University of California
Santa Barbara

# Advances in the Climate-Economy Literature: The Impact of Weather on Homicide, Deforestation, and Mortality

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Drew Behnke

Committee in charge:

Professor Olivier Deschenes, Chair
Professor Paulina Oliva, University of California Irvine
Professor Gary Libecap

June 2017

The Dissertation of Drew Behnke is approved.

Professor Paulina Oliva, University of California Irvine

Professor Gary Libecap

Professor Olivier Deschenes, Committee Chair

May 2017

Advances in the Climate-Economy Literature: The Impact of Weather on Homicide, Deforestation, and Mortality

# Acknowledgements

First and foremost, I'd like to thank my committee chair, Olivier Deschenes. It is not hyperbole to say I could not have done this without you. To the rest of my committee, Paulina Oliva and Gary Libecap, thank you for the many helpful comments and advice throughout this process. David Roland-Holst, thank you for being the best unofficial committee member I could ask for. Your guidance inspired me to pursue my degree, and I value all you have taught me over the years.

To my friends and family, thank you for supporting and encouraging me throughout the years. Through the hardest times you were there for me and encouraged me to continue. Thank you for believing in me and offering unwavering support.

# Curriculum Vitae

Drew Behnke

## Education

| | |
|---|---|
| 2017 | Ph.D. in Economics, University of California, Santa Barbara. |
| 2012 | M.A. in Economics, University of California, Santa Barbara. |
| 2008 | B.A. in Economics, University of California, Berkeley. |

## Professional Employment

| | |
|---|---|
| 2014 - Present | Principal, Berkeley Economic Advising and Research. |
| 2011 - 2014 | Teaching Assistant, University of California, Santa Barbara. |

**Abstract**

Advances in the Climate-Economy Literature: The Impact of Weather on Homicide, Deforestation, and Mortality

by

Drew Behnke

My dissertation considers the role of weather on a variety of economic outcomes, including homicide, deforestation, and mortality. In chapter 1, in order to better understand why weather drives conflict, I focus my analysis on one of the most damaging forms of conflict: homicide. My results suggest that failing to recognize the potential for systematic differences across the homicide rate by race in the United States presents an inaccurate reflection of the true relationship between temperature and homicide. Previous research has found that colder temperatures decrease homicide while warmer temperature increase it. I find that black victims drive much of the negative effect for cold temperatures, while white victims drive the positive effect. While this finding alone is a new addition to the literature, these results are consistent with different mechanisms driving the relationship between temperature and homicide. For white victims, I cannot rule out the hypothesis that heat acts as a physiological stressor leading to violence and aggression. However, for black victims, and especially young black males, this hypothesis has limited support. Instead it appears that for black victims, an interactional mechanism plays a role, as colder temperatures reduce the homicide rate drastically.

In chapter 2, in collaboration with co-authors, we study the relationship between agricultural productivity and deforestation in the Brazilian Amazon. Using annual variation in growing-season temperature, we demonstrate that increased soy yields lead to greater land in agriculture and ultimately more deforestation. We find a delayed effect between

the increase in expansion of planted area and deforestation consistent with patterns of indirect land use change documented in the literature. Our findings suggest there may be negative environmental spillovers from policies that increase productivity for highly developed agricultural industry in the tropics.

Finally, in chapter 3, in joint work with a co-author, we study the relationship between temperature extremes and mortality. Although there is consistent evidence that temperature extremes lead to significant reductions in health, there has been limited analysis focused in the developing world. The few existing studies to date document that rural populations drive much of the mortality response to temperature, but offer little insight into what role adaptation plays in urban areas. This chapter attempts to overcome this gap in the literature by considering the relationship between temperature, mortality, and adaptation in Thailand. However, our results raise serious concerns over data quality issues. We find that extreme high temperature leads to decreased mortality rates, which is the opposite effect that has been documented in the literature. We also consider residential energy use as a possible adaptation strategy and model the relationship between temperature and energy use. Here we find that high temperature days relative to moderate days cause residential energy use to increase, which is consistent with the literature.

# Contents

# Chapter 1

# Temperature and Homicide: The Role of Race and Implications for Different Mechanisms

## 1.1 Introduction

The idea that heat may be directly related to conflict has long been a topic of study and thought dating back thousands of years. As early as the enlightenment, scholars such as Montesquieu claimed that weather helped shape the structure and prosperity of different societies, while artists like Shakespeare included a variety of imagery across his works relating heat and aggression (Montesquieu, 1748). In contemporary times, the connection between conflict and temperature appears to be a foregone conclusion as murders typically spike over summer months and this phenomena has been documented extensively across the media (Lehren and Baker, 2009; Economist, 2015a,b; Kirkos, 2016). Nowhere is this perhaps better illustrated than in America's largest cities, such as Chicago, which in 2016 witnessed an especially violent summer with some 90-people murdered in August alone, one of the highest totals in the city's history (Gorner, 2016).

Of course, correlation does not imply causation, and a recent wave of academic literature has begun to explore precisely how weather impacts aggression and human conflict. These papers build on the climate-economy literature and exploit the plausibly random nature of weather to estimate the causal relationship between weather and a variety of conflicts, both across individuals or groups. Remarkably, across a diverse range of geographies, time periods, spatial scales, and types of conflict there is clear support for a causal association between weather and conflict (Burke et al., 2015). Despite the clear evidence, the underlying mechanism remains difficult to disentangle, due to the fundamental fact that weather impacts so many aspects of society and the economy (Dell et al., 2014).

To better understand why weather drives conflict, I focus my analysis on one of the most damaging forms of conflict: homicide. The most comprehensive study documenting the weather-homicide relationship comes from Ranson (2014), who provides an estimate of the impact of weather on a variety of violent crimes in the US including homicide. He finds that elevated temperatures increase homicide and although this work offers a welcome analysis at the high level, it offers little insight into the more complicated underlying mechanism. Furthermore, the analysis makes no distinction about differential effects based on different homicide types. This is a major shortcoming as homicide in the US is not a simple summary statistic, but instead a far more complicated measure that has drastic differences among race, sex, age, and communities. Specifically, much of the homicide epidemic in America is largely driven by black victims, as the black homicide rate has been nearly 5 to 7 times greater than the white over much of the last 30 years (Federal Bureau of Investigation, 2016). The overwhelming share of black homicide victims has prompted concerns that homicide in certain black communities may be fundamentally different compared to others (Leovy, 2015).

If the motivation for homicide is different across groups, this has direct implications for how weather affects homicide. The leading mechanism from the literature assumes that homicide is associated with a type of aggression that is impulsive where harm is the primary goal. Under this theory, heat acts as a physiological stressor leading to aggression and violence. However, others have argued that in certain black communities with disproportionately high homicide rates, murder is used as a means of obtaining another goal such as justice, revenge, or status. Under these instances of homicide, the role of weather as a physiological stressor would be less clear. By separating the homicide rate by race, I consider if a physiological mechanism applies universally across different homicide types, and provide insight into other mechanisms where weather might affect homicide.

To conduct my analysis, I combine the most comprehensive record of homicide data in the US with high-frequency weather data and use panel methodologies to identify the effect of temperature on a variety of homicide rates. My results suggest that failing to recognize for the potential for systematic differences across the homicide rate by race presents an inaccurate reflection of the true relationship between temperature and homicide. Strikingly, I find that elevated warm temperatures appear to have no effect for black victims or other races, which contrasts with much of the previous literature. Instead, it appears that much of the positive effect between elevated warm temperatures and homicide is driven by white victims alone. I also find differential effects for cold temperatures. Cold temperatures decrease the homicide rate for young, black male victims, but have no effect for young males of other races. If I only consider the total homicide rate, I find that colder temperatures decrease homicide while warmer temperatures increase it, which corroborates findings from the previous literature. However, this effect relies heavily on the victim in question, which demonstrates that failing to separate the homicide rate by race produces misleading results.

The findings of a differential effect across race are unique to the literature and provide insight into the relevant mechanism in which weather affects homicide. For white victims, I cannot rule out the hypothesis that heat acts as a physiological stressor leading to violence and aggression. However, for black victims, and especially young black males, this hypothesis has limited support. Instead, for young black males, it appears that an international mechanism plays a role, as colder temperatures reduce the homicide rate drastically. For young white victims, I find no evidence that cold temperatures suppress homicide, and find that the homicide rate is increasing in temperature. An interactional mechanism is unlikely in this setting as homicide increases with uncomfortably hot temperatures, where increased interactions are less likely. Overall these results are consistent with a physiological mechanism where heat increases aggression and violence.

The rest of the chapter is organized as follows. Section 1.2 highlights the previous literature considering the relationship between interpersonal conflict and weather and discusses several mechanisms in which weather would affect homicide. It also provides a discussion on the nature of homicide in America as well as the data used in the analysis. Section 1.3 presents a theoretical framework that presents several mechanisms in which weather would affect homicide and presents the econometric specification used in my analysis. Section 1.4 presents results followed by a brief discussion in Section 1.5. Section 1.6 concludes.

## 1.2 Background

### 1.2.1 Previous Literature

The relationship between weather and conflict has long been a topic of academic interest, but quantitative research has increased dramatically in recent years as improvements in data quality and methods for causal inference have produced a new wave of rigorous work. For example, in their review of the literature, Burke et al. (2015) identify over 50 quantitative studies that consider the relationship between weather and conflict using modern econometric techniques. Remarkably, across a diverse range of geographies, time periods, spatial scales, and types of conflict, the authors' find strong support for a causal association between weather and conflict.

I focus my analysis on homicide and there has been much work in this field across a variety of disciplines. The modern literature begins with psychologists in the early 1970s, seemingly inspired by a United States Riot Commission paper that found the majority of riots in 1967 began on days when temperature exceeded 80° F (United States Riot Commission, 1968). These early studies were exclusively conducted in a laboratory setting where elevated temperatures were used to evaluate changes in aggression responses (Baron, 1972; Baron and Lawton, 1972; Baron and Bell, 1976). Concerns over external validity however quickly took the discipline to field studies that considered the correlation between heat and various forms of aggression, such as protesting, horn-honking, or crime (Baron and Ransberger, 1978; Carlsmith and Anderson, 1979; Kenrick and MacFarlane, 1986; Anderson, 1987).

By the end of the 1980s, a convergence of findings suggested that elevated temperatures increased a variety of types of aggression (Anderson, 1989). Of all types of aggression, violent crime became a focus of future work given the wide availability of data, and soon researchers in criminology began studying the effect of weather on crime as well. Researchers in both fields pushed forth the next wave of work using more rigorous statistical methods utilizing higher-frequency, time-series data (Anderson et al., 1997; Cohn and Rotton, 1997; Rotton and Cohn, 2000, 2003; Bushman et al., 2005). Once again, these studies demonstrated a strong correlation between heat and violent crime, although causal inference was beyond the scope of the statistical analysis. Instead, the focus of these papers was measuring the returns to temperature and crime and explaining why temperature affects crime. Although many theories were initially discussed, these literatures put forth two competing theories that are still prevalent today and are discussed in the next section (Anderson et al., 1997; Rotton and Cohn, 2003).

The economics literature first approached the relationship between violent crime and weather through indirect means, as Jacob et al. (2007) exploit the correlation between weather and crime to study the short-run dynamics of crime. The authors' find evidence of temporal displacement, as violent crime rates fall following weeks of increased crime from elevated temperatures. However, the relationship between weather and crime is not a central focus of their analysis and instead the most complete analysis comes from Ranson (2014), who is the first to utilize the modern weather-economy methodology with panel data and high frequency weather variables for causal inference. Ranson (2014) takes a 30-year panel of monthly crime and weather data for US counties to identify the effect of weather on a variety of crime rates. Remarkably, across a variety of offensives, he finds that elevated temperatures increase crimes and for most categories of violent crimes, such as homicide, the rate is linear.

### 1.2.2 Understanding the Mechanism

Previous work across the literature has proposed several mechanisms why temperature would increase homicide. The first, and arguably most supported, is that elevated temperature acts as a physiological stressor causing heightened aggression and loss of control leading to violence. Commonly known as the "Heat Hypothesis" in the psychology literature, this mechanism simply suggests that uncomfortably hot temperatures increase aggressive motives and aggressive behavior (Anderson et al., 1997; Anderson, 2001). Psychologists typically define aggression as any behavior directed toward another individual with the immediate intent to cause harm and there is typically a distinction between types of aggression. *Hostile aggression* (also called affective, angry, impulsive, and retaliatory) is often considered impulsive, angry behavior, where the primary goal is to hurt someone. *Instrumental aggression*, on the other hand, is a premeditated, calculated behavior, where harm is used as a means of obtaining another goal (e.g. money, power, dominance, or social status). The heat hypothesis suggests that elevated temperatures would affect hostile aggression as it is primarily annoyance motivated. Instrumental aggression is primarily incentive motivated and thus temperature should not have an effect. Homicide can fall under both categories, but is more commonly associated with hostile aggression.

A second theory from the criminology literature uses an interactional approach and considers how weather changes routine activities. This theory builds on the more general "Routine Activity (RA) theory," which suggests predatory crimes occur when circumstances increase potential interactions between victims and offenders (Cohen and Felson, 1979). RA theory was extended to research on weather and crime by noting that warm weather increases interactions between individuals (Cohn, 1990; Rotton and Cohn, 2003). Proponents of RA theory note that it does a better job explaining why there are low levels of violence recorded on cold days. Additionally, they argue that just as heat acts as a physiological stressor, the same argument could be made for cold, yet fewer violent crimes occur on cold compared to temperate or warm days (Rotton and Cohn, 2003; Ranson, 2014). Furthermore, RA theory speaks to the positive association between temperature and other types of crime such as rape or robbery, where changing patterns (such as leaving a house) are a more rational mechanism than physiological reasons. Rotton and Cohn (2003) also argue that the relationship between the victim and offender are relevant. For individuals with no prior relationship, favorable weather shocks can make the chance of interacting more likely, connecting those who would not normally encounter each other. Based on this, they suggest that temperature plays a smaller role in homicide than assaults, as the majority of homicides are committed by individuals known to the victim while assaults are committed by strangers.[1]

RA theory also has precedence in the economics literature, although weather is not directly considered. Instead it falls under the larger framework of social interactions and crime, which has primarily focused on how individuals influence others. Glaeser et al. (1996) build a framework where agents' decisions about crime are a function of their own attributes as well as their neighbors' decisions about crime. From this model, they demonstrate that social interactions create enough covariance across individuals to explain the high cross-city variance of crime rates. Based on this framework, weather that would increase social interactions could cause increases in crime, although this model would predict it is from network effects rather than increased opportunities for interactions as RA theory suggests. Furthermore, Glaeser et al. (1996) find that there are very low levels of social interactions for homicide, and therefore even if weather increased network effects, their model suggests the effect on homicide would be negligible.

Elsewhere in the economics literature, weather has been considered as an input into Becker's seminal model of crime where the decision to commit a crime is based on the considerations of the costs and benefits (Becker, 1968). Jacob et al. (2007) build on Becker's model by creating a simple two-period model and propose that adverse weather conditions might affect the cost of executing a crime, either by increasing the probability of successfully completing the crime or reducing the chance of arrest. However, their model is concerned with how an exogenous shock to the cost of crime in the current period affects crime in the following period, and thus weather is only used as an instrument. Although weather is discussed as a possible input into the cost of crime, the mechanism why weather would affect crime receives little discussion beyond their initial hypothesis

[1] Existing empirical evidence does not seem to support this claim, as Ranson (2014) finds a similar effect for both homicide and assault.

that it could affect costs.

The most detailed framework in the economics literature comes from Burke et al. (2015) who present a theoretical model that considers multiple mechanisms at once. Their model builds on work from Chassang and Padro-i Miquel (2009) who use a framework to focus on intergroup conflict by illustrating how economic shocks drive organized civil conflict. Burke et al. (2015) enrich the model by adding additional mechanisms and by applying the findings to interpersonal conflict as well. The model presents a condition for no conflict in which the marginal value of peace in the current period is weighed against the discounted marginal expected utility from attacking. The primary intention of the model considers how two agents engage in conflict over a resource, and thus the nature of the model is one of competition between identical agents, where the opportunity of attacking first is weighed against not attacking. Although homicide might involve some competition of resources (i.e. territory dispute among gangs), the model can also be adjusted for situations where there is no competition of resources and harm is the primary goal (i.e. hostile aggression). The flexibility of the model and its inclusion of several mechanisms makes it a useful framework for how weather effects homicide and is presented formally in Section 1.3.

### 1.2.3 The Role of Race and Homicide in America

Race is not explicitly considered in any of the mechanisms, yet it likely plays a significant role. Homicide in America is not a simple summary statistic, but is instead a complex measure that has significant variance across communities, race, and motivations. The US has one of the highest homicide rates in the world among high-income countries, but often less discussed is that much of this is driven by black victims. This has been true throughout much of the past century. Early measurements are problematic due to data concerns, but for cities that reported data in the early 20th century historians have found that the nonwhite homicide rate consistently exceeded the white rate. By the 1960s and 1970s, and the introduction of reliable data, the black homicide rate was roughly ten times higher than the white rate, and has remained between 5 to 7 times greater over the past thirty years (Leovy, 2015).

Take the overall homicide rate in 2014 of 4.1 homicides per 100,000, which is high compared to most other developed countries. However, the white homicide rate of 2.5 is on par with other high-income, non-violent countries such as Taiwan (3.0 in 2011), Malaysia (2.3 in 2012), or Belgium (1.8 in 2014). Conversely, the black homicide rate of 15.4 is closer to the more violent countries of the Dominican Republic (17.4 in 2014), Mexico (15.7 in 2013), and the Democratic Republic of Congo (13.5 in 2012). This discrepancy is even more pronounced in the highly segregated, majority-black communities that are commonly plagued by violence. In some communities in Chicago the homicide rate was nearly 70 in 2016, which is on par with the most dangerous countries in the world such as Honduras (84.6 in 2014) and El Salvador (64.2 in 2014) (Federal Bureau of Investigation, 2016; US Census Bureau, 2016; United Nations Office on Drugs and Crime, 2016; The Chicago Tribune, 2017).

The fact that parts of America are as dangerous as conflict zones while others are as safe as prosperous, highly developed countries, underscores how much homicide varies as a single measure. There are numerous, complicated reasons why certain communities experience such shocking rates of violence, yet it remains constant that young black men are significantly more likely to be killed than young men (or women) of other races. These statistics suggest there is something fundamentally different about the nature of homicide in black communities compared to others. One of the most compelling theories explaining why the American homicide problem disproportionately impacts young black men comes from the Los Angeles Times Reporter Jill Leovy in her book *Ghettoside*. She suggests that when the criminal justice system fails to adequately respond to violent injury and death, homicide becomes endemic. Asserting black Americans suffer from too little application of criminal justice is a controversial statement for a system that is commonly perceived to be unjustly harsh to minorities. Yet in reality, a system that is simultaneously punishes for low-level offenses, while inadequately responding to violent crimes only exacerbates the problem.

Homicide clearance rates are a crude measure of police effectiveness, but nonetheless illustrate that many more murders go unsolved in black communities perpetuating a dark legacy from the Jim Crow era where a legal system was lenient towards killers of black people. Homicide clearance rates in cities with large black populations such as Chicago and Baltimore are below half the national average and are roughly the same as they were in Mississippi nearly 50 years ago - 30% then compared to 26% in Chicago and 36% in Baltimore in 2015 (Leovy, 2015; Economist, 2015a; Kotlowitz, 2016). Low clearance rates coupled with over-policing of minor offenses creates an atmosphere where the police are both feared and not trusted to achieve justice, further alienating the community, decreasing cooperation, and ultimately increasing violence.

If the hypothesis that homicide in black communities is often of a different nature than other types of homicide is correct, this would have implications for both the relevant mechanism and ultimately the relationship between weather and homicide. Specifically, if homicide in black communities is more readily associated with instrumental aggression, this effectively eliminates the heat hypothesis and suggests that elevated temperatures would not increase homicide. Thus, given the substantial differences in the homicide rate across race, the failure to account for race presents a major shortcoming in the literature. This chapter addresses these shortcomings by separating the homicide rate by race, and exploits differences to consider the relevancy of different mechanisms proposed across the literature.

### 1.2.4 Data

To conduct my analysis, I create a 33-year panel from 1981 - 2013 of monthly homicide rates and weather from 74 large, urban counties in the US. This panel is based on three sources: (1) Homicide data comes from the US Federal Bureau of Investigation's Supplementary Homicide Reports (SHR) as part of the Uniform Crime Reporting (UCR) Program, (2) population data from the US Census Bureau, (3) weather data from PRISM. Each of the following data sources is discussed in more detail below.

#### Homicide and Population Data

Homicide data comes from the Supplementary Homicide Reports (SHR) as part of the US Federal Bureau of Investigation's Uniform Crime Reporting (UCR) program. The FBI's UCR compiles monthly crime reports from over 18,000 individual local law enforcement agencies. Participation is voluntary and self-reported, but by 2015, law agencies involved in the UCR program represented 97.7% of the total population making it the most comprehensive record of criminal activity in the US (Federal Bureau of Investigation, 2016).

Although the UCR program includes homicide, it only reports total homicides per agency, and has no further information about the victim, offender, or circumstance. Instead this information is contained in the SHR program, which includes additional details such as victim and offender demographic characteristics, weapon used, the circumstances surrounding the incident, and the relationship between the victim and offender if known. These additional details are of particular interest for this chapter, as they allow me to create a more nuanced measure of homicide that allows for further insight into proposed mechanisms. Much like the UCR program, the SHR are also voluntary and historically 85% to 90% of all homicides reported in the URC program submit a corresponding SHR form (Regoeczi and Banks, 2014).

Given the voluntary nature of the SHR program it is difficult to discern if missing values represent no homicides for that specific agency-month-year, or if data was just not submitted. This creates a challenge in constructing a monthly county-level homicide panel that is consistent across years. One option is to manually add back months that were not reported under the assumption that if an agency reports homicide in one month, missing months must represent a month with no homicide. For smaller counties, this approach is likely valid, although in larger counties there are clear instances where data is simply not reported for the month. A selection rule based on the average number of reported months per county-year would ensure internal consistency, but without a specific way to test if missing months represent no homicides or unreported data, this approach has the possibility to create spurious observations. Therefore, a far better approach is to only include counties in which there are a complete 12 months of reported data for that year. Of the 3,144 counties in the US, only 174 meet these criteria. In effect, this approach focuses on the largest, most populous counties in the US with the highest rates of homicide. This is particularly relevant for this chapter, because it is precisely these counties that are home to the largest cities in America, where differences in homicide rates by race will be the most pronounced. These counties are presented visually in Figure 1.1.[2]

To build the panel of homicide rates, I sum the total number of homicides reported by each agency in each county to generate a county total for each month and year. With monthly county totals, I then create different homicide rates using population data. Population data comes from the US Census Bureau, which reports yearly estimates of each county for different demographics such as sex, race, and age (US Census Bureau,

[2]The selection of 174 counties raises the possibility of sample bias. However, this would only affect the external validity of the results, which is less of a concern than spurious observations affecting internal validity.

