and exploring its robustness to other manufacturing workplaces and other study methodologies will be an important task for future research.

Our research pushes this agenda further by extending such analysis from manufacturing settings to the fundamentally different type of work tasks more pertinent to understanding potential impacts in the *service* sector. According to the OECD, in 2018 the service sector accounted for roughly 61% of world GDP (49% in India, 54% in low and middle income countries collectively) - around three to five times bigger than the contribution of manufacturing in most countries.

The sort of tasks we have in mind might involve interpersonal interaction and the exchange of information between people and the completion of 'paperwork' of the sort that are routine in many service jobs in the private, public and third sector.[1] Examples of the type of work we have in mind would be a bank employee filling in a loan or mortgage application with a client, a college advisor working with a student to put together a study schedule, a store employee trying to understand and service the needs of a customer, or a lawyer gathering information pertinent to a legal case. Typical in many service sector roles these work tasks are inherently two-sided or

[1]The third sector includes community groups, voluntary organizations, faith and equalities groups, charities, social enterprises, co-operatives, community interest companies, mutuals and housing associations. Third sector organizations have a set of distinctive characteristics, which include: being self-governing, independence of both formal structures of government and the profit sector and an significance reliance on volunteers to carry out its work.(FOSS, 2008)

social in character. Our research question is: Is the productivity of an employee engaged in this sort of work affected by outdoor heat?

The outcome data for our main analysis is drawn from the India Human Development Survey (IHDS). The IHDS is a large-scale, multi-wave survey of the health and living conditions executed in a large, representative sample of Indian households. It is executed by the National Council of Applied Economic Research (NCAER) and the University of Maryland and its research practices are highly regarded. We focus on the 41,544 surveys that were conducted in the 2005 wave (IHDS-1). Each survey requires that a project employee sit with a family member and work together through a lengthy series of questions on various topics. The answers to these questions have underpinned numerous published studies on a wide variety of topics including marriage market and associated economic conditions Corno, Hildebrandt, and Voena, 2020, urbanization in rich and poor countries Chauvin et al., 2017 and mobility, growth, and inequality in India Munshi and Rosenzweig, 2009.

However our interest is not in the responses to any of these questions, but rather the performance of the survey employees, referred to as 'field-workers'. The metadata collected incidental to the survey responses includes a record of the minute that the interview started and finished. From this we derive out main outcome measure, the number of minutes taken for completion of the survey. There is a lot of variation in this, even for a particular field-worker. Across the whole sample the mean time taken is a little over 67 minutes, with a standard deviation of 26. This is unsurprising given the social nature of the task and the variation in characteristics of respondents. However survey protocols imply that those characteristics - observed and unobserved - can sensibly be regarded as orthogonal to same-day temperature (something that we verify statistically for a range of observable). In essence we estimate, in a fixed effects framework, correlation between time taken for completion and outdoor temperature on day of interview. Correlation, if found, we will interpret as evidence of a causal impact of temperature on field-worker productivity.

A number of features make this a good setting in which to probe our research question. The task of the field-worker is routine - to elicit information from the respondent and complete a standardized form - and prototypical of a wide range of job contexts. We have data on 29,890 interviews executed by 204 field-workers. As such we observe the work rate of each worker under a wide variety (though typically hot to very hot) outdoor temperature conditions allowing us to report only within-worker estimates. The field-work takes place at the home of the respondent. We know a lot about living circumstances of each respondent, including for example whether or not the home is equipped with air-cooling technology, offering the potential to measure the protective benefit of the most obvious technological mitigation of high outdoor temperatures, at least on this outcome variable.

Our central finding is that we find no evidence of an effect of outdoor temperature on day of interview on the length of time taken to complete a unit of work (an interview). The coefficient on temperature in our preferred fixed effect regression, and all alternative specifications, is a precisely estimated zero. This is the case whether in climate-protected or unprotected homes. We can rule out effects on interview duration bigger than an increase of 1.83 minutes, or 110 seconds, in response to 10°C increase in maximum temperature on the day of the interview with 95% assurance.

While our main results use dry bulb temperature, and control for humidity, for direct comparability with Somanathan et al (2018) we also report a variation using wet-bulb, which is a commonly used single-dimensional measure of heat which combines dry-bulb temperature and relative humidity. We find the upper bound of the 95% confidence interval is around 9 to 11 times smaller than the central estimates of effect size reported by Somanathan for manufacturing settings. The coefficient of 0.012 in our preferred specification implies that a 10°C increase in maximum temperature on the day of the interview increases the interview duration by 0.12 minutes, or 7.2 seconds.

As a test of external validity we also sourced the same meta-data on interview start and finish time from waves 13 through 17 of the Household, Income and Labor Dynamics in Australia (HILDA). HILDA is a large-scale survey conducted by the Melbourne Institute, University of Melbourne. The most recent waves that we use have as an advantage that responses were collected on computer tablets rather than by hand, providing a very precise record of start and finish time. Our results from the HILDA surveys confirm our findings from the IHDS survey.

Our results add to the building blocks of evidence needed to project how economies function at higher outdoor temperatures. In particular, to the nascent body of evidence that explores the link between outdoor temperature and indoor work performance.[2] This is important because mainstream climate models point to an increased frequency of hot and very hot days in much of the world, including India. In an average year between 1957 and 2000 there were 5 days with average daily temperatures above 95 °F in India (Burgess et al., 2017). According to the Hadley model, under a business-as-usual scenario there will be 75 such days per year between 2075 and 2100. The equivalent numbers for the United States are 0 and 29. That our analysis delivers a precisely estimated zero - a 'negative result' - implies that we need to be cautious in assuming that productivity effects uncovered in manufacturing and physical work setting will necessarily carry over into the types of socially interactive work tasks common in the service sector.

## 1.2 Data

Our central analysis links data on interview duration from a large-scale India-wide household survey data, with what we know about temperature at the same day and location of the interview. To probe external validity we also report results from a large Australia-wide household survey data in a subsequent section.

### 1.2.1 Survey Data

**India Human Development Survey (IHDS)**

Our primary setting takes place in India, and our data comes from the first wave of India Human Development Survey (IHDS-I) in 2005. It is a nationally representative survey of 41,554 households in 1503 villages and 971 urban neighborhoods across India. Interviews covered multiple aspects including health, education, employment, economic status, and social capital. IHDS was jointly organized by researchers from the University of Maryland and the National Council of Applied Economic Research (NCAER), New Delhi.[3] We utilize IHDS-I data on field-worker ID, interview date,

[2] In a broader sense the results also add to the wider body of research linking heat to a wider set on human outcomes such as mortality (Burgess et al., 2017), student learning (Goodman et al., 2018), cognitive function (Park, Behrer, and Goodman, 2021), violence (Anderson, 2001) and sentiment (Baylis, 2020).

[3] For a detailed description of the sampling and interview methods the reader is encouraged to consult "Data Sharing for Demographic Research" (DSDR) that can be accessed via https://www.icpsr.umich.edu/web/pages/DSDR/idhs-data-guide.html

location and interview start and end time. We also use selected household characteristics to act as controls.

The survey was designed to be nationally-representative along a number of demographic dimensions including ethnicity, religion, education and economic status. Households were selected according to a standardized sampling criteria within a randomly-selected set of villages and urban neighborhoods. Moreover, various tests for representatives of the sample were conducted.

Each household typically completed two questionnaires; household, and heath and education. Our data on interview duration comes from the former, which was generally conducted with the head of the household and covered questions on socio-economic status including income, employment, educational status and consumption expenditure.

Field work was performed by 25 data collection agencies throughout the country. Locally knowledgeable and linguistically fluent interviewers were hired and a team from NCAER organized eleven two-week training sessions. Once trained, interviewers were organized in teams of five. Two pairs; a male and a female interviewer in each pair, and a team leader. The team leader had the responsibility of overseeing and assisting in conducting household interviews and typically handled interviews in the village, school, and medical facility. Upon reaching a PSU (Primary Sampling Unit), the team would get in touch with local leaders to explain the survey, obtain permissions, and create a map of the area. Once the sample was drawn, interviewer pairs began scheduling interviews. Interviews were not scheduled at a given time with households and depended on the availability of respondents. Once interviewers arrived at a designated household, they would obtain consent and fill out two copies of the household roster. Once the household roster was finished, the two copies were split between the interviewers. The female interviewer then conducted the education and health questionnaire, often with the assistance of a senior woman in the household. The household questionnaire - the completion of which is the focus of our study - was completed in conjunction with the head of the household, who was typically male, by the male interviewer. Some questionnaires needed more than one visit to be completed.

In most cases, field-workers were remunerated per interview completed, though in some cases paid a daily stipend and timetabled a fixed number (no more than three) to complete in a day. Unfortunately we are unable to observe the circumstances of individual field-workers in the data set, and must interpret results on the understanding that the subjects face mixed incentives. The payment of a daily stipend or salary is likely more typical in many of the service roles identified in the motivation above, with the employee often residual claimant over time left 'spare' if assigned work is finished early. Under piece-rate, the employee may be able to monetize the benefit of faster work by conducting an extra interview, but again more likely is that they are residual claimant over the time saved - they finish work early. We recognize the incomplete knowledge of remuneration as a limitation of our setting. However the very precisely estimated zero effects that we obtain suggest little or not impact on performance across any of the remuneration models in place.

For our purpose, we removed observations with missing interview date, location, start or end time or interviewer ID. We also excluded incomplete interviews, interviews that took less than ten minutes, interviews that took more than five hours,[4]

[4]In most cases this seems likely due to input errors in start and end time. Later we report robustness of results to dealing with outliers statistically by trimming and windsorizing the data.

interviews that required more than one visit to be completed, and field-workers who conducted ten interviews or less.

Table 1.1 presents summary statistics. Interview duration ranges from 10 to 300 minutes. After the data cleaning process, there are 29,890 observations in total, with a mean of 67 and a standard deviation of roughly 26 minutes. Furthermore, Figure 1.1 shows a peak around 60 minutes.

### 1.2.2 Weather Data

Our focal independent variable of interest in this study is outdoor temperature in the vicinity of the respondent's home on the day of interview. To obtain this we match micro-data from households' interviews to district-level climate data from weather stations. The location of interviews is drown from 365 districts in India where weather conditions widely differ. For example, maximum temperature ranged from -0.5°C to 53°C during the survey period.

Temperature and other weather data are obtained from The National Centers for Environmental Prediction (NCEP), Climate Forecast System Reanalysis (CFSR).[5] District-level data was downloaded using NCEP's Interactive Map Application for daily data sets by bounding a box of latitude and longitude of each district in our study (*The National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR)*). Data reports maximum and minimum temperature, daily rainfall, wind speed, relative humidity, and solar radiation. The CFSR is produced using cutting-edge data-assimilation techniques (both conventional meteorological gauge observations and satellite irradiances) as well as highly advanced atmospheric, oceanic, and surface-modeling components at 38 km resolution (Saha et al., 2010).

Summary statistics for all weather variables are reported Panel A in Table 1.1. The mean of maximum temperature during the survey period is 32°C. Figure 1.2 plots the histogram of maximum temperature which shows great heterogeneity, with temperatures ranging from -0.5°C to 53°C, and a peak around 30°C. To measure the non-linear effects of temperature, we construct maximum temperature bins of 5°C bins, with the lowest bin including all temperatures below 25 °C and the highest bin including all temperatures above 45 °C.

For a subset of exercises we compute daily wet bulb temperature or WBT (Parson, 1993;ISO, 1989). WBT is a widely-used measure of heat stress that combines temperature with relative humidity. Special instruments are required to directly measure WBT, however, we use the following approximation (Lemke and Kjellstrom, 2012):

$$WBT = 0.567T_A + 0.216\rho + 3.38 \tag{1.1}$$

where, $T_A$ is air temperature in degrees Celsius and $\rho$ is water vapor pressure which is calculated from relative humidity, RH as follows:

$$\rho = \left(\frac{RH}{100}\right) \times 6.105 \exp\left(\frac{17.27T_A}{237.7 + T_A}\right) \tag{1.2}$$

[5]The CFSR global atmosphere resolution is approximately 38 km (T382) with 64 levels extending from the surface to 0.26 hPa. The global land surface model has four4 soil levels and the global sea ice model has 3 layers. Most available in-situ and satellite observation data were included in the CFSR. Satellite-based radiance observations were bias corrected with spin-up runs at full resolution.The CFSR atmospheric, oceanic and land surface output products are available at an hourly time resolution and at a 0.5 deg x 0.5 deg latitude and longitude resolution.

## 1.3 Methods

In this section we detail the empirical approach implemented to estimate the causal impact of weather on productivity expressed by the length of household interviews in minutes. We also thoroughly explain the specifications we deploy.

We make use of the panel structure of the IHDS-I data where interview duration is recorded for each interviewer and adopt a fixed-effects specification. Identification comes from quasi-random assignment of outdoor temperatures to interview days. Such a specification accounts for individual, time invariant heterogeneity between interviewers such as education or gender as it follows the same interviewer over time.

We take three alternative approaches to estimation and in each case various specifications incorporating different sets of controls. Our preferred specification controls for interviewer ID, district, year, month, number of persons in a household, households ownership to air cooler/ air conditioner, access to electricity, non-temperature weather characteristics, total household income and membership to social groups (caste/religion).

Our first and simplest specification is an interviewer fixed effects model estimated by Ordinary Least Squares (OLS):

$$Y_{i,j,d,s} = \beta_0 + \beta_1 Temperature_{d,s} + \alpha_i + \delta_j + \lambda_M + \zeta_Y + \gamma_s + \iota_{d,s} + \epsilon_{i,j,d,s} \tag{1.3}$$

The dependent variable $Y_{i,j,d,s}$ is the duration in minutes of the interview conducted by field-worker $i$ with household $j$ on date $d$ in district $s$. Our parameter of interest is $\beta_1$, the coefficient of outdoor maximum temperature on date $d$ in district $s$. The inclusion of interviewer, month, and district fixed effects ($\alpha i$, $\lambda_M$, and $\gamma_s$ respectively) indicates that the identification comes from within interviewers, within year/month and within district variation.

Even though most observations took place in 2005, around 11% of interviews were done in 2004. Hence, year fixed effects ($\zeta_Y$) control for any common events across districts considerations that vary between 2004 and 2005 that might influence the outcome variable (such as nationwide policies or economy-wide shocks for example). The vector $\iota_{d,s}$ contains non-temperature weather variables. $\delta_j$ is a vector of household controls including number of persons in the interviewed household, access to electricity, ownership of air cooler or/and air conditioner, total household income and membership to social groups (caste/religion). The error term $\epsilon_{i,j,d,s}$ is clustered by district to allow for serial and spatial correlation within each district.

Under our linear measure of temperature, the coefficient of interest $\beta_1$, measures the change in interview length in minutes when outdoor temperature on day of interview increases by 1°C.

To probe possible non-linearity of the temperature effect we estimate two non-linear variants of the model. First, we replace the continuous temperature regressor in equation 1.3 by a series of indicator variables corresponding to bins of width 5°C. Here, the coefficient of each bin measures the percentage change in total interview duration that occurs because of temperature falls into that bin rather than the reference bin of 0-25°C.

Second, by replacing the continuous (dry-bulb) temperature regressor with WBT. For the WBT we follow exactly the specification of Somanathan et al., 2021 who estimate the following piece-wise linear function[6]:

$$Log(Y)_{i,j,d,s} = \beta_0 + \beta_k WBT_{d,s} \times D_k + \alpha_i + \delta_j + \lambda_M + \zeta_Y + \gamma_s + \theta R_{d,s} + \iota_{d,s} + \epsilon_{i,j,d,s} \tag{1.4}$$

We again include a range of fixed effects to control for idiosyncratic interviewer productivity, temporal and seasonal shocks, district fixed effects and household controls. Non-linear effects of heat-stress are allowed for by interacting the maximum wet bulb temperature in district $d$ on date $s$ ($WBT_{d,s}$) with indicator variables $D_k$ for four temperature ranges: (0°C - 20°C), (20°C - 25°C), (25°C - 27°C), (27°C -35°C). We therefore arrive at estimates of the marginal effect of a degree change in WBT on our outcome variable within each of these intervals.

## 1.4 Results

As a suggestive exercise we start with a simple plot of temperature against interview duration in minutes. This is presented in Figure 1.3. Visual inspection of the scatter points to no obvious positive nor negative association between the two variables. However fallibility of human perception and the absence of potentially important controls cautions against over-interpreting this. The rest of the paper is dedicated to uncovering any statistically significant correlation that may exist.

### 1.4.1 Linear

Table 1.2 reports the main results from our central linear specification. Recall that the dependent variable is interview duration in minutes.

Column (1) in Table 1.2 is the sparsest specification, including interviewer, location (district) and time (year and month) fixed effects, and controls for the number of members of the surveyed household. The interview involves collecting socio-demographic and other data from the respondent on each household member, so the last of these is plausibly an important influence on interview length. The estimated effect in column (1) is very small and very far from statistical significance at conventional levels. The 0.028 coefficient implies that a 10°C increase in maximum temperature on the day of the interview increases the interview duration by 0.28 minutes, or 16.4 seconds. Recall that the average interview duration is 67 minutes so this implies an increase of 0.4 percent in the interview length. The confidence interval is centered tightly around zero, with the upper bound of 0.17 and a lower bound of -0.12 of a standard deviation in mean duration.

Moving rightwards the specifications include various permutations of controls. In Column (2) we add a number of household variables that might impact access to refrigeration and cooling technology by controlling for owning an air cooler and/or air conditioner as well as having access to electricity. Column (3) adds controls for a set of non-temperature weather characteristics (relative humidity, precipitation, wind speed and solar exposure). Finally, in column (4), we further control for total household income and and membership to social groups. In each column the estimated coefficient is close to zero. We choose Column (4) as our preferred specification, incorporating what we believe to be the most natural set of controls. The 0.012 coefficient implies that a 10°C increase in maximum temperature on the day of

[6]This is not our preferred approach. Rather we follow the specification of these authors here to allow direct comparison of our results with theirs.

the interview increases the interview duration by 0.12 minutes, or 7.2 seconds. The confidence interval is centered tightly around zero, with the upper bound of only 0.16 and a lower bound of -0.18 of a standard deviation in mean duration.

### 1.4.2 Non-linear

Any relationship between temperature and performance could plausibly be non-linear. Table 1.3 reports the results of investigating possible non linearity by re-estimating the earlier specification but replacing the continuous temperature measure by a series of indicator variables each taking value of one if temperature fell into the associated bin 5°C in width. The 0-25°C bin is the omitted or reference category. Such an approach has been used in a number of recent temperature and social and economic outcome papers, for examples Deschênes and Greenstone, 2011 and (Zivin et al., 2020).

Each column embeds a different set of controls corresponding to the same combinations in Table 1.2. Column (4) is again our preferred specification and contains interviewer, location (district), time (year and month), household size, ownership to heat adaptation devices, access to electricity, non-temperature weather variables, total household income and membership in social groups. The point estimates in each bin shows no significant relationship between interview day temperature and interview duration, with very small coefficient values. The 95% confidence interval for the estimated impact of working on a day with maximum temperature above 45°C - which is extraordinarily hot day in any locale - compared to a day where maximum temperature lies in what would normally be considered the comfortable range, 0-25°C, range between 0.05 to -0.04 standard deviations. Coefficient estimates and associated confidence intervals are plotted in Figure 1.4.

**Comparison with Somanathan et al** Somanathan et al., 2021 We further explore possible non-linearity in a way that permits direct comparison with the results in the seminal paper by Somanathan et al., 2021, already outlined. To do this we change the temperature measure to wet-bulb temperature (WBT) detailed in the method section. The impact of high temperature on human function depends not just on equivalent temperature in dry air, but also the moisture content of the air which determines the efficiency with which a human body can regulate core temperature through sweating. WBT combines dry-bulb temperature and relative humidity in non-linear way and is commonly adopted as a measure of heat stress of the 'experience of heat'. Table 1.4 presents our results. Specifications 1-4 are similar to the previous non-linear results. Estimates are *positive* in sign, very small and never approach significance at conventional levels. Focusing on the results in column (4) of Table 1.4, we see no systematic change in productivity. Above 27 degrees, a one degree change in WBT is associated with an increase in interview duration of 0.2 percent. Moreover, confidence intervals for the estimates are centered close to zero.

For comparison purpose, we plot our results in the top left panel of Figure 1.5 against results from Somanathan et al., 2021 from various manufacturing plants in India.

## 1.5 Robustness

In this section we report various robustness checks to our linear preferred specification, all tables are in the paper Appendix.

### 1.5.1 Time of interview

Table 1.9 in the Appendix looks for difference in the role of temperature in interviews starting in the morning versus in the afternoon. Recall that in terms of interpretation, field-workers may be 'rewarded' for faster work in additional leisure time - in other words by going home early. This might lead us to expect different effects of treatment in morning compared to afternoon interviews. To explore this possibility we interact the temperature regressor with a dummy variable that takes the value 1 if the interview started in the afternoon, zero otherwise. Here, the interpretation of the $\beta_1$ coefficient becomes the effect of a 1°C change in temperature during the afternoon interviews. As shown in Table 1.9 column (2), the coefficient of interest interacted with maximum temperature is very small, insignificant, and tightly centered around zero.

### 1.5.2 Cumulative daily interview time

The interviews are relatively short, which is a contrast with manufacturing sector tasks that lasts for about 8 hours a day. Our objective in this robustness check is to examine the relationship between the maximum temperature on the interview day and the total time spent conducting interviews each day. We aggregate the duration of total interviews in minutes done by each interviewer in one day and estimate Equation (1.3) to examine the effect of temperature on the cumulative working time per day. Results are shown in Table 1.10 in the Appendix. We include the same set of controls as our results for the linear specification in Table 1.2. Our preferred specification is shown in Column (4) that includes interviewer, location (district) and time (year and month) fixed effects, and controls for the number of members of the surveyed household, owning an air cooler and/or air conditioner, having access to electricity, non-temperature weather characteristics, total household income and membership to social groups. In each column the estimated coefficient is close to zero, indicating no significant relationship between temperature and cumulative daily work hours.

### 1.5.3 Weather adoptive devices

Two specifications in Table 1.11 in the Appendix control for household access to air-cooling technology. In Table 1.11 we report the results of further probing the possible protective effect of such technology by running specifications that include a dummy if household owns an air-cooler (column (1)), air conditioner (column (2)), or either an air-cooler or an electric fan (column(3)).

In each case the dummy is included independently and also interacted with the temperature treatment variable. While the coefficients on each of the interaction terms is negative - consistent with mitigative or protective effects - the coefficient values are small and none achieves statistical significance. This is un-surprising given that the effect of temperature itself appears tiny to non-existent. Importantly in terms of robustness of our main specification the various inclusions have little discernible impact on the value of the primary coefficient value, or the precision with which it is estimated.

### 1.5.4 Alternative standard errors

Standard errors reported earlier are clustered by district to allow for spatial correlation across districts. Table 1.12 in the Appendix shows that results are robust to a variety of other approaches; un-clustered standard errors, district and month, and interviewer and month.

### 1.5.5 Additional exercises

In this subsection we examine other tests of falsification by performing our linear preferred specification on different sub-samples.

In Table 1.13 in the Appendix, column (1) shows results from districts in urban areas that make up 33% of the sample. Column (2) shows estimates from households residing in rural areas that represents approximately 65% of the sample. Both coefficients are very small and insignificant. In column (3) we consider observation from households whose kids undertook learning tests during the fieldworkers visit. Learning tests consisted of short reading, writing, and arithmetic knowledge tests that were administered to all available children aged 8 to 11 years old in the household.

In column (4) we only consider relatively smaller districts in India by dropping 36 districts of area larger than 10,000 km. We also performed our analysis on a subset of 50% of the sample where the difference between daily maximum and minimum temperatures is minimal, results are in column (5). Column (6) looks only at districts with more than 200 interviews conducted. Finally in column (7) we use average daily temperature instead of maximum temperature as the independent variable. All results shown in Table 1.13 are statistically insignificant and very small in magnitude.

## 1.6 External Validity: HILDA in Australia

While the IHDS field-workers provided a good venue for testing our research question in a country with large population and climate conditions increasingly prevalent under climate change, we seek to provide at least some insight into possible external validity of the (null) results. Hence, we turn here to the Household, Income and Labor Dynamics in Australia (HILDA).

HILDA is a large-scale, multi-wave, nationally representative household panel survey with particular attention paid to family and household formation, income and work. Fieldwork runs during the same time frame each year (July through February).[7] Starting from wave-13 onward, the process of data collection moved from paper to computer tablet, providing very accurate meta-data on time taken to execute a questionnaire. For this reason we concentrate on waves 13 through 17, concentrating on the main 'Person Questionnaire' that is administered to every member of the household aged 15 years and over.

Our sample consists of 47,560 personal interviews conducted by 159 field-workers. Mean of interview duration is 33 minutes with standard deviation of roughly 12 minutes. Figure 1.6 shows a peak around 30 minutes.

For households surveyed in previous waves, a primary approach letter was sent to the last known address approximately one month prior to when the interview was scheduled. A newsletter is also attached to provide respondents with some results from prior waves. New households were given a "New Entrants Brochure" that provided more information about the purpose of the study. A cash incentive of $35 was provided to each respondent immediately after the face-to-face interview.

[7] 50% of HILDA fieldwork was done in spring, 48% in winter, and 1.5% in summer.

For field-workers, all new and experienced interviewers attended training sessions. With respect to remuneration, interviewers were employed on a causal basis and paid an hourly wage for the hours they work each week during the fieldwork period. There was no limit on the number of interviews performed each day.

For our analysis, we use data on interview date, duration of interview in minutes, interviewer ID and household controls. In terms of location, for privacy reasons HILDA releases only the city where interviews took place. However, temperature across locations within a city are unlikely to vary (compared to, for example, air quality conditions) so this is unlikely to introduce discernible measurement error on the regressor of interest.

As our Indian analysis, interviews missing information about interview date, location, interview length or interviewer ID where removed. We also excluded non-responding persons, phone- interviews, and field-workers who conducted less than ten interviews throughout the survey period.

### 1.6.1 Weather Data

Similar to our analysis in India, the main independent variable of interest is outdoor temperature in the locale of household's residence on day of the interview. We rely on The Australian Data Archive for Meteorology (ADAM), a database that holds weather observations dating back to the mid 1800s.

The meteorological data in ADAM comes from the Bureau of Meteorology at the Australian government. It manages observing systems over mainland Australia and from neighboring islands, the Antarctic, ships and ocean buoys. It also stores few observations from other local and international sources. Their website (www.bom.gov.au) provides access to recent and past weather information through their Climate Data Online (CDO). Data is retrieved by choosing the bureau station of interest.

For our research purpose, we merged city level weather observations (daily maximum temperature and humidity) with HILDA database using the interview date and location.[8]

Summary statics for interview duration, daily maximum temperature, average relative humidity, WBGT, number of persons in a household and and household total gross income (in categories constructed by HILDA) are reported in Table 1.5. As shown, maximum temperature varied widely during the survey period, it ranged from 10°C to around 44°C with mean of 20°C. Figure 1.7 plots the frequency maximum temperature on the days of the HILDA interviews. For our first non-linear analysis, we construct maximum temperature bins of 5°C wide.[9] Likewise our IHDS analysis, we incorporate average relative humidity and compute daily wet bulb temperature (WBT) as ((Parson, 1993);(ISO, 1989)). In our second non-linear analysis.

### 1.6.2 Methods and Results

For visualization purpose, Figure 1.8 plots temperature at the day of interview against interview length in minutes with no additional controls. No apparent correlation between both variables is observed.

[8]The cities identified by HILDA are; Sydney, Melbourne, Brisbane, Adelaide, Perth and Tasmania

[9]India and Australia have distinct climate conditions. For example, the mean of maximum temperature is 32°C and 20°C in India and Australia respectively. And hence, the maximum temperature bins differ in both analyses.

To estimate the causal effect of temperature on productivity expressed by interview duration in minutes, we employ the same empirical strategy used in IHDS analysis described above. We use Equation (1.3) for linear specification and and Equation (1.4) for non-linear specification. In all specifications we cluster standard errors by household area code. HILDA identifies 574 unique area codes for household location.

Our preferred specification have interviewer, city by month and year fixed effects. We also control for the number of persons in the household, whether the person was interviewed in previous waves or not and gross household income. We show results of further controlling for gross household income. Table 1.6 reports the results of the linear specification. Column (3), our preferred specification, includes interviewer, city by month, and year fixed effects. And control variables for number of household members, whether they were interviewed in previous waves or not and household income.

The estimated coefficient of -0.025 in column (23) is small and insignificant. It suggests that a 10°C increase in maximum temperature on the day of the interview shortens its length by 0.25 minutes, or 15 seconds. Given that the average interview duration is 33 minutes, this implies a decrease of 0.7 percent in duration of the interview. Moreover, the upper and lower bounds of the confidence intervals are 0.01 and -0.06 respectively, i.e centered tightly around zero.

To explore the possibility of non-linear relationship between temperature and productivity, we firstly report in Table 1.7 results from the non linear specification with temperature bins. For reference, the omitted weather bin is 5 - 15°C. Column (1) to (3) have the same set of controls as the above results of the linear specification. The estimates are very small and insignificant with only one exception. And the confidence intervals are very close to zero. Column (3) is again our preferred specification, which we plot its estimates and associated confidence intervals in Figure 1.9

Then, we report results from the non-linear specification with wet-bulb temperature in Table 1.8. Results from column (3), incorporating our preferred set of controls, show no evidence of a causal relationship between temperature and performance. In particular, a one degree change in WBT implies a decrease of interview duration of 0.1 percent in all four categories (below 20°C, between 20-25, between 25-27 and above 27°C) with confidence intervals centered firmly around zero. We again plot our results in the top right panel of Figure 1.10 against our results from IHDS in the top left panel and Somanathan et al., 2021's results from different Indian manufacturing plants.

## 1.7 Conclusions

The objective of the research reported here is to contribute to understanding the impacts that increased outdoor temperature, of the sort predicted for much of the world by most mainstream climate models, might affect human productivity. While there is persuasive evidence of a detrimental impact on workers in physical task in the manufacturing sector the extent to which similar effects might be seen in typical service sector roles remains unexplored.

In this paper we apply a fixed effects model to study the causal relationship of outdoor temperature on human productivity measured by the time taken to complete household questionnaire. We utilize data from 204 interviewers conducting 29,890 interviews for India Human Development Survey (IHDS).

The data that we present yields little evidence of any impact of outdoor temperature on our set of field-workers executing a routine, interactive, form-filling task of a type familiar in many service sector settings. Our main results point to a precisely-estimated zero effect of temperature on the speed at which work is completed, and can rule out effect sizes more than a small fraction of the effect sizes observed in physical tasks.

The setting that we studied had a number of features that lent it to this research. However there are two caveats worth holding in mind whenevaluating these results. First, the incentives of field workers to finish work early is not clear. With a survey as big as IHDS, it was necessary to subcontract the field work to existing NGOs in participating states. We know that interviewers were either paid per day or on a per interview basis, but we were unable to track how each fieldworker was remunerated. However, our precisely estimated zero effects that we get propose little or nonexistent impact on performance across any of the remuneration models in place.

Second, the fieldworker task is not office based and requires traveling to work. We observe neither their precise movements during the day of the interview nor the exact indoor setting that they work in. Nevertheless, the IHDS provide us with detailed household characteristics that we use as controls to mitigate these concerns including ownership of weather adoptive devices. In addition, our null results show robustness to interviews conducted in the afternoon versus the morning. The precise estimated zero result pertains for several sub-samples falsification tests.

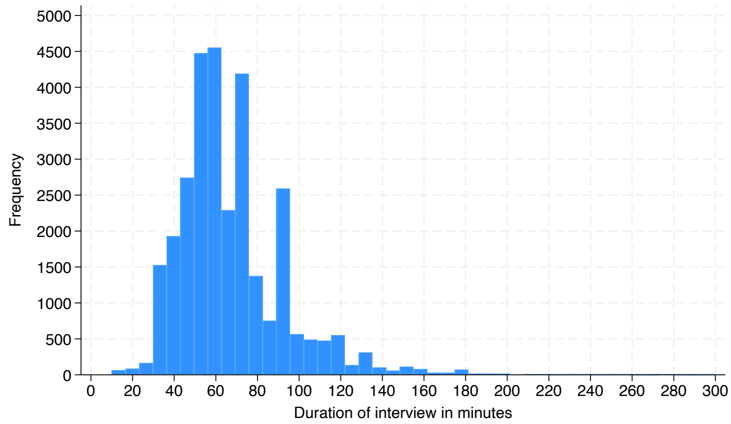
As an external validity exercise, we apply the same fixed effects model on a sample of 47,639 interviews conducted by 163 field workers as part of Household Income and Labor Dynamics survey in Australia (HILDA). Likewise, the coefficient of interest is very small and closely tight around zero.

This paper has fundamental implications for how we should think about the productivity costs of climate change. Recent evidence have called for investments in weather adaptation measures to reduce the responsiveness of productivity to high temperatures in the manufacturing sector. This could plausibly be applied to similar work settings including agriculture and construction jobs.

Yet, our results suggest that we can not simply expect the same substantial productivity cuts on hot days to fundamentally different work tasks. We show that local workers whose tasks involve interpersonal interaction exhibit some kind of resilience to extreme heat and humidity through a capacity of human adaptation to especially hot days.

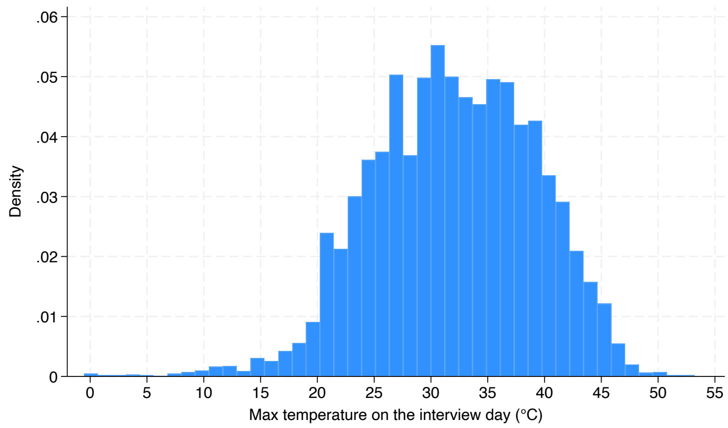
# Figures and Tables

FIGURE 1.1: IHDS: Frequency of Interview Duration



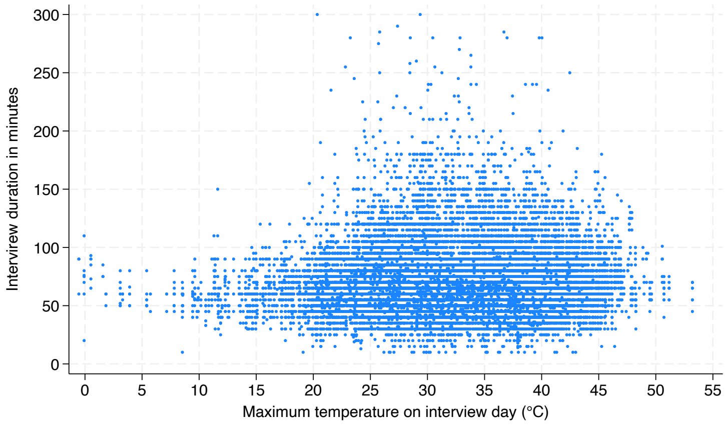

Notes: This Figure plots the frequency of IHDS interview duration

FIGURE 1.2: IHDS: Frequency of Daily Maximum Temperature



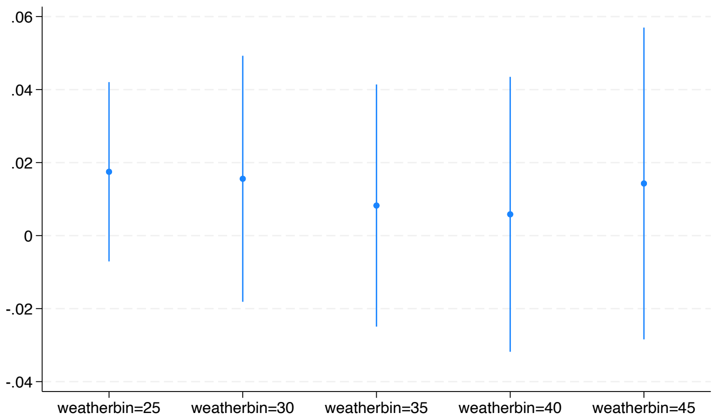

Notes: This Figure plots the frequency maximum temperature on the days of the IHDS interviews.

FIGURE 1.3: IHDS: Maximum Temperature and Interview Duration



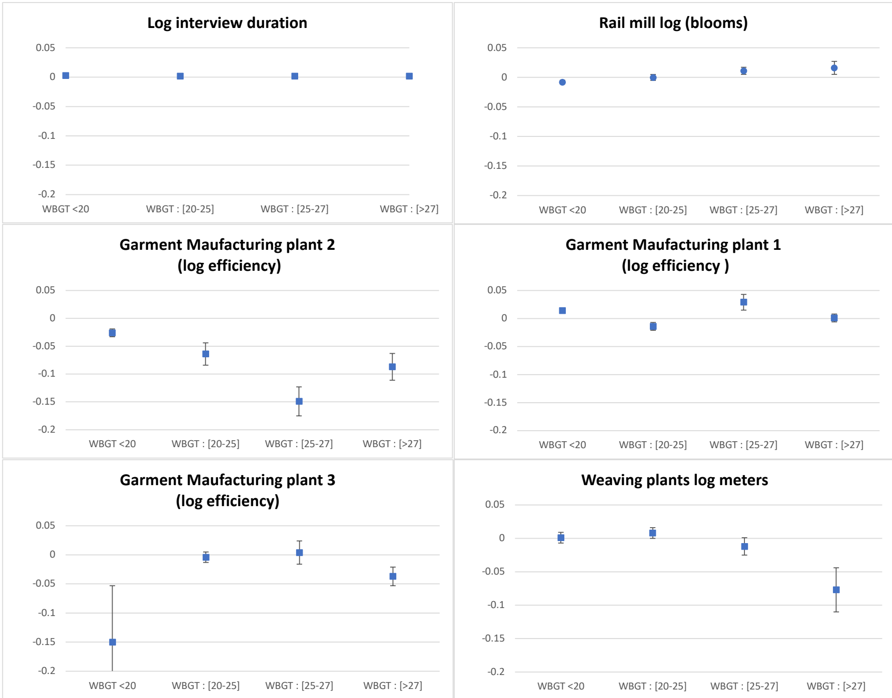

Notes: This Figure visualize the relationship between maximum temperature on the days of the IHDS interviews and interviews duration.

FIGURE 1.4: IHDS: Evidence of a Precise Estimated Zero



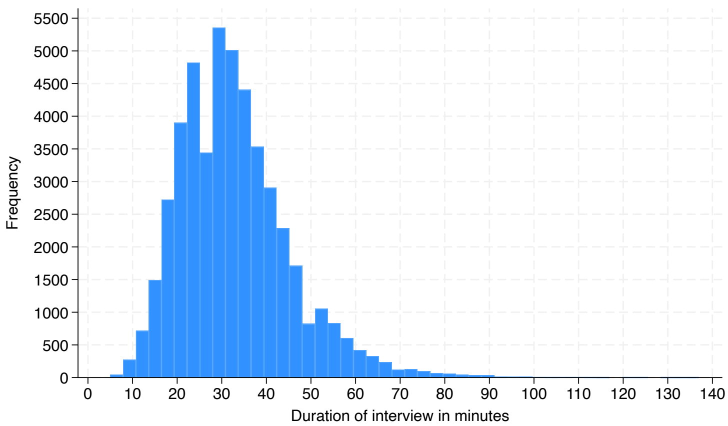

Notes: This Figure shows the coefficient plot of the OLS estimation of Equation (1.3) using the non-linear variant of the model by replacing the continuous temperature regressor in equation 1.3 by a series of indicator variables corresponding to bins of width 5°C. Here, the coefficient of each bin measures the percentage change in total interview duration that occurs because of temperature falls into that bin rather than the reference bin of 0-25°C. Estimates are all centered around zero

FIGURE 1.5: IHDS: Comparison with (Somanathan et al. 2015)



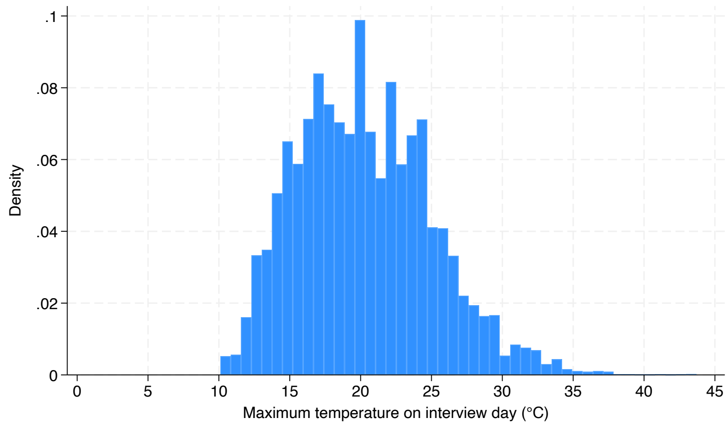

Notes: The top right panel shows the coefficient plot of the OLS estimation of Equation (1.4).WBGT is a widely-used measure of heat stress that combines temperature with relative humidity. The outcome variable is an estimate of the marginal effect of a degree change in WBGT on interview duration within each of the intervals. The rest of the panels reflect the point estimates from Somanathan et al., 2021 where he studied the effect of temperature on productivity in the manufacturing sector including: rail mill, garment manufacturing and weaving plants.

FIGURE 1.6: HILDA: Frequency of Interview Duration



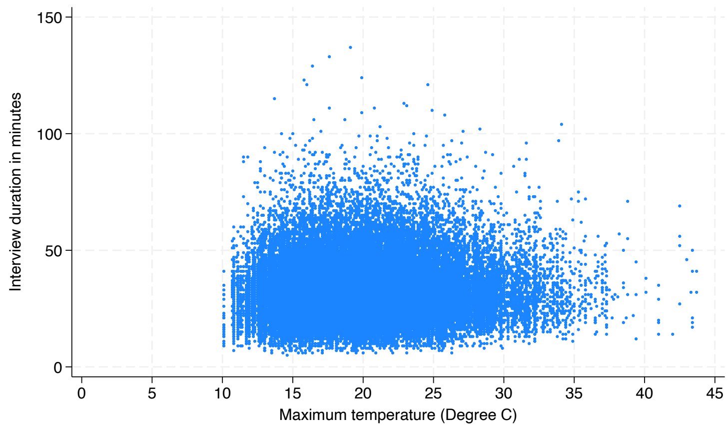

Notes: This Figure plots the frequency of HILDA interview duration

FIGURE 1.7: HILDA: Frequency of Daily Maximum Temperature



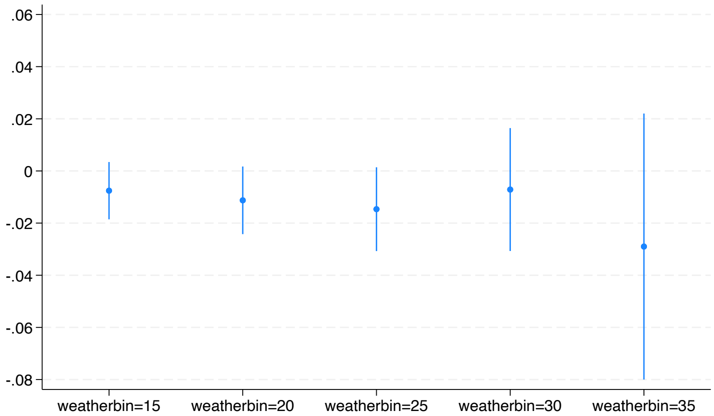

Notes: This Figure plots the frequency maximum temperature on the days of the HILDA interviews.

FIGURE 1.8: HILDA: Scatter Plot of Maximum Temperature and Interview Duration



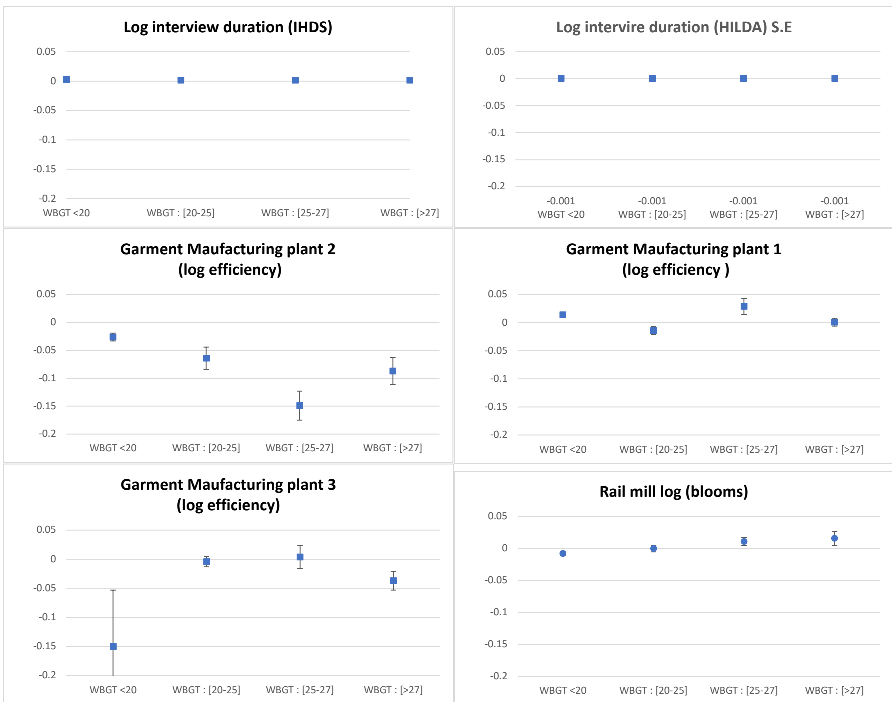

Notes: This Figure visualize the relationship between maximum temperature on the days of the HILDA interviews and interviews duration.

FIGURE 1.9: HILDA: Evidence of a Precise Estimated Zero




Notes: This Figure shows the coefficient plot of the OLS estimation of Equation (1.3) using the non-linear variant of the model by replacing the continuous temperature regressor in equation 1.3 by a series of indicator variables corresponding to bins of width 5°C. Here, the coefficient of each bin measures the percentage change in total interview duration that occurs because of temperature falls into that bin rather than the reference bin of 5-15°C. Estimates are all centered around zero

FIGURE 1.10: HILDA: Comparison with (Somanathan et al. 2015)




Notes: The top right panel shows the coefficient plot of the OLS estimation of Equation (1.4).WBGT is a widely-used measure of heat stress that combines temperature with relative humidity. The outcome variable is an estimate of the marginal effect of a degree change in WBGT on interview duration within each of the intervals. The rest of the panels reflect the point estimates from Somanathan et al., 2021 where he studied the effect of temperature on productivity in the manufacturing sector including: rail mill, garment manufacturing and weaving plants.

TABLE 1.1: IHDS: Summary Statistics

| | Mean | Standard Deviation | Min | Max |
|---|---|---|---|---|
| **Panel A** | | | | |
| Interview duration (Minutes) | 67.325 | 26.335 | 10 | 300 |
| Number of interviews per interviewer | 256 | 135 | 10 | 627 |
| Maximum temperature (°C) | 32.022 | 7.145 | -0.536 | 53.207 |
| Minimum temperature (°C) | 17.614 | 7.131 | -10.283 | 33.783 |
| Relative humidity (fraction) | 0.459 | 0.223 | 0.044 | 0.984 |
| Precipitation (mm) | 1.814 | 5.838 | 0 | 74.86 |
| Wind (m/s) | 2.629 | 1.01 | 0.508 | 8.054 |
| Solar (MJ/m$^2$) | 19.228 | 5.899 | 0.771 | 31.984 |
| Elevation (m) | 186.612 | 1280.157 | -9999 | 3248 |
| WBGT (max temp) | 21.583 | 4.059 | 3.083 | 33.589 |
| WBGT (average temp) | 17.499 | 3.8 | 0.488 | 26.647 |
| **Panel B** | | | | |
| Number of persons in HH | 5.184 | 2.497 | 1 | 38 |
| Number of children | 1.654 | 1.579 | 0 | 17 |
| Number of teens | 0.746 | 0.973 | 0 | 8 |
| Number of adults | 2.783 | 1.372 | 0 | 18 |
| Number of married females | 1.217 | 0.702 | 0 | 8 |
| Number of married males | 1.176 | 0.692 | 0 | 8 |
| Air cooler ownership | 0.128 | 0.334 | 0 | 1 |
| Electric fan ownership | 0.619 | 0.486 | 0 | 1 |
| AC ownership | 0.006 | 0.077 | 0 | 1 |
| Access to electricity | 0.765 | 0.424 | 0 | 1 |
| Total household income | 52,108 | 82,603 | -108,328 | 6,520,261 |
| **Panel C** | | | | |
| Interview starts at AM (share) | | 40% | | |
| Interview starts at PM (share) | | 60% | | |
| Interview ends at AM (share) | | 26% | | |
| Interview ends at PM (share) | | 73% | | |
| Number of observation | | 29,890 | | |

Notes: Total HH income is a constructed variable by IHDS from 50 income sources. Negative income was reported by 9% of the sample and is due to crop failure. The sample incorporates 8 caste /religion groups; Brahmin, High caste, OBC,Dalit, Adivasi, Muslim, Sikh, Jain and Christian.

TABLE 1.2: IHDS: Effect of Maximum Temperature on Interview Duration

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Interview duration in minutes | | | |
| Maximum temperature | 0.028 | 0.028 | -0.012 | -0.012 |
| | (0.073) | (0.073) | (0.087) | (0.087) |
| Lower bound of 95% C.I | -0.116 | -0.116 | -0.182 | -0.183 |
| Higher bound of 95% C.I | 0.172 | 0.172 | 0.158 | 0.159 |
| **FE/Controls:** | | | | |
| Interviewer FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES |
| Ownership of air cooler | | YES | YES | YES |
| Ownership of air conditioner | | YES | YES | YES |
| Access to electricity | | YES | YES | YES |
| Non temp weather controls | | | YES | YES |
| Total HH income | | | | YES |
| Social group (caste/religion) | | | | YES |
| Constant | 86.724*** | 87.314*** | 87.602*** | 88.109*** |
| | (9.157) | (9.037) | (9.171) | (9.184) |
| R-squared | 0.617 | 0.617 | 0.617 | 0.617 |
| Observations | 29,890 | | | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview in °C. The outcome variable reflect interview duration in minutes. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.3: IHDS: Non-linear Effect of Temperature on Interview Duration (Weather Bins)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Log interview duration in minutes | | | |
| **25<Maximum temperature <30** | 0.017 | 0.018 | 0.016 | 0.017 |
| | (0.012) | (0.012) | (0.012) | (0.012) |
| Lower bound of 95% C.I | 0.007 | 0.007 | 0.008 | 0.007 |
| Higher bound of 95% C.I | 0.042 | 0.042 | 0.040 | 0.040 |
| **30< Maximum temperature <35** | 0.016 | 0.016 | 0.012 | 0.013 |
| | (0.017) | (0.017) | (0.017) | (0.017) |
| Lower bound of 95% C.I | 0.018 | 0.018 | 0.021 | 0.020 |
| Higher bound of 95% C.I | 0.049 | 0.049 | 0.045 | 0.045 |
| **35< Maximum temperature <40** | 0.008 | 0.009 | 0.004 | 0.004 |
| | (0.017) | (0.017) | (0.017) | (0.017) |
| Lower bound of 95% C.I | 0.025 | 0.025 | 0.030 | 0.030 |
| Higher bound of 95% C.I | 0.041 | 0.042 | 0.038 | 0.038 |
| **40< Maximum temperature <45** | 0.006 | 0.006 | -0.000 | -0.000 |
| | (0.019) | (0.019) | (0.020) | (0.020) |
| Lower bound of 95% C.I | 0.032 | 0.032 | 0.040 | 0.040 |
| Higher bound of 95% C.I | 0.043 | 0.043 | 0.039 | 0.040 |
| **45< Maximum temperature** | 0.014 | 0.015 | 0.009 | 0.009 |
| | (0.022) | (0.022) | (0.023) | (0.023) |
| Lower bound of 95% C.I | 0.028 | 0.028 | 0.036 | 0.036 |
| Higher bound of 95% C.I | 0.057 | 0.057 | 0.054 | 0.054 |
| **FE/Controls:** | | | | |
| Interviewer FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES |
| Ownership of air cooler | | YES | YES | YES |
| Ownership of air conditioner | | YES | YES | YES |
| Access to electricity | | YES | YES | YES |
| Non temp weather controls | | | YES | YES |
| Total HH income | | | | YES |
| Social group (caste/religion) | | | | YES |
| Constant | 4.563*** | 4.573*** | 4.557*** | 4.573*** |
| | (0.147) | (0.145) | (0.148) | (0.148) |
| R-squared | 0.597 | 0.597 | 0.597 | 0.598 |
| Observations | 29,879 | | | |

Notes: The Table shows results from an OLS estimation of Equation (1.3) using the non-linear variant of the model by replacing the continuous temperature regressor in equation 1.3 by a series of indicator variables corresponding to bins of width 5°C. Here, the coefficient of each bin measures the percentage change in total interview duration that occurs because of temperature falls into that bin rather than the reference bin of 0-25°C. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.4: IHDS: Non-linear Effect of Temperature on Interview Duration (WBGT)

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Log interview duration in minutes | | | |
| WBGT:[<20] | 0.003 | 0.004 | 0.003 | 0.003 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| Lower bound of 95% C.I | -0.002 | -0.002 | -0.004 | -0.004 |
| Higher bound of 95% C.I | 0.009 | 0.010 | 0.010 | 0.010 |
| WBGT:[20-25) | 0.003 | 0.003 | 0.002 | 0.002 |
| | (0.003) | (0.003) | (0.003) | (0.003) |
| Lower bound of 95% C.I | -0.002 | -0.002 | -0.004 | -0.004 |
| Higher bound of 95% C.I | 0.008 | 0.008 | 0.008 | 0.008 |
| WBGT:[25-27) | 0.003 | 0.003 | 0.002 | 0.002 |
| | (0.002) | (0.002) | (0.003) | (0.003) |
| Lower bound of 95% C.I | -0.002 | -0.002 | -0.004 | -0.004 |
| Higher bound of 95% C.I | 0.007 | 0.007 | 0.008 | 0.008 |
| WBGT: (> 27) | 0.002 | 0.002 | 0.002 | 0.002 |
| | (0.002) | (0.002) | (0.003) | (0.003) |
| Lower bound of 95% C.I | -0.002 | -0.002 | -0.004 | -0.004 |
| Higher bound of 95% C.I | 0.007 | 0.007 | 0.007 | 0.007 |
| **FE/Controls:** | | | | |
| Interviwer FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES |
| Ownership of air cooler | | YES | YES | YES |
| Ownership of air conditioner | | YES | YES | YES |
| Acess to electricity | | YES | YES | YES |
| Non temp weather controls | | | YES | YES |
| Total HH income | | | | YES |
| Social group (caste/religion) | | | | YES |
| Constant | 5.281*** | 5.289*** | 5.295*** | 5.307*** |
| | (0.335) | (0.333) | (0.334) | (0.332) |
| R-squared | 0.597 | 0.597 | 0.597 | 0.597 |
| Observations | 29,890 | | | |

Notes: The Table shows results from an OLS estimation of Equation (1.4) following the specification of Somanathan et al., 2021. WBGT is a widely-used measure of heat stress that combines temperature with relative humidity. The outcome variable is an estimate of the marginal effect of a degree change in WBGT on interview duration within each of the intervals. For example, above 27 degrees, a one degree change in WBT is associated with an increase in interview duration of 0.2% The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.5: HILDA: Summary Statistics

| | Mean | Standrad Deviation | Min | Max |
|---|---|---|---|---|
| Interview duration (Minutes) | 32.92 | 12.47 | 5 | 137 |
| Number of interviews per interviewer | 582 | 344 | 14 | 1212 |
| Maximum temperature (°C) | 20.30 | 4.75 | 10.10 | 43.70 |
| Relative humidity (fraction) | 64.93 | 13.42 | 13.88 | 98 |
| WBGT (maximum temp) | 18.25 | 18.25 | 11.05 | 36.24 |
| Number of persons in HH | 2.94 | 1.48 | 1 | 11 |
| HH gross income (categories) | 9.03 | 3.04 | 1 | 14 |
| Number of observation | | 47,560 | | |

Notes: Total HH income categorical variable by HILDA.