2012, 2016). The homicide rate is then calculated under the standard practice by dividing the total number of homicides in each county-month-year by the total population in each county-year multiplied by $100,000$. For different homicide rates, these totals can be further refined. For example, to calculate the black homicide rate, I first sum the total number of homicides with a black victim in each county-month-year, and calculate the homicide rate using the total black population for that county-year. The detail in the SHR data allows even further refinement by including sex, age, and other circumstances.

Some final points are worth noting. Race in the SHR database is separated into five categories: white, black, Asian, Native American, and unknown. The SHR database considers Hispanic as an ethnicity, and thus in addition to race, all victim and offenders are classified as either of Hispanic origin or not of Hispanic origin. Although I could further subdivide race by Hispanic ethnicity, the ethnicity of most victims (56%) is unknown. Furthermore, population data including ethnicity is not available before 1990, further reducing observations and panel length. Therefore, I do not explicitly consider the victim's (or offender's) ethnicity to preserve observations. In the interest of completeness, other races are included in my analysis, although I aggregate Asian and Native Americans into one category. Taken together these races collectively account for 2.33% of all victims in the SHR dataset, thus the clear majority of homicides have either a white or black victim and these races are the primary focus of my analysis.

Summary statistics for demographics and homicide rates between the entire SHR dataset and sample are presented below in Table 1.1. This summary reveals that sample counties are larger, and more diverse compared to counties in the entire SHR dataset. Homicide rates in many cases are in fact larger across the entire SHR dataset, but this comes from the fact that small counties skew the homicide rate higher. A better illustration is that sample counties on average record nearly 7 homicides a month, almost 5 times as many as the average county across the entire SHR dataset. Looking at differences within the sample, I find that the black homicide rate is 4.82 times greater than the white homicide rate, which is largely consistent with national data. Separating the homicide rate for young males reveals an even larger discrepancy as the young black male homicide rate is nearly 7 times greater than the young white male homicide rate, the largest difference across the sample.

### Weather Data

Weather data represents the third major component of my dataset, and observations come from the PRISM Climate Group of Oregon State University. I utilize the AN81d dataset which provides daily weather observations from the continuous US beginning in 1981 (PRISM Climate Group, 2016). Variables include daily maximum and minimum temperature as well as total precipitation, although I follow the convention in the literature and focus on the daily mean temperature. PRISM data offers an improvement to other publicly available data, such as the US National Climatic Data Center's GHCN-Daily database, as it draws on more stations (over 10,000 versus 1,200) and emphasizes important geographical features such as elevation and coastal proximity. As a result, PRISM data offers a more accurate reflection of weather realizations, particularly in the West (Daly et al., 2008).

PRISM weather data consists of daily observations across a 4km gridded resolution. Mapping this grid to counties requires some form of overlapping and county shapefiles were overlapped on the weather grid and the share of the relevant county in each grid cell was then calculated using population weights. These population weights were then used to match the gridded data to the county level by averaging the observations across each county. Summary statistics of weather data between the entire SHR dataset and sample are presented below in Table 1.2. Overall, there is less variation in weather data between the sample and entire dataset, and the sample is slightly warmer than the national data.

Figure 1.1 reveals a slight bias for countries in the warmer temperature distribution of the US as many northern states are excluded. Although this could potentially introduce bias, I consider the role of different temperature distributions explicitly in my analysis and find no difference between warmer and colder locations.

## 1.3 Model and Methodology

### 1.3.1 Framework

I begin with a theoretical model developed by Chassang and Padro-i Miquel (2009) and later modified by Burke et al. (2015) that describes several mechanisms how weather drives interpersonal conflict. I further modify the model by including an additional index $i$ to represent different races. The original intention of this model considers how two agents engage in conflict over land, and thus the nature of the model is one of competition between identical agents, where the opportunity of attacking first is weighed against not attacking. Rather than present the full solution to the model, the framework is presented in brief to illustrate mechanisms in which weather drives homicide and what role race might play.

The initial model considers two agents of race $i$ competing in infinite periods $t$ who share a territory of size 2. Agent 1 is marginally wealthier and controls $1 + \lambda$ units of land, while agent 2 controls $1 - \lambda$. Land is used to produce crops in the original Chassang and Padro-i Miquel (2009) framework, but can be generalized to any resource here. Each agent has an economic output of $\theta_{it}$ and total output is based on the amount of land. Thus, total production in the economy is $2\theta_{it}$.

In addition to production, agents can seize land and economic output from the other. If one agent attacks first, they gain a first strike advantage, and gain the other agent's assets with probability $P_{it} > 0.5$. If both agents attack simultaneously, each wins with probability 0.5. In the case of conflict, both groups divert a fraction of output with cost $c > 0$. The winning agent receives all land and thus captures the total production of the economy. Therefore, the winner's payoff is $2\theta_{it}(1-c)$, while the loser receives 0.

There is only one round of fighting, and the results of the conflict last all remaining periods. Thus, if an agent decides not to attack they obtain $\theta_{it} + \delta V^P$, where $V^P$ is the present discounted value of future equilibrium with no conflict (value of peace) and $\delta$ is the per period discount rate. Conversely, if an agent decides to attack, the expected payoffs are $P_{it}(2\theta_{it}(1-c) + \delta V^V)$, where $V^V$ is the present discounted value of victory. Burke et al. (2015) extend the framework by also including a nonrival psychological consumption value of violence, $\gamma_{it}$. If the agent dislikes violence than $\gamma_{it} < 0$, and if the agent enjoys violence $\gamma_{it} > 0$. Taken together this implies the following condition for no conflict:

$$\underbrace{\theta_{it} + \delta V^P}_{\text{value of peace}} > \underbrace{P_{it}\left(2\theta_{it}(1-c) + \delta V^V\right) + \gamma_{it}}_{\text{value of attacking}}. \tag{1.1}$$

Put another way, an agent will not attack if the value of consuming all economic output with initial assets plus the discounted expected utility under peace exceeds the expected utility of consuming total output minus expenditures of the conflict, plus discounted expected value of all future output and the consumption value of violence. Although Chassang and Padro-i Miquel (2009) apply this result to intergroup conflict, Burke et al. (2015) argue that these results can be extended to intergroup conflict as well, where the value of victory, $V^V$ contains information on the expectation of completing the attack without apprehension. Furthermore, in purely violent crimes where there is no economic motivation, such as some murders, assaults and rapes, then $\theta_{it} = 0$, and the psychological consumption value of violence, $\gamma_{it}$, becomes the primary motivation for the attacker

(Burke et al., 2015).

Equation 1.1 can be rearranged so that the condition for no conflict is alternatively expressed as

$$\theta_{it}(1 - 2P_{it}(1 - c)) - \gamma_{it} > \delta\big[P_{it}V^V - V^P\big]. \tag{1.2}$$

This alternative expression is the fundamental equation of the Chassang and Padro-i Miquel (2009) framework, and compares the marginal value of peace in the current period weighed against the discounted marginal expected utility for attacking on the right-hand side. More importantly for this chapter, this equation presents several mechanisms vulnerable to weather.

Starting first with economic output, $\theta_{it}$, adverse weather shocks are believed to affect economic productivity, and this mechanism has received much of the research focus among economists especially for intergroup conflict. The majority of work considers weather as an income shock that can drive civil conflict in developing countries (see Burke et al. 2015 for further discussion of relevant studies). However, as Burke et al. (2015) highlight, even if it is true that income affects conflict risk, and weather affects income, this is still insufficient to prove weather affects conflict through income alone as weather affects numerous other confounding factors as well. For homicide in a developed country this mechanism has less merit. Homicides can be motivated by capturing a rival's economic output (e.g. territory dispute between gangs), yet it is unclear how local weather conditions would affect this. Furthermore, there is limited support that weather affects income in developed countries (Dell et al., 2012).[3] Finally, in many instances of

[3] In general, climatic conditions are believed to play a limited role in the economic output of developed countries, whose economies are not agrarian and have a variety of adaptation mechanisms (see Dell et al. 2012 for example). However, recent work provides conflicting evidence that elevated temperatures do in fact impact annual economic output in the United States (Deryugina and Hsiang, 2014). This calls into question the previously held belief that developed economies are immune from weather, although there is still no support of impacts on a higher frequency timescale.

homicide, harm is the primary goal and there is no intention to transfer wealth, suggesting $\theta_{it}$ would be completely removed from the model. Taken together, this implies that $\theta_{it}$ is not a key mechanism for the weather-homicide relationship. In regards to race, if specific races are more likely to work in an industry that is reliant on weather, such as agriculture, than there could be differential effects. However, given the limited amount of weather dependent industries across the US, these differences are likely trivial for the purposes of this chapter.

Moving next to the probability of successfully completing an attack, a weather-induced increase in $P_{it}$ would raise the likelihood of conflict as the expected value of attacking is higher. Once again, this mechanism has some precedence across the conflict literature especially with respect to intergroup conflict (see Burke et al. 2015 for further discussion). One way weather could affect the probability of completing a successful attack is by altering the physical environment in a way that makes completing the attack easier or more likely to succeed. With respect to homicide, this mechanism falls under RA theory, as weather can increase the chance of interactions increasing the probability of a successful attack. Race plays a role here as weather shocks presumably affect the probabilistic interactions of racial groups at different rates. For example, the black unemployment rate is over twice the white unemployment rate over the sample period, and this could result in a higher potential for interactions and conflict through favorable weather shocks (Bureau of Labor Statistics, 2017). Furthermore, the baseline probability of successfully completing an attack likely varies among racial groups (i.e. $P_{it}$ takes on different values depending on $i$) and thus weather shocks would have differential impacts on the likelihood for conflict across groups.

Finally, weather can affect the psychological consumption value of conflict, $\gamma_{it}$. This mechanism has received the most support across the interpersonal conflict literature, where heat is believed to act as a physiological stressor that leads to aggression. Translated to Equation 1.2, heat would increase $\gamma_{it}$ by either increasing the utility of acting violently or decreasing the psychological cost of a violent act, decreasing the marginal value of peace and increasing the likelihood for conflict. Of all mechanisms, Burke et al. (2015) believe that evidence of a psychological motivation for intergroup conflict is among the most convincing because economic conditions are unlikely to be affected by climatic conditions on the high frequency timescale of some studies (either weekly for Jacob et al. 2007 or monthly for Ranson 2014). Furthermore, there is considerable evidence of a connection between heat and aggression in a variety of settings where other plausible mechanisms would play no role, such as violent retaliation in professional baseball, road rage, or use of violence in police training (Larrick et al., 2011; Kenrick and MacFarlane, 1986; Vrij et al., 1994). Once again, race plays a role here as well, as different races might have different baseline physiological consumption values for conflict. In communities with especially high homicide rates, the physiological cost of a violent act could be lower, making conflict more likely. Furthermore, in these communities homicide may be more readily associated with instrumental aggression, where heat would be expected to play no role.

Although the above pathways provide varying degrees of support for different mechanisms, none completely rule out alternative hypotheses. These different and competing mechanisms underpin the central problem of how difficult it is to disentangle how weather affects conflict when weather impacts so many other aspects of an economy. Based on this challenge, Burke et al. (2015) propose that the best path forward is to directly test individual hypotheses through the elimination of other pathways and determine if the connection between weather and conflict persists. Although I am unable to directly test for specific mechanisms, by separating the homicide rate across race I am able to see if there are differential effects for weather and if the findings are consistent with proposed mechanisms.

### 1.3.2 Primary Econometric Specification

This chapter's empirical methodology relies on the plausibly exogenous variation in weather outcomes over time to identify the effects of weather on homicide. This approach was first popularized to the economics literature by the seminal work of Deschenes and Greenstone (2007) who recognized that annual fluctuations in weather can be used to measure agriculture profits. The primary advantage of this methodology is identification, as panel methodologies and random weather outcomes can causatively identify the effect of weather on a variety of economic outcomes (Dell et al., 2014).

In the interpersonal conflict literature, Ranson (2014) makes use of this methodology to study the effects of weather on a variety of US crime rates, including homicide. I follow his approach by using a semi-parametric specification for weather and precipitation that does not impose any structural assumptions on the relationship between weather and homicide. The primary econometric specification is listed below in equation (1.3):

$$Y_{iym} = \sum_{j=i}^{10} \beta_j TMEAN_{iym} + \sum_{k=1}^{5} \theta_k PREC_{iym} + \alpha_{sm} + \gamma_{iy} + \varepsilon_{iym}, \tag{1.3}$$

where $Y_{iym}$ represents the log homicide rate in county $i$ in in year $y$ and month $m$. The homicide rate often takes on a value of zero and to preserve the relationship using log, I add 1 to all values before the log transformation. The key variables of interest are those that capture the distribution of daily temperature. Denoted by $TMEAN_{iym}$, this variable represents the number of days in county $i$ in year $t$ and month $m$ that the average temperature falls into one of the 10 following bins: $< 10°F$, $10- < 20°F$, $20- < 30°F$, $30- < 40°F$, $40- < 50°F$, $50- < 60°F$, $60- < 70°F$, $70- < 80°F$, $80- < 90°F$, $\geq 90°F$. Temperature is the primary focus of this analysis, but precipitation is included as a control as it is often correlated with temperature (Auffhammer et al., 2013a). Precipitation is therefore modeled in a similar approach, where $PREC_{iym}$ represents the number of days that the total temperature falls into one of the 5 following bins: 0 mm, $> 0- < 10$ mm, $10- < 20$ mm, $20- < 30$ mm, $\geq 30$ mm.

Following Ranson (2014) I use a comprehensive set of fixed effects, where $\alpha_{sm}$ is a state-by-month fixed effect and $\gamma_{iy}$ is county-by-year fixed effect. The motivation for such an extensive set of fixed effects is twofold. First, there is substantial year to year variation in the number of reporting agencies per county across years. However, an identical set of agencies report for each month within a given year, and therefore county-by-year fixed effects controls for any unobserved differences in reporting across county-years. Second, there are seasonal and regional differences in the homicide rate across the US. In general, homicide spikes during the summer months and tends to be higher in southern states. Therefore state-by-month fixed effects control for any unobserved differences in the homicide rate across months and states.

Furthermore, these extensive fixed effects control for any observed variables that are likely to affect homicide rates. This is particularly important for homicide, as the exact causes and origins are not well defined. Traditional research has focused on socio-economic factors, and the most reliable predictor tends to be age, gender, and race, all of which are explicitly considered in my analysis (Field, 1992).[4] Economic outcomes likely play a role, but the exact correlates are not clear. Outcomes such as unemployment, inequality, and poverty have been shown to affect homicide, yet these factors are not predictors of crime as many locations with lower economic outcomes have lower crime rates (Krueger et al., 2004; Glaeser et al., 1996). These complexities have given rise to a whole group of literature across disciplines that tries to understand why homicide persists in some areas but not others (Sampson et al., 1997; Krueger et al., 2004; Glaeser et al., 1996).

[4]Besides age, gender, and race, weather is one of the most common factors consistently correlated with violent crime, which explains why the topic has been such an active research frontier across numerous disciplines.

Given the difficulty in identifying correlates of homicide, fixed effects offer an ideal methodological solution as they control for unobserved factors that affect homicide rates. Specifically, county-by-year control for unobserved county-specific determinants of homicide across years, which would include economic and demographic outcomes that vary across counties such as education levels, income, poverty, and inequality as well as racial compositions. Additionally, state-by-month fixed effects absorb any unobserved state-specific determinants of homicide that vary across months, which would account for both seasonal and regional factors. Identification comes from the remaining monthly variation in weather and homicide across county-years, controlling for average monthly state patterns.

Although the empirical strategy used in equation (1.3) is common across the weather-economy literature, the interpretation of the results is dependent on a somewhat subjective, reference point. Typically, a bin representing temperate weather (either $50- < 60°F$ or $60- < 70°F$) is dropped, and all coefficients are compared to this omitted bin. The interpretation of results is therefore explained by how an additional day of temperature per month compares to the omitted bin, which is specific to the reference point and might not always be intuitive. In response to this, I also make use of the "degree-day" approach, which still does justice to the non-linear relationship between temperature and homicide but presents results in a single-index. This approach allows a more unambiguous interpretation of results, while still collapsing a year's worth of temperature realizations into a single-measure. This alternative specification is presented below:

$$Y_{iym} = \beta_1 CDD70_{iym} + \beta_2 HDD50_{iym} + \sum_{k=1}^{5} \theta_k PREC_{iym} + \alpha_{sm} + \gamma_{iy} + \varepsilon_{iym}, \tag{1.4}$$

where $CDD70_{iym}$ is the number of cumulative degree-days in county $i$ in in year $y$ and month $m$ that exceed $70°F$. This degree-days approach is standard in the weather-economy literature and creates a measure of the number of days with degrees above a certain temperature threshold. For example, if a county-month-year had 2 days where the average temperature was above $70°F$, $75°F$ and $78°F$, the value of $CDD70_{iym}$ would be 13. Heating degree days work in a similar way except measures cumulative temperature below a threshold of $50°F$. This approach assumes that the effect of temperature on homicide is linear above $70°F$ and below $50°F$, with no effect for days in between. These thresholds are based on findings from the bin analysis as the primary effects of temperature are found outside these thresholds. Rather than estimate 10 coefficients, this approach collapses temperature to two, ensuring more statistical power and allowing easier interpretation.

Finally, consistent with Ranson (2014), I also include a one-month lag for each weather variable in equations (1.3) and (1.4) as a robustness check. The inclusion of lagged variables is in response to Jacob et al. (2007), who find evidence that changes in crime rates from weather shocks may have negative serial correlation. These models are identical to the above with the inclusion of the lagged variables and are presented below. Results are presented in the annex:

$$\begin{aligned} Y_{iym} = &\sum_{j=i}^{10} \beta_j TMEAN_{iym} + \sum_{k=1}^{5} \theta_k PREC_{iym} + \sum_{j=i}^{10} \beta_j TMEAN_{iy,m-1} \\ &+ \sum_{k=1}^{5} \theta_k PREC_{iy,m-1} + \alpha_{sm} + \gamma_{iy} + \varepsilon_{iym} \end{aligned} \tag{1.5}$$

$$
\begin{aligned}
Y_{iym} = {} & \beta_1 CDD70_{iym} + \beta_2 HDD50_{iym} + \beta_3 CDD70_{iy,m-1} + \beta_4 HDD50_{iy,m-1} \\
& + \sum_{k=1}^{5} \theta_k PREC_{iym} + \sum_{k=1}^{5} \theta_k PREC_{iy,m-1} \alpha_{sm} + \gamma_{iy} + \varepsilon_{iym}
\end{aligned} \tag{1.6}
$$

For the above models, there are two additional points worth noting. First, consistent with the weather-economy literature, I assume that the error terms, $\varepsilon_{iym}$, are correlated across counties over time and therefore cluster standard errors at the county level. Second, regressions are weighted by the relevant populations. For example, the white homicide rate regression is weighted by the total white population. The choice to use population weights is also standard in the weather-economy literature and reflects that estimates from counties with higher populations will be more precise and implies that estimates represent the effect on the average person, rather than average county.

### 1.3.3 Adaptation Specification

One shortcoming of the primary empirical strategy is that it fails to account for adaptation to climate across counties. Presumably increases in temperature could have differential impacts on homicide depending on mean temperature of the county. For example, the effect of a 90° F day in Boston might be different than in Miami, as people in Miami experience more 90° F days per year. This is especially relevant given the proposed mechanisms. If individuals in warmer counties experience more extremely hot days per year, then heat might not cause as much physical stress as it would to an individual in a colder climate that experiences an unusually warm day. Furthermore, different mean temperatures across counties could impact the level of interactions. In counties that experience an overall lower mean temperature, interactions in cold weather are still likely as people still routinely brave cold temperatures to be outside. In more temperate locations however, cold days likely change behavior as people postpone going outside until temperatures return to normal levels. Unusually warm days could trigger differential effects as well. In colder counties, a warm day might encourage even more interactions as individuals try to make the most of it, while those in already warm counties might not change their patterns. Conversely, too warm of a day might encourage those in colder climates to want to stay inside. People in warmer climates could be better adapted to warm temperature and thus be unlikely to change their behavior. These examples illustrate the potential role of cross-sectional adaptation. There is no reason to believe that race would play any differential role here as adaptation has more to do with location than demographic characteristics.

State fixed effects will control for any observed differences across populations in different climatic regions, but this does not offer insight into the question of whether adaptation to climate plays a role in the effect of weather on homicide. Therefore, in order to explore this issue further, I present an additional econometric specification that considers adaption in the cross-section. This approach is based on the methodology from Barreca et al. (2015), who consider if heat increases mortality more in areas that experience extreme heat less often. To conduct their analysis, the authors consider the heterogeneity of mortality responses to extreme temperature by considering how the effect varies across ten deciles of the long-term distribution of extreme temperatures across states. I present a modified version of this approach below:

$$
\begin{aligned}
Y_{iym} = &\sum_{d=i}^{10} \beta_d D80_{iym} \times 1(Decile_{iy}^{D80} = d) \\
&+ \sum_{k=1}^{5} \theta_k PREC_{iym} + \alpha_{sm} + \gamma_{iy} + \varepsilon_{iym},
\end{aligned} \tag{1.7}
$$

where the primary variable of interest is represented by $D80_{iym} \times 1(Decile_{iy}^{D80} = d)$ and the other variables continue from previous specifications. Here $D80_{iym}$ represents the total number of days in a county-year-month where the daily average temperature is greater or equal to 80° F. Counties are also placed into one of ten deciles based on the average annual number of D80 days across the sample represented by $1(Decile_{iy}^{D80} = d)$. Thus, taken together, this specification provides a measure of how extremely hot days affect different homicide rates across counties that experience different ranges of hot days per year. Note that I elect to measure an extreme hot day as one whose mean temperature is 80° F or greater and this choice differs from the Barreca et al. (2015) paper who instead use a greater than 90° F threshold. Although this threshold was initially considered, my chapter uses a significantly shorter sample, and the large number of observations whose days do not exceed 90° F prohibits separating counties into deciles.[5]

## 1.4 Results

This section presents the main results of the chapter and finds that weather has differential effects on homicide depending on demographic characteristics. My findings question the previously held belief that heat increases homicide, as I find this effect

[5]For example, 37.83% of counties in my sample have no days that exceed the mean 90° F threshold annually, and as a result the first 3 deciles share the same decile. However, the new threshold of 80° F or greater still does justice to extreme temperatures, as it captures the two most extreme temperature bins, and is well within the range of established effects for temperature on homicide.

depends very much on victim demographics. Although I find that elevated temperatures increase homicide across the general population and for white victims, I find a limited effect for black victims. These findings are consistent with different mechanisms by race. For white victims, I cannot rule out the hypothesis that heat acts as a physiological stressor leading to violence and aggression. However, for black victims, and especially young black males, this hypothesis has limited support. Instead it appears that for black victims, an interactional mechanism plays a role, as colder temperatures reduce the homicide rate drastically. These findings are discussed in greater detail below.