TABLE 1.6: HILDA: Effect of maximum Temperature on Interview Duration

| | (1) | (2) | (3) |
|---|---|---|---|
| | Interview duration in minutes | | |
| Maximum temperature | -0.018 | -0.021 | -0.025 |
| | (0.018) | (0.017) | (0.017) |
| Lower bound of 95% C.I | -0.053 | -0.054 | -0.058 |
| Higher bound of 95% C.I | 0.017 | 0.013 | 0.009 |
| **FE/Controls:** | | | |
| Interviwer FE | YES | YES | YES |
| Greater city # Month FE | YES | YES | YES |
| Year FE | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES |
| Type of interviewee | | YES | YES |
| Total HH income | | | YES |
| Constant | 41.355*** | 40.568*** | 41.806*** |
| | (1.952) | (1.818) | (2.330) |
| R-squared | 0.314 | 0.343 | 0.339 |
| Observations | | 47,560 | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview in °C. The outcome variable reflect interview duration in minutes. The standard errors in parenthesis are clustered at statistical area code. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.7: HILDA: Non-linear Effect of Temperature on Interview Duration (Bins)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Log interview duration in minutes | | |
| **15<Maximum temperature <20** | -0.006 | -0.007 | -0.008 |
| | (0.006) | (0.006) | (0.006) |
| Lower bound of 95% C.I | -0.017 | -0.018 | -0.019 |
| Higher bound of 95% C.I | 0.005 | 0.004 | 0.003 |
| **20< Maximum temperature <25** | -0.008 | -0.010 | -0.011* |
| | (0.007) | (0.007) | (0.007) |
| Lower bound of 95% C.I | -0.021 | -0.023 | -0.024 |
| Higher bound of 95% C.I | 0.005 | 0.003 | 0.002 |
| **25< Maximum temperature <30** | -0.013 | -0.013 | -0.015* |
| | (0.008) | (0.008) | (0.008) |
| Lower bound of 95% C.I | -0.029 | -0.029 | -0.031 |
| Higher bound of 95% C.I | 0.004 | 0.003 | 0.001 |
| **30< Maximum temperature <35** | -0.003 | -0.005 | -0.007 |
| | (0.012) | (0.012) | (0.012) |
| Lower bound of 95% C.I | -0.027 | -0.028 | -0.031 |
| Higher bound of 95% C.I | 0.021 | 0.019 | 0.016 |
| **35< Maximum temperature** | -0.019 | -0.021 | -0.029 |
| | (0.027) | (0.026) | (0.026) |
| Lower bound of 95% C.I | -0.072 | -0.072 | -0.080 |
| Higher bound of 95% C.I | 0.034 | 0.029 | 0.022 |
| **FE/Controls:** | | | |
| Interviwer FE | YES | YES | YES |
| Greater city # Month FE | YES | YES | YES |
| Year FE | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES |
| Type of interviewee | | YES | YES |
| Total HH income | | | YES |
| Constant | 3.691*** | 3.669*** | 3.602*** |
| | (0.056) | (0.052) | (0.052) |
| R-squared | 0.330 | 0.353 | 0.360 |
| Observations | | 47,560 | |

Notes: The Table shows results from an OLS estimation of Equation (1.3) using the non-linear variant of the model by replacing the continuous temperature regressor in equation 1.3 by a series of indicator variables corresponding to bins of width 5°C. Here, the coefficient of each bin measures the percentage change in total interview duration that occurs because of temperature falls into that bin rather than the reference bin of 5-15°C. The standard errors in parenthesis are clustered at statistical area code. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.8: HILDA: Non-linear Effect of Temperature on Interview Duration (WBGT)

| | (1) | (2) | (3) |
|---|---|---|---|
| | Log interview duration in minutes | | |
| WBGT:[<20] | -0.001 | -0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| Lower bound of 95% C.I | -0.003 | -0.003 | -0.004 |
| Higher bound of 95% C.I | 0.001 | 0.001 | 0.001 |
| WBGT:[20-25) | -0.001 | -0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| Lower bound of 95% C.I | -0.003 | -0.003 | -0.003 |
| Higher bound of 95% C.I | 0.001 | 0.001 | 0.001 |
| WBGT:[25-27) | -0.001 | -0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| Lower bound of 95% C.I | -0.003 | -0.003 | -0.003 |
| Higher bound of 95% C.I | 0.001 | 0.001 | 0.000 |
| WBGT: (> 27) | -0.001 | -0.001 | -0.001 |
| | (0.001) | (0.001) | (0.001) |
| Lower bound of 95% C.I | -0.002 | -0.002 | -0.003 |
| Higher bound of 95% C.I | 0.001 | 0.001 | 0.001 |
| **FE/Controls:** | | | |
| Interviwer FE | YES | YES | YES |
| Greater city # Month FE | YES | YES | YES |
| Year FE | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES |
| Type of interviewee | | YES | YES |
| Total HH income | | | YES |
| Constant | 3.698*** | 3.680*** | 3.617*** |
| | (0.060) | (0.056) | (0.055) |
| R-squared | 0.330 | 0.353 | 0.360 |
| Observations | | 47,560 | |

Notes: The Table shows results from an OLS estimation of Equation (1.4) following the specification of Somanathan et al., 2021. WBGT is a widely-used measure of heat stress that combines temperature with relative humidity. The outcome variable is an estimate of the marginal effect of a degree change in WBGT on interview duration within each of the intervals. For example, above 27 degrees, a one degree change in WBT is associated with an increase in interview duration of 0.1%. The standard errors in parenthesis are clustered at statistical area code. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

# Appendix

TABLE 1.9: IHDS: Effect of Maximum Temperature Interacted with Interview Start Time on Interview Duration

| | (1) | (2) |
|---|---|---|
| | Interview duration in minutes | |
| Maximum temperature | -0.012 | -0.019 |
| | (0.087) | (0.087) |
| Lower bound of 95% C.I | -0.183 | -0.190 |
| Higher bound of 95% C.I | 0.159 | 0.151 |
| Interviews started at PM | | -1.479 |
| | | (1.076) |
| Lower bound of 95% C.I | | -3.595 |
| Higher bound of 95% C.I | | 0.637 |
| Interviews started at PM # maximum temperature | | 0.015 |
| | | (0.031) |
| Lower bound of 95% C.I | | -0.047 |
| Higher bound of 95% C.I | | 0.076 |
| **FE/Controls:** | | |
| Interviewer FE | YES | YES |
| District FE | YES | YES |
| Year FE | YES | YES |
| Month FE | YES | YES |
| Number of persons in HH FE | YES | YES |
| Ownership of air cooler | YES | YES |
| Ownership of air conditioner | YES | YES |
| Access to electricity | YES | YES |
| Non temp weather controls | YES | YES |
| Total HH income | YES | YES |
| Social group (caste/religion) | YES | YES |
| Constant | 88.238*** | 90.217*** |
| | (9.182) | (9.385) |
| R-squared | 0.617 | 0.618 |
| Observations | 29,890 | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview in °C interacted with a dummy variable that equal one if the interview started in the PM and zero if the interview started in the AM. The outcome variable reflect interview duration in minutes. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.10: IHDS: Effect of Maximum Temperature on Cumulative Daily Interview Duration

| | (1) | (1) | (2) | (3) |
|---|---|---|---|---|
| | Total daily nterview duration in minutes | | | |
| Maximum temperature | 0.038 | 0.037 | -0.021 | -0.021 |
| | (0.087) | (0.087) | (0.104) | (0.104) |
| Lower bound of 95% C.I | -0.134 | -0.135 | -0.225 | -0.226 |
| Higher bound of 95% C.I | 0.210 | 0.209 | 0.184 | 0.184 |
| **FE/Controls:** | | | | |
| Interviewer FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES |
| Ownership of air cooler | | YES | YES | YES |
| Ownership of air conditioner | | YES | YES | YES |
| Access to electricity | | YES | YES | YES |
| Non temp weather controls | | | YES | YES |
| Total HH income | | | | YES |
| Social group (caste/religion) | | | | YES |
| Constant | 92.175*** | 92.637*** | 92.470*** | 94.048*** |
| | (7.631) | (7.364) | (7.644) | (7.668) |
| R-squared | 0.717 | 0.717 | 0.717 | 0.717 |
| Observations | 9,372 | | | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview. The outcome variable reflect total daily interview duration in minutes. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.11: IHDS: Effect of Maximum Temperature Interacted with Heat Adaptation Device Time on Interview Duration

| | (1) | (2) | (3) |
|---|---|---|---|
| | Air cooler | Air conditioner | Air cooler or/and electric fan |
| | Interview duration in minutes | | |
| Maximum temperature | -0.003 | -0.012 | -0.002 |
| | (0.086) | (0.087) | (0.089) |
| Lower bound of 95% C.I | -0.172 | | |
| Higher bound of 95% C.I | 0.166 | | |
| Air cooler # Maximum temperature | -0.056 | | |
| | (0.080) | | |
| Lower bound of 95% C.I | -0.212 | | |
| Higher bound of 95% C.I | 0.101 | | |
| Air conditioner # Maximum temperature | | -0.063 | |
| | | (0.213) | |
| Lower bound of 95% C.I | | -0.482 | |
| Higher bound of 95% C.I | | 0.356 | |
| Air cooler and/ or Fan # Maximum temperature | | | -0.016 |
| | | | (0.046) |
| Lower bound of 95% C.I | | | -0.107 |
| Higher bound of 95% C.I | | | 0.075 |
| **FE/Controls:** | | | |
| Interviewer FE | YES | YES | YES |
| District FE | YES | YES | YES |
| Year FE | YES | YES | YES |
| Month FE | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES |
| Ownership of air cooler | YES | YES | YES |
| Ownership of air conditioner | YES | YES | YES |
| Access to electricity | YES | YES | YES |
| Non temp weather controls | YES | YES | YES |
| Total HH income | YES | YES | YES |
| Social group (caste/religion) | YES | YES | YES |
| Constant | 87.831*** | 88.220*** | 88.382*** |
| | (9.217) | (9.180) | (9.129) |
| R-squared | 0.617 | 0.617 | 0.617 |
| Observations | 29,890 | | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). In Columns (1),(2) and (3), the independent variable is maximum daily temperature on the day of the interview in °C interacted with a dummy variable that equal one if household owns an air-cooler , air conditioner, or either an air-cooler or an electric fan, respectively. The outcome variable reflect interview duration in minutes. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 1.12: IHDS: Effect of Maximum Temperature on Interview Duration with Alternative Standard Errors

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | District level | Un-clustered | District and month | Interview and month |
| | Interview duration in minutes | | | |
| Maximum temperature | -0.012 | -0.012 | -0.012 | -0.012 |
| | (0.087) | (0.044) | (0.072) | (0.066) |
| Lower bound of 95% C.I | -0.183 | -0.097 | -0.153 | -0.141 |
| Higher bound of 95% C.I | 0.159 | 0.073 | 0.129 | 0.117 |
| **FE/Controls:** | | | | |
| Interviewer FE | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES |
| Ownership of air cooler | YES | YES | YES | YES |
| Ownership of air conditioner | YES | YES | YES | YES |
| Access to electricity | YES | YES | YES | YES |
| Non temp weather controls | YES | YES | YES | YES |
| Total HH income | YES | YES | YES | YES |
| Social group (caste/religion) | YES | YES | YES | YES |
| Constant | 88.109*** | 88.246*** | 88.109*** | 88.109*** |
| | (9.184) | (18.589) | (12.361) | (13.441) |
| R-squared | 0.617 | 0.617 | 0.617 | 0.617 |
| Observations | 29,890 | | | |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview in °C. The outcome variable reflect interview duration in minutes. Columns (1), (2), (3) and (4) reflect standard errors clustered at the district, un-clustered, at district and month, and at interviewer and month, respectively. We also show estimates for the 95 percent confidence intervals. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 1.13: IHDS: Effect of Maximum Temperature on Interview Duration: Robustness Checks

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | Urban districts | Rural districts | learning tests for kids | Small districts | Minimum difference between daily max and min temp | High interview frequency districts | Average temperature as the dependent variable |
| | Interview duration in minutes | | | | | | |
| Maximum temperature | 0.102 | -0.066 | 0.067 | -0.007 | -0.009 | 0.019 | |
| | (0.112) | (0.101) | (0.126) | (0.093) | (0.137) | (0.088) | |
| Lower bound of 95% C.I | -0.119 | -0.265 | -0.180 | -0.190 | -0.278 | -0.155 | |
| Higher bound of 95% C.I | 0.323 | 0.132 | 0.313 | 0.177 | 0.260 | 0.192 | |
| Average temperature | | | | | | | 0.001 |
| | | | | | | | (0.095) |
| Lower bound of 95% C.I | | | | | | | -0.187 |
| Higher bound of 95% C.I | | | | | | | 0.188 |
| **FE/Controls:** | | | | | | | |
| Interviewer FE | YES | YES | YES | YES | YES | YES | YES |
| District FE | YES | YES | YES | YES | YES | YES | YES |
| Year FE | YES | YES | YES | YES | YES | YES | YES |
| Month FE | YES | YES | YES | YES | YES | YES | YES |
| Number of persons in HH FE | YES | YES | YES | YES | YES | YES | YES |
| Ownership of air cooler | YES | YES | YES | YES | YES | YES | YES |
| Ownership of air conditioner | YES | YES | YES | YES | YES | YES | YES |
| Access to electricity | YES | YES | YES | YES | YES | YES | YES |
| Non temp weather controls | YES | YES | YES | YES | YES | YES | YES |
| Total HH income | YES | YES | YES | YES | YES | YES | YES |
| Social group (caste/religion) | YES | YES | YES | YES | YES | YES | YES |
| Constant | 62.590*** | 69.057*** | 99.534*** | 88.252*** | 85.516*** | 86.458*** | 88.720*** |
| | (4.412) | (7.679) | (6.517) | (8.482) | (4.585) | (7.416) | (9.078) |
| R-squared | 0.614 | 0.630 | 0.649 | 0.603 | 0.639 | 0.638 | 0.617 |
| Observations | 9,868 | 19,589 | 7,130 | 27,154 | 14,337 | 26,009 | 29,890 |

Notes: The Table shows results from an OLS estimation of Equation (1.3). The independent variable is maximum daily temperature on the day of the interview in °C. The outcome variable reflect interview duration in minutes. Columns (1) considers only interviews conducted in urban districts, column (2) considers only interviews conducted in rural districts. Column (3) we consider only interviews with kids learning tests. Column (4) reflects interviews conducted in relatively smaller districts. Column (5) uses a subset of 50% of the sample where the difference between daily maximum and minimum temperatures is minimal. Column (6) looks only at districts with more than 200 interviews conducted. Column (7) uses average daily temperature as the independent variable. The standard errors in parenthesis are clustered at the district. We also show estimates for the 95 percent confidence intervals. *** p<0.01, ** p<0.05, * p<0.1.

.

# Chapter 2

# Conditional demand analysis as a tool to evaluate energy policy options on the path to grid decarbonization

## 2.1 Introduction

Pathways to meeting global decarbonization goals typically include increasing the electrification of end-use services in buildings coupled with lowering the carbon intensity of electricity generation (Steinberg et al., 2017; Rogelj et al., 2018). In buildings, which account for over a third of global emissions UN Environment, 2017; Delzendeh et al., 2017, the most promising avenues to support a shift to large-scale electrification include reducing heating and cooling emissions through envelope improvements, high-efficiency technology adoption, and increasing usage of renewable energy (IEA, 2017a; Rogelj et al., 2018; Calderon et al., 2019).

To meet the global community's goal to reach net zero carbon emissions by 2050-2070 (IPCC, 2018), estimates of future end-use energy consumption loads under higher electrification scenarios are needed. In this context, estimates that disaggregate total energy demand into appliance or equipment-level end-uses have been derived using a number of different methodologies.[1] The statistical methods and engineering models developed in this literature can provide a valuable framework to estimate the impact of technological and behavioural changes in buildings, and support policy-making aimed at increasing both electrification and energy conservation in the coming decades.

This paper implements a conditional demand analysis (CDA) using a large dataset of electricity consumers in a Canadian province with a high market share of electric heating technologies. While originally developed using multiple linear regression techniques in economics (Parti and Parti, 1980), CDA has now become widely used in a number of energy-related disciplines. The CDA methodology has also been expanded to include bayesian econometrics and hybrid approaches that incorporate engineering model simulations (Caves et al., 1987; Hsiao, Mountain, and Illman, 1995; Larsen and Nesbakken, 2004).

Our paper provides three contributions to this existing literature. First, we present a unifying review of the breadth of interdisciplinary applications of CDA, beginning

[1]Examples include Parti and Parti, 1980; Bartels and Fiebig, 1990; Hsiao, Mountain, and Illman, 1995; Aydinalp-Koksal and Ugursal, 2008; Swan and Ugursal, 2009; Aranda et al., 2012 and Fumo, 2014. Foucquier et al., 2013 provide a comprehensive review of the breadth of methodological approaches that have been utilized to evaluate this question across several disciplines.

from the earliest studies up to the present. Second, we use econometric techniques that take advantage of a large household-level panel dataset to test for evidence of unobservable variable bias from the random effects model. Unobserved characteristics correlated with both the independent variables of interest and the dependent variable (i.e. consumption at the household level), may result in the consumption *level* effects of variables of interest being estimated with bias. In that case an alternative estimator, the fixed effects or within estimator, is preferable. However, by construction, the fixed effects estimator implies that only the consumption gradient, which measures how consumption for a particular variable changes as outdoor temperature changes, can be estimated. The advantage of the fixed effects model is that it includes fewer identifying assumptions and therefore the model coefficients can be estimated with more confidence (Jessoe, Papineau, and Rapson, 2020). In addition, the gradient itself may be of interest to policymakers as it can offer insight on the peak load demand of individual variables.

Finally, the particular characteristics of our data setting, a Canadian province with a high share of both electrification and low-emission power sources, can help provide insights to researchers and policymakers working on predicting and managing future electrical loads. Canada's electricity grid is among the most decarbonized worldwide: nearly 82 percent of its electricity is derived from non-GHG emitting sources, and in several provinces a significant share of heating demand is met from electricity-using equipment.[2] In New Brunswick, the province studied in this paper, about 60 percent of electricity is produced using non-emitting sources and almost 52 percent of homes heat their homes using electricity (Canada Energy Regulator, 2017; Natural Resources Canada, 2019).[3]

## 2.2 Approaches to estimating appliance energy use

The primary goal of this study is to estimate robust statistical models to evaluate the effects of energy conservation measures (ECMs) in practice, and inform the development of ECM policies in support of climate policy goals. There are three primary methods used to estimate the energy use of individual appliances and savings from specific ECMs. The first is direct sub-metering of individual devices in multiple houses for a representative period. This will yield very high quality data and energy use estimates. However, this method is also very expensive and invasive; as a result its applications have been limited and these typically utilize small sample sizes, which may limit their generalizability (e.g., Isaacs et al., 2010; De Almeida et al., 2011; Parker, 2003; Nelson, Berrisford, and Xu, 2014).

A second approach is engineering estimates. In this case, the basic power draw of an appliance at the important points in its operational cycle are measured (or estimated), and this is conflated with an assumed usage pattern. The usage pattern may come from a defined standard household definition, or may be estimated by other means NRCan, 2019; Larsen and Nesbakken, 2004. This method has relatively low data requirements, but whereas the power measurements may be made unambiguously, the usage patterns may not accurately describe the range of possible patterns in real households.

[2]There are large variations in the share of non-emitting sources across provinces, from about 12 percent in Alberta to close to 100 percent in Québec.

[3]Another 10 percent use dual electric and wood or dual electric and natural gas systems.

The third method, and the one used in this paper, is CDA. This is a statistical technique that relies on household appliance ownership data combined with energy use data, collected via survey or directly from utility billing records. More sophisticated models, as in this paper, may be derived with additional survey data on household socio-economic variables and behaviours. It is relatively inexpensive and non-invasive for householders, and was particularly attractive in this research because the necessary data already existed, having been collected by a utility for a different purpose. In the rest of this section we describe the CDA method and evolution of the literature in more detail.

### 2.2.1 The evolution of conditional demand analysis

#### Early work and the basic model

The CDA model was first introduced in the peer-reviewed literature thirty years ago by Parti and Parti, Parti and Parti, 1980. Their analysis used linear regression techniques on cross-sectional data to estimate average electricity consumed for 16 appliance types. This early work spurred a number of both peer-reviewed studies and utility-sponsored reports (Aigner, Sorooshian, and Kerwin, 1984; Lawrence and Parti, 1984; Archibald, Finifter, and Moody, 1982; Florida Power & Light, 1986; Caves et al., 1987; Berg, 1988; Larsen and Nesbakken, 2004; Tiedemann, 2007; Aranda et al., 2012).

The specification for estimating appliance-specific consumption in most of these studies is similar to the general formulation

$$Y_{it} = \sum_{j=1}^{k} d_{ij} N_{ij} \beta_j + \beta_{k+1} + \varepsilon_{it}, \tag{2.1}$$

where $Y_{it}$ and $\varepsilon_{it}$ are random variables measuring monthly electricity consumption in kWh and an error term for household $i$ in month $t$, respectively.[4] The variable $d_{ij}$ is an indicator for ownership of the $j$-th appliance by household $i$, and $N_{ij}$ is the number of appliance $j$ present in household $i$. The coefficients $\beta_j$ are interpreted as the estimated consumption of the $j$-th appliance. Finally, $\beta_{k+1}$ is the average consumption of the set of all other unspecified appliances.[5]

Some studies have extended this model to allow $\beta_j$ to vary with socioeconomic, demographic, economic or physical variables such as weather (Parti and Parti, 1980; Lafrance and Perron, 1994; Larsen and Nesbakken, 2004). However, an alternative approach is to directly control for other observable explanatory variables as separate coefficients in the regression or interacted with weather variables (Fiebig, Bartels, and Aigner, 1991; Jessoe, Papineau, and Rapson, 2020). In our empirical specification described below we will estimate different extensions of equation (2.1) and incorporate the effect of weather on consumption.

Another version of these foundational models utilizes cross-sectional data and assumes that the appliance dummy variables are a function of a random error term to account for variations in the intensity of appliance use, size or capacity among different households. The random error results in a heteroskedastic regression error variance that must be estimated using feasible generalized least squares (Judge et al.,

[4]Some studies have used a variation of this specification with indoor temperature as a dependent variable Krüger and Baruch, 2004.

[5]If time-series data on energy consumption for each household is not available, equation (2.1) can simply be re-written without the $t$ subscripts.

1985). This is the approach taken by Fiebig, Bartels, and Aigner, 1991 in a study of 380 households in New South Wales, Australia.

Previous literature reviews have synthesized the available estimates from early CDA studies. Sebold and Parris, 1989 reviewed 30 separate US CDA studies conducted in the 1980s, and Lawrence and Parti, 1984 reviewed 15 separate US CDA studies conducted prior to these, with sample sizes typically of several thousand households. It is also noteworthy that the well-established US Residential Energy Consumption Survey (RECS) uses CDA to generate individual end-use estimates Battles, 1994.

In Canada, Newsham and Donnelly, 2013 used data from a comprehensive national energy use survey from 9,773 households to derive estimates of the energy savings associated with upgrading appliances, lighting, and enacting thermostat setbacks. Aydinalp-Koksal and Ugursal, 2008 applied CDA to an earlier vintage of the national Canadian dataset used by Newsham and Donnelly, 2013, although electricity and natural gas data were available for only 2,050 and 1,012 households, respectively. Tiedemann, 2007 used CDA on data from 791 households in British Columbia to develop a model of electricity use. Lafrance and Perron, 1994 used CDA to study the change in major energy end-uses in Quebec using three large samples of households (42,000, 24,000, and 46,000) five years apart.

#### Hybrid modeling and other innovations

Researchers also developed hybrid CDA models that incorporate engineering and direct metering approaches. The first example of this approach is Caves et al., 1987, who used a Bayesian updating model combining engineering estimates of appliance usage as the prior beliefs and CDA estimates to transform the priors into a posterior distribution of appliance usage. Two other Bayesian analyses, Bauwens, Fiebig, and Steel, 1994 and Hsiao, Mountain, and Illman, 1995, integrated direct appliance metering data and CDA estimates. Bartels and Fiebig, 1990; Bartels and Fiebig, 2000 also combined metering and CDA analysis using an econometric framework that was shown by Bartels and Fiebig, 1990 to be equivalent to the Bayesian approach. The authors found that this combined approach led to statistical efficiency gains whereby the appliance coefficients could be estimated with more precision even in relatively small samples.

While the existing CDA literature has expanded to include more advanced statistical techniques such as Bayesian updating, hybrid approaches, machine learning and neural network modeling (Foucquier et al., 2013), one as yet unstudied but important question is the extent of coefficient bias due to omitted variables (Hausman, 1978). We turn to this question in the following section.

### 2.2.2 Panel data estimators for more robust statistical inference

The now-ubiquitous availability of repeated cross-sections of consumption data from a sample of utility customers, also known as panel data, has enabled researchers to model statistical relationships using an 'error components' framework (Baltagi, 2008). The error components model allows for research designs whereby individual units differ in unobservable, time-constant ways that affect outcomes of interest. For example, in the short-run when consumer income levels are constant, unobserved customer income in any given sample may be correlated with both appliance use and total energy consumption, resulting in biased coefficient estimates. The same can be said for many other persistent consumer or household-level characteristics such as political ideology or environmental awareness (Costa and Kahn, 2013a).

The error components model is a simple modification of equation (2.1):

$$Y_{it} = \sum_{j=1}^{k} d_{ij} N_{ij} \beta_j + \beta_{k+1} + c_i + u_{it}, \tag{2.2}$$

where the error $\varepsilon_{it} = c_i + u_{it}$ now includes both an individual-specific, time-constant component ($c_i$) and an individual and time-varying term ($u_{it}$). Panel data estimates in the current CDA literature typically adopt the identifying assumptions of the random effects model, in which the $d_{ij}$ and $N_{ij}$ terms are assumed uncorrelated with the composite error term $\varepsilon_{it}$. This is a strong assumption that is unlikely to be met in most samples that are derived from realized energy consumption and survey data.[6]

An alternative panel model that allows researchers to relax this uncorrelatedness assumption is the fixed effects model. This model allows for correlations between the independent variables of interest and the time-constant error component $c_i$, and therefore assume that only $u_{it}$ is independent of the included right hand side variables. In other words, any time-invariant unobserved variables over the sample period are accounted-for in the fixed effects model. The question of whether the random effects or fixed effects model is most appropriate to obtain unbiased coefficient estimates can be answered using the Hausman (Hausman, 1978) and Mundlak tests (Mundlak, 1978). We will test this hypothesis in Section 2.5.1.

One potential drawback of the fixed effects model is that the procedure to eliminate bias from the unobserved $c_i$ results in the inability of the statistical model to identify the average effect of an appliance on the *level* of energy consumption, $Y_{it}$, as time-invariant variables cannot be identified. However, the consumption gradient, or how consumption changes as temperature changes, can be identified using fixed effects panel estimators. The gradient effects themselves are important from a policy perspective as they indicate potential electricity demand during peak demand periods, and potential peak savings from different conservation measures Papineau, 2019.

## 2.3 Data

### 2.3.1 Energy and survey data

The raw data for this analysis came from the 2017 administration of an Energy Planning Survey conducted by NB Power, the primary electric utility in the Canadian province of New Brunswick. Respondents from 6,941 households completed the survey. The survey consisted of more than 80 questions on dwelling and occupant characteristics, household appliances, space conditioning and water heating equipment, fuel usage, and behaviours related to energy consumption. Since we are interested in electricity-using appliances and measures that affect total household energy use, the analysis in this paper focuses only on households reporting their primary heating fuel as being electricity, close to 60 percent of surveyed households. New Brunswick has cold winters and warm but relatively mild summers. In this climate the biggest single energy use is for space heating, and space heating effects will only be apparent in the dependent variable if the space heating is via electric equipment.

[6] The unobserved income and consumer characteristics from the previous section offer some examples, though there are many other unobserved characteristics this could apply to, including number of occupants or household 'comfort' preferences.

The primary electric heating equipment in the sample includes both electric baseboards and heat pumps.

Total household electricity use data were obtained from utility records for each billing period spanning approximately one year for most households (12 billing periods). As is common with utility data, the billing periods were approximately one month long but were not typically aligned with the start and end of calendar months, nor were they all the same between households. We divided the energy use in a given billing period by the number of days in that billing period to account for (small) variations in the number of days per billing period between households. The full sample period spans April 1, 2017 to March 31, 2018.

The survey was conducted by the utility for its own planning purposes, and the authors had no role in designing data collection methods. Anonymized data were shared with geographical information denoting the first three digits of the provincial postal code, also known as the "forward sortation area", enough geographical information to facilitate matching with appropriate weather data. Given that over 99 percent of the customers in our sample are served either directly or indirectly from a single utility we assume that all residential customers were exposed to the same prices, and thus price is not a factor in the model.[7]

TABLE 2.1: Summary Statistics

| | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|
| Annual electricity usage (kWh) | 18847.28 | 6807.31 | 3,106 | 41,320 |
| Single detached home | 0.75 | 0.43 | 0 | 1 |
| House size (ft2) | 1514.86 | 589.44 | 600 | 2,700 |
| House age (years_ | 3.41 | 1.55 | 1 | 7 |
| Number of occupants | 2.25 | 1.04 | 1 | 6 |
| Electric water heater | 0.96 | 0.18 | 0 | 1 |
| Number of dehumidifiers | 0.59 | 0.57 | 0 | 3 |
| Well pump | 0.37 | 0.48 | 0 | 1 |
| Pool pump | 0.07 | 0.26 | 0 | 1 |
| Electric baseboard | 0.59 | 0.49 | 0 | 1 |
| Local heat pump | 0.29 | 0.46 | 0 | 1 |
| Central heatpump | 0.11 | 0.32 | 0 | 1 |
| Number of window AC | 0.27 | 0.58 | 0 | 4 |
| Thermostat setback | 0.54 | 0.50 | 0 | 1 |
| Energy related renovations | 1.09 | 1.53 | 0 | 6 |
| Observations | | 3,214 | | |

Notes: House age was measured by the following seven year built ranges: before 1961, 1961–1974, 1975–1989, 1990–1999, 2000–2009, 2010–2015, and after 2015. The house age variable was calculated by taking the average of each customer's year built range, subtracting that from 2017, and averaging those values. Values above 2015 were recorded as 2016 and values before 1961 recorded as 1955.

Our final data set after omitting non-electrically heated homes and cleaning the data includes 3,214 households.[8] The data summary statistics are presented in Table 2.1. The annual electricity consumption (in kWh) is calculated from utility billing

[7]Close to 90 percent of the customers in our sample are served directly by NB Power, and two other utilities comprising about nine and one percent of the sample, respectively, purchase their wholesale electricity from the same large utility, resulting in almost identical pricing across the sample.

[8]Appendix A provides further details on our data cleaning approach.

records by aggregating twelve billing periods (April 2017-March 2018) for each household. The remaining variables on house age, size, appliance ownership, energy-related renovations and thermostat setback practices are obtained from the household survey data. Central heat pumps use an internal duct system to distribute conditioned air around the the entire house volume, whereas local heat pumps have an outlet in one or more zones.[9]

The survey asked households whether they turn down their thermostat in the winter while sleeping; when no one is home; both; or never. The thermostat setback variable in Table 2.1 indicates the share of households reporting any one of the setback practices. Almost three quarters of households report undertaking some form of thermostat setback.

Energy related renovations represents the response to a series of survey questions, "Please indicate if you have recently completed the following: insulate your basement, add insulation to your attic, add insulation to your walls, replace exterior doors or windows, install LED lights, install weather-stripping or caulking." The reported variable is simply the total number of measures the homeowner reported completing. It take on values ranging from 0 to 6 depending on the number of renovations undertaken. On average households in the sample had undertaken about one energy-related renovation.

### 2.3.2 Climate data

For household energy use related to space heating and cooling, we introduced a dependency on local climate data. The standard indices used in CDA and in building energy studies more generally are heating-degree-days (HDD) and cooling-degree days (CDD), typically to a base temperature of 18°C, which is considered the typical outdoor temperature at which there is a transition from space heating to space cooling ASHRAE, 2009.[10]

HDD, which measure demand for space heating services, are calculated by subtracting the average Celsius temperature on a given day from 18 on days with temperatures below 18°C. CDD measure demand for space cooling services and are calculated by subtracting 18 from the average Celsius temperature on a given day with temperatures above 18°C.[11] Our data source for these variables was daily weather data from Environment Canada Environment Canada, 2020. Further details on our construction of the heating and cooling degree day variables can be found in Appendix A.

## 2.4 Empirical framework

In this study, monthly energy use and climate data were available, enabling panel regression. With data on a large sample of houses, equation (2.2) can be solved using

[9]While we have no data on how many zones were served in each house, other studies of similar heat pumps in North America (e.g., Baylon, Storm, and Robison, 2013 Table 6), a large majority service only one zone directly.

[10]Prior studies have used various base temperatures, and there is recognition that there is a wide range of appropriate base temperatures between buildings, depending on their specific construction and operational characteristics. Nevertheless, for large building populations where specific characteristics are not known, 18°C persists in being the most common choice for base temperature, and HDD and CDD to this base are the most widely available in published climate data **Zhang2019**.

[11]Temperatures of 18 degrees would be recorded as having zero HDD and CDD, though our temperature data includes decimals and no temperatures of exactly 18 degrees were recorded during our sample period.

either random effects or fixed effects estimators. Regression relies on variability in the explanatory variables. For example, considering data on whether a household has a refrigerator or not, since almost every modern household has a refrigerator of some kind, there is almost no variability in this input. The consequence is that it will be mathematically impossible to separate out refrigerator energy use from other end uses. One way around this, for some variables, is to have additional data, such as total number of refrigerators in the household, which introduces variability through the variable $N_{ij}$ in equations (2.1) and (2.2).

Assuming sufficient variability is observed for a given set of variables, the coefficient estimates will depend on the quality of the raw data, and as with most survey-based data sources, householders might not be completely accurate in their responses to the survey questions, or might not interpret the questions in a universal manner. This may introduce noise into the data and if present will reduce the absolute value of any affected coefficients towards zero.

### 2.4.1 Model development

We followed the process outlined in Aydinalp-Koksal and Ugursal, 2008 by first considering logical approaches to modelling individual appliances from engineering principles and the data available. Then we combined these together into a CDA model. Space heating/cooling appliances, energy efficient behaviours of interest to policy makers (thermostat setback, renovations) and household characteristics shown to be influential in prior studies (e.g. house size, number of occupants) were included. Since the fixed effects model does not separately identify time-invariant variables, we interacted most of our variables of interest with climate data (heating and cooling degree days).

We began with a model that included all potential predictors from the survey data, and then began to sequentially remove predictors from the model where predictors offered little variability, or where the predictor did not significantly contribute to the explanatory power of the model, as measured by $R^2$. This process yields the most parsimonious and interpretable model, and is common practice in CDA studies. For example, clothes washers, clothes driers, and electric cookers were dropped as their ownership was almost universal, and was limited to one per house, whereas central air conditioning was dropped because very few households reported ownership (only 35 households).

### 2.4.2 Empirical specifications

In this section we present our econometric specifications to estimate the electricity consumption effect of specific house characteristics, appliance ownership, space thermal conditioning equipment and behaviours intended to reduce energy use. We also explain the different model identifying assumptions.

#### Traditional CDA random-effects specification

The traditional panel data CDA model consists of estimating equation (2.1) using the random effects estimator (Wooldridge, 2010), in which the coefficients of interest are estimated without bias if both the strict exogeneity and uncorrelated effects assumptions hold. Strict exogeneity assumes that after conditioning on all the independent variables of interest available to the researcher ($d_{ij}$ and/or $N_{ij}$), the error term $\varepsilon_{it}$ is uncorrelated with these variables as well as any unobserved variables correlated with the variables of interest. Uncorrelated effects assumes that after conditioning on the available variables, any unobserved individual-level and time-constant variables that affect energy consumption (represented by $c_i$), such as income in the short-run, are uncorrelated with the variables of interest.

With a rich set of explanatory variables to control for, the strict exogeneity assumption will frequently be assumed to be satisfied. However, the uncorrelatedness assumption effectively assumes that any included variables are orthogonal to the composite error term $\varepsilon_{it} = c_i + u_{it}$, which includes an individual-level, time-constant unobserved variable $c_i$ that is likely to be correlated with electricity consumption, as previously noted in Section 2.2.2. If this is the case, the estimated variables in the random effects model will be biased and a fixed effects estimator, which is identified with less stringent assumptions, should be used instead.[12]

Since the fixed effects estimator explicitly controls for all time-invariant factors that could affect electricity consumption, the coefficients on variables that do not vary over time cannot be separately identified or estimated. However, an alternative approach that has been taken in prior studies to overcome this constraint is to estimate how variables of interest vary in response to weather changes, by interacting these variables with heating- and cooling degree-days (Jessoe, Papineau, and Rapson, 2020; Kotchen, 2017). We proceed below using this approach, first by estimating weather-interacted variables using a random effects estimator, followed by estimating the same model using the fixed effects estimator. With the coefficient estimates for both sets of estimators in hand, we are then able to conduct two different statistical tests to evaluate the evidence for the presence of omitted variable bias in the random effects specification.

We estimate the following random-effects specification:

$$Y_{it} \quad = \quad \alpha + \sum_{j=1}^{k} \beta_{1j} d_{ij} N_{ij} * AHDD_{zt} + \sum_{j=1}^{k} \beta_{2j} d_{ij} N_{ij} * ACDD_{zt} + \eta_t + \varepsilon_{it} \quad (2.3)$$

The outcome variable $Y_{it}$ in equation (2.3) is electricity consumption in kWh for household i in billing period t, and $\alpha$ is the regression constant term. As explained in section 2.2.1 of the paper, the variable $d_{ij}$ is an indicator for the j-th appliance, behavior or characteristic of household i, while $N_{ij}$ is the number of units of $d_j$ present in household i.

Our reported model incorporates a total of seven primary variables of interest. Four indicator variables take the value of 1 if household i owns the following appliances and zero otherwise: pool-pump and three options for households' primary heating systems. The three heating systems are electric baseboard, local heat pump (also known as ductless or minisplit heat pumps), and central air source heat pump. Since their ownership are mutually exclusive, we omit electric baseboard heating systems from the analysis so that the econometric results are interpreted relative to this omitted category. This is the logical choice for results interpretation because heat pumps are often proposed as a more efficient replacement for baseboards. In addition, the model has three independent variables that reflect the number of the following appliances: dehumidifiers, window air-conditioners, energy-related renovations.

[12]This is the approach taken by Azam, 2016 in the context of evaluating links between economic growth and environmental degradation.

The variables $AHDD_{zt}$ and $ACDD_{zt}$ are average daily cooling and heating degree days for a household in billing month t and located in weather region z. These two variables were obtained by dividing our heating and cooling degree day variables ($HDD_{zt}$ and $CDD_{zt}$, respectively) by the number of days in the monthly billing period. We interact our regressors of interest with $AHDD_{zt}$ and $ACDD_{zt}$ to estimate how electricity usage changes in response to a 1 unit change in $AHDD_{zt}$ or $ACDD_{zt}$. $\beta_1$ and $\beta_2$ when $d_{ij} = 1$ and $N_{ij} = 1$ are interpreted as the change in monthly usage (demand) of electricity for households who possess a certain appliance relative to those who do not, in response to a one unit increase in $AHDD_{zt}$ and $ACDD_{zt}$, respectively. If $d_{ij} = 1$ and $N_{ij} > 1$, $\beta_1$ and $\beta_2$ are interpreted as the additional electricity demand from one extra unit of the appliance in response to a one-unit change in $AHDD_{zt}$ and $ACDD_{zt}$, respectively.

We control for billing period as denoted by $\eta_t$, and $\varepsilon_{it} = c_i + u_{it}$ is the error term that incorporates the household-specific random effects $c_i$, which are not explicitly estimated in the random effects model.

### Fixed-effects model

We also estimate a fixed-effects model where the uncorrelatedness assumption of the random-effects specification is relaxed. In other words, the fixed-effects estimator accounts for the impact of any time-invariant characteristics that may be correlated with the independent variables.

Statistically, the key distinction between fixed-effects and random-effects specifications is whether we model the correlation between the individual effects, $c_i$, and the covariates ($d_{ij}$ and/or $N_{ij}$), or whether we assume that they are independent. In fixed-effects models we explicitly estimate $c_i$, a household-specific coefficient that captures the effect of unobserved, time-invariant household characteristics.

Hence, to estimate the effect of energy conservation measures such as appliance ownership, space thermal conditioning equipment and behaviours, which are all time invariant variables, we estimate the fixed-effects specification by interacting these variables with heating and cooling degree days:

$$Y_{it} = \alpha + \sum_{j=1}^{k} \beta_{1j} d_{ij} N_{ij} * AHDD_{zt} + \sum_{j=1}^{k} \beta_{2j} d_{ij} N_{ij} * ACDD_{zt} + \eta_t + c_i + u_{it}, \quad (2.4)$$

The regressors of interest and interpretation of $\beta_1$ and $\beta_2$ are similar to the random effect specification defined above in equation (2.3).

To evaluate whether the fixed-effects or random-effects model is most appropriate in our setting, we use the Mundlack test (Mundlak, 1978) as well as the Hausman endogeneity test (Hausman, 1978). These results are reported in the following section.

## 2.5 Results

For our empirical analysis we have a balanced panel data set of 12 billing periods among 3,214 households, for a total of 38,568 observations. In this section we report the results of estimating models using both fixed effects and random effects estimators, then implement the Mundlak and Hausman tests to evaluate whether the random effects specification is likely to be estimated without bias. With these results about model choice in hand, we then move on to discuss specific coefficient estimates in more detail as well as their implications for energy policy in the context of grid decarbonization.

Table 2.2 presents results from estimating the change in electricity usage that results from home characteristics and energy conservation measures such as appliance ownership, space thermal conditioning equipment and household behaviours in response to cooler/warmer temperature. The two different panel estimators introduced in Section 2.4.2 are implemented here: column (1) of Table 2.2 presents results from the estimation of equation (2.3), a random-effects model, and column (3) displays estimates from the estimation of equation (2.4), a fixed effects model. Both models also control for billing period as well as occupancy interactions with cooling and heating degree days, where the latter are coded as dummy variables interactions for each occupancy value. In the random effects model we also incorporate a house size control, which is implicitly accounted for in the fixed effects model and has a significant effect on consumption when included in the random effects model. Standard errors are two-way clustered by household and billing period.

The annual electricity consumption implied by these estimated coefficients is presented in Columns (2) and (4). Since the coefficients are in units of average daily heating and cooling degree days per billing month, we obtain annual consumption by multiplying the coefficient estimates from columns (1) and (3) by 147.6 and 3.7 respectively, which are annual $AHDD_{zt}$ and $ACDD_{zt}$ (i.e. $\sum_{t=1}^{12} AHDD_t, \sum_{t=1}^{12} ACDD_t$) averaged across all climate regions.[13]

We interact all variables of interest with both $AHDD_{zt}$ and $ACDD_{zt}$ in our main analysis, as presented in equations (2.3) and (2.4). However, as New Brunswick is a relatively cold province, we only report variables that are both policy relevant and make sense from the physical point of view. For example, we do not report the coefficient on pool pumps interacted with heating degree days as pool pumps are not used in the heating season, and the coefficient is insignificant. Moreover, most variables interacted with $ACDD_{zt}$ are insignificant and these are not reported in the main results in Table 2.2. Table 2.2 therefore shows a subset of the coefficients from our complete specification. Other included variables that are not reported in the Table are electric water heater, well pump and number of low flow shower heads interacted with heating and cooling degree days, as their coefficients were statistically insignificant.[14]

### 2.5.1 Model choice: random or fixed effects?

To evaluate whether the random effects model results in biased coefficient estimates we implement the Hausman (Hausman, 1978) and Mundlak (Mundlak, 1978) tests. Each of these tests seek to determine whether the time-invariant unobservable error component $c_i$ is uncorrelated with $u_{it}$ and the right-hand side regressors in equation (2.4). If the null hypothesis of uncorrelatedness is rejected, this suggests the random effects model will lead to biased coefficients and the fixed effects model is preferred. Otherwise, the random effects assumptions are satisfied. The advantage

[13]More specifically, average annualized daily AHDD and ACDD are obtained by summing $AHDD_{zt}$ and $ACDD_{zt}$ across billing months for each customer (where $AHDD_t$ and $ACDD_t$ for a given $t$ are the same for each customer in climate zone $z$), to obtain total AHDD and total ACDD per customer, then calculating the average total AHDD and average total ACDD across customers. For comparison purposes, the annual mean daily AHDD for 2016 and 2019 (the two years before and after our sample period) are 152 and 166, respectively. The annual mean daily ACDD for 2016 and 2019 are 4.3 and 3.2, respectively. Hence, the weather variables used in the analysis are typical in our province of interest.

[14]The full results Table is available from the authors upon request.

TABLE 2.2: Random effects and fixed effects model results

| | Random Effects | | Fixed Effects | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Independent Variables | Coefficients | Annual Usage | Coefficients | Annual Usage |
| Number of dehumidifiers # AHDD | 6.584*** | 976.04 | 6.123*** | 907.70 |
| | (1.214) | | (0.708) | |
| Poolpump # ACDD | 378.994*** | 1400.38 | 280.019*** | 1034.67 |
| | (42.597) | | (32.954) | |
| Local heat pump # AHDD | -16.596*** | -2460.26 | -17.599*** | -2608.95 |
| | (1.516) | | (0.898) | |
| Local heat pump # ACDD | 14.440 | 53.36 | -2.839 | -10.49 |
| | (15.164) | | (14.129) | |
| Central heatpump # AHDD | 0.072 | 10.67 | 0.753 | 111.63 |
| | (2.628) | | (1.648) | |
| Central heatpump # ACDD | 173.687*** | 641.77 | 185.113*** | 683.99 |
| | (22.301) | | (22.397) | |
| Number of window AC #ACDD | 55.386*** | 204.65 | 49.773*** | 183.91 |
| | (11.168) | | (11.137) | |
| Energy related renovations #AHDD | 0.161 | 23.87 | 0.012 | 1.78 |
| | (0.435) | | (0.259) | |
| Energy related renovations #ACDD | -4.859 | -17.95 | -7.791* | -28.79 |
| | (4.366) | | (4.172) | |
| Thermostat setback #AHDD | -4.587*** | -680.00 | -3.508*** | -520.04 |
| | (1.552) | | (0.903) | |
| Year Built (1961-1974) #AHDD | -6.644** | -984.93 | -8.860*** | -1313.44 |
| | (2.696) | | (1.528) | |
| Year Built (1975-1989) #AHDD | -13.452*** | -1994.18 | -14.948*** | -2215.95 |
| | (2.436) | | (1.387) | |
| Year Built (1990-1999) #AHDD | -11.000*** | -1630.68 | -11.827*** | -1753.28 |
| | (2.614) | | (1.490) | |
| Year Built (2000-2009) #AHDD | -14.047*** | -2082.38 | -14.991*** | -2222.33 |
| | (2.564) | | (1.492) | |
| Year Built (2010-2015) #AHDD | -21.950*** | -3253.96 | -22.759*** | -3373.89 |
| | (2.976) | | (1.808) | |
| Year Built (2015-2018) #AHDD | -31.677*** | -4695.93 | -31.778*** | -4710.90 |
| | (4.340) | | (2.891) | |
| Billing period FE | YES | | YES | |
| Household FE | | | YES | |
| House size control | YES | | | |
| Occupancy control | YES | | YES | |
| Observations | 38,568 | | 38,568 | |
| R-squared | 0.62 | | 0.83 | |
| Number of Households | 3,214 | | 3,214 | |
| **Mundlak test:** | **Chi-Sq. Statistic** | | **Prob.** | |
| | **464.690** | | **0.00** | |
| **Hausman test:** | **Chi-Sq. Statistic** | | **Prob.** | |
| | **244.400** | | **0.00** | |

Notes: The dependent variable in columns (1) and (3) is monthly electricity usage in kWh. Columns (2) and (4) show the implied annual consumption. The four primary heating system options from the survey are electric baseboard, local heat pump and central heat pumps and the omitted category is electric baseboard. The omitted category for Year Built is houses built before 1961. Dummy variables for the number of occupants are interacted with heating and cooling degree days in both models but are not reported for space purposes. Full regression results are available by request from the authors. Standard errors are clustered by household and billing period are in parentheses, ***p<0.01, ** p<0.05, * p<0.1.

of the Mundlak test relative to the Hausman test is that it permits the use of robust standard errors when the regression errors are heteroskedastic.

The results from these two tests will guide our preference between the random effects or fixed effects model coefficients. The bottom two rows of Table 2.2 report the chi-squared ($\chi^2$) statistics for each of the Mundlak and Hausman tests. As shown, the p-value of each test is less than 0.00, which therefore strongly rejects the null hypothesis of no bias affecting our independent variables of interest. This implies that the fixed-effects specification from column (3) is preferred since it controls for time-constant unobservable variable bias.

Despite the findings of biased inference from the random effects model, in general the results from Table 2.2 indicate relatively small variations in coefficient values between the random and fixed effects models. For most coefficients, there is less than a 15 percent difference between the estimates for individual variables. Two exceptions are the energy consumption effects of a pool pump and energy-related renovations.

A plausible potential reason for the larger differences between the random and fixed effects models for these variables is the impact of income level, which is unobserved in this sample. For example, higher income households may undertake energy-related renovations at different rates, utilize pool pumps with different intensities and own larger pools that require larger pumps, relative to lower income households. To the extent that income level is constant over the one-year period in our sample, the fixed effects model will net out the effect of income, as well as the mean effect of all other time-constant variables, such as thermal comfort preferences).

### 2.5.2 Coefficient estimates

Results from column (3) in Table 2.2 show that owning an additional unit of a dehumidifier raises electricity consumption by 6.1 kWh per month in response to 1 unit increase in average daily HDD (AHDD). Column (4) indicates this translates to an annual increase of 908 kWh/yr in the colder months. These values are within the range of estimates shown in Table 2.1, where we compare the findings of the same effects reported in other studies. We focus these comparisons on studies conducted in cool or cold climates in North America and Europe, as these are most relevant to the sample in our analysis.[15] Houses with pool pumps use substantially more electricity, as expected; the estimated coefficient indicates 1,035 kWh of additional annual electricity use than a house without a pool.

About 32 percent of survey respondents use local heat pumps as a primary source of heating, which is more energy efficient in both colder and hotter months. The other heating sources included in Table 2.2 are electric baseboards (used by 54 percent of households), and central heat pumps (used by 13 percent of households).[16] Table 2.2 shows that houses heating primarily with local heat pumps consume about 2,600 kWh less electricity annually than houses with electric baseboards, and all of this excess electric baseboard consumption occurs during the heating season. Local heat pumps do not consume statistically different quantities of electricity compared to baseboards during the cooling season. On the other hand, statistically central heat pumps use the same amount of electricity as electric baseboards during the heating season, but 684 kWh more electricity during the cooling season. Taken together, these heat pump estimates suggest local heat pumps are not often used for cooling in summer (although they are capable), whereas central heat pumps are.

[15] Nevertheless, local climate variations, potential differences in appliance use practices among sample households, and other methodological differences in data collection and analysis, can also lead to variations in effect estimates.

[16] As previously noted, electric baseboard heating is the omitted heating system category in Table 2.2, so that the coefficients on local and central heat pumps are interpreted as their consumption effect relative to electric baseboards.

An additional window air conditioner has a statistically significant effect on monthly electricity consumption in the cooling season, resulting in an increase of about 184 kWh per year. Investing in an extra energy related renovation has a borderline significant effect on reducing cooling demand but does not have a statistically significant effect on heating demand. The associated savings from these behaviours are modest and total about 29 kWh per year.

The effect of house age in Table 2.2 is measured in the construction year ranges as reported in the survey: Before 1961, 1961-1974, 1975-1989, 1990-1999, 2010-2015 and 2015-2018. In Table 2.2 homes built before 1961 are the omitted category so the home age coefficients are interpreted relative to the oldest homes. With the exception of 1990-1999, homes of progressively newer vintages tend to consume less electricity, and this effect is largest for the two most recent age categories. On average, the two newest vintages save about 340 kWh per year relative to their preceding age category.[17]

We formally assessed if the differences in the coefficients from vintage to vintage are statistically significant by testing individual hypotheses evaluating whether each coefficient is statistically different from the preceding vintage (using Wald tests). All of our tests strongly reject their null hypotheses. The highest p-value is 1.8% for the test between coefficients on the vintage years of 1990-1999 and 2000-2009. Most of the other p-values are lower than 0.02%.

These results suggest that since the 1990s there has been a clear and statistically significant trend towards reductions in energy consumption that can be attributed to progressively newer homes. This could be a result of greater emphasis being placed on house energy performance over time among home builders, and possibly reinforced by the adoption of building energy codes from 2012 onwards.