### 1.4.1 Total Homicide

I begin by presenting regression results from equation (1.3) and (1.4) for different homicide rates by victim race. These results are listed in Table 1.3 and Table 1.4 below. For the bin analysis in Table 1.3, the temperature coefficients are interpreted as the percentage change in homicide per month caused by one extra day in that temperature bin relative to a day in the omitted $60- < 70°F$ temperature bin. This bin represents a temperate day and is commonly used in the weather-economy literature, including the Ranson (2014) work. The interpretation of the precipitation coefficients is the same for both tables and represents the percentage change in homicide per month caused by one extra day in that precipitation bin relative to a day with no rainfall (0 mm precipitation bin). Finally, in Table 1.4, the temperature coefficients represent the percentage change in homicide for one additional degree day above $70°F$ (or below $50°F$).

My results indicate that there are differential effects between weather and homicide depending on the race of the victim. In Table 1.3, I present regression results for four different homicide categories: all victims, white victims only, black victims only, and other race victims. For all victims (column 1), I find a similar effect as Ranson (2014), with additional days in warmer bins causing an increase in homicide, and some evidence that colder days decrease homicide. Some of these effects are meaningful as well, considering the coefficients represent one additional day of weather per month. For example, a usually warm month with 10 additional warmer days between $80- < 90°F$ compared to $60- < 70°$ F, would cause the homicide rate to increase by nearly 2%. When I separate the effect across race however, I begin to find some differential effects. If I restrict my analysis to white victims only (column 2), I find that much of the positive effect seen in warmer bins for the general population appears to be driven by white victims. The effect of an additional day in either the $70- < 80°$ F or $80- < 90°$ F bin is both highly significant and larger than what is seen across the entire population. For colder temperature bins I find no effect, and instead the effect observed in the general population appears to be driven by black victims. In column 3, I find that an additional day in either the $50- < 60°$ F or $20- < 30°$ F bin decreases the homicide rate for black victims, and this effect also appears in the general population albeit at a smaller coefficient estimate and lower significance. An additional warm day only increases the homicide rate for the $80- < 90°$ F bin, while the $70- < 80°$ F bin has no effect. Finally, for victims of other races (column 4), I find little evidence that temperature impacts homicide rates, with a small positive effect found for the $80- < 90°$ F bin. In the degree-day approach, I find a persistent effect for degree days over 70° F for the total population, white, and black victim categories, but no effect for victims of other races. Degree days under 50° F only have an effect for black victims (Table 1.4).

These initial results weakly suggest black victims drive much of the negative effect for cold temperatures, while white victims drive the positive effect for warm temperatures. General population effects are consistent with the previous literature, but this chapter is the first to demonstrate the differential effects of race. Although race is not a perfect measure of different homicide circumstances, the results from Table 1.3 demonstrate that heat has differential impacts on homicide depending on the victim. For certain black communities, the homicide epidemic appears to be driven by motivations that would not be classically associated with heat and annoyance. These results are somewhat illustrative of this, yet reflect the fact that simply looking at differences between white and black victims in major counties is not a prefect proxy for different types of aggression.

### 1.4.2 Young Male Homicides

A better illustration of the differential impacts of heat comes from looking at homicide rates between young men aged 15-34. Young men, and particularly young black men, represent the majority of homicide victims in the US. Of the nearly 290,000 black victims in the SHR, 54% of them are men between 15 - 34.[6] Furthermore, the divide between the white and black homicide rates are even more pronounced in this group, as the homicide rate for young, black men is nearly seven times higher than white men of the same age across my sample (the largest difference across comparison groups). Given that the discrepancies between young male homicide rates are even more pronounced than across the entire population, this group offers a more useful comparison of the differential effects of temperature on homicide.

Indeed, across these groups the differential effect between temperature and homicide is even more pronounced. In Table 1.5, I present regression results from four different homicide categories for young males: all victims, white victims only, black victims only, and victims of other races. For all young male victims (column 1), the general population effects (column 1 Table 1.3) are seen once again, with larger coefficient estimates and a greater range of significant bins. However, only looking at all young male victims is misleading as there are differential effects depending on race. For young white victims (column 2), I find that elevated temperatures increase the homicide rate, but colder days

[6] Young, white males represent 38% of all white homicides.

have no effect. Conversely, for young black victims (column 3), the opposite relationship holds as colder days decrease the homicide rate, while warm days have no effect. These effects are larger than what is demonstrated across the general population as well. Furthermore, there appears to be increasing returns to these effects for young white males and decreasing returns for young black males. For example, the effect for young white males increases with temperature and 1 additional day in the $> 90°$ F relative to a day in the $60- < 70°F$ omitted bin, causes the young white male homicide rate to increase by approximately 0.9%. For young black males, mild colder days have a larger effect than extreme cold days. One extra day in a mild cold bin (between $30- < 60°F$) causes the monthly homicide rate to decrease by over 0.5%. This size of the effect decreases in more extreme bins (between $10- < 30°F$) before ultimately having no effect for temperatures below 10° F. Although I am primary interested in the differential effects between the white and black populations, other races are included for completeness and suggest that weather plays no meaningful role in determining homicide rates. Note that the non-effect could simply be a reflection of how rare homicides of other races are, as 92% of observations for this category have a homicide rate of zero.

The degree day analysis also finds differential effects. In Table 1.6, I present results for the four categories and find that elevated temperatures increase the homicide rate for the entire population of young males and white victims, but have no effect for young black males or victims of other races. Conversely, I find that colder days decrease the homicide rate for young black males, but have no effect for white victims, victims of other races, or the general population.

### 1.4.3 Homicides including Offender

A discussion on the role of aggression surrounding homicide is not complete without considering the offender who is responsible for committing the homicide. This was not an oversight in previous sections, but rather a reflection of the fact that over a third of homicides in the SHR database have unknown offenders. Furthermore, the clear majority of homicides occur within the same race. For example, in instances where the offender is known in the SHR database, over 80% of victims and offenders are of the same race. This is logical due to the fact most people are killed by someone they know, and the prevalence of racial segregation in America suggests they would be of the same race. Thus, although the above analysis does not explicitly consider the race of the offender, it implicitly assumes the offender is likely to be of the same race. This section takes the analysis one step further by explicitly considering the offender as well.

Introducing offenders to the analysis removes much of the previous variation as observations with an unknown offender no longer count towards the homicide rate. Furthermore, in instances where the offender and victim are of a different race, most observations are zero as these homicides are rare (see data section for descriptive statistics). Considering this, I opt to use less extensive fixed effects for this analysis to preserve some variation to identify an effect. Specifically, I no longer make use of state-by-month fixed effects and instead only use monthly fixed effects to control for seasonal trends across the entire country.

In Table 1.7, I present regression results from estimating equation (1.3) for four different victim-offender categories: white offender-white victim, black offender-black victim, white offender-black victim, and black offender-white victim. I elect not to include other race victims here because previous sections have failed to establish meaningful results for this group. In the cases where the offender is of the same race as the victim, the findings from the previous section are weakly supported. For example, for a white victim and white offender (column 1), I find that elevated temperatures increase the homicide rate while colder temperatures have no effect. However, only an additional day in the $80- < 90°F$ bin has an effect and at a smaller magnitude compared to all white victims (column 2 Table 1.3) or young white male victims (column 2 Table 1.5). Similarly, for black victim and black offender (column 2), I find that colder temperatures decrease homicide, and this effect more closely matches those seen in the previous section. Once again, I find no support that elevated temperatures increase the black homicide rate. These results are supported by the degree day findings in columns 1 and 2 in Table 1.8.

What is particularly striking about these results is that in the rare instances when the victim and offender are of a different race, I find an effect that is like those seen of white victims. Specifically, I find that elevated temperatures increase homicide, while colder temperatures have no effect. For example, for a white offender and black victim (column 3), an additional day in the $70- < 80°F$ or $80- < 90°F$ relative to a day in the omitted $60- < 70°F$ bin, increases the homicide rate by roughly 0.01%. For a black offender and white victim, I only find an effect in the most extreme bin, where one additional day above 90° increases the homicide rate for white victims by 0.03%. Although these effects are small and these types of homicide are rare, it adds further support to the claim that black-on-black homicide is unique compared to other types of homicide in the US.

### 1.4.4 Role of Adaptation

As a final robustness check I consider the role that regional adaptation to localized climate plays across counties. This analysis is in response to the concern that extreme temperatures may have differential effects to homicide in locations with different mean temperatures. Results from equation 1.7 are listed below in Table 1.9.

These tables demonstrate that adaptation appears to play no role in the impact of extreme temperatures on homicide rate, and furthermore there do not appear to be differential effects across race. In Table 1.9, each decile represents the impact of an additional day greater or equal to 80° F across the distribution of annual days greater or equal to 80° for each sample county. Thus, this analysis provides a simple test for differential impacts for relatively colder counties (lower deciles) or warmer counties (higher deciles). For all victims (column 1), I find that an additional warm day has no effect across many deciles. Furthermore, this effect holds for white victims (column 2) and black victims (column 3). For victims of other races (column 4), some deciles are significant, but the effect is relatively constant and is neither decreasing or increasing.

This finding contrasts with the work of Barreca et al. (2015), who using a similar methodology find decreasing returns to mortality for extreme temperature across increasingly warmer deciles. Their findings are consistent with the temperature-mortality literature at large that suggests long-term adaptation to mortality exists through a combination of physical acclimatization, adaptations of technologies (such as air-conditioning), and changes in behavior. In theory, these adaptations should affect the homicide rate as well, yet the relatively short panel length of my analysis appears to suggest that adaptation plays no role over this time frame.

## 1.5 Discussion

The strong differential effects of race are new to the literature, but these findings also offer insights into the relevant mechanisms in which heat increases homicide. These results suggest that for white victims, and especially young white males, I cannot rule out the heat hypothesis while an interactional approach (RA theory) has less support. Two aspects of my results favor the heat hypothesis over an interactional mechanism. First, there is no effect in colder months, where an interactional mechanism would suppress homicides. Second, the returns to warm temperature bins are increasing, and the largest effect is seen in the warmest temperature bin. This is indicative of an annoyance based mechanism as extremely warm temperatures are the most physiologically stressing and at some point, interactions decrease as individuals seek refuge. Therefore, given that colder temperatures do not decrease homicide for white victims, and the largest effect is seen in warmest bin, this supports the heat hypothesis. For black victims, and especially young black males, the evidence for the heat hypothesis is limited as elevated temperatures have no effect on homicide. Conversely, the strong negative effect seen for colder days is instead illustrative of an interactional mechanism where cold weather suppresses interactions leading to less conflict and homicide. Although monthly fixed effect control for the fact that interactions decrease across colder months, these results suggest that when temperatures fall within months the homicide rate decreases.

To generalize theses findings back to the types of aggression used would suggest that homicides with white victims would be associated with hostile aggression while black victims would be associated with instrumental aggression. This is of course a generalization, as there are clearly homicides motived by both types of aggression regardless of race, yet these regression results do highlight that there are clear differential effects depending on race, and motivations may play a role. Furthermore, this claim does fit the narrative previously discussed where the disproportionately high black homicide rate is explained by a inadequate criminal justice system that has made homicide endemic. If homicide in these communities is more closely associated with instrumental aggression than elevated temperatures would not play a role, which is what these results weakly indicate. The much stronger non-effect for young black victims for elevated temperatures could be reflective of the fact that the deficiencies in the criminal justice system are disproportionately burdened on this demographic. Meanwhile, the stronger effect for colder temperatures could be reflective of the fact that the young male black demographic has more free time or flexibility across schedules, and thus more subject to weather changing interactional patterns.

Regardless of the mechanism implications, these results demonstrate how important it is to separate the homicide rate by race. If I simply consider all young male victims, it demonstrates that colder temperatures decrease homicide, while warmer temperatures increase it (which is largely consistent with what the previous literature has found). However, the more accurate story is more nuanced as these effects appear to be driven by different races. Thus, any future work must explicitly separate the homicide rate by race.

## 1.6 Conclusion

Across the entire body of work that studies the relationship between temperature and homicide there has been surprisingly little to no discussion about the role of race. Homicide is not a simple summary statistic, but a nuanced measure that varies widely across different demographic and socioeconomic backgrounds. The importance of race cannot be overstated. Despite representing only 13% of the total population, over half of all homicides are black victims. There are many complicated reasons why the homicide epidemic in America is disproportionately black, but what cannot be denied is that the sheer number of black deaths suggest there is something fundamentally different about homicide in black communities. Since 1981, nearly 240,000 black men have been killed, which is over four times the number of Americans who died in the Vietnam War. For many, particularly those in the most affected communities, these deaths are unacceptable and lack of public outrage is often incomprehensible. It is no wonder that moments such as "Black Lives Matter" have gained traction in recent years, when the deaths of so many goes unnoticed.

By failing to recognize the potential for systematic differences across the homicide rate by race, much of the previous work presents an inaccurate reflection of the true relationship between temperature and homicide. As my results suggest, it appears that black victims drive much of the negative effect for cold temperatures, while white victims drive the positive effect for warm temperatures. While this finding alone is a new addition to the literature, these results also provide insights into the mechanism in which temperature affects homicide. Race is far from a perfect measure of different homicide types, yet these results reveal the differential effects for weather across victim race. For white victims, I cannot rule out the hypothesis that heat acts a physiological stressor leading to violence and aggression. However, for black victims, and especially young black males, the support for this hypothesis is limited. Instead it appears that for black victims, an interactional mechanism plays a role as colder temperatures reduce the homicide rate drastically.

Relating this finding back to types of aggression suggests that homicide for white victims would be more closely associated with hostile aggression, while black victims (especially young black males) would be associated with instrumental aggression. This claim is controversial, as clearly not every homicide can be so easily described by race, yet my results do reveal strong differential effects that may be explained by different types of aggression. Although homicide rates are at some of the lowest levels over the past 30 years, one cannot escape the obvious fact that young black men remain disproportionately victimized. There are many currently working to helps these rates continue to fall, and it is unclear if the correct remedy lies in more prevention, responsiveness, or a combination of both. What cannot be ignored is the urgency of the problem and to truly understand the homicide epidemic in the US, researchers must be very explicit about the differences across race.

Figure 1.1: Sample Counties

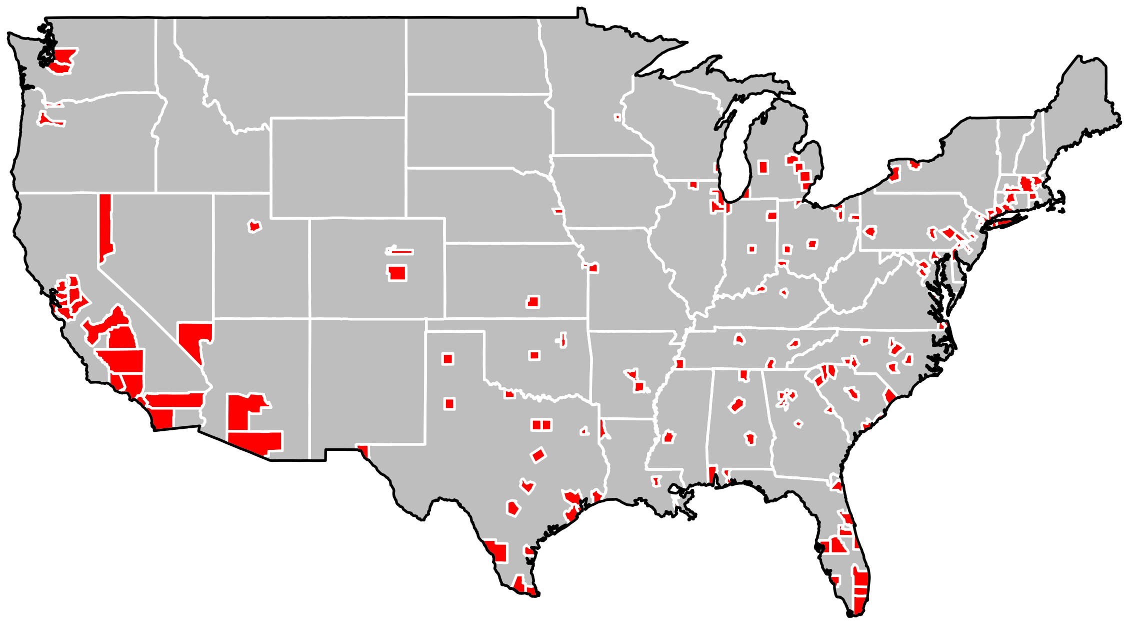


Table 1.1: Summary Statistics of Sample

| | Entire SHR Dataset | Sample |
|---|---|---|
| **County Characteristics** | | |
| Number of Counties | 2997 | 174 |
| Population | 88,921 (310,407) | 770,636 (1,044,631) |
| Pct White | 0.88 (0.15) | 0.76 (0.14) |
| Pct Black | 0.09 (0.15) | 0.19 (0.15) |
| Pct Other Race | 0.03 (0.06) | 0.05 (0.05) |
| **Homicide Rates** | | |
| Homicides per Month | 1.52 (3.31) | 6.99 (12.52) |
| Homicide Rate (Total) | 0.96 (1.29) | 0.87 (0.65) |
| White Homicide Rate | 0.82 (1.29) | 0.47 (0.26) |
| Black Homicide Rate | 6.55 (130.46) | 2.27 (1.12) |
| Other Race Homicide Rate | 1.73 (21.71) | 0.36 (0.33) |
| Young Male Homicide Rate | 2.09 (4.87) | 2.67 (2.39) |
| White Young Male Homicide Rate | 1.63 (4.84) | 1.14 (0.71) |
| Black Young Male Homicide Rate | 5.44 (77.36) | 7.74 (4.89) |
| Other Race Young Male Homicide Rate | 2.12 (26.02) | 0.70 (0.85) |
| White-on-White Homicide Rate | 0.59 (1.01) | 0.22 (0.12) |
| Black-on-Black Homicide Rate | 4.75 (128.38) | 1.17 (0.54) |
| Black-on-White Homicide Rate | 0.01 (0.45) | 0.02 (0.01) |
| White-on-Black Homicide Rate | 0.02 (0.07) | 0.03 (0.02) |

*Note: The table shows county level averages for the entire SHR dataset and sample between 1981 - 2013. Standard Deviations are in parenthesis.*

Table 1.2: Weather Summary Statistics of Sample

| | Entire SHR Dataset | Sample |
|---|---|---|
| **Monthly Number of Days in Weather Bin** | | |
| Number of Counties | 2997 | 174 |
| Temp: $< 10°$ F | 0.49 (0.82) | 0.14 (0.30) |
| Temp: $10- < 20°$ F | 0.86 (0.91) | 0.44 (0.59) |
| Temp: $20- < 30°$ | 2.00 (1.49) | 1.28 (1.28) |
| Temp: $30- < 40°$ | 3.74 (1.67) | 2.86 (1.96) |
| Temp: $40- < 50°$ | 4.70 (1.19) | 4.40 (1.74) |
| Temp: $50- < 60°$ | 5.04 (0.99) | 5.47 (2.05) |
| Temp: $60- < 70°$ | 5.66 (0.90) | 6.09 (1.57) |
| Temp: $70- < 80°$ | 5.72 (2.32) | 6.51 (2.22) |
| Temp: $80- < 90°$ | 2.20 (2.44) | 3.15 (3.13) |
| Temp: $> 90°$ | 0.03 (0.23) | 0.10 (0.51) |
| Precip: 0 mm | 16.57 (3.23) | 16.81 (3.21) |
| Precip: $> 0- < 10$ mm | 11.49 (3.33) | 11.22 (2.89) |
| Precip: $10- < 20$ mm | 1.40 (0.52) | 1.38 (0.47) |
| Precip: $20- < 30$ mm | 0.54 (0.25) | 0.55 (0.22) |
| Precip: $\geq 30$ mm | 0.45 (0.29) | 0.46 (0.26) |
| **Monthly Degree Days** | | |
| Degree Days Above 70° F | 56.75 (44.64) | 75.87 (56.34) |
| Degree Days Below 50° F | 179.35 (118.53) | 114.45 (91.98) |

*Note: The table shows county level weather averages for the entire SHR dataset and sample between 1981 - 2013. Standard Deviations are in parenthesis.*

Table 1.3: Log Total Homicide Rate by Race: Temperatutre Bins

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temp : $< 10°$ F | 0.000919 | 0.000262 | 0.00125 | 0.00231 |
| | (0.00118) | (0.00111) | (0.00248) | (0.00251) |
| Temp: $10- < 20°$ F | -0.00269 | -0.00170 | -0.00511 | -0.00292 |
| | (0.00191) | (0.00138) | (0.00315) | (0.00178) |
| Temp: $20- < 30°$ F | -0.00133* | -0.000531 | -0.00341** | 0.00216 |
| | (0.000777) | (0.000777) | (0.00158) | (0.00205) |
| Temp: $30- < 40°$ F | -0.00182* | -0.000918 | -0.00331 | 0.000379 |
| | (0.00101) | (0.000773) | (0.00204) | (0.00129) |
| Temp: $40- < 50°$ F | -0.000513 | -0.000210 | -0.000979 | 0.000124 |
| | (0.000511) | (0.000483) | (0.00105) | (0.000658) |
| Temp: $50- < 60°$ F | -0.000800** | -0.000345 | -0.00187** | 0.000417 |
| | (0.000401) | (0.000385) | (0.000918) | (0.000626) |
| Temp: $70- < 80°$ F | 0.000907** | 0.00116*** | 0.000484 | -0.000257 |
| | (0.000364) | (0.000374) | (0.000639) | (0.000415) |
| Temp: $80- < 90°$ F | 0.00188*** | 0.00191*** | 0.00196** | 0.00151** |
| | (0.000486) | (0.000565) | (0.000929) | (0.000729) |
| Temp: $\geq 90°$ F | 0.00133 | 0.00103 | 0.00357 | -0.000425 |
| | (0.000811) | (0.000794) | (0.00332) | (0.00208) |
| Precip: $> 0- < 10$ mm | 0.000204 | -0.000386 | 0.00153*** | 0.000429 |
| | (0.000285) | (0.000386) | (0.000584) | (0.000778) |
| Precip: $10- < 20$ mm | 0.0000186 | 0.000358 | 0.000658 | 0.00187 |
| | (0.000892) | (0.000925) | (0.00173) | (0.00184) |
| Precip: $20- < 30$ mm | 0.000318 | 0.00147 | -0.000196 | -0.00220 |
| | (0.00136) | (0.00146) | (0.00269) | (0.00260) |
| Precip: $\geq 30$ mm | -0.00123 | -0.00209 | -0.000693 | 0.00280 |
| | (0.00141) | (0.00146) | (0.00278) | (0.00312) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.796 | 0.612 | 0.609 | 0.252 |

Notes:
Regressions weighted by relevant populations
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.4: Log Total Homicide Rate by Race: Degree Day

| | (1)<br>Homicide Rate:<br>All Victims | (2)<br>Homicide Rate:<br>White Victims | (3)<br>Homicide Rate:<br>Black Victims | (4)<br>Homicide Rate:<br>Other Race |
|---|---|---|---|---|
| Temperature | | | | |
| (Degree Days Over 70F) | 0.000108*** | 0.000104*** | 0.000138** | 0.0000529 |
| | (0.0000301) | (0.0000345) | (0.0000536) | (0.0000492) |
| Temperature | | | | |
| (Degree Days Under 50F) | -0.0000271 | -0.0000163 | -0.0000721** | 0.0000133 |
| | (0.0000217) | (0.0000186) | (0.0000354) | (0.0000401) |
| Precip: $> 0- < 10$ mm | 0.000124 | -0.000432 | 0.00147** | 0.000545 |
| | (0.000292) | (0.000381) | (0.000579) | (0.000808) |
| Precip: $10- < 20$ mm | -0.000109 | 0.000332 | 0.000420 | 0.00180 |
| | (0.000861) | (0.000896) | (0.00170) | (0.00178) |
| Precip: $20- < 30$ mm | 0.000205 | 0.00145 | -0.000405 | -0.00207 |
| | (0.00138) | (0.00145) | (0.00269) | (0.00260) |
| Precip: $\geq 30$ mm | -0.00137 | -0.00217 | -0.000879 | 0.00278 |
| | (0.00142) | (0.00147) | (0.00277) | (0.00300) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.796 | 0.612 | 0.609 | 0.252 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.5: Log Homicide Rate for Young Males (15-34) by Race: Temperature Bins

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temp : $< 10°$ F | -0.00118 | -0.00142 | -0.00262 | -0.000484 |
| | (0.00293) | (0.00329) | (0.00508) | (0.00369) |
| Temp: $10- < 20°$ F | -0.00531 | -0.00281 | -0.0109** | -0.00736* |
| | (0.00366) | (0.00282) | (0.00518) | (0.00400) |
| Temp: $20- < 30°$ F | -0.00474*** | -0.00114 | -0.0109*** | 0.00176 |
| | (0.00147) | (0.00187) | (0.00251) | (0.00323) |
| Temp: $30- < 40°$ F | -0.00335* | -0.00196 | -0.00650** | -0.000434 |
| | (0.00175) | (0.00180) | (0.00280) | (0.00274) |
| Temp: $40- < 50°$ F | -0.00250*** | -0.000782 | -0.00551*** | 0.000483 |
| | (0.000781) | (0.000919) | (0.00174) | (0.00164) |
| Temp: $50- < 60°$ F | -0.00222*** | -0.000938 | -0.00565*** | 0.00122 |
| | (0.000730) | (0.000741) | (0.00145) | (0.00160) |
| Temp: $70- < 80°$ F | 0.00112* | 0.00214*** | -0.000796 | 0.000326 |
| | (0.000659) | (0.000729) | (0.00132) | (0.00109) |
| Temp: $80- < 90°$ F | 0.00251*** | 0.00361*** | 0.00124 | 0.00207 |
| | (0.000930) | (0.00114) | (0.00194) | (0.00172) |
| Temp: $\geq 90°$ F | 0.00651** | 0.00894*** | 0.00305 | 0.000809 |
| | (0.00280) | (0.00314) | (0.00679) | (0.00520) |
| Precip: $> 0- < 10$ mm | 0.0000311 | -0.00132 | 0.00254** | -0.0000239 |
| | (0.000599) | (0.000812) | (0.00110) | (0.00138) |
| Precip: $10- < 20$ mm | 0.00000249 | -0.000279 | 0.00417 | -0.000831 |
| | (0.00216) | (0.00246) | (0.00337) | (0.00529) |
| Precip: $20- < 30$ mm | 0.00227 | -0.0000103 | 0.00462 | 0.00847 |
| | (0.00291) | (0.00315) | (0.00485) | (0.00721) |
| Precip: $\geq 30$ mm | -0.00345 | -0.00280 | -0.00510 | 0.00745 |
| | (0.00313) | (0.00378) | (0.00501) | (0.00904) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.678 | 0.534 | 0.513 | 0.233 |

Notes:
Regressions weighted by relevant populations
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.6: Log Homicide Rate for Young Males (15-34) by Race: Degree Days

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temperature (Degree Days Over 70F) | 0.000201*** | 0.000269*** | 0.000133 | 0.0000529 |
| | (0.0000516) | (0.0000691) | (0.000105) | (0.0000492) |
| Temperature (Degree Days Under 50F) | -0.0000754 | -0.0000268 | -0.000168** | 0.0000133 |
| | (0.0000460) | (0.0000473) | (0.0000679) | (0.0000401) |
| Precip: $> 0 - < 10$ mm | -0.000334 | -0.00153* | 0.00213* | 0.000545 |
| | (0.000635) | (0.000799) | (0.00115) | (0.000808) |
| Precip: $10 - < 20$ mm | -0.000539 | -0.000428 | 0.00327 | 0.00180 |
| | (0.00215) | (0.00246) | (0.00335) | (0.00178) |
| Precip: $20 - < 30$ mm | 0.00179 | -0.000123 | 0.00365 | -0.00207 |
| | (0.00291) | (0.00315) | (0.00487) | (0.00260) |
| Precip: $\geq 30$ mm | -0.00378 | -0.00293 | -0.00585 | 0.00278 |
| | (0.00316) | (0.00377) | (0.00501) | (0.00300) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.678 | 0.534 | 0.513 | 0.252 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.7: Log Homicide Rate by Offender and Victim Race: Temperature Bins

| | (1) Homicide Rate: White-White | (2) Homicide Rate: Black-Black | (3) Homicide Rate: White-Black | (4) Homicide Rate: Black-White |
|---|---|---|---|---|
| Temp : $< 10°$ F | -0.000158 | -0.00313 | -0.000170 | 0.0000955 |
| | (0.000876) | (0.00264) | (0.000269) | (0.000597) |
| Temp: $10- < 20°$ F | -0.00130 | -0.00572** | -0.000147 | -0.000454 |
| | (0.000931) | (0.00241) | (0.000337) | (0.000312) |
| Temp: $20- < 30°$ F | 0.000306 | -0.00357** | -0.000291 | -0.000156 |
| | (0.000523) | (0.00171) | (0.000188) | (0.000232) |
| Temp: $30- < 40°$ F | -0.000118 | -0.00409*** | -0.0000752 | -0.000151 |
| | (0.000405) | (0.00146) | (0.000102) | (0.000155) |
| Temp: $40- < 50°$ F | 0.0000712 | -0.000254 | -0.000132* | -0.00000445 |
| | (0.000299) | (0.000860) | (0.0000691) | (0.000111) |
| Temp: $50- < 60°$ F | -0.000339 | -0.00235*** | -0.0000168 | -0.0000788 |
| | (0.000270) | (0.000733) | (0.0000734) | (0.0000914) |
| Temp: $70- < 80°$ F | 0.000381* | 0.000268 | 0.000135** | 0.000124 |
| | (0.000207) | (0.000524) | (0.0000582) | (0.0000869) |
| Temp: $80- < 90°$ F | 0.00105*** | 0.000957 | 0.000150** | 0.000166 |
| | (0.000226) | (0.000772) | (0.0000713) | (0.000110) |
| Temp: $\geq 90°$ F | -0.000236 | 0.0000677 | 0.000103 | 0.000292** |
| | (0.000298) | (0.00206) | (0.000130) | (0.000134) |
| Precip: $> 0- < 10$ mm | -0.000174 | 0.000964 | -0.00000261 | 0.0000159 |
| | (0.000254) | (0.000654) | (0.0000655) | (0.0000848) |
| Precip: $10- < 20$ mm | 0.000255 | -0.000525 | 0.0000668 | -0.000245 |
| | (0.000633) | (0.00159) | (0.000228) | (0.000302) |
| Precip: $20- < 30$ mm | 0.000434 | -0.00116 | -0.000125 | -0.000372 |
| | (0.00119) | (0.00314) | (0.000449) | (0.000487) |
| Precip: $\geq 30$ mm | -0.000788 | 0.000891 | -0.000116 | -0.000199 |
| | (0.00112) | (0.00267) | (0.000404) | (0.000453) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.445 | 0.518 | 0.166 | 0.183 |

Notes:
Regressions weighted by relevant populations
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.8: Log Homicide Rate by Offender and Victim Race: Degree Days

| | (1)<br>Homicide Rate:<br>White-White | (2)<br>Homicide Rate:<br>Black-Black | (3)<br>Homicide Rate:<br>White-Black | (4)<br>Homicide Rate:<br>Black-White |
|---|---|---|---|---|
| Temperature | | | | |
| (Degree Days Over 70F) | 0.0000421** | 0.0000296 | 0.00000535 | 0.00000720 |
| | (0.0000191) | (0.0000416) | (0.00000470) | (0.00000496) |
| Temperature | | | | |
| (Degree Days Under 50F) | 0.00000350 | -0.0000944*** | -0.00000513* | -0.00000468 |
| | (0.00000953) | (0.0000279) | (0.00000310) | (0.00000285) |
| Precip: $> 0- < 10$ mm | -0.000301 | 0.000756 | -0.0000245 | -0.000000212 |
| | (0.000226) | (0.000669) | (0.0000650) | (0.0000809) |
| Precip: $10- < 20$ mm | 0.000414 | -0.000768 | 0.0000521 | -0.000240 |
| | (0.000616) | (0.00160) | (0.000229) | (0.000307) |
| Precip: $20- < 30$ mm | 0.000529 | -0.00126 | -0.000126 | -0.000349 |
| | (0.00119) | (0.00310) | (0.000449) | (0.000485) |
| Precip: $\geq 30$ mm | -0.000990 | 0.000865 | -0.0000950 | -0.000192 |
| | (0.00112) | (0.00264) | (0.000408) | (0.000450) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.445 | 0.518 | 0.166 | 0.183 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 1.9: Estimated Effect of Days with Temperature Greater Than or Equal to 80° F on Log Mortality Rate by Race, by Decile of its Distribution from 1981 - 2013

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| 1st Decile | 0.00391 | -0.000191 | 0.0125 | 0.0185*** |
| | (0.00371) | (0.00259) | (0.0132) | (0.00485) |
| 2nd Decile | -0.00108 | 0.0000808 | -0.00561 | 0.00561 |
| | (0.00363) | (0.00288) | (0.0117) | (0.00468) |
| 3rd Decile | -0.0000459 | 0.00168 | 0.000561 | -0.00403 |
| | (0.00164) | (0.00178) | (0.00295) | (0.00260) |
| 4th Decile | 0.00248 | 0.00196 | 0.00259 | 0.00300*** |
| | (0.00159) | (0.00206) | (0.00249) | (0.000710) |
| 5th Decile | 0.00116 | 0.000532 | 0.00153 | 0.00108 |
| | (0.000866) | (0.000993) | (0.00191) | (0.00175) |
| 6th Decile | 0.000310 | -0.000390 | 0.000736 | 0.00122 |
| | (0.000868) | (0.00116) | (0.00116) | (0.00220) |
| 7th Decile | 0.00121 | 0.00157 | 0.000721 | 0.00218*** |
| | (0.000824) | (0.00102) | (0.00142) | (0.000469) |
| 8th Decile | 0.00184** | 0.00187 | 0.00158 | -0.000500 |
| | (0.000826) | (0.00119) | (0.00116) | (0.000938) |
| 9th Decile | 0.000835 | 0.000925* | 0.00147 | 0.000801 |
| | (0.000677) | (0.000556) | (0.00184) | (0.00198) |
| 10th Decile | 0.00112 | 0.000873 | 0.00202 | 0.00191 |
| | (0.000709) | (0.000578) | (0.00148) | (0.00157) |
| Precip: $> 0- < 10$ mm | 0.0000757 | -0.000448 | 0.00135** | 0.000484 |
| | (0.000291) | (0.000381) | (0.000592) | (0.000798) |
| Precip: $10- < 20$ mm | -0.000123 | 0.000294 | 0.000392 | 0.00187 |
| | (0.000856) | (0.000895) | (0.00169) | (0.00176) |
| Precip: $20- < 30$ mm | 0.000238 | 0.00143 | -0.000291 | -0.00205 |
| | (0.00140) | (0.00146) | (0.00272) | (0.00259) |
| Precip: $\geq 30$ mm | -0.00136 | -0.00222 | -0.000829 | 0.00286 |
| | (0.00141) | (0.00148) | (0.00277) | (0.00295) |
| No Obs | 30456 | 30456 | 30456 | 30456 |
| $R^2$ | 0.796 | 0.612 | 0.609 | 0.253 |

Notes:
Regressions weighted by relevant populations
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# Chapter 2

# The Dynamics of Weather, Yields and Deforestation: Evidence from Soy in the Brazilian Amazon

## Attributions

The content of chapter 2 is the result of a collaboration with Ryan Abman and Sam Heft-Neal. Ryan Abman is an Assistant Professor in the Department of Economics at San Diego State University. Sam Heft-Neal is a Postdoctoral Fellow in the Department of Earth System Science at Stanford University.

## 2.1 Introduction

The dual threats of carbon emissions and biodiversity loss make tropical deforestation one of the world's most pressing environmental challenges. Deforestation accounts for approximately 15 percent of global greenhouse gas emissions, more than the entire global transportation sector. Tropical forests are also home to over half of the world's species, many of which are found in microhabitats and are vulnerable to extinction. In light of these risks, research considering the drivers of deforestation has increased dramatically in recent years (Kaimowitz and Angelsen, 1998; Barbier and Burgess, 2001).

A consensus has emerged that land use conversion to agriculture is a leading driver of tropical deforestation yet the relationship between agricultural productivity and deforestation is ambiguous. The classical view, typically called the "Borlaug Hypothesis," is that increased agriculture productivity reduces the amount of land needed for agriculture and reduces deforestation. Named after one of the leaders of the Green Revolution, Norman Borlaug, this hypothesis argues that the agriculture productivity improvements seen in the Green Revolution had a positive effect on forest cover as less cropland was needed to feed rising populations. However, others have argued that agricultural productivity can increase deforestation pressures, as the return to clearing land increases. If profits, rather than total output are the ultimate goal of the farmer, than agriculture productivity improvements would encourage more land cleared to capture rents. Therefore, the relationship between agriculture productivity and deforestation very much depends on the context and production system (Angelsen and Kaimowitz, 2001). Empirical evidence considering the relationship between agriculture productivity and deforestation is limited. As noted by Villoria et al. 2014 in a recent review, much of the existing literature focuses on cross-country differences in productivity and land in agriculture. There are relatively few within-country studies that effectively answer this question due to a lack of exogenous spatial and temporal variation in agricultural productivity.

In this chapter, we study the dynamic relationship between agricultural productivity and deforestation in the Brazilian Amazon. We begin with a conceptual model that considers how an exogenous shock to productivity may influence a farmers decision to clear land in the following year. This model demonstrates the conditions where a positive relationship between productivity shocks and land clearing exist and serves as the motivation for our empirical specification. Our empirical analysis uses annual variation in growing-season temperature as a source of exogenous productivity variation in soy. Using a panel of municipalities in the Brazilian Amazon, we demonstrate that increased temperatures during the growing season lead to higher soy yields in that year, increased planted area for soy in the following year, and increased deforestation two years after. While the delayed effect of temperature on deforestation may at first appear puzzling, it is consistent with findings of pastureland displacement. The majority of deforestation in the Brazilian Amazon is for pastureland, and although there is some forest conversion to soy occurring, soy frequently moves into pastureland, displacing pasture to forest frontiers (Morton et al., 2006; Barona et al., 2010; Arima et al., 2011; Macedo et al., 2012). Our results provide evidence that short-run impacts of improved productivity may increase deforestation, refuting - at least in part - the '"Borlaug Hypothesis".

Our results, while consistent with static models of deforestation (such as Angelsen 1999), do stand in contrast to findings of some recent work. In a working paper, Assunção et al. (2016) study the impact of electrification in the Brazilian Amazon on subsequent deforestation from 1970 to 2006. They find that electrification led to more agricultural intensification and, over time, less deforestation. Electrification allows farmers to make extensive-margin adjustments in crop composition and agricultural activity that can be studied over longer time periods. We focus our efforts on weather fluctuations during one crop's growing season which may lead to intensive margin adjustments but likely not extensive margin adjustments from year to year. In another working paper, Abman and Carney (2016) study the impact of fertilizer subsidies on deforestation in Malawi, finding a negative yield-deforestation relationship from fertilizer. In Malawi, most maize farming is done at the subsistence level and fertilizer may delay or reduce the need to shift cultivation to a new, forested plot of land. In Brazil, soy is farmed at commercial scales, thus labor and input constraints present at the Malawian household-level are unlikely to bind for Brazilian soy farming.

Existing literature studying agricultural productivity and deforestation often relies on cross-country differences in agricultural productivity. Notable examples include Ewers et al. (2009) and Rudel et al. (2009) which look at country-level increases in agricultural productivity and changes in total agricultural area, finding little evidence that increased productivity led to land-sparring. By leveraging within-country variation in productivity we are able to avoid many of the inherent issues of country-level governance and land use regulations that may also be correlated with agricultural productivity. Furthermore, panel data allows us to examine the dynamic relationship between productivity and deforestation.

We merge two existing literatures to study the relationship between agricultural productivity and deforestation in Brazil. The first literature examines the relationship between weather and agriculture and the second examines the relationship between agriculture, land use, and deforestation. While both literatures are further discussed in the following section, we apply the insights from the weather and agriculture literature, namely that transitory weather shocks can have significant impacts on agricultural yields. By merging spatially explicit data on weather and soy yields, we model the growing-season temperature relationship with soy yields in the Brazilian Amazon. We then consider how yields relate to planted area and planted area to deforestation. For both, we find evidence of a positive relationship with a one year lag. That is, a one percent increase in soy yields is associated with a 0.4% increase in soy planted area in the following year, while a one percent increase in planted area is associated with a 0.1% increase in deforestation in the following year. Together this suggests that a favorable weather shock would increase deforestation after two years.

We then examine the relationship between lagged growing-season temperature and deforestation using satellite data on forest loss. We find increased growing-season maximum temperatures (which increase soy yields and subsequent planted area) lead to significantly higher deforestation two years later (the year after we observe increases in soy planted area). The finding of the lagged effect is consistent with literature on indirect land use changes in the Brazilian Amazon. This literature finds that as soy moves into the Amazon, it will often be grown on land previously dedicated to pasture land and push cattle pastures further into the Amazon forest. This chapter is the first to our knowledge to use weather-induced yield variation to study land use change and deforestation from agriculture. Furthermore, much of the indirect land use change literature in Brazil uses long run changes in agricultural composition in distant municipalities to show evidence of indirect land use change. Our chapter provides evidence of short-run local effects driven by soy expansion displacing cattle pasture land.

The rest of the chapter is organized as follows. Section 2.2 presents the background and motivation for our work. Section 2.3 presents our conceptual model. Section 2.4 presents data and our empirical specification. Section 2.5 discuses our empirical results. Section 2.6 concludes with final remarks.

## 2.2 Background

The foundation of this chapter's empirical methodology falls within the growing field of the climate-economy literature as profiled by Dell et al. (2014). This new wave of empirical research uses panel methodologies and high frequency weather variables to identify the impact of weather on a variable of interest. This literature overcomes identification issues as the plausibly random variation in weather outcomes over time within a given spatial area are able to identify the effects of weather on the outcome of interest. Specifically, the most closely related work to our own analysis is from Schlenker and

Roberts (2009), who find that soy yields in the United States increase gradually until 30° C, but temperatures above this threshold are highly damaging to yields. This finding documents the direct relationship between temperature and soy yields, and serves as the starting point for our analysis when we attempt to document a similar relationship between temperature and soy in the Brazilian Amazon.

Moving next to the deforestation literature, we focus our analysis on Brazil where numerous studies have explored the relationship between pasture, soybean, and deforestation in the Brazilian Amazon. Historically, along the "arc of deforestation" in the southern and eastern Brazilian Amazon, forest conversion began with small-scale exploration for timber or subsistence agriculture, followed by larger-scale cattle ranching. However, more recently the rise of large-scale mechanized agriculture for export crops have introduced new pressures on the forest frontier. Soybean cultivation in particular has increased dramatically, and is the largest cultivated crop in terms of harvested area since the 1990s (Barona et al., 2010). This rapid rise of intensive agriculture, matched with the large remaining potential for expansion has caused much alarm about the potential for future forest loss. Indeed, this concern appears warranted as the Brazilian Amazon lost some 18.9 million hectares of forest from 2000 - 2006, while soybean cultivation reached new highs (Barona et al., 2010).

Although many hypothesized that the rapid increase in soy production was to blame for the large forest loss, research suggests a more complicated picture. In particular, there have been several recent works that take advantage of high quality satellite remote sensing data to empirically consider how pasture and soybean cultivation impact deforestation rates in the Brazilian Amazon. Brown et al. (2005) and Morton et al. (2006) suggest that although soy is responsible for some forest conversion, the majority of deforestation that occurred in the first half of the 2000s was destined for cattle pasture. A second wave of papers considers the role of indirect land use change, in which mechanized agriculture encroaches on existing pastures, displacing those to new forest frontiers. This body of research became particularly relevant in the latter half of the 2000s, as deforestation rates across the Brazilian Amazon decreased dramatically, but soybean production continued to grow to all-time highs. These studies find evidence that the increased soybean production in the latter half of the 2000s moved predominately into previous pastureland, which in turn caused pasture to move to new forest frontiers (Barona et al., 2010; Arima et al., 2011; Macedo et al., 2012).

Another closely related work from the Brazilian deforestation literature comes from Garrett et al. (2013), who consider how land institutions and supply chain configurations are determinants of soybean planted area and yields in the Brazilian Amazon. Although the authors' provide some evidence that supply chain infrastructure is positively associated with planted area and yields, their analysis is only a cross-sectional study, which does not offer a causal analysis into what role these variables play in determining yields and planted area. Most importantly for our work, they also find evidence that soy yields are positively associated with planted area, although this finding is only cross sectional.

Although much of the previous research touches on similar topics as our work, our analysis advances the literature in several ways. First, our empirical strategy and data allow us to overcome issues of exogeniety that have adversely affected much of the previous literature. For example, while Garrett et al. (2013) also considers how biophysical conditions such as weather affect soy yields in the Brazilian Amazon, that analysis is only conducted on a cross-sectional level. Conversely, our empirical strategy produces unbiased estimates of the effect of weather on soy yields through the use of fixed effects and the plausibly random year-to-year variation in temperatures. Second, and more importantly, our analysis considers how changes in agriculture productivity through weather shocks impact deforestation rates, which is unique to the deforestation literature.