### 2.5.3 Energy conservation measures discussion

#### Heat pumps

An important question for this study and for policy guidance is: do heat pumps in practice use substantially less electricity than electric baseboards? The annual consumption levels of heat pumps and electric baseboard heaters have been estimated in prior studies, as shown in Table 2.1. The simple average of annual electric baseboard estimates across these studies is 10,585 kWh. Two studies from Quebec measured heat pump annual consumption levels, although these did not distinguish between centrally-ducted and local heat pump systems, and found average annual consumption values of 11,805 kWh. Our study disaggregates between local and central heat pumps and finds that central heat pumps (but not local heat pumps) consume more electricity than electric baseboards, and this higher consumption is driven by cooling season usage.

For houses that heat primarily with central heat pumps, our estimated fixed effects model coefficient finds that these systems consume the same amount of electricity compared to electric baseboards during the heating season, and more than

[17] This value results from the calculation [(4710.9-3373.89)/3 + (3373.89-2222.33)/5]/2, using the annual savings values from the two newest year built categories from the fixed effects model in Table 2.2.

| | Newsham and Donnelly (2013) | Manitoba Hydro (2011) | Bernard and Lacroix (2005) | Tiedemann (2007) | Lafrance and Perron (1994) | Sebold and Parris (1989) | Burlington Hydro (2020) | EREN (2020) |
|---|---|---|---|---|---|---|---|---|
| | Canada | Manitoba | Quebec | British Columbia | Quebec | N. Midwest, USA | S. Ontario | USA |
| Dehumidifier | | | | | | | 504-3,024 | 700 |
| Pool Pump | | 4,898 | 2,114 | 3,912 | 1,800 | | 1,944 | |
| Electric Baseboard Heating | 8,600 | 10,518-16,556* | 12,926 | 5,037* | 9,873 | | | |
| Window AC units | 396 | 675 | | 207 | | 500 | | |
| Heat Pumps Heating | | | 13,643 | | 9,966 | | | |

FIGURE 2.1: Estimated end use effects of appliances from prior studies and published guidance (kWh/yr)

Notes: * above denotes it is not clear whether this is baseboard heating only or includes other delivery mechanisms.

electric baseboards during the cooling season.[18] More precisely, the annual consumption results from Table 2.2 column (4) indicate central heat pumps consume about 684 kWh more annually than electric baseboards. On the other hand, local heat pumps consume 2,609 fewer kWh annually compared to electric baseboards, and 3,293 fewer kWh than central heat pumps.

Heat pumps are relatively expensive components of a space conditioning system, and many utilities have incentives to encourage their adoption in both new construction and in retrofits. This is because they are, in principle, much more efficient than other heating systems. In our study, the traditional incumbent heating system was electric baseboards, which have a coefficient of performance (COP) of 1, that is, for every one unit (in kWh) of electricity of input, one unit of heat is output. Air-to-air heat pumps, the most common alternative to electric baseboards in a residential context, are typically considered, in a simplified analysis, to have a COP of approximately 3: for every one unit of electricity input, three units of heat energy are extracted from the outdoor air and brought indoors. In the context of heat pumps replacing baseboards, simple engineering estimates often assume that the heat pumps will assume all of the heat load previously met by baseboards, and therefore suggest that heat pumps will lower heating energy use by 2/3, or 67 percent. For a typical electrically-heated house in a cool climate, this might equate to a saving of 30-35 percent of total annual electricity use. Our results show that houses with local (or ductless) heat pumps as their primary heating system do use substantially less electricity than those with electric baseboards as the primary system, all else being equal. However, the conservation effect is much smaller than the simple engineering estimate: 14 percent of total energy relative to electric baseboards.[19]

Other studies of the field performance of heat pumps in heating-dominated climates have found similar results. In Ontario, Marshall et al., 2015 evaluated a heat pump incentive pilot involving 100 homes that had formerly used electric baseboards as their sole heating system. Estimated energy use, based on 12 months of measured consumption, declined by only 11.2 percent compared to total pre-pilot energy use. Similarly, and also in Ontario, Marshall et al., 2017 evaluated a heat pump incentive pilot involving 100 homes that had formerly used electric baseboards or electric furnaces as their sole heating system. Estimated energy savings, based on 12 months of measured consumption, were only 7.9 percent compared

[18] As noted in section 2.3.2, the heating season here is measured as periods when the outdoor temperature falls below 18 degrees celsius, and the cooling season when the temperature is above 18 degrees celsius.

[19] Based on the calculation 2,609/18,847, where the denominator is annual average in-sample electricity use.

to total pre-pilot energy use. Hamelin, Michaud, and Chartrand, 2018 conducted a multiple regression-based analysis using billing data of the effectiveness of heat pump retrofits in Nova Scotia. Focusing on the subset where minisplit heat pumps were installed in 118 exclusively electrically-heated houses, annual savings of 3,504 kWh were reported. They concluded that this saving was well below estimates from standard methods pre-installation. Baylon, Storm, and Robison, 2013 report on a large study in the northwest USA where ductless minisplit heat pumps replaced electric heating. For houses with no self-reported supplemental fuel use, a billing data analysis indicated average annual energy savings were 2,718 kWh (N=2,295), or about 15 percent of pre-installation total energy use, very similar to the results in our study. However, a CDA on these data suggested about a third of savings were given back for increased space temperature and occupancy. Finally, Halvorsen and Larsen, 2013, using a CDA method to analyze data from $\geq$ 1000 households in Norway, found that the energy saving potential of heat pumps was offset by other changes in energy consumption behaviour.

There are several reasons in the previously cited literature for lower performance than simple engineering estimates:

- COP declines as outdoor temperature goes down, especially for older heat pumps that might not have been specifically designed for cold climates Korn et al., 2016. Therefore, the actual average operating COP may be substantially lower than the value of 3 often assumed.

- Some electric baseboards are retained as supplementary heat sources and are used even after the heat pump is installed. In some cases, this permits a greater fraction of the house volume to be heated, thus increasing comfort and space utilization.[20]

- People exchange savings for increased comfort. This is a classic rebound effect: when the heating bill is lower some people increase their thermostat setpoint.

Another potential energy penalty is that people who did not have air conditioning prior to obtaining a heat pump will then use the heat pump in cooling mode in the summer, thus adding to total electricity use. Again, this brings a tangible comfort benefit to the occupants, and is thus seen as a good thing by heat pump owners, even with an energy penalty. Our analysis shows that a central heat pump is associated with an air-conditioning load of 684 kWh/yr, higher than the estimated use of a standard window air conditioner in our sample, of about 184 kWh/yr.

Our results that central, ducted heat pumps use significantly more energy than local heat pumps may arise because they are used to service the entire house volume all of the time, whereas local heat pumps might be used in specific zones at least some of the time. Overall, local heat pumps generally realize substantial energy savings and improve comfort. However, measured savings are substantially lower than typically forecast, which may impact the cost-benefit equation for homeowners. It also has implications for program design on the part of policy-makers. From a policy perspective, our finding that central heat pumps do not save electricity compared to electric baseboards may call into question whether program dollars are best spent incentivizing their adoption.

[20] In our estimated model this scenario would inflate the estimated consumption for heat pumps as households were asked about their primary heating system.

### Thermostat setback

Houses in which the householder indicated employing thermostat setbacks used statistically-significantly less electricity during the heating season, as expected; the estimated annual saving is 520 kWh/yr. Newsham and Donnelly, 2013 also used CDA to estimate this effect at 390 kWh/yr per degree of setback, for electric heating systems, and this is broadly consistent with the effect reported in the current analysis. In an experimental study in full-scale test houses with natural gas heating, Manning et al., 2007 found winter heating savings of about 2.2-2.5 percent per degree of setback, when setbacks were implemented for approximately seven hours per day. Assuming a 2 °C setback that persists for similar periods, a 2.35 percent energy saving per degree of setback, the middle of the savings range in their study, results in savings of approximately 780 kWh/yr.[21] This is somewhat higher than our estimate, however the study setting was Ottawa, Canada, a region with colder winters than our jurisdiction under study, which may account for the higher savings. A basic engineering calculation results in savings within a very similar range as our estimate. Heat loss (and thus the energy to maintain the desired indoor temperature) is proportional to the temperature difference between inside and outside; in a cold climate this temperature difference may be estimated at approximately 20 °C. A setback of 2 °C that persists for half of the time, suggests a saving of 5 percent (2/20 * 0.5), which, on a total heating use of 11,200 kWh/yr (the average of the electric baseboard and heat pump annual consumption estimates in Table 2.1) yields a saving of 560 kWh/yr.

Overall, a synthesis of research results, including this current study, indicate that thermostat setbacks do yield annual heating energy savings of approximately 500-800 kWh/yr, and may often be realized by utilizing equipment the homeowners already has. As such, continued policy support for this energy saving strategy is justified.

### Renovations

Houses reporting having recently completed energy-efficiency retrofits used less electricity during the cooling season; the estimated coefficient was a saving of about 29 kWh/yr per renovation. However, the coefficient is borderline significant at the 10 percent significance level. This may be due in part to the high heterogeneity in the savings associated with the different categories of energy renovations. For example, insulation retrofits typically save significantly more energy than sealing air leaks (Giandomenico, Papineau, and Rivers, 2020; Liang et al., 2018). Due to the way this variable was assessed and developed in the analysis, it is difficult to compare to data collected elsewhere. Energy related renovations completed represents the response to a series of survey questions, "Please indicate if you have recently completed the following: insulate your basement, add insulation to your attic, add insulation to your walls, replace exterior doors or windows, install LED lights, install weather-stripping, caulking, etc." Each of these measures is unlikely to have an equal effect. Further "recently" in the survey was not defined and may have been interpreted differently by different respondents. We also do not know the state of the house prior to the renovation, nor the extent of the renovation; e.g., if a respondent reported adding insulation to their attic we do not know how much insulation was there before the renovation nor how much insulation was added. Given the importance of

[21] These savings were converted to kWh from MJ of natural gas.

renovations in future code development, better data on previous renovations and their effects should be a priority.

These measures have different costs and implementation challenges, and, given that we cannot parse out the effects of each measure, a cost-benefit analysis for policy purposes is not possible. Nevertheless, the fact that we estimated a modest savings effect is encouraging and supportive of further research on these measures.

## 2.6 Conclusion

We conducted a conditional demand analysis on electricity use data from approximately 3,200 electrically-heated dwellings in a cold climate in North America that exhibits high electrification of end-usage. The results provide estimates of the effectiveness of various energy efficiency measures on electricity usage post-installation, at a time when the international community looks towards increasing the electrification and decarbonization of energy end-uses in order to reach climate policy targets.

Evaluating efficiency investments on the basis of realized energy use is important because such analyses estimate their effects including secondary interactions among appliances and energy-using equipment, as well as behavioural responses such as the rebound effect. In the end, it is not theoretical savings that matter, but rather the net savings on actual energy consumption using robust statistical models with low coefficient bias. This information can guide policy-makers in deciding which energy efficiency measures are deserving of the most support, and what appropriate incentive levels might be with respect to the cost of a specific measure and its expected, i.e. realized, energy reduction effect.

A further contribution of our work is to evaluate the evidence for coefficient bias in the random effects models that are frequently used in conditional demand analyses. While on the basis of statistical tests we find that the random effect estimator is rejected in favour of the fixed effects estimator, most of the individual coefficient estimates exhibit only small variations between models, with a few exceptions. Overall, the estimates of energy use by specific appliances and conservation measures were quite robust across different statistical estimators, and reasonable when compared to relevant prior studies using various estimation methods.

Finally, this analysis reinforces the value of CDA as a technique to derive policy-guiding estimates. As relevant energy use data become potentially more available (e.g. smart meter data, smart thermostat data), and survey data become easier to collect via on-line methods, CDA may become a more attractive research tool for climate policy guidance going forward.

## 2.7 Acknowledgments

The authors gratefully acknowledge NB Power for generously providing the anonymized survey data on which the analysis was based, and National Research Council (NRC) colleagues Ghassan Marjaba and Heather Knudsen for support via NRC-Construction's Net Zero Energy Research project.

## 2.8 Appendix

### A. Data cleaning and variable coding

#### A1. Consumption data

Our general approach to data cleaning was to remove, or cap, extreme values for the major variables in the model. In common with prior work in this area, we considered univariate outliers only. For electrical energy use, we dropped cases where total annual energy use was more than 3 standard deviations from the mean, or if usage in any billing period was >10000 kWh, as high outliers. Given that we had a sample of houses in a cold climate using electric heat, we also dropped cases where total annual energy use was <3000 kWh, or if usage in any billing period was <100 kWh, as low outliers. We also excluded a small number of cases that had reported ownership of an electric vehicle or local renewable generation because their monthly grid-sourced electricity consumption would be considerably higher or lower (respectively) than their counterparts.

#### A2. Survey data

Some variables that were reported in categories were recoded as numeric, where appropriate. For example, on the survey the variable house area (finished area without garage) was originally reported as $< 601$ square feet (s.f.), 601–1200 s.f., 1201–1800 s.f., 1801–2400 s.f., $> 2401$ s.f., and these were recoded to 600, 900, 1500, 2100, and 2700, respectively; responses of "Don't know" were coded as missing and therefore excluded from the analysis. The survey also collected data on year of home construction in the following year built categories: before 1961, 1961–1974, 1975–1989, 1990–1999, 2000–2009, 2010–2015, and after 2015. Each year built range was recoded to an indicator (or dummy) variable to reflect the seven age range categories.

The survey also asked households about their thermostat usage in winter time. Specifically, households responded whether they turn down their thermostat while sleeping, when no one is home, both or never. We coded the thermostat setback variable as a dummy variable that takes the value of one if the household turns down their thermostat anytime of the day and zero otherwise.

#### A3. Weather data

Using postal code data and weather zones in New Brunswick, we identified six major cities/towns that are representative of the different weather zones of the province, allocated each household location to one of these six geographical regions, and obtained daily weather data from Environment Canada for each of the six cities from 2016–2018 to calculate HDD and CDD. Daily HDD and CDD were calculated and then totalled for each billing period and each household in a specific region. Finally, to simplify coefficient interpretation, we divided cumulative HDD and CDD by the number of days in that billing period to arrive at average daily HDD and CDD by billing month (hereafter AHDD and ACDD respectively). Table A1 presents an example of weather data for one of the six regions for the 2017 calendar year. It illustrates that heating days are by far more prevalent than cooling days, which is typical for this coastal province with cold winters and mild summers.

| Month (2017) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CDD18 | 0 | 0 | 0 | 0 | 4 | 30 | 53 | 37 | 31 | 2 | 0 | 0 | 157 |
| HDD18 | 748 | 691 | 690 | 373 | 215 | 69 | 15 | 33 | 59 | 205 | 503 | 800 | 4401 |

Table A1: Example CDD and HDD data for one example region for the 2017 calendar year

# Chapter 3

# Peer effects in Residential Energy Consumption

## 3.1 Introduction

The study of individual behavior in economics has traditionally focused on rational decision-making based on personal preferences and available information. However, humans are inherently social beings, and their choices are profoundly influenced by interactions with others. Peer effects, also known as social interactions or social networks, have emerged as an innovative area of research that seeks to understand how individuals' economic decisions are shaped by the actions and characteristics of their peers. It provides valuable insights into the complexities of economic behavior in many settings ranging from schooling (Sacerdote, 2001; Angrist and Lang, 2004; Duflo, Dupas, and Kremer, 2011; Lavy and Schlosser, 2011) to health (Lundborg, 2006; Fowler and Christakis, 2008; Trogdon, Nonnemaker, and Pais, 2008; Carrell, Hoekstra, and West, 2011) to labor and program participation (Card et al., 2012; Dahl, Løken, and Mogstad, 2014; Cornelissen, Dustmann, and Schönberg, 2017; Carrell, Hoekstra, and Kuka, 2018), to agriculture (Foster and Rosenzweig, 1995; Conley and Udry, 2010), and investment, spending, and consumption decisions (Kuhn et al., 2011; Bursztyn et al., 2014; Agarwal, Qian, and Zou, 2021).

For example, in education, Sacerdote, 2001 shows that peers have an impact on grade point average and on decisions to join social groups. In health, Carrell, Hoekstra, and West, 2011 explore the contagious nature of poor fitness among randomly assigned friends, revealing the presence of peer effects in physical fitness outcomes. Additionally, Dahl, Løken, and Mogstad, 2014 show that employees whose coworkers or brothers previously participated in a paid paternity leave program exhibit about 13% higher likelihood of taking paternity leave themselves.

In the domain of environment and energy, previous studies have documented the presence of peer effect in the uptake of solar photovoltaic panels (Bollinger and Gillingham, 2012; Noll, Dawes, and Rai, 2014; Graziano and Gillingham, 2015), consumption behavior (Bollinger, Burkhardt, and Gillingham, 2020; Wolske, Gillingham, and Schultz, 2020), and hybrid vehicles (Narayanan and Nair, 2013; Heutel and Muehlegger, 2015). This paper is the first to identify casual peer effects of home energy efficiency retrofits on energy consumption.

This study aligns with broader efforts to address the challenges of climate change, and meeting the net zero carbon emission goal by 2050 (UNEP, 2021). Residential buildings are a huge energy consumer with around 20% of global energy consumption and around 17% of total greenhouse gas (GHG) emissions in 2021 (Delmastro, 2022). Since over two-thirds of the current building stock will still be operational in 2050, the Net-Zero by 2050 agenda is front loaded with home energy efficiency measures such as increasing the insulation, upgrading home's window and doors, and upgrading the heating and/or cooling systems (IPCC, 2018; International Energy Agency, 2021).

Previous studies show that residential energy efficiency measures are proven to successfully reduce energy consumption of retrofitted homes using dis-aggregated house level data (Fowlie, Greenstone, and Wolfram, 2018; Burlig et al., 2020; Christensen et al., 2022; Chuang, Delmas, and Pincetl, 2022; Papineau, Rivers, and Yassin, 2023).[1] In this paper, I provide a causal estimate of peer effects of energy efficiency retrofits on residential energy consumption. Specifically, I use monthly energy use data, house characteristics data, and program participation of Canada's largest home energy efficiency program, from over 20,000 households in Medicine Hat, Alberta to test whether a household's energy consumption is influenced by the retrofit decisions of their peers in the previous periods.[2]

Causal peer effects are inherently difficult to identify using a linear-in-means model for three main reasons; reflection, correlated unobservables, and endogenous group membership (Manski, 1993).[3] In this study, I address these challenges through an identification strategy that relies on the quasi-experimental variation from the staggered nature of program participation surrounding individual households. I utilize an exhaustive list of audit data from the city of Medicine Hat from 2007-2012 and I directly compare neighbors to retrofitted houses, that completed the pre-retrofit and the post retrofits audits, to neighbors to homes that completed the pre-retrofit audit only, which I refer to as "Almost retrofitted" homes. This setting is attractive for at least two reasons. First, the identification assumption is that one's neigbour decision to retrofit (versus only completing the pre-retrofit audit) can be considered as plausibly exogenous. This assumption seems reasonable given that the sample of household-month observations of neighbors to retrofitted and "Almost retrofitted" homes is balanced across a wide range of characteristics (see Section 3.3).

[1] These studies also show that the realization rates of these measures ranges from 40 to 60%. Realization rate is calculated by dividing actual realized savings from energy efficiency measures by the models' predicted savings

[2] In a separate study utilizing consumption, program participation, and house characteristics data sets from the same city (Papineau, Rivers, and Yassin, 2023), estimated the actual energy savings resulting from the designated home energy efficiency program, ecoENERGY Retrofit for Homes, to be approximately 16% of the total energy consumption. Furthermore, the study observed that these savings persisted throughout the entire ten-year time frame examined.

[3] linear-in-means model is commonly used in peer effects studies (e.g., Bayer, Ross, and Topa, 2008; Bollinger and Gillingham, 2012; Towe and Lawley, 2013; Carrell, Hoekstra, and Kuka, 2018; Fletcher and Ross, 2018; Bollinger, Burkhardt, and Gillingham, 2020).

Second, I observe house-level monthly energy data and house-level tax assessment variables for all houses in the city. Additionally, I utilize detailed house-level audit data for all houses that undertook one or more audits as part of the program. This rich dataset, characterized by its high frequency and detailed nature, enables me to analyze the peer effects on energy consumption at the house-level. Analyzing the consumption impacts at this geographical level provides a better opportunity to test empirically the channels through which energy retrofits affects peers. The empirical strategy is also aided by the considerable variation across households in the timing of retrofits.

Using this identification strategy, I find a positive impact on energy savings for households who are close neighbors to a retrofitted home. I also find that easily visible retrofits to neighbors, such as windows and doors, and exterior wall insulation, exhibit a greater impact on energy savings among peers when compared to less conspicuous retrofits such as natural gas furnace upgrade. Finally, I find that as the distance between individuals and their neighbors increases, the strength of the effect gradually decreases until it eventually fades away.

The term "peers" encompasses a diverse range of individuals, including friends, family members, co-workers, and neighbors. In my study, neighbors hold particular significance as they form a distinct and measurable peer group within my research context. It is worth noting that my energy savings estimates represent the lower bound impact of the overall peer effect, as I focus primarily on this measurable group rather than considering alternative definitions of peers.

The ecoENERGY Retrofit for Homes program was announced by the Canadian federal government in 2007 with over 1 billion dollars in investments. Retrofits in the program occur in a three-stage process. First, households who want to retrofit their homes complete a pre-retrofit audit by an accredited auditor, who provides recommended rebate-eligible efficiency investments for their home. Second, home owners choose which retrofits to adopt. Third, households have to complete a post-retrofit audit in order to receive their rebate amount (maximum.$5,000). The most popular retrofits in the program were air sealing, natural gas furnace upgrades, and attic/ceiling insulation upgrades.

Participants in the ecoENERGY Retrofit for Homes program are self-selected to join. While neighbors of retrofitted homes are not, however, my estimates may be biased if I believe that self-selected participants share certain house characteristics such as furnace age, environmental attitudes or income level that increases their likelihood of joining the program, and that they live in a homogeneous neighborhood where their neighbors share these features as well. This could lead me to potentially overestimate energy savings (Boomhower and Davis, 2014). To address this challenge, I show that my results are robust using multiple control groups.

My baseline control group leverages the unique design of the program and distinguishes between houses that did the pre and post retrofit audits (i.e. was retrofitted), and houses that did the pre-retrofit audit only (i.e. was not retrofitted). I refer to the latter group as "Almost retrofitted" homes. Hence the control group in my baseline specification is the closest neighbors to "Almost retrofitted" homes. I also implement my analysis with two additional control groups. The second control group is a matched control group derived from the nearest neighbor matching with no replacement based on tax assessment data variables such as house type, lot size, building size, year of construction, building condition, and assessment value. The third control group consists of all nearby neighbors of non-retrofitted houses in the city, referred to as the "naive" control group.

The study utilizes four main sources of data. Firstly, the EnerGuide for Homes database (EGH) provides information on program participation, home audits, and retrofit details for a sample of 2,067 houses. Secondly, property assessment data from the City staff offers tax assessment variables at the property level. Thirdly, monthly natural gas and electricity consumption data from the municipal utility cover a 13-year period for approximately 20,000 single-family properties. Lastly, a city-wide geo-location database is used to identify the ten closest neighbors to each house in the data set. By address level matching, I use these data sources to construct a comprehensive panel dataset.

My baseline results, using neighbors to "Almost retrofitted" homes as the comparison groups, show that close neighbors to retrofitted homes reduce their monthly gas consumption by a statistically significant 2.8% in their monthly gas consumption in the short run and about 3% in the long run. However, the impact on electricity consumption is largely insignificant, which is reasonable considering that gas is the predominant energy source in Medicine Hat, and the retrofit program primarily focuses on envelope-related improvements. Furthermore, an event study illustrates a delayed but persistent trend in peer savings, indicating that the effects endure over an extended period. These findings remain consistent when employing the other control groups mentioned earlier.

I highlight two additional results regarding mechanisms of peer effect. First, I propose that visibility and neigbour-to-neigbour conversations are key mechanisms driving peer effects in energy savings. I find that the visibility of retrofits is key, particularly when neighbors can visually observe the process such as windows and door upgrades, as well as exterior wall insulation. In contrast, gas furnace upgrades are less visible due to their swift installation process that occurs inside the home. My results show that more visible retrofits lead to significantly higher energy savings among neighbors compared to less visible retrofits. The magnitude difference is about 3.5 times.

Second, I propose two primary mechanisms to explain the observed reduction in peer consumption: behavioral changes and energy-efficiency investments. Behavioral changes involve purposefully modifying actions and habits to align with energy-saving norms, driven by social influence and motivation for similar outcomes. This includes adjusting thermostat settings, adopting energy-conscious lighting practices, and utilizing energy-saving modes on appliances and equipment. Energy-efficiency investments entail making similar upgrades to the retrofitted neighbors, such as upgrading to more efficient gas furnaces and appliances.

This paper contributes to several literatures. Most directly, it provides robust evidence to the extensive and expanding field of peer effects in the diffusion of consumer behaviors. Additionally, it contributes to the literature on how information dissemination, particularly through social networks, can impact consumer choices regarding energy consumption. Numerous studies have examined the effectiveness of social norm-based messages in promoting energy conservation and reducing energy use, and this work aligns with and builds upon those investigations (e.g., Allcott, 2011; Ayres, Raseman, and Shih, 2013; Costa and Kahn, 2013b; Allcott and Rogers, 2014; Dolan and Metcalfe, 2015; Gillingham and Tsvetanov, 2018; Bollinger, Gillingham, and Gullo, 2020; Papineau and Rivers, 2022). The observed peer energy savings of approximately 2.7% from retrofits is in line with other behavioral interventions employing social comparisons mechanism. For example, Allcott, 2011 conducted a large-scale randomized experiment that involved sending Home Energy Report (HER) letters to residential utility customers, providing comparisons of their electricity usage with that of their neighbors. This intervention resulted in a reduction of energy consumption by 2.0%.

The remainder of the paper is organized as follows: Section 3.2 provides some background on the ecoENERGY Retrofit for Homes program; Section 3.3 describes the data sources used in the analysis and presents summary statistics of the variables used; Section 3.4 discusses the empirical strategy; Section 3.5 presents the results; Section 3.6 explores potential mechanisms; Section 3.7 presents the robustness checks; Section 3.8 concludes.

## 3.2 Program Description

The ecoENERGY Retrofit for Homes program took place between 2007-2012, and aimed at enhancing the energy efficiency of existing low-rise housing in Canada, and thereby reducing emissions generated from energy consumption and contributing to a cleaner environment. The primary goal of the program was to empower property owners with valuable information to make informed decisions about energy retrofits for their homes, while also offering financial incentives in the form of grants for energy-saving measures (IEA, 2017b).

Initially, the program had an original budget of 170 million to provide an average grant of CAD 1,070 to approximately 140,000 homes. In 2009, as part of the Government of Canada's 2009 Economic Action Plan, an additional investment of CAD 300 million was made, allowing an additional 200,000 homeowners to participate from 2009 to 2011. A funding expansion of CAD 285 million was announced in 2010, and further investment of CAD 400 million was allocated in the Government of Canada's 2011 Budget (Natural Resources Canada, 2012). In total, the ecoENERGY Retrofit for Homes initiative had over 1 billion dollars in budget and has assisted over 640,000 Canadians in improving their homes' energy efficiency.[4]

To deliver the program, a third-party system was employed, wherein organizations licensed by Natural Resources Canada engage energy assessment services. These organizations recruit and train energy advisors and quality control personnel, who conduct local energy audits and provide homeowners with pre-retrofit assessments and a checklist of recommended upgrades. Alongside this, an EnerGuide home energy rating is provided. The homeowner then selects and implements the recommended retrofits within an 18-month time frame. Energy advisors conduct post-retrofit assessments and verify the completion of retrofits. The grant amount is determined by aggregating the eligible retrofit costs with a maximum rebate amount of $5,000. Based on the engineering simulations, it was projected that the program would result in an average of 20% consumption savings and a reduction of 0.75 mega-tonnes of greenhouse gas emissions in 2011-2012.