## 2.3 A Simple Model of Weather, Agricultural Productivity and Land Clearing

The central hypothesis of our chapter is that weather shocks affect deforestation through their impact on soy yields and soy planted area. To better illustrate why a farmer would make planting decisions based on yield changes from weather shocks we present a two-period, partial equilibrium model. In this model, a price-taking individual is endowed with a certain amount of cleared land ($H_0$), whose productivity ($\alpha_0$) depends on some exogenous factor (weather). We abstract away from labor and capital decisions for simplicity. Total production in both periods, $F(\alpha, H)$, is increasing in both land and and productivity ($F' > 0, F'' < 0$). After the growing season, the individual can choose to clear additional land ($h$) to put into production in the following year ($H_1 = H_0 + h$). The productivity of the land in the following year ($\alpha_1$) is unknown to the individual and is a random variable distributed log-normal. The individual may also choose to invest in a liquid asset $a$ which earns a return of $R$ the following year with certainty. The price of the agricultural good produced is normalized to 1. Consumption in year 0 is everything produced minus total clearing costs and investments, and consumption in year 1 is everything produced in that year plus the total return from investments. The individual's utility function is increasing and concave in consumption ($u' > 0, u'' < 0$). The individual discounts future consumption by $\beta$.

We can define the individual's problem as:

$$\max_{h,a} \mathbb{E}[u(c_0, c_1)] = u(F(\alpha_0, H_0) - ch - a) + \beta\,\mathbb{E}[u(F(\alpha_1, H_1) + Ra)] \tag{2.1}$$

subject to:

$$F(\alpha_0, H_0) - ch - a > 0$$

$$h \geq 0$$

$$a \geq 0$$

$$H_1 = H_0 + h$$

The lagrangian is as follows:

$$L = u(F(\alpha_0, H_0) - ch - a) + \beta\,\mathbb{E}[u(F(\alpha_1, H_0 + h) + Ra)] + \lambda_1 h + \lambda_2 a + \lambda_3 (F(\alpha_0, H_0) - ch - a) \tag{2.2}$$

As above, the concavity of the utility function implies that $\lambda_3 = 0$ and the budget constraint will not bind. The FOCs with respect to $h$ and $a$ are:

$$L_h = -u'(F(\alpha_0, H_0) - ch - a)(c) + \beta\,\mathbb{E}\left[u'(F(\alpha_1, H_0 + h) + Ra)\frac{\partial F}{\partial H}\right] + \lambda_1 = 0$$

$$L_a = -u'(F(\alpha_0, H_0) - ch - a) + \beta R\,\mathbb{E}[u'(F(\alpha_1, H_0 + h) + Ra)] + \lambda_2 = 0$$

The optimal pair of land clearing and liquid asset investment, $\{h^*, a^*\}$, solves the following system of equations

$$u'(F(\alpha_0, H_0) - ch^* - a^*)(c) = \beta\,\mathbb{E}\left[u'(F(\alpha_1, H_0 + h^*) + Ra^*)\frac{\partial F}{\partial H}\right] + \lambda_1$$

$$u'(F(\alpha_0, H_0) - ch^* - a^*) = \beta R\,\mathbb{E}[u'(F(\alpha_1, H_0 + h^*) + Ra^*)] + \lambda_2$$

### 2.3.1 Characterization of Possible Outcomes

In the above model, the optimal choice of $a^*$ and $h^*$ can be classified by the following four cases.

1. $a^* = h^* = 0$, All first year production consumed in first year.
2. $a^* > 0,\ h^* = 0$, Landowner only invests in liquid asset, no land clearing.
3. $a^* = 0,\ h^* > 0$, Landowner does not invest in liquid asset, but chooses to clear land.
4. $a^* > 0,\ h^* > 0$, Landowner invests in liquid asset and clears land.

All production is consumed in year 0 if the following two conditions hold:

$$u'(F(\alpha_0, H_0))(c) > \beta\, \mathbb{E}\left[u'(F(\alpha_1, H_0))\frac{\partial F}{\partial H}\right] \tag{2.3}$$

$$u'(F(\alpha_0, H_0)) > \beta R\, \mathbb{E}[u'(F(\alpha_1, H_0))] \tag{2.4}$$

In this case the realization for $\alpha_0$ is sufficiently low relative to the expected realization of $\alpha_1$ that the marginal utility of consumption in year 0 is greater than the discounted expected marginal utility of consumption in year 1.

For higher realizations of $\alpha_0$, the landowner's choice of how to smooth consumption will be decided by the relationship between the return on the liquid asset $R$ and the expected marginal product of land clearing. The landowner will clear no land if:

$$R > \left(\frac{1}{c}\right) \mathbb{E}\left[\frac{\partial F}{\partial H}(\alpha_1, H_0)\right] \tag{2.5}$$

Conversely, if the expected returns to clearing land are sufficiently large, the landowner may choose to invest all excess consumption into clearing land, so long as the expected marginal product of land remains above the return on the liquid asset. That is, $a^* = 0$ if:

$$R < \left(\frac{1}{c}\right) \mathbb{E}\left[\frac{\partial F}{\partial H}(\alpha_1, H_0 + h^*)\right] \tag{2.6}$$

Finally, the landowner will choose to invest both in land clearing and the liquid asset if the expected marginal product of land clearing is above the return on the liquid asset for the initial endowment of land ($H_0$) but is sufficiently diminished when more land is put into production such that becomes equal to the return on the liquid asset before the desired level of total investment has been made. Thus both $a^*$ and $h^*$ will be positive if:

$$R = \left(\frac{1}{c}\right) \mathbb{E}\left[\frac{\partial F}{\partial H}(\alpha_1, H_0 + h^*)\right] \tag{2.7}$$

and

$$u'(F(\alpha_0, H_0) - ch^*) < \beta R\, \mathbb{E}[u'(F(\alpha_1, H_0 + h^*))] \tag{2.8}$$

### 2.3.2 Productivity and Land Clearing

While simple, the above model does provide some insight into the relationship between weather during the current growing season (productivity) and the choice of land clearing for the future. The model provides justification regarding why we might expect current weather to increase future land in agriculture even when landowners have access to alternative savings technology. This mechanism does not rely on assumptions of landowner's perceptions of serial correlation in weather patterns.[1] An important takeaway from this

[1]These beliefs could be modeled using this framework and an assumption of how $\mathbb{E}[\alpha_1]$ differs from $\mathbb{E}[\alpha_1|\alpha_0]$ from the perspective of the landowner.

model is that under certain conditions, $\frac{\partial h^*}{\partial \alpha_0} > 0$ and under other conditions, $\frac{\partial h^*}{\partial \alpha_0} = 0$.

Our model suggests positive growing season weather shocks should have no impact on agricultural expansion if the expected marginal return to land in agriculture is sufficiently below the return on investment for savings (or other liquid assets available to landowners) - which corresponds to case 2 above. As for case 3 above, our model suggests that $\frac{\partial h^*}{\partial \alpha_0} = 0$ if landowners are already choosing to both clear land and save, because the expected marginal return to land clearing is already driven down to the rate of return on the liquid asset and thus additional income in year 0 will be passed forward via the liquid asset.[2] Our model suggests a positive relationship between weather shocks and land clearing is possible only for cases 1 and 4 above. Case 4 is clear. The expected marginal returns to clearing are above the rate of return on the liquid asset and additional production in year 0 will lead to additional land clearing, with no liquid asset investment. In case 1, it must be that the expected return to clearing land at $H_0$ is greater than the returns to saving and an increase in $\alpha_0$ pushes marginal utility of consumption at $F(\alpha_0, H_0)$ beyond the expected discounted marginal utility of consumption at $F(\alpha_1, H_0)$, thus an increase in $\alpha_0$ may move the landowner from case 1 to case 3 or 4. In either of these settings, an increase in productivity from growing season weather would lead to an increase in cleared land for agriculture. It is this finding which motivates the empirical work that follows.

[2] We acknowledge that allowing landowner beliefs about $\alpha_1$ to be shaped by realizations of $\alpha_0$ could change this relationship. If landowners believe weather to be serially-correlated, an increase in $\alpha_0$ will increase the expected return to clearing land (which is a function of $\mathbb{E}[\alpha_1]$) in addition to increasing output in year 0. Unfortunately, without data on savings by landowners, we cannot empirically test this mechanism.

## 2.4 Empirical Analysis of Growing Season Weather, Soy Yields and Deforestation in the Brazilian Amazon

The model above predicts that weather could influence deforestation in the Brazilian Amazon through its impacts on soy yields and planted area. Elevated yields from beneficial growing season temperatures could lead landowners to increase the area dedicated to growing soy for the following season. This may increase deforestation either directly by clearing land to plant soy or indirectly through moving into already cleared forest. To investigate this hypothesis, we combine several datasets on weather, deforestation, and agriculture in Brazil. By leveraging plausibly exogenous within-municipality, year-to-year variation in growing season temperatures we can estimate the impact of temperature yields and how this might contribute to deforestation pressures.

### 2.4.1 Data

We combine three unique data sources to create a municipal-level panel of weather, agriculture outcomes, and deforestation rates across the Brazilian Amazon from 2001 - 2012.[3] First, we utilize a high-resolution gridded weather data published by the Climatic Research Unit at the University of East Anglia (Harris et al., 2014) to generate municipal-level weather observations. Second, we use data from yearly agricultural surveys from the Brazilian Institute of Geography and Statistics (IBGE) for municipal-level soy data. Finally, we use remotely sensed satellite data from Hansen et al. (2013) for our deforestation observations.

[3]The municipalities of Brazil are the second level administrative subdivisions below states and are analogous to the US county level.

### Weather Data

Obtaining high quality weather data across the Brazilian Amazon presents some challenges due to the limited number of weather stations. Given the remote nature of regions that are heavily forested, this requires high levels of interpolation to generate a daily gridded panel of observations. Therefore, in lieu of processing our own data, we utilize version 3.23 of the high-resolution global gridded data published by the Climatic Research Unit (CRU) at the University of East Anglia (Harris et al., 2014). Although this data set also relies on weather station data, results are published at the *monthly* level, which alleviates much of the concern about data quality as monthly estimates are easier to accurately measure than daily. However, using monthly data presents a tradeoff. Even though this data provides consistency at a very high spatial resolution and quality, the monthly resolution limits much of the daily variation in observations and do not allow use of the common "bin" analysis in the climate-economy literature. We argue that it is better to have higher quality data overall than a higher frequency of low quality data since the types of errors often associated with interpolation from sparse stations can systematically bias impact estimates (Auffhammer et al., 2013b).

The CRU weather data consists of monthly total precipitation and minimum, maximum, and mean temperature at a gridded 0.5° resolution (approximately 55.5 km). Mapping this grid to the relevant administrative unit requires some form of overlapping and municipal shapefiles were overlapped on the weather grid and the share of the relevant municipal in each grid cell was then calculated. These area weights were then used to match the gridded data to the municipal level by averaging the observations across each municipal.

Merging monthly weather observations to yearly deforestation also presents different specifications to consider. Given we are interested how weather affects deforestation through soy yields, we chose to focus our analysis on the primary soy growing months in Brazil, which span approximately from planting in November to harvest in April. To do so, we first calculate the average temperature and precipitation for every municipality by month and year. Then to calculate the average growing season temperature we calculate the mean temperature across months November through April for each municipality. For precipitation, we take the total sum of precipitation across months for each municipality. These two measures give us our primary units of analysis and we consider higher order specifications to account for any non-linearities. Summary statistics are presented in Table 2.1.

### Agriculture Data

The Brazilian Institute of Geography and Statistics publishes estimates of municipal-level agriculture activity through their annual Municipal Agricultural Production (MAP) survey (IBGE, 2014). This high level of disaggregation is rare for a developing country and we take full advantage of observations at the municipal level by matching agriculture outcomes to relevant weather and deforestation data. The surveys provide annual estimates of planted area, harvested area, total harvested quantity and total market value for soy along with a wide variety of other crops.[4] Although other crops are included in our dataset, we choose to focus our analysis on soy, as it has shown to be the primary crop responsible for deforestation in the Brazilian Amazon (Morton et al., 2006; Barona et al., 2010; Arima et al., 2011; Macedo et al., 2012).

In addition to crops, IBGE also provides annual estimates on the number of cattle in each municipal in each year. Along with soy, forest conversion to pasture is a leading driver of deforestation. Unfortunately, heads of cattle are not sufficient to allow cattle

[4]These crops include: pineapple, cotton, garlic, peanuts, rice, sweet potato, english potato, sugar cane, onion, rye barley, peas, fava beans, tobacco, sunflower, jute, flax, malva, castor, cassava, watermelon, melon, corn, rami, sorghum. tomatoes, wheat, and triticale.

to be considered in our analysis as we have no way to measure cattle/pasture density without estimates of cleared land dedicated to pasture. Without being able to determine how weather impacts cattle productivity and any subsequent changes in pasture area, we cannot make use of the heads of cattle data.

The MAP data are estimated from surveys of farmers, exporters and input providers and meant to provide annual estimates of the production from full-time, professional farmers. For municipality-crop-years in which there is less than 1 hectare planted or harvested or less than 1 ton produced, no data are recorded. While these estimates may be more prone to measurement error than the agricultural census taken every ten years, they offer the best annual production data at the municipality-level for Brazil, which is necessary to study year-to-year variation in productivity driven by annual variation in growing-season weather. As long as the potential measurement error in the MAP is uncorrelated with the annual growing-season variation, this should increase the standard errors of our estimates, but not induce systematic bias that would undermine our identification.[5]

Figure 2.1 displays the geographic distribution of forest, soy, and cattle in the Brazilian Amazon. This figure reveals that cattle production is the much more dominant agriculture activity in the Brazilian Amazon, with soy being more localized. Additionally, we see how soy is concentrated in areas of lower forest density while cattle encroaches on denser forest. This is consistent with the theory that soy expansion creates indirect land use pressures on forest through displaced pasture.

[5]Classical measurement error using MAP estimates as explanatory variables would lead to attenuation bias, but we strictly use these measures as dependent variables. We use predicted variation in yields rather than yields themselves in our second-stage regressions. Under the assumption of classical measurement error, this should increase our standard errors undermining the consistency of our estimates.

### Deforestation Data

The proliferation of remote-sensed satellite data in recent years has vastly improved deforestation data availability and quality. This chapter utilizes a recent dataset from researchers at the University of Maryland who use satellite images to create a measure of global forest coverage, losses, and gains (Hansen et al., 2013).[6] Specifically, forest coverage estimates are derived from a 30-meter spatial resolution grid of Landsat images from 2000, where forest cover is defined as an area covered by vegetation greater than 5 meters high. While Hansen et al. (2013) count any cell with forest cover to be classified as forest, we restrict our measure of forest to only include cells with at least 30 percent of forest coverage to reduce noise. In other words, our classification of baseline forest is for any cell in 2000 with at least 30 percent of area covered by vegetation of 5 meters or more.

In addition to baseline tree cover, Hansen et al. (2013) provide indicators for changes in forest coverage from 2001 - 2012. These indicators serve as our primary measure of deforestation, as we count any cell as deforested if it transitions from a baseline forested cell in 2000 to zero forest coverage in any subsequent year and deforestation rates are visually presented in Figure 2.2. For example, if a baseline forest cell in 2000 loses forest coverage in subsequent years, it will only count as a deforested cell when the total vegetation drops to zero coverage in a given year. Therefore, if a cell has 50 percent forest coverage in 2000, and drops to 30 percent in 2004 and 0 percent in 2005, we count the cell as deforested only in 2005 despite the fact there was forest loss in the previous year. Once a cell is counted as deforested, it will never be counted again as the baseline forest measure is never updated and thus reforestation is not observed in the data set. In total,

[6] At the time of this chapter, version 1.1 of this data is available, increasing the time covered to 2014. This data is not used, however because deforestation estimates in years 2011 - 2014 are not comparable to earlier years in the sample.

our deforestation measure sums the number of baseline forest cells the reach zero forest coverage across each municipality in any given year.[7] This measure of deforestation is not perfect, but it represents the highest quality data available. Accurate deforestation data is notoriously difficult to obtain due to the often-illegal nature of removing forest cover in remote areas. Although we would welcome further descriptive statistics on who precisely is responsible for deforestation, there simply is not data available.

**Other Data**

Our analysis chooses not to include other variables that likely affect deforestation due to data quality concerns and methodological constraints. For example, soy, cattle, and timber prices undoubtedly affect deforestation rates yet we intentionally elect not to include them in our analysis. District level prices are not available across our sample, and the closes measurement in our dataset is soy value.[8] Using both soy value and quantity we could create a back of the envelope calculation or farm-gate prices at the district level. However, including prices will not strengthen our analysis as the primary intention of our work is to document the effect of temperature on deforestation. In the best-case scenario, controlling for prices would create a more accurate measure of the size of the effect, but will have no impact on the identification of an effect in the first place. Moreover, including prices has the potential to introduce bias in the coefficients as prices are endogenously determined and very likely affected by weather variation. This concern is of particular note given that our identification strategy relies on the exogenous nature

[7] It should be noted that this definition of deforestation is not universally accepted. Tropek et al. (2014) point out that classifying forest as vegetation taller than 5 meters can lead to classification of different plantations as forest. Harvesting of these plantations may result in observed deforestation when, in reality, the land had been cleared of native vegetation and planted prior to the beginning of the study period.

[8] While global commodity prices for soy or beef also present an option, these would not reflect the localized weather effects. Furthermore, year fixed effects do offer a control for global prices that vary across years but are constant across municipalities.

of weather outcomes. Given that weather can affect so many other variables typically included as controls (prices, income, population via migration or mortality, etc.) we restrict our analysis to the variables that are relevant to our analysis.

There are other variables that we would like to include that are likely not correlated with weather, and impact deforestation, but once again data availability is an issue. For example, institutional conditions that affect deforestation such as road access or land tenure would strengthen our analysis, but data is not available. A high-quality panel data set on roads would be a welcome addition, but given the informal nature of many logging roads, no dataset was identified. In regard to land tenure, there is no dataset that identifies if land is publicly or privately owned. However, this of limited concern for our analysis as the tenure setting across the Amazon is the largely the same for all agents. Given all agents face the same constraint to land conversion, there is no different governance or enforcement structures to consider.

### 2.4.2 Empirical Methodology

Our central hypothesis is that elevated temperatures increase deforestation in the Brazilian Amazon through the impact on soy productivity. Specifically, we propose that if a weather shock increases yields in the current period, it can lead to an increase in planted area in the following period. This increase in planted area can subsequently increases deforestation contemporaneously (via direct conversion) or over a several year period (via displaced pasture). Taken together, we predict that a favorable weather shock in the current period would increase deforestation rates in the following period via direct conversion or over the following several periods from indirect land use pressures.

This distinction across periods requires explicit consideration given that our data is observed in yearly intervals. Suppose that a weather shock in period $t$ increases soy yields. Our model suggests that soy planted area would increase in the following period, $t+1$. This increase in planted area can then affect deforestation in one of two ways. First, it can move directly into existing forest and this effect should occur contemporaneously. This suggests that a weather shock in period $t$ would increase deforestation in period $t+1$. Second, increased planted area can move into existing pasture, which displaces cattle to forest frontiers. Although cattle displacement might immediately increase deforestation, we propose that additional periods might be necessary for cattle farmers to relocate animals and begin the process of converting forest to pasture. Previous research suggests that between 72%-86% of pasture clearings were identified in the year after deforestation while the remaining pasture clearings needed 2 years or more to develop a grass phenology (Morton et al., 2006). We therefore expect that a weather shock in period $t$ might not cause increases deforestation until period $t + 2$ or even $t + 3$. The timing of theses mechanisms is presented visually in Figure 2.3. Based on these findings we propose that any model considering the impact between weather and deforestation must account for at least three years of lagged weather observations.

To empirically model this approach, we begin with the relationship between growing season weather and soy yields. With this model in place, we then consider how yields relate to planted area and how planted area affects deforestation. Finally, we consider how weather directly impacts deforestation using the identical temperature specification used in the yield model. We propose that elevated temperatures increase yields, in turn increasing planted area, which increase land use pressures, ultimately increasing deforestation. The methodology for each approach is listed below in detail.

### Yield Model, Planted Area, and Deforestation

The first step in our empirical strategy is establishing a yield model that captures the relationship between temperature and soy yields. Fortunately, there has been considerable work in this field, and we build on the seminal work of Schlenker and Roberts (2009), who document the importance of nonlinear effects between weather and yields. Rather than model a fully flexible weather-yield specification, we follow the findings of Lobell et al. (2011), who discover that a quadratic maximum temperature specification approximates the more flexible growing degree-day specification for maize yields in Africa. This finding is important as we do not have daily temperature observations allowing use of a growing degree-day measure. Furthermore, we are ultimately interested in how weather impacts deforestation, and therefore argue a simpler model is adequate to establish that weather impacts yields. Taking the previous literature into account, our model reflects the non-linear relationship between maximum temperature and yields and only considers maximum temperature over the primary soybean growing season (running from November to April). We also consider how periods of relative dryness or wetness impact yields.

We also consider the relationship between yields and planted area, and then planted area and deforestation. Here we attempt to simply demonstrate that yields affect planted area, and in turn planted area affects deforestation. Formally, our empirical strategy is listed below:

$$
\begin{aligned}
YIELD_{it} = {} & \beta_1 TMAX\_GROW_{it} + \beta_2 TMAX\_GROW_{it}^2 \\
& + \beta_3 PREC\_GROW_{it} \{\text{in lowest tercile } i\} \\
& + \beta_4 PREC\_GROW_{it} \{\text{in highest tercile } i\} \\
& + \alpha_i + \gamma_t + \lambda_s t + \varepsilon_{it} \qquad (2.9)
\end{aligned}
$$

$$
PLANT_{it} = \sum_{s=1}^{3} \theta_s YIELD_{i,t-s} + \alpha_i + \gamma_t + \lambda_s t + \varepsilon_{it} \qquad (2.10)
$$

$$
DEFOR_{it} = \sum_{u=0}^{2} \eta_u PLANT_{i,t-u} + \alpha_i + \gamma_t + \lambda_s t + \varepsilon_{it} \qquad (2.11)
$$

In equation (2.9), $YIELD_{it}$ represents the natural logarithmic average soy yield in municipality $i$ in year $t$, $TMAX\_GROW_{it}$ represents average growing season maximum temperature. We choose not to model precipitation in average terms as we do for temperature given that rainfall is easier to store and extreme events such as droughts or floods are of more interest than averages. Therefore, we place each administrative unit-year rainfall observation into an upper, middle, or lower tercile that is determined from the municipal-average across the entire sample. We omit the middle tercile in our specification and therefore $PREC\_GROW_{it}\,\{\text{in lowest tercile } i\}$ and $PREC\_GROW_{it}\,\{\text{in highest tercile } i\}$ represents the effect that rainfall in each tercile has on yields. This modeling approach appears justified as responses to temperature appear to be the more significant driving factor in determining yields (see results for further discussion).