In a separate study that looks at the energy savings from the ecoENERGY Retrofit for Homes program in the same city, Papineau, Rivers, and Yassin, 2023 show that participants experienced an average reduction of approximately 15% in house energy consumption. Most of the energy savings occurred in natural gas consumption, which declined by 21% for up to 10 years. Meanwhile, electricity consumption showed a more modest decline of 0% - 5%. The study pointed out that the most effective retrofits in generating significant savings were the adoption of new furnaces and wall insulation. Considering that the average annual gas bill in Medicine Hat is around $850, their data indicated that a retrofitted house saved $269 on average from completing energy-efficiency retrofits through the program .

## 3.3 Data

### 3.3.1 Data Sources

To study peer effects of home energy efficiency retrofits, I utilize data sets from four distinct sources pertaining to the city of Medicine Hat in Alberta. Medicine Hat is located in the South Saskatchewan River Valley with about 60,000 residents.[5]

The first data set I use is the EnerGuide for Homes (EGH) database, which provides comprehensive information on program participation and home audits for all houses enrolled in the program. A total of 2,067 houses underwent one or more audits between 2007 and 2012. Among these, 1,684 houses completed two detailed

[4] Budget 2021 provided $4.4 billion for a new loan program called Greener homes to help homeowners and affordable housing providers complete deep home retrofits. This ongoing program has a very similar in structure to the ecoENERGY Retrofit - Homes (Government of Canada, 2021).

[5] Based on the 2011 federal census, the official population count during the study period was 60,005 (The city of Medicine Hat: City Clerk Department, 2012). The most recent population count as of 2021 was 63,271.

retrofit audits, indicating that these homes undertook renovations, while 383 houses completed only a pre-retrofit audit. The EGH database includes variables related to house characteristics, recommended retrofits, predicted energy consumption, predicted energy savings from recommended retrofits, and completed retrofits (available only for retrofitted homes). The predicted energy consumption and savings, both before and after retrofitting, were generated using the HOT2000 model, a building energy consumption simulation software. The exact retrofit dates are not observed, however, I employ a variable called "pre-retrofit creation date" that indicates when the household file was created in the database to proxy for the retrofit date. Based on the program design and discussions with program administrators, the actual retrofit date typically occurs a few weeks after the recorded "pre-retrofit creation date" in the database.

Secondly, I obtained property assessment data provided by the City staff, which includes various tax assessment variables at the property level. These variables encompass information such as house type, house size, lot size, year of construction, building condition, and an evaluation of the building's assessment value.

Thirdly, I acquired monthly natural gas and electricity consumption data from the municipal utility, covering a 13-year period from 2007 to 2019. This data set includes consumption information for approximately 20,000 single-family properties, representing all houses within the city. By address-level matching, I merged the three data sets to create a comprehensive panel data set. Out of the total of 2,067 houses from the EGH database, 200 observations remained unmatched due to incorrect recording of addresses in the EGH database, absence of property assessment data, or EGH addresses located outside the boundaries of Medicine Hat city and hence no observed utility data for these houses.

The fourth and final data source consists of a city-wide geo-location database. Leveraging the address-level panel data from the combined EGH, tax assessment, and consumption data sources, I generated the latitudes and longitudes of all single-family homes. This geo-location database enabled us to identify the ten closest neighbors to each of house.

### 3.3.2 Peer Definition

The term "peers" encompasses a wide range of individuals, including friends, family, co-workers, and neighbors. In the context of my study, neighbors are particularly relevant and measurable as a peer group.[6] This is because the spillover effect of energy efficiency programs propagate through word of mouth and/or by observing the retrofit process through the presence of insulation trucks, installation of upgraded windows and doors, and the activity of laborers at the retrofitted house.

[6]Numerous research papers have employed the term "peer effect" to denote neighbor influence in situations where neighbors constitute the most pertinent category of peers (Bollinger and Gillingham, 2012; Noll, Dawes, and Rai, 2014; Bollinger, Burkhardt, and Gillingham, 2020; Wolske, Gillingham, and Schultz, 2020; Agarwal, Qian, and Zou, 2021)

Given this context, an appropriate definition of peers revolves around spatial proximity to the retrofitted house. In my baseline sample, I define peers as the two nearest neighbors using geographic coordinates for all addresses of single-family homes in the city (Kuhn et al., 2011).

Hence, the treated group consists of the two neighbors spatially closest to the retrofitted home. As for my baseline control group, I consider the two neighbors spatially closest to houses that have only undergone a pre-retrofit audit, which I refer to as "Almost retrofitted" houses. These households had the intention to retrofit and completed a paid pre-retrofit audit but did not proceed for ambiguous reasons such as time and financial constraints. For robustness purposes I estimate peer energy savings with two other control groups; a matched control group based on tax assessment data variables (house type, lot size, building size, year of construction, building condition and assessment value), and all close neighbor to non-retrofitted houses in the city, hereafter "naive" control group.

To accurately measure the causal effect of proximity to an energy retrofitted home, I conduct further data cleaning. Firstly, I exclude houses whose two closest neighbors have different retrofit status. For instance, if the neighboring house on the right had completed a retrofit while the neighboring house on the left had only undergone a pre-retrofit assessment. In this case, I remove this particular house' from my treated group. This approach ensures a clear distinction between the treatment and control groups. Secondly, I encounter situations where two spatially close houses eventually joined the energy efficiency program. To capture the potential peer effect in program participation, I allow for this occurrence, except for one specific scenario. If my proxy variable for the retrofit date indicates that the neighbor underwent retrofits on the same day or the same month. If this is the case, I exclude this house from the treatment group because the program structure requires a time lag between the decision to retrofit through the program and the pre-retrofit audit date. In other words, it typically takes over a month to schedule an energy audit. Finally, if a household is considered a neighbor to two retrofitted homes with different proxy for retrofit dates, this neighbor is considered treated at the earliest of both retrofits dates.

After implementing the additional data cleaning procedures described above, I end up with 1,438 retrofitted homes, and 248 "Almost retrofitted" homes. The treated group comprises 2,287 households residing near retrofitted homes, while the control group consists of 387 households residing near "Almost retrofitted" homes. Table 3.1 provides summary statistics. The mean of my dependent variables, the monthly natural gas and electricity consumption is 292 $m^3$ and 805 kWh respectively.

## 3.4 Empirical Strategy

In this section, I discuss my main empirical strategy that utilizes the quasi-experimental variation created by staggered adoption of energy efficiency retrofits between 2008-2012 in the city of Medicine Hat. I first discuss the baseline specification, employing our preferred sample, and then proceed to explore two alternate samples. My main outcome measures household energy consumption. My set up implies that the pre-program trends in energy consumption as well as house characteristics can sensibly be regarded as orthogonal to the treatment status (something that I verify statistically for pre-program consumption and a range of house observable). In essence I estimate, in a fixed effects framework, the causal impact of residing close to an energy retrofitted home on natural gas and electricity consumption.

### 3.4.1 Baseline specification

The baseline model compares the effect of being a close neighbor to a retrofitted home relative to being a neighbor to an "Almost retrofitted" home ( i.e., a house that underwent a pre-retrofit energy consumption audit but did not receive a retrofit) on energy consumption. To estimate this, I employ a Difference-in-Differences (DiD) approach, using households near either retrofitted homes or "Almost retrofitted" homes. I refer to the sample of houses I use in my baseline model as the "Almost retrofitted" sample. The model includes monthly natural gas and electricity consumption observations, and it is specified as follows:

$$C_{iym} = \theta \text{Post Audit}_{iym} + \beta \text{Post Neighbor Retrofit}_{iym} + \gamma_i + \delta_{ym} + \epsilon_{iym}. \tag{3.1}$$

where $C_{iym}$ is the monthly natural gas or electricity consumption of household $i$ in year $y$ and month $m$. Post Audit$_{iym}$ is a dummy variable equal to one for all months after the neighbor had completed the pre-retrofit audit and zero otherwise. Post Neighbor Retrofit$_{iym}$ is a dummy variable equal to one if the neighbor completed the pre and post-retrofit audit (i.e, if the neighbor retrofitted her home) and zero before the neighbor completed the retrofit or if the neighbor completed the pre-retrofit audit only. $\gamma_i$ and $\delta_{ym}$ are household fixed effects and month-year fixed effects, respectively. I include household-month observations up to seven years after and two years before being treated and I show two-way cluster standard errors by household and month-year.

$\theta$ measures how the energy consumption changes in households who are neighbors to retrofitted homes and "Almost retrofitted" homes after the pre-retrofit audit date. $\beta$ is the main parameter of interest, which captures how the energy consumption changes in households whose neighbors retrofitted their homes, after the proxy retrofit date, relative to households whose neighbor "Almost retrofitted" their homes.

To identify the causal impact represented by $\beta$, it is essential that the pre-trends in energy consumption are similar between households in the treatment and control group. This is crucial for a reliable comparison of changes in energy consumption between neighbors of retrofitted and "Almost retrofitted" homes. To show results in favor of identification, I compute differences in energy consumption and house characteristics between houses in the treatment and the comparison groups. Table 3.1 displays the balance statistics, indicating that the two groups exhibit similar characteristics regarding pre-treatment energy consumption and other observable house features. Furthermore, the normalized differences between the treatment and control groups are significantly below the threshold of 0.25, recommended by Imbens and Wooldridge, 2009.

To study the dynamic effect of being a neighbor to retrofitted home relative to being a neighbor to an "Almost retrofitted" home, I estimate a Difference-in-Differences model by including leads and lags of the Post Audit and Post neighbor Retrofit variables. Specifically, I estimate the following:

$$C_{iy} = \sum_{h=-2}^{h=7} \kappa_h \text{Post Audit}_{i,y-h} + \sum_{h=-2}^{h=7} \lambda_h \text{Post Neighbor Retrofit}_{i,y-h} + \gamma_i + \delta_y + \epsilon_{iy}. \tag{3.2}$$

In this specification, I aggregate monthly energy consumption data to calendar year. The outcome variable ($C_{iy}$) is the natural logarithm of annual gas consumption for household $i$ in calendar year $y$. Post Audit$_{iym}$ (Post Neighbor Retrofit$_{iym}$) is a dummy variable equal to one for all years after the neighbor had completed the pre-retrofit audit and zero otherwise (when a household is neighbor to a retrofitted home, and zero before becoming a neighbor to a retrofitted home or if the house is part of the comparison group). Similarly Post Audit$_{i,y-h}$ (Post Neighbor Retrofit$_{i,y-h}$) s a dummy equal to one for year $h$ before ($h > 0$) or after ($h < 0$) her neighbor retrofitted her home (when a household is neighbor to a retrofitted home, and zero before becoming a neighbor to a retrofitted home or if the house a neighbor to an "Almost retrofitted" home). I include household-year observations up to seven years after and two years prior being treated. As standard, I normalize the estimates by omitting the year before the neighbor's retrofit year.

### 3.4.2 Other Control Groups

Participants of the ecoENERGY Retrofit for Homes program are not randomly selected but rather self-select into the program. This self-selection process may introduce a potential bias in estimating energy savings from retrofits, as certain characteristics like older furnaces, higher incomes, or a strong environmental awareness could make some households more likely to join the program and be correlated with these other unobserved factors that affect energy consumption. Consequently, the estimated energy savings for participants might be biased upward (Boomhower and Davis, 2014). Neighbors are not self selected, however, I can be challenged by an estimate bias if I believe that self-selection into the program are based on characteristics that are shared between neighbors such as income level or environmental awareness (the Tiebout sorting theory (Tiebout, 1956)).[7] I address this challenge by considering two other control groups; a matched control group based on the tax assessment data, hereafter matched sample, and all close neighbor to non-retrofitted houses in the city, hereafter "naive" sample.

First, I create a matched control group through the nearest neighbor matching with no replacement based on the house observable tax assessment variables; lot size, building size, assessment value, building condition, building type and year built. The matched sample consists of 2,553 treated houses and 2,177 control houses. Second, the "naive" sample consists of 2,554 treated houses that are the two closest neighbors in proximity to energy retrofitted homes and 14,529 houses in the control group whose close neighbors never took part in the program.

To check the validity of comparing outcomes from the treatment and comparison groups in both samples, I conduct balance tests based on the normalized differences with regards to pre-treatment gas and electricity consumption, and house features (lot size, building size, assessment value and year built). Results are presented in Table 3.9 and Table 3.10 in the Appendix for the matched and "naive" samples respectively. The Tables show that the treatment and comparison groups are similar in both samples. More specifically, the normalized differences between the treatment and control groups are all significantly lower than 0.25, which is the recommended threshold by Imbens and Wooldridge, 2009.

To estimate the causal peer effect on energy consumption using the matched and "naive" samples, I use a generalized Difference-in-Differences (DiD) approach under certain assumptions explained below. This approach compare energy consumption of treated households before and after their neighbor completed the retrofits against control households whom neighbor has not performed energy efficient retrofit or retrofit audit. Specifically, I employ the following two-way fixed-effect (TWFE) model for estimation:

$$C_{iym} = \nu \text{Post Neighbor Retrofit}_{iym} + \gamma_i + \delta_{ym} + \epsilon_{iym}. \tag{3.3}$$

where $C_{iym}$ is the monthly natural gas or electricity consumption of household $i$ in year $y$ and month $m$. Post Neighbor Retrofit$_{iym}$ is a dummy variable equal to one if the neighbor completed the energy efficiency retrofit and zero before the neighbor

[7]the Tiebout sorting theory Tiebout, 1956 is applicable to neighborhoods within a city, suggesting that households have the freedom to choose their preferred neighborhoods based on factors such as safety, proximity to amenities, school quality, access to public transportation, and demographic composition. Consequently, individuals with similar preferences tend to cluster together, resulting in the formation of homogeneous communities within specific neighborhoods.

completed the retrofit or if the household is part of the comparison group. $\gamma_i$ and $\delta_{ym}$ are household fixed effects and month-year fixed effects, respectively. I estimate (3.3) using Ordinary Least Square (OLS) and show two-way cluster standard errors by household and month-year.

Under the assumption that, in the absence of the retrofit program, the energy consumption of treated and control groups would have followed similar trends, and assuming that the average treatment effects at the household level are consistent over time, the coefficient $\nu$ represents the average treatment effect on energy consumption for neighbors to retrofitted homes (ATT).

The use of the two-way fixed-effect (TWFE) model, under the two previous assumptions helps address several concerns that could otherwise hinder the interpretation of my findings as causal. Firstly, I can eliminate the possibility that the results are influenced by constant differences in energy consumption across households. By incorporating households fixed effects, I can account for all time-invariant characteristics of households such as home characteristics, house size, number of floors and room, along with stable occupant attributes like family size, political views, or environmental attitudes, that can influence the energy consumption and efficiency of homes. Secondly, I can rule out the influence of common monthly trends in consumption over time that impact all houses equally such as weather conditions or energy prices, by adding the month-year fixed effects.

To examine parallel trends and investigate the dynamics of treatment effects, I employ an event-study approach within the Two-Way Fixed Effects (TWFE) model. This involves estimating the model with additional indicators that capture the distance to or from neighbor's retrofit date. This approach allows us to analyze the impact of the treatment over time and assess any variations in effects across different time periods. Specifically, I estimate the following specification:

$$C_{iy} = \alpha_i + \sum_{h=-2}^{h=7} \sigma_h \text{Post Neighbor Retrofit}_{i,y-h} + \gamma_i + \delta_y + \epsilon_{iy} \tag{3.4}$$

Again, I aggregate monthly energy consumption data to calendar year and estimate the leads and lags of a retrofit (the index h). $C_{iy}$ is the natural logarithm of annual gas consumption for household $i$ in calendar year $y$. $PostNeighborRetrofit_{i,y}$ is a dummy variable equal to one if the household's neighbor retrofitted her home in the year y and zero otherwise. Similarly, $treated_{i,y-h}$ is a dummy equal to one for year $h$ before ($h > 0$) or after ($h < 0$) her neighbor retrofitted her home. I include household-year observations up to seven years after and two years prior being treated, and I normalize the estimates by omitting the year before the neighbor retrofitted their home.

While two-way fixed-effects (TWFE) regressions, similar to equation (3.4), are commonly used models for research designs involving staggered adoption, it has been demonstrated that these models yield consistent estimated only when certain assumptions regarding the homogeneity of treatment effects are satisfied. A growing literature shows that if treatment effects vary across different groups or over time, the two-way fixed-effects (TWFE) estimator does not provide consistent estimates for the average treatment effect on the treated (ATT) (De Chaisemartin and d'Haultfoeuille, 2020; Borusyak, Jaravel, and Spiess, 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). To address concerns regarding the reliability of the TWFE estimator, I replicate my results using alternative estimators introduced in Callaway and Sant'Anna (2021), and Sun and Abraham (2021).

## 3.5 Results

In this section, I present the results of estimating the average peer effect of residential energy efficiency retrofits. I then discuss the results from the event-study analysis, followed by robustness checks to my baseline specification.

### 3.5.1 Average Peer Effect on Energy Consumption

#### Baseline specification

In this subsection, I present the results from estimating the DiD specification demonstrated in equation (3.1) using my baseline sample; "Almost retrofitted" sample. Table 3.2 presents the causal effect of being a neighbor to a retrofitted home relative to being a neighbor to an "Almost retrofitted" home on energy consumption in levels and natural logarithms of monthly natural gas and electricity consumption.

The main independent variable is *Post Neighbor Retrofit*, which is equal to one after the neighbor completed the energy efficiency retrofit and zero otherwise. Panel A in Table 3.2 shows the short-term results using consumption data up to the program termination year; 2012. While panel B shows the long-term energy results for up to seven years after the program ended. Column (1) shows that close neighbors to a retrofitted home consume approximately 7 cubic meters less gas per month after their neighbor undergone an energy retrofit. This translates to 2.8% reduction in monthly gas consumption as shown in column (2). Panel B shows that the long term estimates are a little higher at 3.1% reduction per month. Gas results are statistically significant at the 1% level using both consumption ranges. The coefficients associated with $PostAudit$ variable are positive and statistically significant for gas, indicating an increase in gas usage over time for both groups within our sample.

For the electricity results, panel B shows a significant reduction of around 10 kWh/month for neighbors to retrofitted home which translates to 1% monthly electricity savings in the long run. However, I do not find consistency in the electricity results with respect to significance and magnitude of the effect in the short and the long run. This is not surprising for two reasons. First, the city of Medicine Hat is officially labeled "The Gas City", and the majority of houses use gas for heating, which represents the largest share of their energy consumption. Second, the main focus of the ecoENERGY Retrofits for Homes program was primarily on improving space heating and thermal envelope efficiency.[8]

#### Other Control Groups

In this subsection, I present the results of estimating equation (3.3) using the other samples defined above; matched and "naive" samples. Additionally, I compare the results to my baseline sample; "Almost retrofitted" as shown in Table 3.4 and Table 3.5.

Columns (1) and (2) in Table 3.4 and Table 3.5 are copied from Table 3.2 to reflect the DiD estimation of equation (3.1) using the "Almost retrofitted" sample. Columns (3), (4) in both tables show results of the TWFE estimation of equation (3.3) using the matched sample. While columns (5) and (6) reflect the TWFE estimation of equation (3.3) using the "naive" sample. For each sample, I show the natural logarithm of monthly gas consumption in the short and the long term. The detailed results for the matched and the naive samples are in the Appendix (Tables 3.12 and 3.11).

Notably, the sample size grows considerably between the different samples due to the size of the control group in each sample. The gas estimates in Table 3.4 using the three samples are consistent with respect to direction and statistical significance. The magnitude of the effect using the matched sample is considerably lower at about 1.7% compared to 2.8% using the "Almost retrofitted" and "naive" samples in the short run. On average, the effect observed in all three samples indicates a reduction of approximately 2.4% in the short run and 3% in the long run. Examining the electricity results in Table 3.5, there is more variability in terms of magnitude and statistical significance. The findings reveal a significant average decrease of 1.5% in long-term electricity consumption among neighbors in the "Almost retrofitted" and "naive" samples. Whereas no statistical significant effect on electricity usage is observed in the matched sample.

### 3.5.2 Dynamic Peer Effects on Energy Consumption

Next, I study the dynamic peer effect residential energy efficiency retrofits. I aggregate monthly energy consumption data to calendar year and I include household-year observations up to seven years after and two years prior the treatment year. As standard in the literature, the year before the treatment year is the omitted category.

#### Baseline specification

In this subsection, I look at the dynamic peer effect using my baseline sample; "Almost retrofitted", by estimating equation (3.2). The event study in Figure 3.1 shows

[8]The ecoENERGY Retrofit for Homes program offered rebates for installing heat pumps, however, only 3 houses completed a heat pump installation compared to approximately 1,000 households upgrading their gas furnace.

that it takes approximately one period for the peer effect of home energy efficiency retrofits to appear. This is aligned with the fact that there is a time lag between information treatment (through word of mouth and observing neighbor's retrofits) and the decisions to retrofit. I further document that the effect does not fade over time but remains between -2.5% to -4% seven years after the treatment. Appendix Figure 3.6 replicates Figure 3.1using quarterly data.

Additionally, the Figures illustrate that households in the treatment and comparison groups had similar gas consumption before the treatment. The estimate is -1.2 and is statistically insignificant at the 10% significance level.

#### Other Control Groups

In this subsection, I study the dynamic treatment effect of being a neigbour to a retrofitted home using my other samples; matched and "naive", by estimating equation (3.4) for both samples. Figures, 3.2, and 3.4 show that the estimates are consistent with the parallel trends assumption. Regardless of the sample used, the coefficients associated with the pre-treatment year are consistently close to zero and show no discernible pretrends. As expected, there is no significant change in the natural gas use during the neighbor's retrofit year. Savings starts at the first year following the retrofit and progressively increase during the post-periods.

In accordance with Sun and Abraham (2021), the fully dynamic version of the TWFE model, estimated using OLS, provides consistent estimates only when strong assumptions regarding the homogeneity of treatment effects are satisfied. To account for potential heterogeneity in treatment effects across both time and treated units, I present event study figures generated by a set of recently proposed estimators that are robust to treatment effect heterogeneity (Sun and Abraham, 2021; Callaway and Sant'Anna, 2021).

Figures 3.3 and 3.5 show that the TWF event study specification is robust to other dynamic specifications and that treatment effects increase over time starting from the year following the retrofit year of the neighbor and continuing onward. Together, these results show that home energy efficiency retrofits result in a long-term decrease in peer energy consumption. The effects last for at least seven years after the treatment. Last, the effect is robust to three different samples associated with different control groups discussed above.

In the following two subsections of the paper, we reinforce our identification strategy that our primary estimate accurately captures peer effect of residential homes energy efficiency retrofits. To achieve this objective, we demonstrate the pivotal role played by the visibility of retrofits in determining the extent of this effect. Furthermore, we illustrate that as the distance from the retrofitted home increases, the impact gradually attenuates until it eventually dissipates.

## 3.6 Mechanisms

### 3.6.1 Visibility and Word-Of-Mouth

In the context of energy efficiency retrofits, peer effects can operate through various mechanisms such as visibility, word-of-mouth communication, social comparisons, competition, and community engagement initiatives. In my context I believe the two first two are the main operating mechanism for reasons I explain below.

Firstly, the visibility of retrofit processes plays a significant role. When neighbors can visually observe these retrofits, it creates a tangible demonstration of energy-saving actions, which can trigger a sense of curiosity, social influence and motivation among others to inquire and possibly follow suit. The two most visible retrofits in the Eco-energy program are windows and door upgrades, and exterior walls insulation because the installation process is timely, and very visible to close neighbors. In contrast, gas furnace upgrades are less visible due to their swift installation process, which can be completed by a single person.

To test the hypothesis of visibility, two sub-samples were created: one comprising neighbors to homes that underwent visible retrofits (windows or door upgrade, and exterior walls insulation), and another comprising neighbors to homes with less visible retrofits (gas furnace upgrade). The results presented in Table 3.8 support the visibility hypothesis, showing that neighbors to more visibly retrofitted homes achieve significantly higher energy savings of approximately 8% compared to around 1.7% for other retrofits. This magnitude difference is 4.8 times larger, emphasizing the importance of visibility in driving peer energy savings.

Secondly, the word-of-mouth communication between close neighbors, which I believe is another mechanism through which peer effects can occur. Close neighbors often have direct interactions, discussions, and they share their experiences. Conversations among neighbors about the energy savings, comfort improvements, or cost reductions achieved through retrofits can create a ripple effect, motivating more households to undertake similar measures and practices.

### 3.6.2 Behavioral Changes and Energy-Efficiency Investments

I identify two primary mechanisms that contribute to explaining the observed reduction in peer consumption: behavioral changes and energy-efficiency investments. The first mechanism, behavioral changes, involves purposefully modifying one's actions, habits, or routines to achieve desired outcomes or align with specific energy-saving norms. Social influence, the desire to conform, and the motivation to achieve similar energy-saving outcomes drive this mechanism. Behavioral changes encompass a range of actions, such as adjusting thermostat settings by reducing heat during winter or minimizing air conditioning usage during summer. Additionally, adopting energy-conscious lighting practices, such as switching to energy-efficient light bulbs and diligently turning off lights when leaving a room, and utilizing energy-saving modes on appliances like washing machines, dishwashers, and computers, contribute to reducing energy consumption.

The second mechanism, energy-efficiency investments, pertains to individuals making similar investments as their retrofitted neighbors. This could involve upgrading their gas furnace or investing in high-efficiency appliances. However, I find limited evidence of program participation among neighbors, with only 7% of the treated group joining the program at a latter date (at least one month after their neighbor retrofit date). This low participation rate could be attributed to the perceived hassle and associated costs of joining the program. Factors such as the need to schedule, book, and pay for pre and post retrofit audits may deter some from participating, making them more likely to undertake retrofits outside the program to avoid these additional steps.

## 3.7 Robustness Checks

### 3.7.1 Further Neighbors

In this subsection, I expand my analysis beyond the closest two neighbors and examine the effects when considering the 3rd to the 5th closest neighbors and then the 6th to the 10th closest neighbors in spatial proximity. These robustness checks aim to investigate whether the observed peer effects persisted or diminished as I included neighbors further away from the retrofitted homes. If peer effects are at work, I would anticipate that the influence would gradually wane as I move progressively away from the retrofitted home.

Column 1 in Table 3.6 repeats the main result from estimating my TWFE specification in (3.3) using the closest two neighbors as the treated group. In column 2, I use the closest 3rd to 5th neighbor to define the treated and control groups, and column 3 shows the results from considering the 6th to the 10th closest neighbors. I find that the magnitude of the effect decreased from 2.8% to zero and that I lose significance as I look at further neighbors. This suggests that the influence of peer effects on energy consumption attenuates when considering a slightly larger neighborhood radius. Interestingly, when I extended my analysis and consider the 6th to the 10th closest neighbor in proximity, I find that the coefficient of interest is a well-estimated zero. This finding provides reassurance to the identification strategy that the observed peer effects are specific to the closest neighbors and are not present among neighbors residing further away from the retrofitted homes.