We also include year and municipal fixed effects well as a state linear time trend depending on the specification. The year fixed effects, $\gamma_t$, absorb any time-varying differences in the dependent variable that are common across all municipalities (such as improved agriculture technology over the entire sample period) while municipal fixed effects, $\alpha_i$, control for unobserved municipal heterogeneity that is fixed over time. The state linear time trend, $\lambda_s t$, controls for specific time trends that are linear across states Finally, $\varepsilon_{it}$ is the stochastic error term. We follow the climate-economy literature by choosing to cluster standard errors at the municipal level. This is in response to the assumption that the error terms are likely correlated across spatial areas over time. We also weight our regressions by the amount of soy grown in each district.

For equation (2.10), $PLANT_{it}$ represents the observed natural logarithmic average of soy planted area. Our model suggests that an exogenous shock to yields can affect planted area decisions in the following year and to be consistent with the temperature-deforestation analysis in the following section we include two additional lags. In equation 2.11, $DEFOR_{it}$ represents the natural logarithmic of deforested cells in municipality $i$ in year $t$. Given that planted area could affect deforestation directly or through indirect channels, we model both the contemporaneous and lagged relationship. Here we only include two lags reflecting the fact that planted area decisions are based on the yields from the previous year. This allows internal consistency across specifications.

### Temperature and Deforestation

The final piece of our empirical methodology directly considers how weather relates to deforestation. We hypothesize that elevated temperatures affect soy yields, which in turn impact planted area, ultimately affecting deforestation rates through a combination of direct and indirect land use pressures. As previous work in the region has demonstrated, the expansion of soy can move directly into forest, but more commonly moves into previously cleared pasture land, displacing pasture to the forest frontiers. In order to test this hypothesis, we consider the relationship between weather and deforestation using the identical weather variables that are used in the weather-yield model.

The primary empirical specification is listed below:

$$
\begin{aligned}
DEFOR_{it} = & \sum_{s=0}^{3} \beta_s TMAX\_GROW_{i,t-s} + \sum_{v=0}^{3} \theta_v TMAX\_GROW^2_{i,t-v} \\
& + \sum_{u=0}^{3} \eta_u PREC\_GROW_{i,t-u} \{\text{in lowest tercile } i\} \\
& + \sum_{z=0}^{3} \phi_z PREC\_GROW_{i,t-z} \{\text{in highest tercile } i\} + \alpha_i + \gamma_t + \lambda_s t + \varepsilon_{it}
\end{aligned}
\tag{2.12}
$$

In equation (2.12), $DEFOR_{it}$ represents the natural logarithmic of deforested cells in municipality $i$ in year $t$ and $TMAX\_GROW_{it}$ represents average soy growing season maximum temperature as used in equation (2.9). We model precipitation in the identical way as the weather-yield model in equation (2.9), and therefore $PREC\_GROW_{it}$ {in lowest tercile $i$} and $PREC\_GROW_{it}$ {in highest tercile $i$} represents indicators for rainfall in a comparatively dry and wet year respectively. Once again, we include year and municipal fixed effects, a linear state time trend, and cluster standard errors at the municipal level. We also weight our regressions by the amount of soy grown in each district. In regard to lagged weather variables, we present three years of lagged weather based on the timing considerations discussed above. Given growing season weather will not impact land use decisions until the following period we do not consider the contemporaneous relationship between weather and deforestation.

## 2.5 Results

Our results support the hypothesis that weather shocks effect deforestation through their impact on agriculture productivity. We find that elevated temperatures increase yields in the Brazilian Amazon and that increases in yields are positively associated with increases in planted area in the following year. Furthermore, we find evidence that increases in planted area are positively associated with deforestation in the following year. Taken together these findings suggest that elevated temperatures increase yields, which in turn increase planted area, ultimately increasing deforestation. When we directly consider how temperature affects deforestation we find support of this hypothesis. Using the same temperature specification for deforestation as our temperature-yield model we find no effect in the following year but do find evidence of a positive effect with a two-year lag, which is consistent with the previous findings. A detailed discussion of the results follows.

### 2.5.1 Yields, Planted Area, and Deforestation

Starting first with the yield model, the results of equation (2.9) are listed in Table 2.2 below and column (2) presents our preferred specification. Our results suggest that elevated maximum growing season temperatures largely increase soy yields in the Brazilian Amazon. Specifically, we find that yields decrease with elevated temperatures until a temperature threshold after which temperature becomes beneficial. However, for each specification the temperature threshold falls on the lower end of our observed temperature distribution, and thus across the relevant range of temperatures we primarily find increasing returns. This finding is in contrast to the well-known results of Schlenker and Roberts (2009), who find that yields have decreasing returns to temperature before a threshold in which temperature ultimately becomes damaging. We expect that temperatures would ultimately become damaging to yields, but do not see this effect across our observed temperature distribution or even across a variety of more flexible temperature specifications.[9] It therefore appears likely that there is a unique relationship between

[9] Figure B.1 in the Annex presents the marginal effects of higher order specifications and we find a persistent effect for increasing returns.

temperature and soy yields in the Amazon that is distinct from the rest of the country. This is further illustrated by the fact that when we consider soy grown outside of the Amazon, we find a similar relationship as Schlenker and Roberts (2009).[10]

For precipitation, we find some support that dry years lead to a reduction in yields and stronger evidence that wetter years increase yields. In order to better determine how much of the explanatory power comes from precipitation versus temperature, we consider a variety of different precipitation specifications to see if the underlying temperature relationship changes. Additional specifications suggest that much of the explanatory power of our yield model is driven by temperature rather than precipitation as the temperature effect is persistent while precipitation results vary. For example, when we consider a quadratic precipitation specification as done by Schlenker and Roberts (2009), we find a similar temperature effect but no significance for precipitation.[11]

After establishing that elevated temperatures increase yields, we next consider how yields affect planted area. Our model suggests that increases in yields from weather shocks in the current period would increase planted area in the following period. We find support of this theory in Table 2.3. Our results suggest that a one percent increase in yields increase planted area in the following year by approximately 0.37 - 0.41%. Although our model makes no claims about how additional lags should impact planted area, we find support that elevated yields from two and three years before may also drive increases in planted area. We are not the first to document this finding in the Brazilian Amazon (see Garrett et al., 2013 for example), but our analysis supports the hypothesis that a productivity shock can increase planted area.

[10]Results for the relationship between growing season temperature and soy outside of the Amazon are listed in Table B.1 in the Annex. Here we find the similar relationship to Schlenker and Roberts (2009) demonstrating decreasing returns.

[11]Table B.2 in the Annex presents results with a quadratic growing season precipitation measure. Additionally, Figure B.1 in the Annex considers higher order precipitation specifications and we see that the temperature effect remains persistent.

Moving next to the relationship between deforestation and planted area, we find that increases in planted area increase deforestation in the following year. Results for equation (2.11) are listed in Table 2.4. Here we find that a one percent increase in planted area increases deforestation the following year by approximately 0.1%. We find no evidence that planted area affects deforestation contemporaneously or in the following two years. These results are indicative of soy creating indirect land use pressures on forest frontiers. The fact that planted area only increases deforestation in the following year is suggestive of mechanism where soy moves into previously cleared forest for pasture, displacing pasture to forest frontiers. Since we do not find a contemporaneous effect, there is less evidence that soy production is moving directly into existing forest over our sample. This finding is supported across the Brazilian deforestation literature, which suggests the majority of deforestation was for cattle pasture, while soy typically moves into existing pasture Morton et al. (2006); Macedo et al. (2012).

### 2.5.2 Temperature and Deforestation

In the previous section, we find that elevated temperatures increase yields and increased yields are positively associated with planted area in the following year. We also find that increased planted area is positively associated with deforestation in the following year. This suggests that a positive weather shock can increase yields, increasing planted area, and ultimately act as a driver of deforestation. In this section, we explicitly consider if temperature is a direct driver of deforestation. We find support for this hypothesis in Table 2.5, which presents results from equation (2.12). Given that temperature drives much of the effect for yields, we only present temperature results for brevity. What is immediately striking about these results is we find a persistent effect only for two-year lagged temperatures. Specifically, we find that elevated two-year lagged temperature increases deforestation until approximately 89° F, after which elevated temperatures decrease deforestation. These decreasing returns to two-year lagged temperature are persistent across different precipitation specifications as well as different lag structures.

The evidence of a two-year lag for growing season temperature driving deforestation has significant implications in regard to our proposed mechanism. Our model predicts that a productivity shock from weather would increase deforestation through changes in planted area. However, we hypothesize that any effect on deforestation would not be immediate as a productivity shock from weather would not impact planted area decisions until the following year after yields are realized. Although we might expect to find some effect in the following year, the Brazilian deforestation literature suggests that increased soy production is more likely to move into existing pasture, and displaced pasture then acts as a direct driver of deforestation rather than soy (Barona et al., 2010; Arima et al., 2011; Macedo et al., 2012). Indeed, our own analysis appears to support the literature as the only effect we find is for a two-year lag. The previous section supports this finding as well. We find that soy yields are positively associated with planted area in the following year, and planted area is positively associated with deforestation in the following year. Therefore, a favorable weather shock in the current year would impact deforestation after two years, which is what this section supports. Furthermore, given that the relationship between soy and deforestation is a complicated, multifaceted process, this two-year timeline seems an appropriate amount of time for a productivity shock from temperature to impact indirect land use decisions and ultimately deforestation. Overall, these results offer new insight into the relationship between soy productivity and deforestation and highlight how indirect land use pressures can drive deforestation.

The fact that we find decreasing returns to temperature for deforestation while we find increasing returns to yields does present a bit of a paradox. In theory, we would expect that these two effects should have the same relationship as we believe that changes in yields are driving deforestation. That being said, we do not believe the increasing returns to temperature relationship for yields is entirely accurate. Despite considering a variety of different specifications, our results suggest that higher temperatures will always increase yields. Clearly this cannot happen as there is some physiological threshold for soy in which temperature becomes damaging, but we did not find it across our observed temperature distribution. It is possible that this threshold falls outside of our distribution or we need a larger sample to document this effect. Thus, it is encouraging that we find the expected effect of decreasing returns to temperature for deforestation, as we believe this more accurately reflects the fact that yields (and land use pressures) will decrease after some temperature threshold.

## 2.6 Conclusion

Our chapter is the first to document that weather is a previously overlooked driver of tropical deforestation. As satellite images have drastically improved data quality, a growing body of work has identified key drivers, yet weather remains an undiscussed factor. We believe this to be a major shortcoming, especially since land use conversion to agriculture is often cited as one of the primary drivers of tropical deforestation. Given that weather has been shown to be an important factor into agriculture yields, and agriculture yields impact planted area and land use decisions, this lack of discussion of the weather represents an important gap in the literature.

We find that elevated temperatures during the soy growing season increase yields, and increased yields are positively associated with planted area in the following year. Furthermore, we find that increased planted area is positively associated with deforestation in the following year. Taken together this would suggest that a positive weather shock should increase deforestation after two years. To test this hypothesis, we consider how weather during the soy growing season directly affects deforestation. Using the identical weather specification as our temperature-yield model we find that elevated temperatures increase deforestation with a two-year lag. The fact we only find an effect with a two-year lag is consistent with our preceding results and is suggestive of a mechanism in which temperature increases deforestation through indirect land use pressures. Although we are unable to directly test this, the deforestation literature strongly supports this claim. These indirect land use implications call into question the efficacy of the 2006 'Soy Moratorium,' which excludes soy cultivated from deforested areas from the supply chains of the leading exporters. This policy is believed to have an effect on limiting soy's contribution to deforestation, but it fails to address the more complicated land use linkages.

Although deforestation rates in Brazil have fallen dramatically in recent years, the pressures surrounding land use decisions have not been eliminated. Policies and falling soy prices have no doubt helped to reduce pressures, but as land that is suitable for soy production becomes scarce, pressures will only increase. Worryingly, the significant progress Brazil has made appears to have taken a major step back as deforestation rates have once again started to increase. Recent data shows that deforestation has increased 29% this year, following a 24% increase the previous year and the current rate is the highest since 2008 (NPR, 2016). As temperatures continue to climb, and land becomes even more scarce, these issues will only intensify.

Our findings highlight the need for further research in this field. Specifically, a closer look at the relationship between cattle physiology and pasture land would be welcome. Temperature presumably impacts cattle physiology given the significant evidence linking temperature and human health (Deschenes, 2014). Much as soy yields are related to planted area, we assume cattle physiology is also related to pastureland, as cattle growth is plausibly linked to the amount of land needed. Thus, an extension of our analysis would consider the relationship between temperature, cattle physiology, and pastureland. We expect that elevated temperatures are harmful to cattle physiology, reducing the amount of pastureland needed, decreasing deforestation pressures. Presumably there is a temperature threshold that exists that is both damaging to cattle production and soy yields, which would explain why deforestation rates fall at extreme temperatures.

Table 2.1: Summary Statistics of Weather

| Sample | Mean Temp (F) | Min Temp (F) | Max Temp (F) | Precip (mm) |
|---|---|---|---|---|
| Brazil (Total) | 74.45 | 65.01 | 83.98 | 1429.84 |
| | (5.12) | (5.29) | (5.26) | (525.88) |
| Brazilian Amazon | 80.29 | 70.71 | 89.95 | 1950.97 |
| | (1.40) | (2.37) | (1.41) | (507.11) |
| Amazon Growing Season | 80.21 | 71.78 | 88.72 | 1476.52 |
| | (1.20) | (1.72) | (1.18) | (330.42) |

Note: Standard Deviations are in Parenthesis.

Figure 2.1: Brazil Area Maps

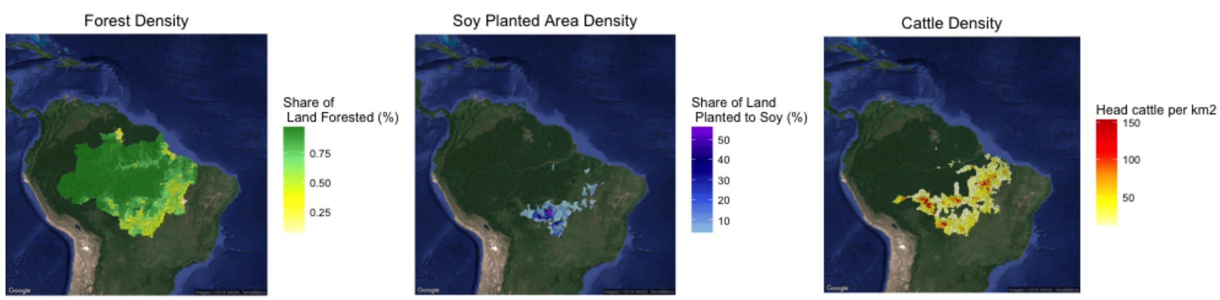



Figure 2.2: Average Conditions in Study Areas

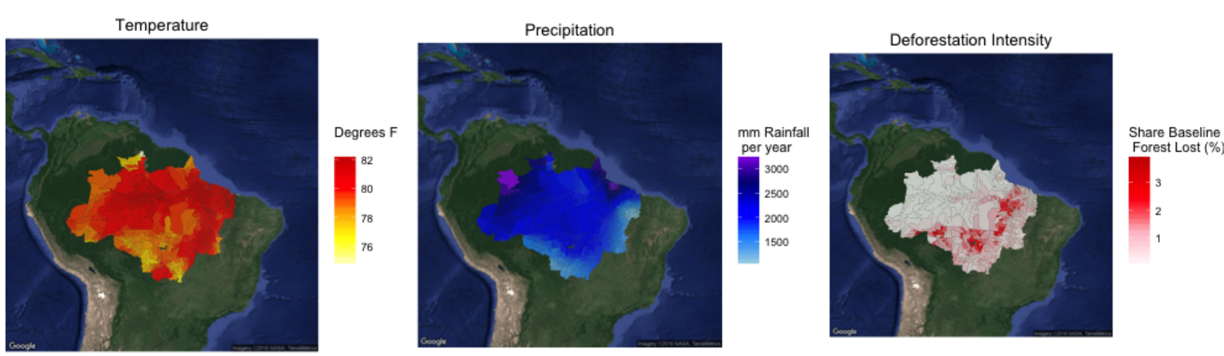


Figure 2.3: Timing of Mechanisms

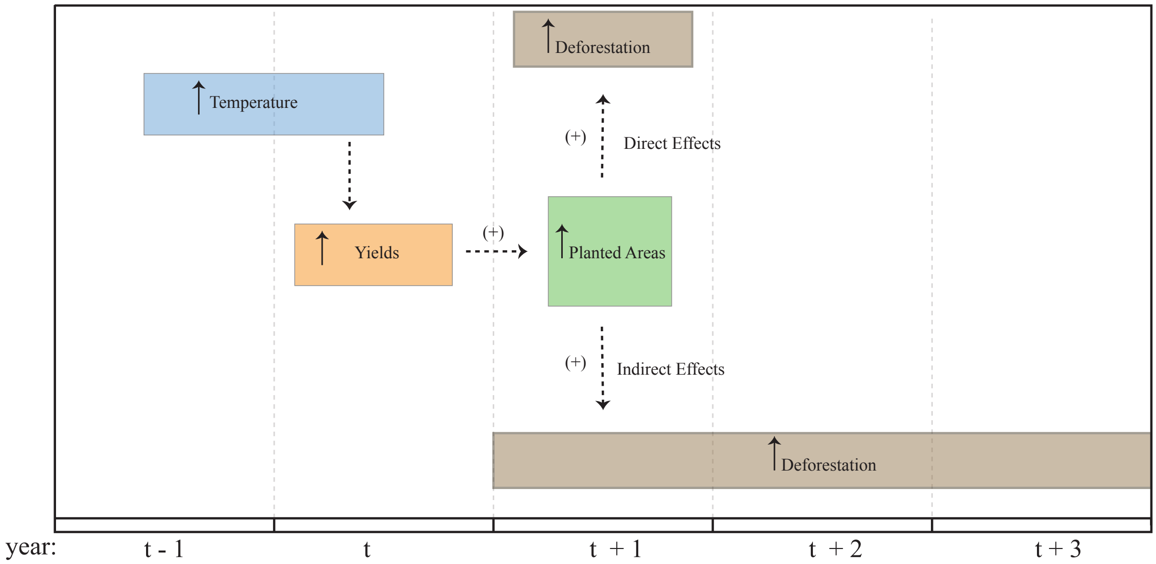



Table 2.2: Contemporaneous Effects of Growing Season (Nov - April) Temperature on Soy Yields in Brazilian Amazon

| | (1) Log Yield | (2) Log Yield |
|---|---|---|
| Average Maximum Temperature | -0.654 | -0.932** |
| | (0.412) | (0.370) |
| Average Max Temperature Squared | 0.00387* | 0.00542** |
| | (0.00233) | (0.00209) |
| Indicator for Rainfall Shock in Lowest Tercile | -0.0178*** | -0.0127* |
| | (0.00672) | (0.00646) |
| Indicator for Rainfall Shock in Highest Tercile | 0.00842 | 0.0115** |
| | (0.00563) | (0.00528) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 2791 | 2791 |
| $R^2$ | 0.525 | 0.559 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.3: Lagged Effect of Actual Soy Yields on Planted Area in the Brazilian Amazon

| | (1) Log Planted Area | (2) Log Planted Area |
|---|---|---|
| 1 lag Log Yield | 0.371*** | 0.411*** |
| | (0.123) | (0.125) |
| 2 lag Log Yield | 0.255** | 0.300*** |
| | (0.106) | (0.111) |
| 3 lag Log Yield | 0.523*** | 0.507*** |
| | (0.124) | (0.132) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 1810 | 1810 |
| $R^2$ | 0.963 | 0.964 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.4: Lagged Effect of Actual Soy Yields on Planted Area in the Brazilian Amazon

| | (1) Log Deforestation | (2) Log Deforestation |
|---|---|---|
| Log Planted Area | -0.0257 | -0.0285 |
| | (0.0908) | (0.0886) |
| 1 lag Log Planted Area | 0.122** | 0.110** |
| | (0.0563) | (0.0552) |
| 2 lag Log Planted Area | 0.0510 | 0.0160 |
| | (0.0617) | (0.0611) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 1875 | 1875 |
| $R^2$ | 0.872 | 0.882 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.5: Impact of Lagged Growing Season (Nov - April) Temperature on Deforestation

| | (1) Log Deforestation | (2) Log Deforestation |
|---|---|---|
| Average Max Temp Last Year | 2.283 | -0.0293 |
| | (4.318) | (4.257) |
| Average Max Temp Squared Last Year | -0.0116 | 0.00103 |
| | (0.0243) | (0.0239) |
| Average Max Temp 2 Year Ago | 16.88*** | 13.94*** |
| | (3.958) | (3.511) |
| Average Max Temp Squared 2 Year Ago | -0.0934*** | -0.0781*** |
| | (0.0223) | (0.0198) |
| Average Max Temp 3 Year Ago | 1.731 | 0.0674 |
| | (4.861) | (5.116) |
| Average Max Temp Squared 3 Year Ago | -0.00892 | -0.00108 |
| | (0.0273) | (0.0287) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 3828 | 3828 |
| $R^2$ | 0.872 | 0.881 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# Chapter 3

# Temperature, Human Health, and Adaptation in Thailand

## Attributions

The content of chapter 3 is the result of a collaboration with Sam Heft-Neal. Sam Heft-Neal is a Postdoctoral Fellow in the Department of Earth System Science at Stanford University.

## 3.1 Introduction

Across studies the evidence that temperature extremes lead to significant reductions in health is consistent (Deschênes, 2014). However, virtually all the previous research has been discussed in the context of the United States. Therefore, a logical extension of this body of work is to consider how temperature affects mortality rates in developing countries. This is not simply an extension of an existing body of literature, but instead the developing country context fundamentally changes the research question. First and foremost, developing country economies are more weather dependent in the sense they are comprised of large, rural populations engaged in agriculture production. Whereas, labor and livelihoods are protected from most weather shocks in a developed country, the opposite is true in developing countries. A subsistence farmer must tend his or her fields regardless of weather outcomes. Second, there are less adaptation options available in developing countries. In the United States, both the ability to move within a country and the widespread use of air conditioning (AC) appear to limit excess mortality from extreme temperature events (Deschênes and Moretti, 2009; Deschênes and Greenstone, 2011). Once again, in the developing country context these adaptation strategies are limited. Although AC penetration rates are increasing throughout the developing world, they remain well below those found in a developed country setting. Furthermore, migrants face more constraints in a developing country setting as the ability to move is more likely to be tied to employment outcomes (rather than quality of life factors).

This chapter attempts to fill this gap in the literature by considering the relationship between temperature, mortality, and adaptation in Thailand. To do this, nearly 20 years of annual observations of mortality and residential energy use are linked to daily district level rainfall and temperature records. Our findings however are inconsistent with the literature. We find that temperature extremes lead to *decreased* mortality, which is the opposite effect that has been demonstrated in the literature. Furthermore, the effect is persistent across a variety of specifications. Even more puzzling is the expected effect is found when the mortality data is matched to following year's weather data. This could indicate there is a timing issue in the reporting of the mortality data. A lack of other key data prevents us from exploring other possible factors. Regardless, our findings raise serious concerns over the quality of the mortality data.

We also consider the impacts of weather on residential electricity use. Originally the intent of this analysis was to determine if residential electricity use may serve as a potential adaptation strategy to extreme temperatures. However, the evidence for electricity use as an adaptation strategy rests on the findings that extreme temperatures lead to increased mortality rates, which we are unable to support. We do find that extreme temperatures lead to increases in residential electricity use, but this finding on its own offers little advancement in the literature. If anything, the expected results for electricity add support to the data quality concerns for mortality. As both findings rely on the same empirical methodology, the counterintuitive findings for mortality are likely best explained by poor data quality. Given these major shortcomings, further analysis has been suspended.

The rest of the chapter is organized as follows. Section 3.2 provides an overview and literature review that inspires the chapter. Section 3.3 provides information on the data used in our empirical analysis. Section 3.4 presents our empirical strategy. Section 3.5 presents our findings. Section 3.6 concludes.

## 3.2 Literature Review

Human health is related to climate and weather through numerous pathways. Direct effects from extreme weather events such as heat waves, droughts, or weather events such as tropical cyclones are often the most visible and discussed, but climate affects health through numerous indirect channels as well. First, climate alters natural systems that impact human health, such as disease vectors, water-borne diseases, and air pollution. Second, climate impacts other economic and social outcomes that provide direct inputs to human health such as food systems, occupational income, and mental stress (Smith et al., 2014)

Given the inexorable link between human health and climate, climate change presents a significant risk to human health and has been recognized as a global research priority. Much of the previous work has been focused in the fields of public health and epidemiology, and typically considers how morbidity and mortality are affected by extreme temperatures. Although these studies are useful in relating extreme temperature events to human health, causal inference is typically beyond their scope. As a result, economists have recently begun to use modern econometric techniques to consider the causal relationship between weather and a variety of health outcomes. This growing body of literature makes use of high frequency weather data and panel methodologies to exploit the plausibly random nature of weather occurrences to identify the effect of weather on health outcomes.

Typically, after an effect between weather and health outcomes is established, researchers then consider how changes in the climate will affect future outcomes. However, using a historical relationship to predict future outcomes is only part of the true affect, as the total impact of climate change must also consider possible adaptation measures. Adaptation in the context of climate change and human health is typically defined as the set of actions taken to reduce health impacts from exposure to extreme weather events or other changes in climate. Adaptation measures include both household-level, short-run decisions such as self-protection from air-conditioning usage, as well as long-term decisions like migration. Additionally, adaptation can refer to the community-level, long-term decisions such as urban planning and infrastructure decisions (Deschênes, 2014)

The most comprehensive review of the empirical relationship between health outcomes, weather, and adaptation comes from Deschênes (2014), who focuses on extreme temperature. Although other weather events such as drought or tropical cyclones impact human health, extreme temperature events have received the majority of research focus due to the frequency of events and concerns of heat-related mortality spikes from such events. Across a diverse range of studies, the author finds consistent evidence that temperature extremes lead to reductions in health, typically measured by excess mortality.

Regarding adaptation, the empirical evidence linking adaptation responses to weather is much more limited due to the more difficult research design and lack of credible data. Whereas measuring the response in health outcomes such as mortality and weather is relatively straightforward, adaptation responses are difficult to measure and long-term responses are best related to changes in long-term climate rather than short-term weather. Thus, many studies that consider adaptation typically do so where adaptation is used as an additional dependent variable to supplement models with health outcomes. Self-protection in the form of air-conditioning measured through residential energy consumption is the most common adaptation measure across the existing literature. Deschênes (2014) finds consistent evidence across studies that residential energy consumption increases significantly at temperature extremes. Furthermore, the effect seen for energy consumption is often larger than that seen for mortality, which is consistent with the hypothesis that residential energy consumption acts a form of self-protection from extreme weather events.

What is noticeably missing from the extensive Deschênes (2014) review article are studies that consider the relationship between human health, weather, and adaptation in a developing country context. This is not a criticism of the work, but rather a reflection of an overall larger gap across the literature. For example, of the five articles reviewed in economics journals, all are focused in the US. Of the fourteen articles from the public health and epidemiology literature, only two include data from developing countries. The lack of research in developing countries is not surprising given the extraordinary data challenges that exist, both in collecting high quality weather data and accurate mortality rates at a fine enough scale. Nonetheless, this gap presents an important research opportunity as the human health response to weather is likely different in a developing country context where individuals are often more reliant on the weather. Furthermore, adaptation options are often limited in a developing country, where self-protection costs can be out of reach, and migration options are limited to employment prospects.

There are some studies that use modern econometric techniques and consider the relationship between human health and weather in a developing country context. Papers by Compeán (2013) and Cohen and Dechezlepretre (2017) focus on Mexico and offer conflicting results. Compeán (2013) considers the mortality-weather response at the municipal level in Mexico and finds extreme heat increases the crude mortality rate while extreme cold has a limited effect. Conversely, Cohen and Dechezlepretre (2017) find that vulnerability to cold drives much of the effect, while heat has a more modest impact. Although the papers use a similar methodology, Cohen and Dechezlepretre (2017) use a higher quality dataset, which could explain these differences.

Despite these studies offering slightly conflicting findings, both show the mortality response to temperature is higher in Mexico than the US, demonstrating the importance of this question in a less developed country. A true comparison between the Compeán (2013) study and those in the US are difficult because of different reference categories, but the results suggest the most extreme hot days have a larger effect on mortality rates in Mexico than the US. Cohen and Dechezlepretre (2017) allow a direct comparison to the US as they include an identical model of the US work of Deschênes and Moretti (2009) and find a universally higher magnitude of effects in Mexico.

Both studies also consider how the response of mortality to temperature differs between urban and rural municipalities. This is of particular importance in developing countries as rural economies are among the most reliant on weather inputs and self-protection options are limited. Compeán (2013) finds a large differential effect between urban and rural areas, with temperature having virtually no effect on mortality in urban areas, suggesting that rural communities drive much of the effect seen across the overall mortality rate. However, Cohen and Dechezlepretre (2017) finds no noticeable difference between urban and rural areas. Cohen and Dechezlepretre (2017) also considers if income contributes to a heterogeneous impact of temperature on mortality, and finds that vulnerability to extreme weather is negatively correlated with income.

Turning to India, Burgess et al. (2017) considers the weather-mortality response and supports the results found in Compeán (2013). Much like Cohen and Dechezlepretre (2017), the authors compare the effect in India to the US using an identical panel and reference categories allowing a true comparison. Here they find that a single day above 95° F annually relative to a 70° - 74° F day increases the mortality rate in India by 0.74 percent, a factor that is nearly 25 times greater than the effect seen in the US. The Burgess et al. (2017) study also explores results across urban and rural populations and adds further support that rural populations drive much of the effect seen across the general population in developing countries. Again, much like Compeán (2013), the authors find that the mortality response to temperature is driven almost entirely by the rural population. In fact, urban India's mortality response is more closely related to that of the US, despite the prevalence of extreme poverty across urban India. Thus, the authors find that even in a developing country setting, living in urban areas offers a significant degree of protection against warm weather.

The key results of both the Compeán (2013) and Burgess et al. (2017) studies are the large divide between the mortality response in urban and rural locations in developing countries. Both papers consider the mechanism by which the rural mortality responses are more dependent on weather, but offer less of a discussion on the role of adaptation in urban settings. Instead, the focus of these papers is to understand the channels in which elevated temperatures increase the mortality rate in rural areas, as opposed to understanding why urban areas are better protected. Burgess et al. (2017) make a strong case that in rural areas weather shocks to agriculture drive the strong differential impacts. They find that hot weather decreases both agriculture yields and wages, but has no effect on manufacturing wages and output. Furthermore, the authors find that the rural mortality response to heat is entirely driven by temperature changes within the agriculture growing season, even though the hottest days typically fall outside of the growing season.

Thus, it appears that the rural mortality response to elevated temperatures in India (and to a lesser degree Mexico) can be explained by the corresponding effect on agricultural outcomes. However, it is unclear if the differential effect between urban and rural populations is solely explained by this, or if populations in urban areas have more adaptation options. This gap in the literature is not trivial, as understanding the role of adaptation will be necessary to reduce the unequal burden of climate change between urban and rural populations in developing countries. As previously discussed, most adaptation research has considered how temperature affects residential energy consumption as a proxy for self-protection via air-conditioning. Much like mortality, a lot of this work has been focused in developed countries, even though residential air conditioning adoption is growing rapidly in less developed countries. Furthermore, given the warmer locations of many developed countries, the scale of increased energy costs will dwarf effects seen in developed countries where rising temperatures may have an ambiguous effect as less heating is required in winter months.

Davis and Gertler (2015) is one of the few studies that considers the relationship between temperature and residential air conditioning in a developing country setting. Their paper focuses on Mexico, and along with income data, allow the authors to consider both the intensive margin (i.e. how weather affects energy consumption given current equipment stock) and extensive margin (i.e. how income and climate affect future energy consumption decisions). This paper is unique in the sense it compares both margins (most others do intensive or extensive only) and focuses in a country where air conditioning growth will be the most robust.

Starting first with the intensive margin, the authors find large increases in residential energy consumption for hot days. For example, one additional day per year whose temperature exceeds 90° F relative to a 65° - 70° F, increases residential energy consumption by 3.2 percent. Although their empirical approach does not directly measure air conditioning use, the authors also find that the energy response function is steeper in states with higher air conditioning adoption rates. Regarding the extensive margin, the authors find a strong relationship between the interaction of income and climate. In cool areas, air conditioner adoption is virtually zero across all income groups and income plays no role for ownership rates. However, in warm areas there is evidence of a strong relationship between income and air conditioning adoption, with ownership rates increasing 2.7 percentage points per $1,000 of annual household income. Taken together, these findings suggest air conditioning penetration rates will increase from 13 percent to more than 70 percent by the end of the century leading to an 83 percent increase in residential energy consumption (Davis and Gertler, 2015).

The Davis and Gertler (2015) study highlights the enormous potential impacts of air conditioning adoption in developing countries as the climate warms. However, the expected increase in air conditioning also presents an opportunity for reductions in mortality in the form of self-protection. A true comparison between Compeán (2013), Cohen and Dechezlepretre (2017), and Davis and Gertler (2015) is not possible due to the different reference categories, but it is reasonable to conclude that self-protection in Mexico likely already plays a role. For example, Davis and Gertler (2015) finds a larger effect for residential energy consumption than for mortality as seen in Compeán (2013) and Cohen and Dechezlepretre (2017). Again, this claim is slightly misleading due to different temperature comparisons, but would be consistent with findings from work in the US (Deschênes and Greenstone, 2011).

Reviewing the existing literature to date it becomes clear that a new empirical study is needed that compares the effect of temperature on mortality and residential energy rates in a developing country. Although some studies provide insight into both relationships, there is currently no paper that considers both simultaneously. As Deschênes and Greenstone (2011) demonstrate in the US, elevated temperatures increase residential energy consumption at a greater rate than mortality, suggesting a form of self-protection. In Mexico, the mortality response to temperature appears to be smaller than seen in other less developed countries such as India. Although Cohen and Dechezlepretre (2017) argue that this could be a reflection of the fact Mexico is less reliant on subsistence agriculture, it could also reflect higher self-protection. Our chapter intends to fill this gap in the literature by considering the relationship between temperature, mortality, and residential energy in Thailand.

## 3.3 Background and Data

This chapter presented considerable data challenges as there is no publicly available data for mortality or residential electricity in Thailand. Instead, mortality and electricity datasets were built from scratch utilizing yearly provincial statistics reports published by the National Statistics Office (NSO) of Thailand. The provincial statistic reports are comprised of tables of data for the previous year, often broken down to district or sub-district level.[1] The NSO maintains hard copies of every published provincial statistics report in their Bangkok main office library. Additionally, some individual tables from these reports have been digitally converted to excel spreadsheets and are kept in an online database where individual tables can be downloaded.

The presence of district level observations for both mortality and residential energy consumption is encouraging, yet creating a workable panel dataset was not without chal-

[1] That is the 2012 series of provincial statistic reports contains data for the 2011 year.

lenges. First, many province/years are missing or tables are mislabeled and contain the wrong data. This required checking every downloaded table by hand and supplementing missing or mislabeled tables with scanned and digitally converted files from the NSO Bangkok library. Second, each table's format changed from year to year and contained a large amount of Thai script. This prohibited automating the merging process and therefore the datasets were built by hand. This time-consuming process ultimately produced a completely unique dataset that was previously unavailable to researchers. Unfortunately, while we initially believed the dataset was of high quality, this no longer appears certain. Regardless, this dataset represents the most detailed and comprehensive district-level observations on mortality and electricity currently available in the region.

### 3.3.1 Data on Mortality and Population

To calculate the mortality rate both observations on total number of deaths and population are required. In some tables this was straightforward as both the total number of deaths and population are reported at the district level. In other tables, only the number of deaths are reported and therefore must be matched to the corresponding district level population. We followed the convention in the literature by calculating the number of deaths per 1,000 population (this is commonly known as the crude death rate). We hoped to account for different age distributions as infants and elderly are likely to be the most at risk from extreme temperature events but unfortunately disaggregated mortality data does not exist at the district level.

Observations are typically reported for the entire district level. However, for some years of data, certain provinces report data at the sub-district level as well which allow the construction of rural and urban mortality rates. Therefore, the urban/rural mortality rate is a subset of the larger data set. For example, over the sample period there are approximately 14,000 observations for total mortality, but only 7,300 for urban and rural. Over the 1991-2011 period the total mortality rate is approximately 5.9 deaths per 1,000 population, while the rural mortality rate is 6.4 and urban is 7.1. These rates compare favorablely to the official rates from the decennial survey of population change, which reports detailed population characteristics for the entire country. Across the relevant sample period are two surveys in 1995 - 1996 and 2005 - 2006, which found total mortality rates of 6 and 6.8 respectively (National Statistics Office of Thailand, 2006).

Although it is encouraging that the rates we calculate are in line with official statistics our data still reflects the inherent problem of trying to record the number of deaths in a developing country setting. For example, both the rural and urban mortality rate are higher than the total mortality rate despite being a subset. This apparent contradiction can likely be explained by the fact that data collection efforts were stronger in years for which urban and rural rates are reported. Furthermore, the urban rate is the highest, which is counterintuitive given the increased access to medical care and higher incomes. Once again however, this is likely a reflection of the better data collection efforts in urban areas since many rural deaths go unreported. While this certainly raises concerns over the data quality, the under-reporting of deaths might not be an issue if it is exogenous across all districts.

Most concerning of all is the fact we were unable to ascertain exactly how the NSO records deaths at the district level. Our mortality data is taken from individual tables listing the total number of deaths at the district level in each province. However, there is no explanation as to how deaths are recorded. Therefore, we do not know if deaths are reported from death certificates, taken from hospitals, or from informal discussions with constituents in each district. All we have been able to determine through our repeated discussions with the NSO is that each individual provincial branch of the NSO is responsible for collecting the data published in the provincial statistic reports, but any detail beyond that is unknown.

### 3.3.2 Electricity Data

Electricity data includes a variety of sales information separated by category. Categories include total, residential, business and industry, government and public utility, and others. The total number of users is also given, although this is not disaggregated across categories.

Electricity data presented its own unique challenges. Another problem was that measurement units typically varied across provinces between kWh, MWh, or GWh. This means that units had to be hand checked and converted to ensure the entire national panel was consistent. Another problem is that in some years certain districts are aggregated. This commonly occurs between neighboring districts where one produces power and supplies it to surrounding districts.

Despite these data difficulties we have created a panel spanning 1991-2011 of nearly 13,000 observations. Furthermore, we believe the data to be of extremely high quality as most power in the country is supplied by the Electricity Generating Authority of Thailand (EGAT), a state-owned power utility.

### 3.3.3 Weather Data

Weather data comes from the Global Surface Summary of the Day (GSOD) of the National Oceanic and Atmospheric Administration (NOAA). This data set provides daily weather realizations from over 9000 weather stations across the world and includes daily average temperature and total precipitation. For Thai weather stations, we included those that had 80% or more of daily observations, resulting in 98 stations across the country. The spatial distribution of stations can be seen in Figure 3.1. Since our primary analysis is focused on temperature this distribution of stations appears reasonable.

For stations with missing daily observations, two techniques were used to fill in missing observations. If there were less than three consecutive days of missing observations, the average was taken between days. However, if there were more than three consecutive days of missing observations the daily average was taken from the two nearest stations. In total, 13% of observations were missing and had to be interpolated. With a complete panel of daily weather station observations, we then calculate district level observations for the 928 districts of Thailand using inverse distance-weighted averages from the four nearest stations.

As seen in Figure 3.2, the average temperature in Thailand is typically very hot and densely distributed with the majority of days having an average temperature between 79-85 degrees Fahrenheit. Cold temperatures in Thailand are rare with only a handful of days falling below 64 degrees Fahrenheit.

## 3.4 Empirical Methodology

The foundation of this chapter's methodology comes from Deschênes and Greenstone (2011), who consider the relationship between daily temperatures on annual mortality rates and annual residential energy consumption in the United States. This framework exploits the plausibly random year-to-year variation in temperature and precipitation to determine what happens to a variable of interest in years that are warmer and/or wetter. By conditioning on fixed effects, variations in weather from district averages control for all shocks common to a district, and the causal effect can be determined.

Relating daily weather observations to yearly observations of mortality and residential energy consumption present several specifications to consider. The first option is to simply relate average annual temperature to our variable of interest. Although this approach does not account for non-linearities in the relationship between extreme temperatures and variables of interest, it can be useful as a first pass in determining if there is an effect.

A second approach creates a single-index of degrees above and below a temperature threshold. This "degree-days" measure collapses a years worth of temperature measures into a single index, while preserving the non-linear effects of temperature. For our purposes we consider the number of degrees above 88 and below 65 degrees Fahrenheit. For example, a 90-degree day in district $i$ would count as 2, while an 86-degree day would be 0. The total number of "degree-days" is then summed over each district year combination.[2]

The final approach places each district's daily observations into one of 11 equally spaced temperature categories or "bins," as seen in figure 2. These bins are defined as less than 64 degrees F, greater than 91 and the nine 3-degree F wide bins in between. This approach is a preferred method of estimating the non-linear relationship between temperature and variables of interest, as the bins fully preserve the daily variation in temperatures across a year.

The three primary empirical specifications can be seen in equations (3.1)-(3.3) below:

[2]Related to this is a measure of the number of days in a particular district where the temperature is 10 degrees warmer than the yearly average for that district. Again, this collapses a full years' worth of weather realizations into one measure but is more sensitive to the effects of extreme-temperature occurrences.

$$Y_{dt} = \beta TMEAN_{dt} + \sum_{k=1}^{2} \delta_k 1\{PREC_{dt}\text{in tercile } k\}$$
$$\alpha_d + \gamma_t + \lambda_p t + \varepsilon_{dt} \tag{3.1}$$
$$Y_{dt} = \beta_1 CDD88_{dt} + \beta_2 HDD65_{dt} + \sum_{k=1}^{2} \delta_k 1\{PREC_{dt}\text{in tercile } k\}$$
$$\alpha_d + \gamma_t + \lambda_p t + \varepsilon_{dt} \tag{3.2}$$

$$Y_{dt} = \sum_{j=1}^{8} \theta_j TMEAN_{dtj} + \sum_{k=1}^{2} \delta_k 1\{PREC_{dt}\text{in tercile } k\}$$
$$\alpha_d + \gamma_t + \lambda_p t + \varepsilon_{dt} \tag{3.3}$$

where $Y_{dt}$ is the log mortality rate in district $d$ in year $t$ (or log residential electricity use in alternative specifications), $\alpha_d$ are district fixed effects, $\gamma_t$ are year fixed effects, $t_p$ is a linear province time trend, and $\varepsilon_{dt}$ is the error term.

The key variables of interest and the difference between each equation are the weather variables that estimate the impact of weather on the dependent variable. In equation (3.1), $TMEAN_{dt}$ represents the average temperature in district $d$ in year $t$. In equation (3.2), $CDD88_{dt}$ sums the total number of degrees in district $d$ in year $t$ above 88° F, while $HDD65_{dt}$ sums the total number of degrees below 65° F. In equation (3.3), $TMEAN_{dtj}$ represents the number of days the average temperature in district $d$ in year $t$ falls into one of $j$ equally spaced bins. In other words, for each district the 365 daily realizations of mean temperature in a given year are placed into one of $j$ equally spaced bins. We considered a variety of different bin specifications in our analysis, but choose a specification of 9 bins in 3 degree intervals. For example, intervals range from less than 67° F, 67°-70° F, 70°-73° F, ... , 88°$F$ and above. As the total sum of the bins will always equal

365, one of the bins cannot be identified, and therefore must be dropped. This dropped bin serves as the reference to which the separate coefficients of interest $\theta_j$ are related. In our analysis, we select the temperature bin of 76°-79° F, which is in the middle of the distribution and represents a temperate day in Thailand.

Equations (3.2) and (3.3) capture the non-linearities associated with responses to weather that have been documented extensively in the literature. By placing a full year's realization of temperature observations into a single index or equally spaced bins, we capture the full distribution of annual fluctuations in weather, and their impact on the dependent variable. That being said, there are some assumptions in this approach worth noting. First, this approach assumes the impact is constant for temperatures above 88° F (or below 65° F) in equation (3.2) and constant within the 3° F intervals in equation (3.3). Second, this approach does not consider the effect that a sequence of concurrent hot days might have. Finally, this estimation only considers the effect of average mean temperature, which masks some important variation between minimum and maximum temperature.

The variable $PREC_{dt}$ captures the variation in rainfall in district $d$ in year $t$. We follow Burgess et al. (2017) and place each district-year rainfall observation in an upper, middle, or lower tercile computed from the district average across the entire sample. The logic behind this different modeling approach is that rainfall is much easier to store, and therefore we consider the differential effects of the sums of daily accumulations, rather than equally spaced precipitation bins.