### 3.7.2 Different Definition of Neighbors

Up to this part of the paper, peers are considered the two spatially nearest neighbors using geographic coordinates. As a robustness check, I use in this subsection a slightly different definition of neighbors proposed by Kuhn et al., 2011. According to this definition, neighbors should share the same street name, have house numbers directly above or below one's own house number, and possess identical postal codes and street names. The sole distinction between this definition and the previous one is the exclusion of neighbors residing at the front or back of the residence. This analysis serves to reinforce the validity of the earlier results, with the expectation that there won't be substantial variations, given that both cases involve examining the closest two neighbors.

Using this definition, I re-estimate equation (3.3) using my baseline sample; "Almost retrofitted" sample. Results are presented in Table 3.7. Panel A in Table 3.7 shows the short-term gas and electricity results using consumption data up to the program termination year; 2012. While panel B shows the long-term energy results for up to seven years after the program ended. Column (2) shows that close neighbors to a retrofitted home consume approximately 2.3% less in their monthly gas consumption. This estimate is slightly lower compared to a reduction of 2.8% in gas consumption using my baseline peer definition in the previous subsection. However, this is reasonable given the sample size shrinks to almost half in size when we employ the new definition of peers.

## 3.8 Conclusion

Peer effects challenges the assumption of individual autonomy and rationality in Economics, recognizing that people's choices are not made in isolation but are deeply intertwined with the choices and actions of those around them. This paper explores peer effects in home energy efficiency retrofits on energy consumption, contributing to the growing field of peer effects in consumer behavior diffusion. By analyzing data from Canada's largest home energy efficiency program, encompassing over 20,000 households in Medicine Hat, Alberta, I provide robust evidence of the influence of peers' retrofit decisions on household's energy consumption.

Employing a quasi-experimental approach and a generalized difference-in-differences empirical strategy, I find a positive impact on energy savings for households residing in close proximity to retrofitted homes. Positive experiences can inspire others to consider similar retrofits or energy-saving practices, fostering a sense of social norm and peer influence. My paper observe that informal communication network serves as a powerful mechanism for disseminating information, building social norms, and amplifying the impact of energy efficiency retrofits within the community. Notably, as the distance between individuals and their retrofitted neighbors increases, the strength of the documented effect gradually diminishes.

Furthermore, easily visible retrofits such as windows, doors, and exterior wall insulation have a greater influence on energy savings among neighbors compared to less conspicuous retrofits like natural gas furnace upgrades. This paper highlights that visibility of energy-saving practices or upgrades plays a crucial role in facilitating the diffusion of energy-efficient behaviors within a community. Conventionally, when homeowners undertake significant upgrades to their homes, like improving windows and doors or enhancing landscaping, it's common to display signs or indicators in their yards, showcasing the changes they've made. This visible manifestation of energy-saving actions serves as a mechanism of social influence, raising neighbors' awareness and motivating them to contemplate similar enhancements.

One implication arising from this research is the potential benefits of extending this practice to incorporate visible indicators for less conspicuous energy-efficient investments, such as heat pumps or new furnaces. This approach could emerge as an effective strategy to foster the adoption of these technologies within the community. By heightening the visibility of such upgrades, it becomes simpler for neighbors to learn about and emulate these energy-saving practices, further amplifying the potential for peer effects and generating a positive ripple effect throughout the neighborhood.

Overall, these results enhance the understanding of peer effects in the diffusion of consumer behaviors and provide valuable insights for policymakers and practitioners in promoting energy efficiency measures. By leveraging peer influence and harnessing social networks, energy conservation initiatives can be designed to encourage widespread adoption of sustainable practices, contributing to the collective efforts to combat climate change.

## Figures and Tables


FIGURE 3.1: Event Study Using "Almost retrofitted" Sample

Notes: This Figure shows results from the estimation of Equation (3.2) using the neighbors to "Almost retrofitted" homes as the comparison group. This Figure plots the gas peer consumption at yearly intervals around their neighbor's retrofit year. To maintain a balance data set, I choose to include only two pre-periods and seven post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households and year of sample level.

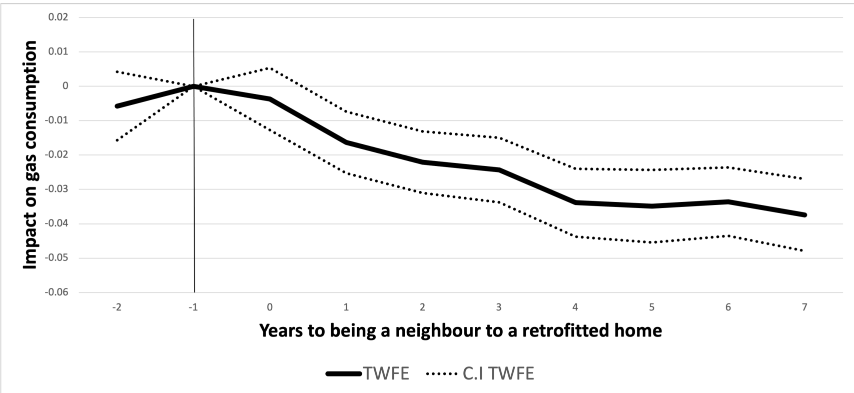



FIGURE 3.2: Event Study Using Matched Sample

Notes: This Figure shows results from an the estimation of Equation (3.4) using the matched comparison group.This Figure plots the gas peer consumption at yearly intervals around their neighbor's retrofit year. To maintain a balance data set, I choose to include only two pre-periods and seven post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households and year of sample level.

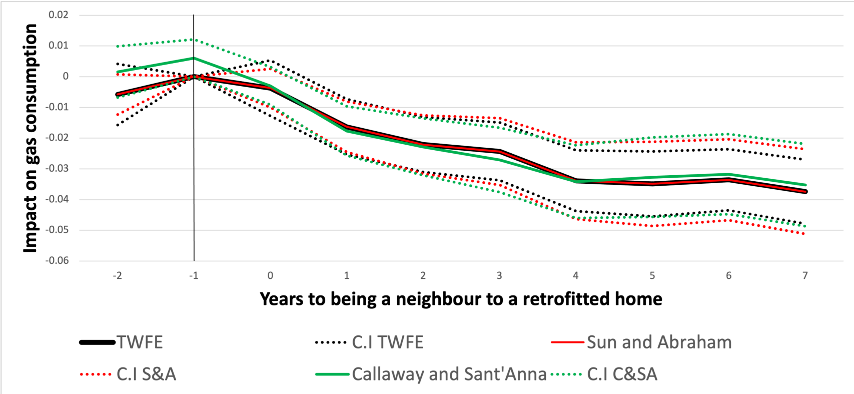



FIGURE 3.3: Event Study Using the Matched Sample: Robust Estimators

This figure overlays the event study plots constructed using three different estimators using the matched comparison group: a dynamic version of the TWFE model, equation ((3.4)), estimated using OLS (in black); Sun and Abraham (2021) (in red) and Callaway and Sant'Anna (2021) (in green). The outcome variable is peer gas consumption. The time variable is the annual gas consumption and the treatment group variable is given by the year in which the neighbor did the energy efficiency retrofits. To maintain a balance data set, I choose in include only two pre-periods and seven post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households level.

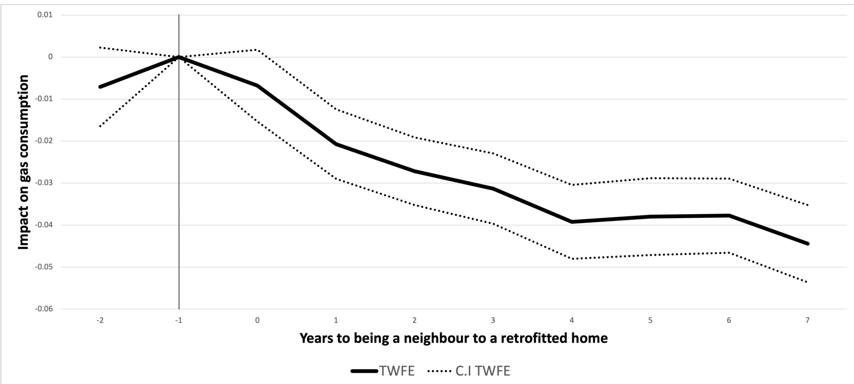



FIGURE 3.4: Event Study Using the "Naive" Sample

Notes: This Figure shows results from an the estimation of Equation (3.4) using the "Naive" comparison group.This Figure plots the gas peer consumption at yearly intervals around their neighbor's retrofit year. To maintain a balance data set, I choose to include only two pre-periods and seven post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households and year of sample level.

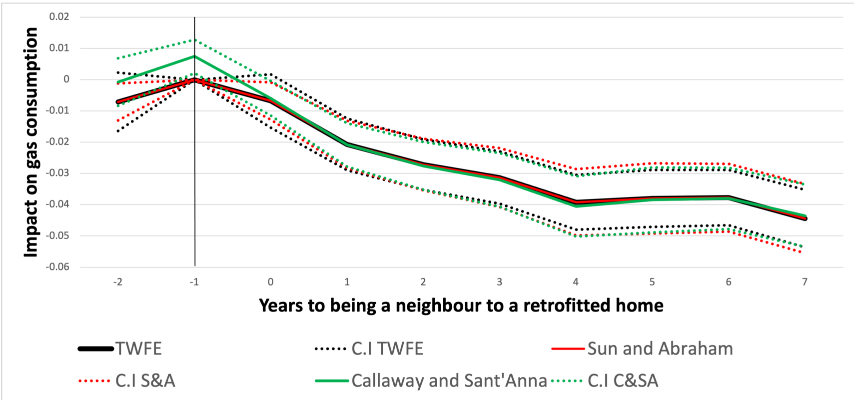



FIGURE 3.5: Event Study Using the "Naive" Sample: Robust Estimators

This figure overlays the event study plots constructed using three different estimators using the "Naive" comparison group: a dynamic version of the TWFE model, equation ((3.4)), estimated using OLS (in black); Sun and Abraham (2021) (in red) and Callaway and Sant'Anna (2021) (in green). The outcome variable is peer gas consumption. The time variable is the annual gas consumption and the treatment group variable is given by the year in which the neighbor did the energy efficiency retrofits. To maintain a balance data set, I choose in include only two pre-periods and seven post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households level.

TABLE 3.1: Summary and Balance Statistics: "Almost retrofitted" Sample

| | Pre-Treatment balance | |
|---|---|---|
| | Control group | Treated group |
| **Monthly gas consumption (m3)** | | |
| Mean | 290.31 | 293.17 |
| Standard deviation | 243.58 | 256.20 |
| Normalize difference | | 0.01 |
| **Monthly electricity consumption (kWh)** | | |
| Mean | 810.96 | 799.93 |
| Standard deviation | 472.80 | 522.56 |
| Normalize difference | | -0.02 |
| **Lot size ($km^2$)** | | |
| Mean | 779.95 | 635.23 |
| Standard deviation | 1,033.57 | 241.73 |
| Normalize difference | | -0.14 |
| **Building size ($m^2$)** | | |
| Mean | 131.15 | 116.50 |
| Standard deviation | 62.00 | 41.37 |
| Normalize difference | | -0.20 |
| **Total assessment value (CAD)** | | |
| Mean | 314,696 | 275,920 |
| Standard deviation | 183,855 | 96,683 |
| Normalize difference | | -0.19 |
| **Year built** | | |
| Mean | 1984 | 1981 |
| Standard deviation | 21 | 20 |
| Normalize difference | | -0.09 |
| Number of households | 387 | 2,287 |

Notes: The Table shows the mean and standard deviation for energy consumption and house characteristics for the treated and control group in the "Almost retrofitted" sample. The gas and electricity figures reflect the pre-treatment period. Additionally, we provide the normalized difference balance statistics for the treated group relative to the control group, indicating the level of overlap in covariates between the two samples. A normalized difference less than 0.25 is typically considered good overlap (Imbens and Wooldridge, 2009).

TABLE 3.2: Peer Energy Savings: "Almost retrofitted" Sample

| Panel A | Short term: up to 2012 | | | |
|---|---|---|---|---|
| | Monthly gas (m3) | Log monthly gas | Monthly elec (kWh) | Log monthly elec |
| | (1) | (2) | (3) | (4) |
| Post Neigbour Retrofit | -7.138*** | -0.028*** | 2.457 | -0.001 |
| | (1.750) | (0.005) | (4.065) | (0.006) |
| Post Audit | 5.060*** | 0.022*** | -3.849 | -0.003 |
| | (1.945) | (0.006) | (4.344) | (0.007) |
| Constant | 298.283*** | 5.284*** | 756.270*** | 6.453*** |
| | (0.717) | (0.002) | (1.542) | (0.002) |
| R-squared | 0.745 | 0.839 | 0.606 | 0.548 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 199,973 | | | |
| Number of houses | 2,674 | | | |
| Panel B | Long term up: to 2019 | | | |
| Post Neigbour Retrofit | -9.157*** | -0.031*** | -10.218*** | -0.010** |
| | (1.421) | (0.004) | (3.573) | (0.005) |
| Post Audit | 6.859*** | 0.027*** | 1.276 | 0.001 |
| | (1.671) | (0.005) | (4.042) | (0.006) |
| Constant | 282.409*** | 5.212*** | 743.997*** | 6.443*** |
| | (0.927) | (0.003) | (2.043) | (0.003) |
| R-squared | 0.763 | 0.856 | 0.571 | 0.515 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 431,731 | | | |
| Number of houses | 2,678 | | | |

Notes: The Table shows results from an OLS estimation of Equation (3.1). The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to "Almost retrofitted" homes, that undertook only the pre-retrofit audit and did not complete the retrofit process. Panel A presents the short term results up to the program termination year; 2012. And Panel B presents the long term results up to 2019. The dependent variable is a dummy that equals to one if a household's neigbour retrofitted his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the comparison group. The outcome variables in Columns (1) to (4) reflect monthly consumption for gas in levels, gas in natural logarithm, electricity in levels and electricity in natural logarithm, respectively. The standard errors in parentheses are two-way clustered at the household and month of sample. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 3.3: Baseline Specification Using Different Control Groups: Gas Results

| | Almost retrofitted sample | | Matched sample | | Naive sample | |
|---|---|---|---|---|---|---|
| | Short-term | Long-term | Short-term | Long-term | Short-term | Long-term |
| | | | Log monthly gas | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated | -0.028*** | -0.031*** | -0.017*** | -0.022*** | -0.027*** | -0.037*** |
| | (0.005) | (0.004) | (0.003) | (0.002) | (0.002) | (0.002) |
| Constant | 5.284*** | 5.212*** | 5.265*** | 5.202*** | 5.221*** | 5.159*** |
| | (0.002) | (0.003) | (0.001) | (0.001) | (0.000) | (0.000) |
| R-squared | 0.839 | 0.856 | 0.838 | 0.856 | 0.838 | 0.855 |
| House fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 199,973 | 431,731 | 342,104 | 739,741 | 2,109,967 | 4,689,803 |
| Number of houses | 2,674 | 2,678 | 4,730 | 4,730 | 17,083 | 17,651 |

Notes: The Table shows results from an OLS estimation of Equation (3.1) in Columns (1) and (2) and (3.3) in Columns (3), (4), (5) and (6), using the three proposed samples in the paper;"Almost retrofitted" sample, "Naive" sample and matched sample. The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to untreated houses in each of the three samples. Columns (1), (3), and (5) present the short term results up to the program termination year of 2012, while columns (2), (4), and (6) present the long term results up to 2019. The dependent variable is a dummy that equals to one if a household's neigbour retrofitted his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the comparison group. The outcome variable is monthly gas consumption in natural logarithm. The standard errors in parentheses are two-way clustered at the household and month of sample. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 3.4: Baseline Specification Using Different Control Groups: Gas Results

| | Almost retrofitted sample | | Matched sample | | Naive sample | |
|---|---|---|---|---|---|---|
| | Short-term | Long-term | Short-term | Long-term | Short-term | Long-term |
| | Log monthly gas | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated | -0.028*** | -0.031*** | -0.017*** | -0.022*** | -0.027*** | -0.037*** |
| | (0.005) | (0.004) | (0.003) | (0.002) | (0.002) | (0.002) |
| Constant | 5.284*** | 5.212*** | 5.265*** | 5.202*** | 5.221*** | 5.159*** |
| | (0.002) | (0.003) | (0.001) | (0.001) | (0.000) | (0.000) |
| R-squared | 0.839 | 0.856 | 0.838 | 0.856 | 0.838 | 0.855 |
| House fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 199,973 | 431,731 | 342,104 | 739,741 | 2,109,967 | 4,689,803 |
| Number of houses | 2,674 | 2,678 | 4,730 | 4,730 | 17,083 | 17,651 |

Notes: The Table shows results from an OLS estimation of Equation (3.1) in Columns (1) and (2) and (3.3) in Columns (3), (4), (5) and (6), using the three proposed samples in the paper;"Almost retrofitted" sample, "Naive" sample and matched sample. The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to untreated houses in each of the three samples. Columns (1), (3), and (5) present the short term results up to the program termination year of 2012, while columns (2), (4), and (6) present the long term results up to 2019. The dependent variable is a dummy that equals to one if a household's neigbour retrofitted his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the comparison group. The outcome variable is monthly gas consumption in natural logarithm. The standard errors in parentheses are two-way clustered at the household and month of sample. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 3.5: Baseline Specification Using Different Control Groups: Electricity Results

| | Almost retrofitted Sample | | Matched Sample | | Naive Sample | |
|---|---|---|---|---|---|---|
| | Short-term | Long-term | Short-term | Long-term | Short-term | Long-term |
| | Log monthly elec | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treated | -0.001 | -0.010** | 0.001 | 0.000 | -0.026*** | -0.020*** |
| | (0.006) | (0.005) | (0.003) | (0.002) | (0.002) | (0.002) |
| Constant | 6.453*** | 6.443*** | 6.439*** | 6.422*** | 6.437*** | 6.427*** |
| | (0.002) | (0.003) | (0.001) | (0.001) | (0.000) | (0.000) |
| R-squared | 0.548 | 0.515 | 0.539 | 0.508 | 0.515 | 0.492 |
| House fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Number of observations | 199,973 | 431,731 | 342,104 | 739,741 | 2,109,967 | 4,689,803 |
| Number of houses | 2,674 | 2,678 | 4,730 | 4,730 | 17,083 | 17,651 |

Notes: The Table shows results from an the estimation of Equation (3.1) in Columns (1) and (2) and (3.3) in Columns (3), (4), (5) and (6), using the three proposed samples in the paper;"Almost retrofitted" sample, "Naive" sample and matched sample. The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to untreated houses in each of the three samples. Columns (1), (3), and (5) present the short term results up to the program termination year of 2012, while columns (2), (4), and (6) present the long term results up to 2019. The dependent variable is a dummy that equals to one if a household's neigbour retrofitted his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the comparison group. The outcome variable is monthly electricity consumption in natural logarithm. The standard errors in parentheses are two-way clustered at the household and month of sample. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 3.6: Mechanism: Further Neighbors

| | 1st - 2nd neighbor | 3th - 5th neighbor | 6th - 10th neighbor |
|---|---|---|---|
| | Log (monthly) gas | | |
| | (1) | (2) | (3) |
| Post Neighbor Retrofit | -0.028*** | -0.002 | -0.007** |
| | (0.005) | (0.006) | (0.004) |
| Post Audit | 0.022*** | -0.001 | 0.003 |
| | (0.006) | (0.006) | (0.004) |
| Constant | 5.284*** | 5.283*** | 5.276*** |
| | (0.002) | (0.002) | (0.001) |
| R-squared | 0.839 | 0.839 | 0.835 |
| Number of observations | 199,973 | 201,139 | 585,351 |
| Number of houses | 2,674 | 2,639 | 6,277 |

Notes: The Table shows results from an OLS estimation of Equation (3.1) for different sets of neighbors with respect to spatial proximity. The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to "Almost retrofitted" homes. The dependent variable is dummy equals to one when a household's neigbour retrofit his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the control group. The outcome variable is monthly gas consumption in natural logarithm. Column (1) considers the closest 2 neighbors, column (2) shows results reflecting the 3th, 4th and 5th neigbour, and column (3) shows results reflecting the 6th, 7th, 8th, 9th and 10th neigbour, in spatial proximity to the retrofitted home. The standard errors in parentheses are two-way clustered at the household and month of sample. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 3.7: Peer Energy Savings: "Almost retrofitted" Sample Using Same Street Neighbors

| | Short term: up to 2012 | | | |
|---|---|---|---|---|
| Panel A | Monthly gas (m3) | Log monthly gas | Monthly elec (kWh) | Log monthly elec |
| | (1) | (2) | (3) | (4) |
| Post Neigbour Retrofit | -2.766 | -0.023*** | 13.291** | -0.009 |
| | (2.390) | (0.007) | (5.450) | (0.007) |
| Post Audit | -0.003 | 0.014* | -4.694 | 0.012 |
| | (2.596) | (0.008) | (5.788) | (0.008) |
| Constant | 302.857*** | 5.299*** | 757.789*** | 6.458*** |
| | (0.916) | (0.003) | (1.916) | (0.003) |
| R-squared | 0.746 | 0.839 | 0.585 | 0.549 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 110,400 | | | |
| Number of houses | 1,517 | | | |
| Panel B | Long term up: to 2019 | | | |
| Post Neigbour Retrofit | -4.931** | -0.022*** | 9.703** | 0.006 |
| | (1.930) | (0.006) | (4.778) | (0.006) |
| Post Audit | 2.945 | 0.015** | -2.396 | 0.002 |
| | (2.233) | (0.007) | (5.376) | (0.007) |
| Constant | 286.482*** | 5.229*** | 735.624*** | 6.436*** |
| | (1.186) | (0.003) | (2.600) | (0.004) |
| R-squared | 0.760 | 0.857 | 0.554 | 0.513 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 238,231 | | | |
| Number of houses | 1,519 | | | |

Notes: The Table shows results from an OLS estimation of Equation(3.1). The treated group is the 2 closest neighbors to a retrofitted home, considering only same street neighbors, and the comparison group is the 2 closest neighbors to "Almost retrofitted" homes, that undertook only the pre-retrofit audit and did not complete the retrofit process. Panel A presents the short term results up to the program termination year; 2012. And Panel B presents the long term results up to 2019. The dependent variable is a dummy that equals to one if a household's neigbour retrofitted his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the comparison group. The outcome variables in Columns (1) to (4) reflect monthly consumption for gas in levels, gas in natural logarithm, electricity in levels and electricity in natural logarithm, respectively. The standard errors in parentheses are two-way clustered at the household and month of sample. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

TABLE 3.8: Visible vs Less Visible Retrofits

| | Visible upgrades | Less visible upgrades |
|---|---|---|
| | Log (monthly) gas | Log (monthly) gas |
| | (1) | (2) |
| Post Neigbour Retrofit | -0.078*** | -0.022*** |
| | (0.014) | (0.006) |
| Post Audit | 0.011 | 0.013** |
| | (0.009) | (0.006) |
| Constant | 5.294*** | 5.291*** |
| | (0.005) | (0.003) |
| R-squared | 0.846 | 0.841 |
| Number of observations | 32,197 | 125,027 |
| Number of houses | 458 | 1,713 |

Notes: The Table shows results from an OLS estimation of Equation (3.1). The treated group is the 2 closest neighbors to a retrofitted home, and the comparison group is the 2 closest neighbors to "Almost retrofitted" homes. We split the sample into two groups, neighbors to houses that did a visible retrofit (windows and door upgrades, or/and exterior walls insulation) in Column (1), and neighbors to houses that only did a less visible retrofit (natural gas furnace upgrade) in Column (2). The dependent variable is dummy equals to one when a household's neigbour retrofit his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the control group. The outcome variable is monthly gas consumption in natural logarithm. The standard errors in parentheses are two-way clustered at the household and month of sample. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

# Appendix

## Figures and Tables

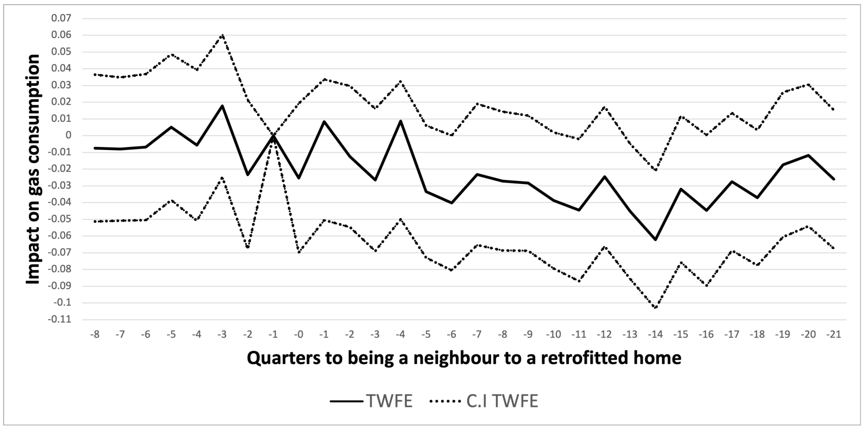



FIGURE 3.6: Event Study Using "Almost retrofitted" Sample: Quarter Data

Notes: This Figure shows results from an the estimation of Equation (3.2) using the neighbors to "Almost retrofitted" homes as the comparison group. This Figure plots the gas peer consumption at quarterly intervals around their neighbor's retrofit quarter. To maintain a balance data set, I choose to include only eight pre-periods and twenty one post periods. The dotted lines represent 95 percent confidence intervals. Standard errors are clustered at the households and quarter of sample level.

TABLE 3.9: Balance Statistics: Matched sample

| | Control group | Treated group |
|---|---|---|
| **Monthly gas consumption (m3)** | | |
| Mean | 278.35 | 294.73 |
| Standard deviation | 254.38 | 257.91 |
| Normalize difference | | 0.05 |
| **Monthly electricity consumption (kWh)** | | |
| Mean | 769.65 | 805.81 |
| Standard deviation | 487.76 | 519.15 |
| Normalize difference | | 0.05 |
| **Lot size($km^2$)** | | |
| Mean | 802.34 | 655.53 |
| Standard deviation | 1,635.03 | 700.53 |
| Normalize difference | | -0.08 |
| **Building size($m^2$)** | | |
| Mean | 118.31 | 117.73 |
| Standard deviation | 48.17 | 43.88 |
| Normalize difference | | -0.01 |
| **Total assessment value (CAD)** | | |
| Mean | 279,226 | 280,847 |
| Standard deviation | 100,086 | 84,816 |
| Normalize difference | | 0.01 |
| **Year built** | | |
| Mean | 1982 | 1982 |
| Standard deviation | 20.80 | 19.93 |
| Normalize difference | | -0.01 |
| Number of households | 2,553 | 2,177 |

Notes: The Table shows the mean and standard deviation for energy consumption and house characteristics for the treated and control group in the matched sample. The gas and electricity figures reflect the pre-treatment period. Additionally, we provide the normalized difference balance statistics for the treated group relative to the control group, indicating the level of overlap in covariates between the two samples. A normalized difference less than 0.25 is typically considered good overlap (Imbens and Wooldridge, 2009).

TABLE 3.10: Balance Statistics: "Naive" Sample

| | Control group | Treated group |
|---|---|---|
| **Monthly gas consumption (m3)** | | |
| Mean | 271.85 | 294.68 |
| Standard deviation | 242.39 | 257.88 |
| Normalize difference | | 0.06 |
| **Monthly electricity consumption (kWh)** | | |
| Mean | 779.47 | 805.64 |
| Standard deviation | 487.74 | 519.13 |
| Normalize difference | | 0.04 |
| **Lot size ($km^2$)** | | |
| Mean | 121.84 | 655.48 |
| Standard deviation | 57.05 | 700.40 |
| Normalize difference | | -0.10 |
| **Building size ($m^2$)** | | |
| Mean | 123.24 | 117.72 |
| Standard deviation | 70.09 | 43.87 |
| Normalize difference | | -0.06 |
| **Total assessment value (CAD)** | | |
| Mean | 293,250.60 | 280,823.70 |
| Standard deviation | 111,248.10 | 84,808.03 |
| Normalize difference | | -0.09 |
| **Year built** | | |
| Mean | 1986 | 1982 |
| Standard deviation | 21.34 | 19.94 |
| Normalize difference | | -0.16 |
| Number of households | 2,554 | 14,529 |

Notes: The Table shows the mean and standard deviation for energy consumption and house characteristics for the treated and control group in the "Naive" sample. The gas and electricity figures reflect the pre-treatment period. Additionally, we provide the normalized difference balance statistics for the treated group relative to the control group, indicating the level of overlap in covariates between the two samples. A normalized difference less than 0.25 is typically considered good overlap (Imbens and Wooldridge, 2009).

TABLE 3.11: Peer energy savings: Matched sample

| | Short term: up to 2012 | | | |
|---|---|---|---|---|
| Panel A | Monthly gas (m3) | Log monthly gas | Monthly elec (kWh) | Log monthly elec |
| | (1) | (2) | (3) | (4) |
| Treated | -4.782*** | -0.017*** | 1.461 | 0.001 |
| | (0.861) | (0.003) | (1.895) | (0.003) |
| Constant | 292.989*** | 5.265*** | 747.039*** | 6.439*** |
| | (0.318) | (0.001) | (0.701) | (0.001) |
| R-squared | 0.740 | 0.838 | 0.591 | 0.539 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | | 342,104 | | |
| Number of houses | | 4,730 | | |
| Panel B | | Long term up: to 2009 | | |
| Treated | -6.827*** | -0.022*** | -3.357** | 0.000 |
| | (0.700) | (0.002) | (1.605) | (0.002) |
| Constant | 279.304*** | 5.202*** | 729.697*** | 6.422*** |
| | (0.322) | (0.001) | (0.750) | (0.001) |
| R-squared | 0.760 | 0.856 | 0.556 | 0.508 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | | 739,741 | | |
| Number of houses | | 4,730 | | |

Notes: The Table shows results from an OLS estimation of Equation (3.3) using the matched sample that we created through the nearest neigbour matching on house characteristics with no replacement. The treated group is the 2 closest neighbors to a retrofitted home, and the control group is the 2 closest neighbors to houses that never take part of the retrofit program. Panel A presents the short term results up to the program termination year of 2012. And Panel B presents the long term results up to 2019. The dependent variable is dummy equals to one when a household's neigbour retrofit his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the control group. The outcome variables in Columns (1) to (4) reflect consumption for gas is levels, gas in natural logarithm. electricity in levels and electricity in natural logarithm, respectively. The standard errors in parenthesis are two-way clustered at the household and month of sample. *** p<0.01, ** p<0.05, * p<0.1.

TABLE 3.12: Peer energy savings: "Naive" Sample

| | Short term: up to 2012 | | | |
|---|---|---|---|---|
| Panel A | Monthly gas (m3) | Log monthly gas | Monthly elec (kWh) | Log monthly elec |
| | (1) | (2) | (3) | (4) |
| Treated | -7.990*** | -0.027*** | -9.749*** | -0.026*** |
| | (0.751) | (0.002) | (1.714) | (0.002) |
| Constant | 278.007*** | 5.221*** | 745.428*** | 6.437*** |
| | (0.125) | (0.000) | (0.296) | (0.000) |
| R-squared | 0.744 | 0.838 | 0.578 | 0.515 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 2,109,967 | | | |
| Number of houses | 17,083 | | | |
| Panel B | Long term up: to 2009 | | | |
| Treated | -11.352*** | -0.037*** | -10.626*** | -0.020*** |
| | (0.589) | (0.002) | (1.387) | (0.002) |
| Constant | 264.759*** | 5.159*** | 730.110*** | 6.427*** |
| | (0.083) | (0.000) | (0.209) | (0.000) |
| R-squared | 0.765 | 0.855 | 0.548 | 0.492 |
| House fixed effects | Yes | Yes | Yes | Yes |
| Calender month fixed effects | Yes | Yes | Yes | Yes |
| Number of observations | 4,689,803 | | | |
| Number of houses | 17,651 | | | |

Notes: The Table shows results from an OLS estimation of Equation (3.3) using the "Naive" sample. The treated group is the 2 closest neighbors to a retrofitted home, and the control group is the 2 closest neighbors to all houses in the city that never take part of the retrofit program. Panel A presents the short term results up to the program termination year of 2012. And Panel B presents the long term results up to 2019. The dependent variable is dummy equals to one when a household's neigbour retrofit his home, and zero before becoming a neigbour to a retrofitted home or if the house is part of the control group. The outcome variables in Columns (1) to (4) reflect consumption for gas is levels, gas in natural logarithm. electricity in levels and electricity in natural logarithm, respectively. The standard errors in parenthesis are two-way clustered at the household and month of sample. *** p<0.01, ** p<0.05, * p<0.1.

# Bibliography

Agarwal, Sumit, Wenlan Qian, and Xin Zou (2021). “Thy neighbor’s misfortune: Peer effect on consumption”. In: *American Economic Journal: Economic Policy* 13.2, pp. 1–25.

Aigner, Dennis J., Cynts Sorooshian, and Pamela Kerwin (1984). “Conditional Demand Analysis for Estimating Residential End-Use Load Profiles”. In: *The Energy Journal* 0.Number 3, pp. 81–98.

Allcott, Hunt (2011). “Social norms and energy conservation”. In: *Journal of Public Economics* 95.9-10, pp. 1082–1095.

Allcott, Hunt and Todd Rogers (2014). “The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation”. In: *American Economic Review* 104.10, pp. 3003–3037.

Anderson, Craig A (2001). “Heat and violence”. In: *Current directions in psychological science* 10.1, pp. 33–38.

Angrist, Joshua D and Kevin Lang (2004). “Does school integration generate peer effects? Evidence from Boston’s Metco Program”. In: *American Economic Review* 94.5, pp. 1613–1634.

Aranda, Alfonso et al. (2012). “Multiple regression models to predict the annual energy consumption in the Spanish banking sector”. In: *Energy and Buildings* 49, pp. 380–387.

Archibald, R.G., D.H. Finifter, and C.E. Moody (1982). “Seasonal Variation in Residential Electricity Demand: Evidence form Survey Data”. In: *Applied Economics* 14, pp. 167–181.

ASHRAE (2009). *ASHRAE Handbook Fundamentals*. American Society of Heating, Refrigeration and Air-Conditioning Engineers.

Aydinalp-Koksal, Merih and V Ismet Ugursal (2008). “Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector”. In: *Applied Energy* 85.4, pp. 271–296.

Ayres, Ian, Sophie Raseman, and Alice Shih (2013). “Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage”. In: *Journal of Law, Economics, and Organization* 29.5, pp. 992–1022.

Azam, Muhammad (2016). “Does environmental degradation shackle economic growth? A panel data investigation on 11 Asian countries”. In: *Renewable and Sustainable Energy Reviews* 65, pp. 175–182.

Baltagi, B (2008). *Econometric analysis of panel data, 4th edition*. New York, NY: Wiley & Sons.

Bartels, Robert and Denzil G. Fiebig (1990). "Integrating Direct Metering and Conditional Demand Analysis for Estimating End-Use Loads". In: *The Energy Journal* 11.4, pp. 79–97.

— (2000). "Residential End-Use Electricity Demand: Results from a Designed Experiment". In: *The Energy Journal* 21, pp. 51–81.

Battles, Stephanie J (1994). "Validation of conditional demand estimates: does it lead to model improvements?" In: *Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings (Pacific Grove, CA, USA)* 7, pp. 13–22.

Bauwens, L, D Fiebig, and M Steel (1994). "Estimating end-use demand: a bayesian approach". In: *Journal of Business and Economic Statistics* 12.2, pp. 221–231.

Bayer, Patrick, Stephen L Ross, and Giorgio Topa (2008). "Place of work and place of residence: Informal hiring networks and labor market outcomes". In: *Journal of Political Economy* 116.6, pp. 1150–1196.

Baylis, Patrick (2020). "Temperature and temperament: Evidence from Twitter". In: *Journal of Public Economics* 184, p. 104161.

Baylon, D., P. Storm, and D. Robison (2013). *Ductless Heat Pump Impact & Process Evaluation: Billing Analysis Report*. Ecotope Inc., for the Northwest Energy Efficiency Alliance, NEEA Report No. 13-262.

Berg, Sanford V. (1988). *Survey of Household End-Use Consumption*. Public Utility Research Center Working Paper.

Bollinger, Bryan, Jesse Burkhardt, and Kenneth T Gillingham (2020). "Peer effects in residential water conservation: Evidence from migration". In: *American Economic Journal: Economic Policy* 12.3, pp. 107–133.

Bollinger, Bryan and Kenneth Gillingham (2012). "Peer effects in the diffusion of solar photovoltaic panels". In: *Marketing Science* 31.6, pp. 900–912.

Bollinger, Bryan, Kenneth Gillingham, and Kelley Gullo (2020). *Making pro-social social: The effectiveness of social norm appeals for energy conservation using social media*.

Boomhower, Judson and Lucas W Davis (2014). "A credible approach for measuring inframarginal participation in energy efficiency programs". In: *Journal of Public Economics* 113, pp. 67–79.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021). "Revisiting event study designs: Robust and efficient estimation". In: *arXiv preprint arXiv:2108.12419*.

Burgess, Robin et al. (2017). "Weather, climate change and death in India". In: *University of Chicago*.

Burlig, Fiona et al. (2020). "Machine learning from schools about energy efficiency". In: *Journal of the Association of Environmental and Resource Economists* 7.6, pp. 1181–1217.

Bursztyn, Leonardo et al. (2014). "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions". In: *Econometrica* 82.4, pp. 1273–1301.

Calderon, Carlos et al. (2019). "An area-based modelling approach for planning heating electrification". In: *Energy Policy* 131, pp. 262–280.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Canada Energy Regulator (2017). *Canada's Renewable Power Landscape 2017 - Energy Market Analysis*.

Card, David et al. (2012). "Inequality at work: The effect of peer salaries on job satisfaction". In: *American Economic Review* 102.6, pp. 2981–3003.

Carrell, Scott E, Mark Hoekstra, and Elira Kuka (2018). "The long-run effects of disruptive peers". In: *American Economic Review* 108.11, pp. 3377–3415.

Carrell, Scott E., Mark Hoekstra, and James E. West (2011). "Is poor fitness contagious?: Evidence from randomly assigned friends". In: *Journal of Public Economics* 95.7, pp. 657–663. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2010.12.005. URL: https://www.sciencedirect.com/science/article/pii/S0047272710002070.

Caves, Douglas W. et al. (1987). "A Bayesian Approach to Combining Conditional Demand and Engineering Models of Electricity Usage". In: *The Review of Economics and Statistics* 69.3, pp. 438–448.

Chauvin, Juan Pablo et al. (2017). "What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States". In: *Journal of Urban Economics* 98, pp. 17–49.

Christensen, Peter et al. (2022). *Energy Efficiency Can Deliver for Climate Policy:Evidence from Machine Learning-Based Targeting*. NBER Working Paper 30467. DOI: https://doi.org/10.3386/w30467.

Chuang, Yating, Magali A Delmas, and Stephanie Pincetl (2022). "Are Residential Energy Efficiency Upgrades Effective? An Empirical Analysis in Southern California". In: *Journal of the Association of Environmental and Resource Economists* 9.4, pp. 641–679.

Conley, Timothy G and Christopher R Udry (2010). "Learning about a new technology: Pineapple in Ghana". In: *American Economic Review* 100.1, pp. 35–69.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg (2017). "Peer effects in the workplace". In: *American Economic Review* 107.2, pp. 425–456.

Corno, Lucia, Nicole Hildebrandt, and Alessandra Voena (2020). "Age of marriage, weather shocks, and the direction of marriage payments". In: *Econometrica* 88.3, pp. 879–915.

Costa, Dora L and Matthew E Kahn (2013a). "Do Liberal Home Owners Consume Less Electricity? A Test of the Voluntary Restraint Hypothesis". In: *Economic Letters* 119.2, pp. 2010–2012.

— (2013b). "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment". In: *Journal of the European Economic Association* 11.3, pp. 680–702.

Dahl, Gordon B, Katrine V Løken, and Magne Mogstad (2014). "Peer effects in program participation". In: *American Economic Review* 104.7, pp. 2049–2074.

De Almeida, Anibal et al. (2011). "Characterization of the household electricity consumption in the EU, potential energy savings and specific policy recommendations". In: *Energy and Buildings* 43.8, pp. 1884–1894.

De Chaisemartin, Clément and Xavier d'Haultfoeuille (2020). "Two-way fixed effects estimators with heterogeneous treatment effects". In: *American Economic Review* 110.9, pp. 2964–2996.

Delmastro, Chiara (2022). *Buildings: Sectorial Overview*. IEA. https://www.iea.org/reports/buildings.

Delzendeh, Elham et al. (2017). "The impact of occupants' behaviours on building energy analysis: A research review". In: *Renewable and Sustainable Energy Reviews* 80, pp. 1061–1071.

Deschênes, Olivier and Michael Greenstone (2011). "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US". In: *American Economic Journal: Applied Economics* 3.4, pp. 152–85.

Dolan, Paul and Robert Metcalfe (2015). "Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation". In: *Becker Friedman Institute for Research in Economics Working Paper* 2589269.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011). "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya". In: *American Economic Review* 101.5, pp. 1739–1774.

Environment Canada (2020). *Historical Data*. Available at https://climate.weather.gc.ca/historical_data/search_historic_data_e.html.

Fiebig, Denzil G., Robert Bartels, and Dennis J. Aigner (1991). "A random coefficient approach to the estimation of residential end-use load profiles". In: *Journal of Econometrics* 50, pp. 297–327.

Fletcher, Jason M and Stephen L Ross (2018). "Estimating the effects of friends on health behaviors of adolescents". In: *Health Economics* 27.10, pp. 1450–1483.

Florida Power & Light (1986). *A Model of Residential Energy Consumption and Appliance Ownership*. Florida Power & Light Company.

FOSS (2008). *Working with The Third Sector*. Front Offices Shared Services/Cabinet Office/Local Government Association. DOI: https://www.local.gov.uk/sites/default/files/documents/working-third-sector-pdf--1ab.pdf.

Foster, Andrew D and Mark R Rosenzweig (1995). "Learning by doing and learning from others: Human capital and technical change in agriculture". In: *Journal of Political Economy* 103.6, pp. 1176–1209.

Foucquier, Aurélie et al. (2013). "State of the art in building modelling and energy performances prediction: A review". In: *Renewable and Sustainable Energy Reviews* 23, pp. 272–288.

Fowler, James H and Nicholas A Christakis (2008). "Estimating peer effects on health in social networks: a response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, Pais". In: *Journal of Health Economics* 27.5, p. 1400.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram (2018). "Do energy efficiency investments deliver? Evidence from the weatherization assistance program". In: *Quarterly Journal of Economics* 133.3, pp. 1597–1644.

Fumo, Nelson (2014). "A review on the basics of building energy estimation". In: *Renewable and Sustainable Energy Reviews* 31, pp. 53–60.

Giandomenico, Lauren, Maya Papineau, and Nicholas Rivers (2020). *A systematic review of energy efficiency home retrofit evaluation studies*. Working Paper.

Gillingham, Kenneth and Tsvetan Tsvetanov (2018). "Nudging energy efficiency audits: Evidence from a field experiment". In: *Journal of Environmental Economics and Management* 90, pp. 303–316.

Goodman, Joshua et al. (2018). *Heat and learning*. Tech. rep. National Bureau of Economic Research.

Goodman-Bacon, Andrew (2021). "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2, pp. 254–277.

Government of Canada (2021). *Budget 2021*. Tech. rep. URL: https://www.budget.canada.ca/2021/report-rapport/p2-en.html#chap6.

Graziano, Marcello and Kenneth Gillingham (2015). "Spatial patterns of solar photovoltaic system adoption: the influence of neighbors and the built environment". In: *Journal of Economic Geography* 15.4, pp. 815–839.

Halvorsen, Bente and Bodil Merethe Larsen (2013). *How do investments in heat pumps affect household energy consumption?* Tech. rep. Discussion Papers.

Hamelin, M-C, N. Michaud, and E. Chartrand (2018). *Establishing the Energy Performance of Mini-Split and Central Air-Source Heat Pumps Through Billing Analysis*. International Energy Policy & Programme Evaluation Conference (Vienna, Austria).

Hausman, Jerry A (1978). "Specification tests in econometrics". In: *Econometrica: Journal of the econometric society*, pp. 1251–1271.

Heutel, Garth and Erich Muehlegger (2015). "Consumer learning and hybrid vehicle adoption". In: *Environmental and Resource Economics* 62.1, pp. 125–161.

Hsiao, Cheng, Dean C. Mountain, and Kathleen Ho Illman (1995). "A Bayesian Integration of End-Use Metering and Conditional-Demand Analysis". In: *Journal of Business & Economic Statistics* 13.3, pp. 315–326.

IEA (2017a). *Energy Technology Perspectives 2017: Catalyzing Energy Technology Transformations*. International Energy Agency, Paris, France.

— (2017b). *World Energy Outlook 2017*. International Energy Agency. URL: https://www.iea.org/policies/1345-energuide-for-houses.

Imbens, Guido W and Jeffrey M Wooldridge (2009). "Recent developments in the econometrics of program evaluation". In: *Journal of Economic Literature* 47.1, pp. 5–86.

International Energy Agency (2021). *World Energy Outlook 2021*. www.iea.org/weo.

IPCC (2018). "Summary for Policymakers". In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels* *and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Ed. by V. Masson-Delmotte et al.

Isaacs, N.P et al. (2010). "Energy use in New Zealand Households: Final Report on the Household Energy End-use Project (HEEP)". In: *BRANZ Study* Report 221.

ISO (1989). *Hot environments - estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature)*. Tech. rep. ISO 7243 Standard. Geneva, Switzerland: International Organization for Standardization.

Jessoe, Katrina, Maya Papineau, and David Rapson (2020). "Utilities Included: Split Incentives in Commercial Electricity Contracts". In: *The Energy Journal* 41.5.

Judge, George et al. (1985). *The theory and practice of econometrics, 2nd ed*. New York, NY: Wiley.

Korn, D. et al. (2016). *Ductless Mini-Split Heat Pump Impact Evaluation*. Cadmus Group Inc., for The Electric and Gas Program Administrators of Massachusetts and Rhode Island.

Kotchen, Matthew J (2017). "Longer-Run Evidence on Whether Building Energy Codes Reduce Residential Energy Consumption". In: *Journal of the Association of Environmental and Resource Economists*.

Krüger, Eduardo and Givoni Baruch (2004). "Predicting thermal performance in occupied dwellings". In: *Energy and Buildings* 36, pp. 301–307.

Kuhn, Peter et al. (2011). "The effects of lottery prizes on winners and their neighbors: Evidence from the Dutch postcode lottery". In: *American Economic Review* 101.5, pp. 2226–2247.

Lafrance, Gaétan and Dominique Perron (1994). "Evolution of residential electricity demand by end-use in Quebec 1979-1989: a conditional demand analysis". In: *Energy Studies Review* 6.2, pp. 164–173.

Larsen, Bodil Merethe and Runa Nesbakken (2004). "Household electricity end-use consumption: results from econometric and engineering models". In: *Energy Economics* 26.2, pp. 179–200.

Lavy, Victor and Analia Schlosser (2011). "Mechanisms and impacts of gender peer effects at school". In: *American Economic Journal: Applied Economics* 3.2, pp. 1–33.

Lawrence, AG and M Parti (1984). *Survey of conditional energy demand models for estimating residential unit energy consumption coefficients*. Angel Economic Reports, Cupertino, CA (USA); Applied Econometrics, Inc.

Lemke, Bruno and Tord Kjellstrom (2012). "Calculating workplace WBGT from meteorological data: a tool for climate change assessment". In: *Industrial Health* 50.4, pp. 267–278.

Liang, Jing et al. (2018). "Do energy retrofits work? Evidence from commercial and residential buildings in Phoenix". In: *Journal of Environmental Economics and Management* 92, pp. 726–743.

Lundborg, Petter (2006). "Having the wrong friends? Peer effects in adolescent substance use". In: *Journal of Health Economics* 25.2, pp. 214–233.

Manning, M. M. et al. (2007). "The effects of thermostat setting on seasonal energy consumption at the CCHT twin house facility". In: *ASHRAE Transactions* 113, pp. 1–12.

Manski, Charles F (1993). "Identification of endogenous social effects: The reflection problem". In: *Review of Economic Studies* 60.3, pp. 531–542.

Marshall, A et al. (2015). *EnWin Utilities Ltd. Residential Ductless Heat Pump Pilot Evaluation*. Report for the Independent Electricity System Operator.

Marshall, A et al. (2017). *Hydro One Networks, Inc. Air-Source Heat Pump Advantage Pilot Evaluation*. Report for the Independent Electricity System Operator.

Mundlak, Y. (1978). "On the pooling of time series and cross section data". In: *Econometrica* 46.1, pp. 69–85.

Munshi, Kaivan and Mark Rosenzweig (2009). *Why is mobility in India so low? Social insurance, inequality, and growth*. Tech. rep. National Bureau of Economic Research.

Narayanan, Sridhar and Harikesh S Nair (2013). "Estimating causal installed-base effects: A bias-correction approach". In: *Journal of Marketing Research* 50.1, pp. 70–94.

Natural Resources Canada (2012). *Natural Resources Canada Briefing Note on the ecoEnergy Retrofit Homes Program*. Tech. rep.

— (2019). *Comprehensive Energy Use Database, Natural Resources Canada*.

Nelson, D.J, A.J Berrisford, and J Xu (2014). "MELs: What have we found through end-use metering?" In: *Proceedings of ACEEE Summer Study on Energy Efficiency in Buildings (Pacific Grove, USA)* 9, pp. 274–285.

Newsham, Guy R and Cara L Donnelly (2013). "A model of residential energy end-use in Canada: Using conditional demand analysis to suggest policy options for community energy planners". In: *Energy policy* 59, pp. 133–142.

Noll, Daniel, Colleen Dawes, and Varun Rai (2014). "Solar community organizations and active peer effects in the adoption of residential PV". In: *Energy Policy* 67, pp. 330–343.

NRCan (2019). *The EnerGuide Label Are Established*. Tech. rep. URL: https://www.nrcan.gc.ca/energy-efficiency/energuide-canada/energuide-label/13609.

Papineau, Maya (2019). "Big Data Meets Local Climate Policy: Energy Star Time-of-Day Savings in Washington, DC's Municipal Buildings". In: *IAEE Energy Forum Montreal Special Issue*, pp. 13–14.

Papineau, Maya and Nicholas Rivers (2022). "Experimental evidence on heat loss visualization and personalized information to motivate energy savings". In: *Journal of Environmental Economics and Management* 111, p. 102558.

Papineau, Maya, Nicholas Rivers, and Kareman Yassin (2023). "Estimates of long-run energy savings and realization rates from a large energy efficiency retrofit program". In: *Available at SSRN 4348993*.

Park, R Jisung, A Patrick Behrer, and Joshua Goodman (2021). “Learning is inhibited by heat exposure, both internationally and within the United States”. In: *Nature human behaviour* 5.1, pp. 19–27.

Parker, Danny S (2003). “Research highlights from a large scale residential monitoring study in a hot climate”. In: *Energy and Buildings* 35.9, pp. 863–876.

Parson, K (1993). *Human Thermal Environments: The Effects Of Hot, Moderate And Cold Environments On Human Health, Comfort and performance (1st ed.)* CRC Press.

Parti, Michael and Cynthia Parti (1980). “The total and appliance-specific conditional demand for electricity in the household sector”. In: *The Bell Journal of Economics* 11, pp. 309–321.

Rogelj, Joeri et al. (2018). “Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development”. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. IPCC. Chap. 2, pp. 93–174.

Sacerdote, Bruce (2001). “Peer effects with random assignment: Results for Dartmouth roommates”. In: *Quarterly Journal of Economics* 116.2, pp. 681–704.

Saha, Suranjana et al. (2010). “The NCEP climate forecast system reanalysis”. In: *Bulletin of the American Meteorological Society* 91.8, pp. 1015–1058.

Sebold, FD and KM Parris (1989). *Residential end-use energy consumption: a survey of conditional demand estimates*. Electric Power Research Institute, Palo Alto, CA.

Somanathan, Eswaran et al. (2021). “The impact of temperature on productivity and labor supply: Evidence from Indian manufacturing”. In: *Journal of Political Economy* 129.6, pp. 1797–1827.

Steinberg, Daniel et al. (2017). *Electrification & Decarbonization: Exploring U.S. Energy Use and Greenhouse Gas Emissions in Scenarios with Widespread Electrification and Power Sector Decarbonization*. National Renewable Energy Laboratory, U.S. Department of Energy.

Sun, Liyang and Sarah Abraham (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. In: *Journal of Econometrics* 225.2, pp. 175–199.

Swan, Lukas G. and V. Ismet Ugursal (2009). “Modeling of end-use energy consumption in the residential sector: A review of modeling techniques”. In: *Renewable and Sustainable Energy Reviews* 13, pp. 1819–1835.

Texas A&M Argulife research Texas A&M University, Cornell University International Water Management Institute (IWMI) National Oceanic and Virginia Tech Atmospheric Administration (NOAA) ICARDA. *The National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR)*. URL: https://globalweather.tamu.edu. last accessed 2019-01-12.

The city of Medicine Hat: City Clerk Department (2012). *Medicine Hat CENSUS REPORT 2012*. Tech. rep. URL: https://www.medicinehat.ca/en/government-and-city-hall/resources/Documents/City-Clerk/Census-Report-2012.pdf.

Tiebout, Charles M (1956). "A pure theory of local expenditures". In: *Journal of Political Economy* 64.5, pp. 416–424.

Tiedemann, K H (2007). *Using conditional demand analysis to estimate residential energy use and energy savings*. European Council for an Energy Efficient Economy Summer Study.

Towe, Charles and Chad Lawley (2013). "The contagion effect of neighboring foreclosures". In: *American Economic Journal: Economic Policy* 5.2, pp. 313–335.

Trogdon, Justin G, James Nonnemaker, and Joanne Pais (2008). "Peer effects in adolescent overweight". In: *Journal of Health Economics* 27.5, pp. 1388–1399.

UN Environment (2017). *Towards a zero-emission, efficient, and resilient buildings and construction sector*. UN Environment and International Energy Agency.

UNEP (2021). *Emissions Gap Report 2021*. United Nations Environment Program.

Wolske, Kimberly S, Kenneth T Gillingham, and P Wesley Schultz (2020). "Peer influence on household energy behaviours". In: *Nature Energy* 5.3, pp. 202–212.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data, Second Edition*. Cambridge, MA: The MIT Press.

Zivin, Joshua Graff et al. (2020). "Temperature and high-stakes cognitive performance: Evidence from the national college entrance examination in China". In: *Journal of Environmental Economics and Management* 104, p. 102365.