Our empirical strategy relies on fixed effects for identification, and we considered a variety of different options. We include district fixed effects, $\alpha_d$, which capture all unobserved district-specific time invariant determinants of the different dependent variables. In other words, district fixed effects absorb the differences in mortality rates that change across districts but are constant over time (such as wealthier districts with better access to medical facilities). We also include year fixed effects, $\gamma_t$, that absorb any time-varying differences in the dependent variable that are common across all districts (such as improved medical technology over the entire sample period). Finally, we include a linear province time-trend term, $t_p$, that allows differential time effects across provinces, but captures any linear trend within a specific province. The use of fixed effects, the linear province trend-term and the plausibly random year-to-year variation in temperature, ensures that our empirical strategy produces unbiased estimates of the parameters of interest, $\beta$, $\theta_j$, and $\delta_k$. These parameters are identified from district-specific deviations from district averages, controlling for both unobserved time invariant trends that vary across districts as well as individual province time trends.

Finally, we follow Burgess et al. (2017) by choosing to cluster standard errors at the district level and weight our regressions by the number of population (or users of electricity) by district. Regarding district clustering, this is in response to the assumption that the error terms are likely correlated within districts over time. Weighting is used to reflect both the fact that districts with larger populations likely have more accurate data, and reflect the impact on the average individual instead of the average district.

## 3.5 Results

### 3.5.1 Mortality

Our initial results suggest that elevated temperatures lead to reductions in mortality, which is in stark contrast to virtually every finding across the literature. This puzzling effect led us to consider the possibility of a timing issue with our mortality data. To test this, we regressed log mortality rates matched to the previous, current, and next year's weather outcomes. Table 3.1 presents these results for equation (3.1), and demonstrates our concerns with mortality data. In column 1, we find no relationship between mortality and temperature when regressing mortality rates to the previous year's weather outcomes (e.g. 2011 mortality rates are regressed on 2010 weather data). Conversely, in column 2 and 3, we find both a negative and positive relationship between mortality and temperature when regressing mortality to the current year's weather and next year's weather respectively.

This unexpected result holds across the different specifications as well. The results from equation (3.2) and (3.3) are presented in Tables 3.2 and 3.3 respectively. We also present our bin regression results in graphical form for easier interpretation in Figures 3.3 - 3.5. Regardless of what specification is chosen the results are clear. When matched to the previous year's weather realizations, no effect is found. However, when matched to the current year's weather, we find strong evidence of a negative relationship between extreme temperature and mortality. Intuitively, this makes little sense as extreme heat waves are routinely associated with excessive deaths. One possible explanation might be related to disease vectors through insects, particularly mosquitoes. If mosquitoes are less active during extreme temperature events, than the associated disease vectors they represent would fall, potentially causing a decrease in mortality. However, malaria is not a major source of death in Thailand with less than 1,000 deaths a year and therefore this explanation seems unlikely (WHO, 2014).

We also find constant evidence of a positive relationship between high temperatures and increased mortality when matched to the next year's weather realization. This result is consistent with the existing literature and led us to question if the timing of our mortality data was incorrect. Recall that our data comes from a yearly provincial statistics report that reports data from the previous year. In other words, the 2010 provincial statistics report will list data for the total number of deaths from 2009. We have been working under the assumption that the year the report was published corresponds to the previous calendar year. However, given no direct timing information about when the data is recorded we suggest it might be reasonable to presume that the timing of the mortality and weather data might be off.

There are other possibilities for our seemingly non-sensible results as well. One limitation of our data set is that we only obtain observations on the crude death rate. As others have shown, other demographics (specifically young and elderly) are most at risk to temperature induced increases to mortality (Deschênes, 2014). Therefore, it remains a possibility that we would still find the intended effects if we focused on a more vulnerable portion of the population. However, data separated by any type of demographic does not exist in Thailand, and thus all we can study is the crude death rate.

Another limitation lies with our weather data as we only observe average temperatures. Our results could be affected by mean temperatures obscuring much of the variation in temperature by not observing minimum and maximum temperatures. Specifically, maximum daytime temperatures could be among the most damaging to vulnerable populations. For example, if we had maximum temperature along with the elderly or infant mortality rates, the expected effect of elevated temperatures increasing mortality might be observed. Another missing weather observation is humidity, which has also been shown to affect mortality rates (Barreca, 2012). Thailand has high levels of humidity overall and failing to account for this may also impact our findings although fixed effects would control for variations in humidity levels across districts.

Finally, our results here only consider the overall mortality rate across entire districts which span both rural and urban populations. As seen in Compeán (2013) and Burgess et al. (2017), there are large differential effects for urban and rural mortality rates. By failing to separate the mortality rate across populations we could potentially be finding misleading results. For example, it is possible that elevated temperatures in urban populations drive much of the negative effect while rural populations could demonstrate the expected positive mortality response. That being said, this explanation appears the least plausible. First, the negative effect between populations would have to be so strong that it would drive the overall effect seen in the preliminary results. Even if this effect was restricted to one population, it would not explain the non-sensible result. Furthermore, we have limited data on mortality rates in urban and rural areas and exploratory analysis suggests these effects are consistent across both populations.

Given the counterintuitive finding for contemporaneous temperature and opposing effects for temperature in the following period, it is clear there are serious concerns over either the quality or timing of the mortality data. Unfortunately, with no direct communication with the NSO we are unable to address these concerns further. Additionally, given the data limitations of our dataset we are unable to explore other issues that may be driving the non-sensible findings. Therefore, despite the lengthy data building efforts it appears we cannot rely on the mortality data for our analysis and any results should not be considered valid at this stage. Further analysis using this dataset has been suspended.

### 3.5.2 Electricity

Despite our disappointing results for mortality, we find strong evidence that extreme temperatures lead to significant increases in residential energy use. This finding is consistent with the existing literature, demonstrating our empirical strategy is sound. Starting first with the simple average temperature specification of equation (3.1), results are presented in Table 3.4, Column 1. Here we find that an increase in 1° F causes residential electricity to increase by 7.7%. This finding provides a useful first pass to establish the link between residential energy use and temperature, but masks any non-linearities in the response. A better illustration of the energy response is instead seen in Table 3.4, column 2 where we present findings for the degree day specification of equation (3.2). These findings demonstrate that the energy response is driven purely by warm days. We find that each additional degree day above 88° F causes residential energy consumption to increase by 0.094%, while degree days below 65° F are not found to have a meaningful effect. For both specifications we find that precipitation has no effect.

Turning to the most flexible specification using the temperature bin specification in equation (3.3), we find the energy response is entirely driven by days with an average temperature greater than 91° F. This can be seen clearly in both Table 3.5 and Figure 3.6, where the only significant result corresponds to the warmest temperature bin. These results indicate that one day with an average temperature of 91° F or greater relative to a $76° - 79°$ F day, causes residential energy use to increase by 0.6%. This is a sizable effect, as it indicates that just two additional hot days per year opposed to temperate ones will cause residential energy use to increase by over 1%.

These results appear logical. Given that average temperatures in Thailand are warm and residential heating systems are virtually unknown, we do not expect cold temperatures to be a large driver of residential energy use. Conversely however, residential fan ownership is nearly universal and AC penetration rates are high for a developing country and steadily increasing. Although this finding is encouraging that our empirical strategy is sound, without mortality results to relate them to, it does not add much to the literature. It corroborates findings from Davis and Gertler (2015) who also find that warm temperatures drive much of the residential energy response in Mexico. However, their paper has more detailed household-level microdata, which contains information on energy use, air conditioning ownership, and income. This detailed household level data enables Davis and Gertler (2015) to measure how air conditioning ownership rates vary the residential energy response to temperature. Furthermore, detailed household level income data allows them to consider extensive margin effects, which is something we are unable to do with our dataset. Therefore, at best our results support their findings, but do not advance the literature on their own.

## 3.6 Conclusion

Despite the growing body of research that suggests temperature extremes lead to significant reductions in health, very little empirical work has focused in developing countries. This presents a significant gap in the literature. Developing country economies are often characterized by economic activities that rely on their surrounding environment and weather shocks can significantly reduce livelihoods. Furthermore, there are less adaptation options available for individuals in developing countries. Taken together, both suggest there would be a larger effect between extreme temperature and mortality than has been demonstrated in developed countries. Indeed, the few existing studies that consider the weather-mortality relationship in a developing country support this claim (Compeán, 2013; Cohen and Dechezlepretre, 2017; Burgess et al., 2017).

Furthermore, the existing studies to date focus their attention on explaining why the mortality response to temperature is entirely driven by rural populations. Less consideration is given to what role adaptation plays in urban areas. This is a subtle difference but understanding the role of adaptation is important to understanding the highly unequal impacts that climate change will have within developing countries. Much attention has been given to the unequal impacts between developed and developing countries, but these papers demonstrate even within countries the damages will likely disproportionately burden the rural poor. Furthermore, adaptation options will be some of the most effective climate change mitigation strategies in the immediate future, and as common forms of self-protection such as air conditioning adoption and usage increase, understanding the role of adaptation in a developing country setting is a new addition to the literature.

Our chapter intends to overcome this gap in the literature by considering the role of temperature, mortality, and adaptation in Thailand. Thailand was chosen as it is the largest, and most economically diverse country in the rapidly industrializing Greater Mekong Subregion. Furthermore, based on our prior experience working in the region, we believed we could assemble a high-quality dataset that was previously unavailable to researchers. This led to an extensive data collection effort, that unfortunately appears to have been in vain. Our results for the relationship between temperature and mortality are mostly non-sensible, and without access to additional mortality and weather data we are unable to explore what drives these results further. We do find that extreme temperatures lead to increased residential energy use, but without the related mortality analysis we are unable to comment on the role of adaptation. Instead these results on their own do not constitute an improvement in the existing literature.

Since the first draft of our chapter was written new studies focused on Mexico have emerged that in retrospect provide a better strategy to conduct our analysis. Compeán (2013) and Cohen and Dechezlepretre (2017) consider the temperature-mortality relationship while Davis and Gertler (2015) consider the temperature-residential energy relationship. With access to their data sources we could replicate the Deschênes and Greenstone (2011) paper in a developing country setting. Much like Burgess et al. (2017) we could then compare our results in Mexico to findings from the US by using consistent reference categories across countries. This would allow a useful comparison between the relationship between temperature, mortality, and adaptation between a developed and developing country that is currently missing in the literature.

Figure 3.1: Spatial Distribution of GSOD Stations

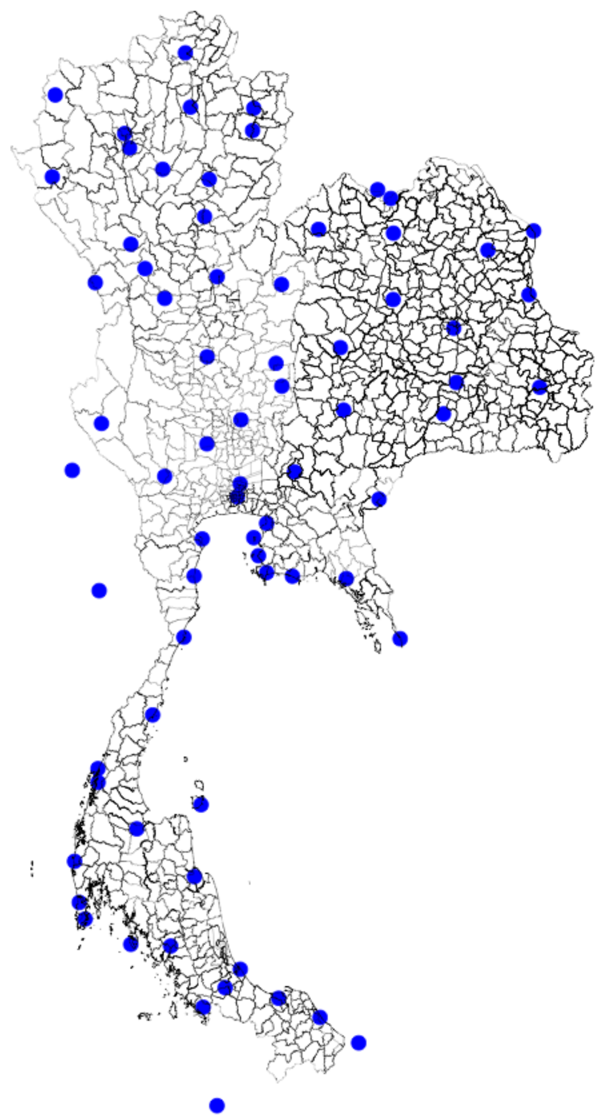


Figure 3.2: Distribution of Annual Daily Mean Temperatures (F), 1991-2011

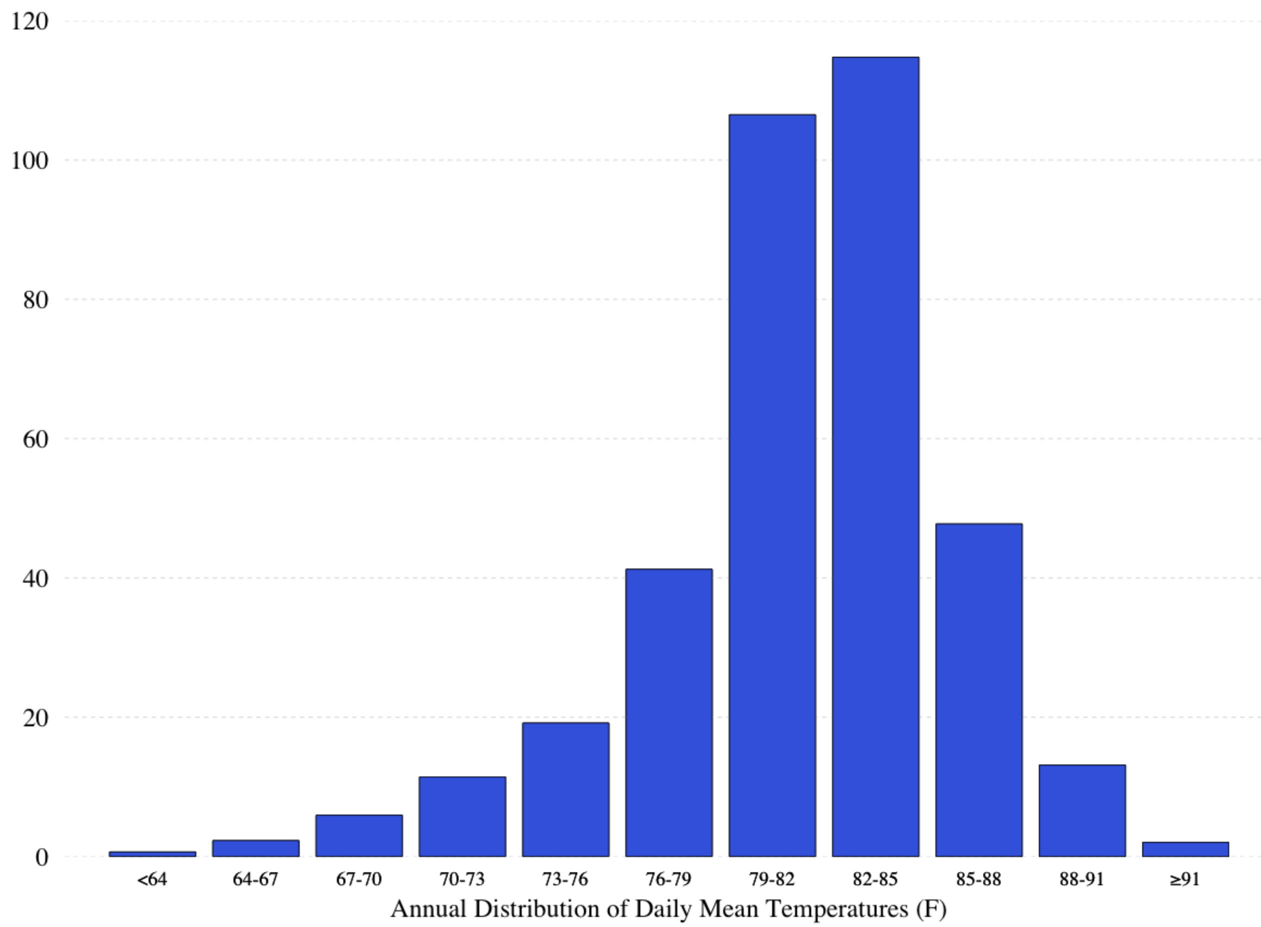
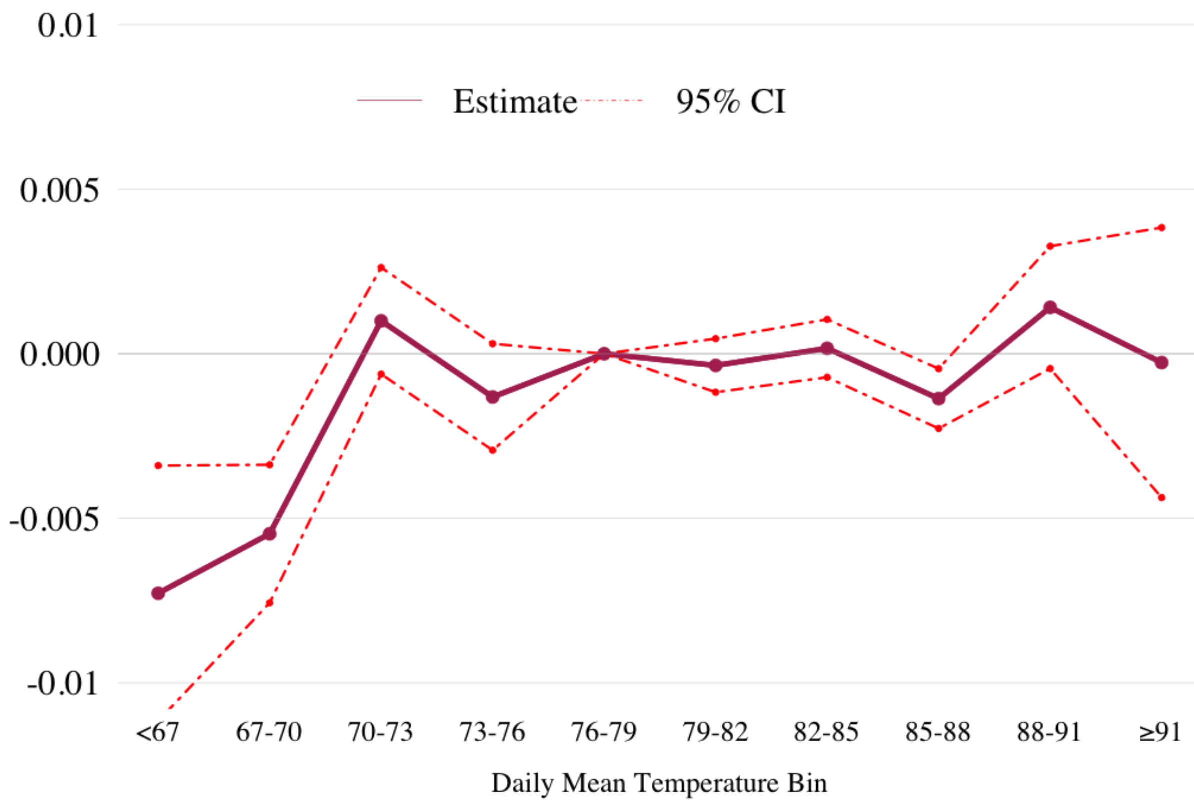


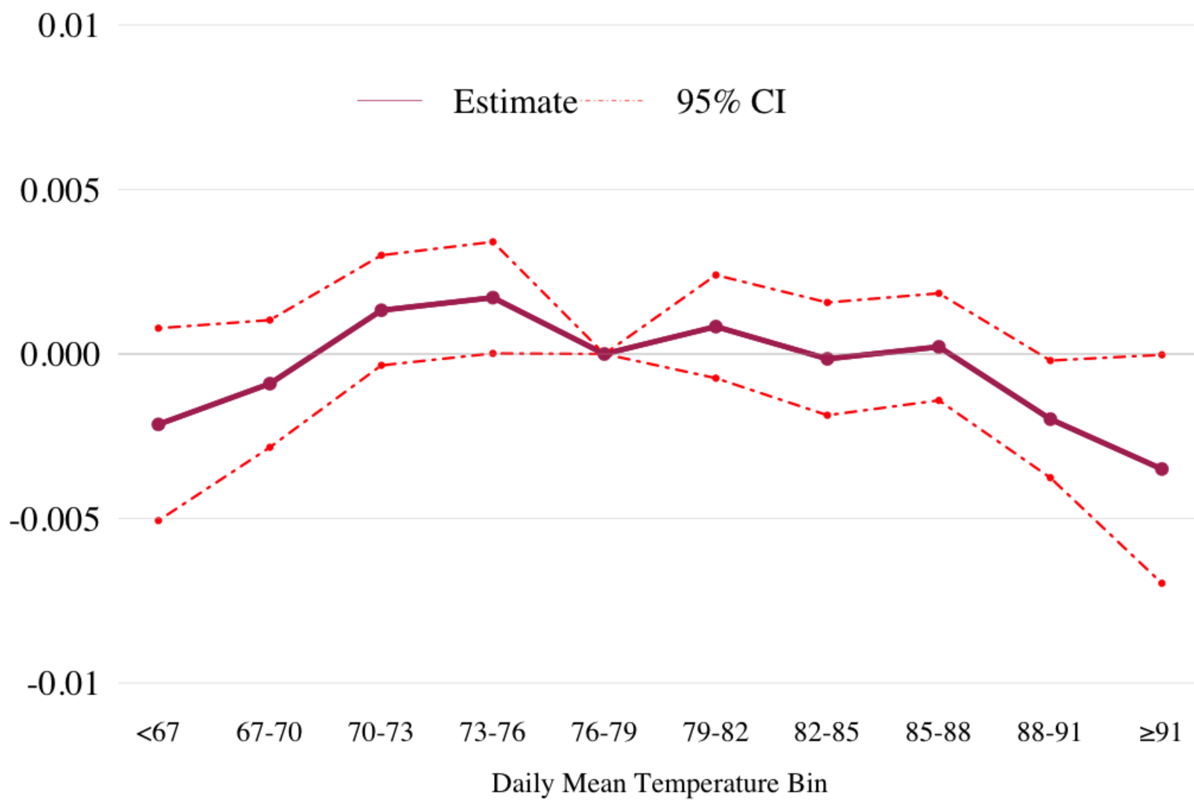
Figure 3.3: The Effect of Daily Average Temperature on Log Mortality




Estimated Impact of a Day in 9 Temperature-Day (Previous Year) Bins on Log Mortality, Relative to a Day in the 76-79 Farenheit Bin

Notes: The solid line is the graphical representation of the coefficient estimates in Table 3.3 , Column 1. The dashed line is representative of the 95% confidence intervals for the coefficient estimates.

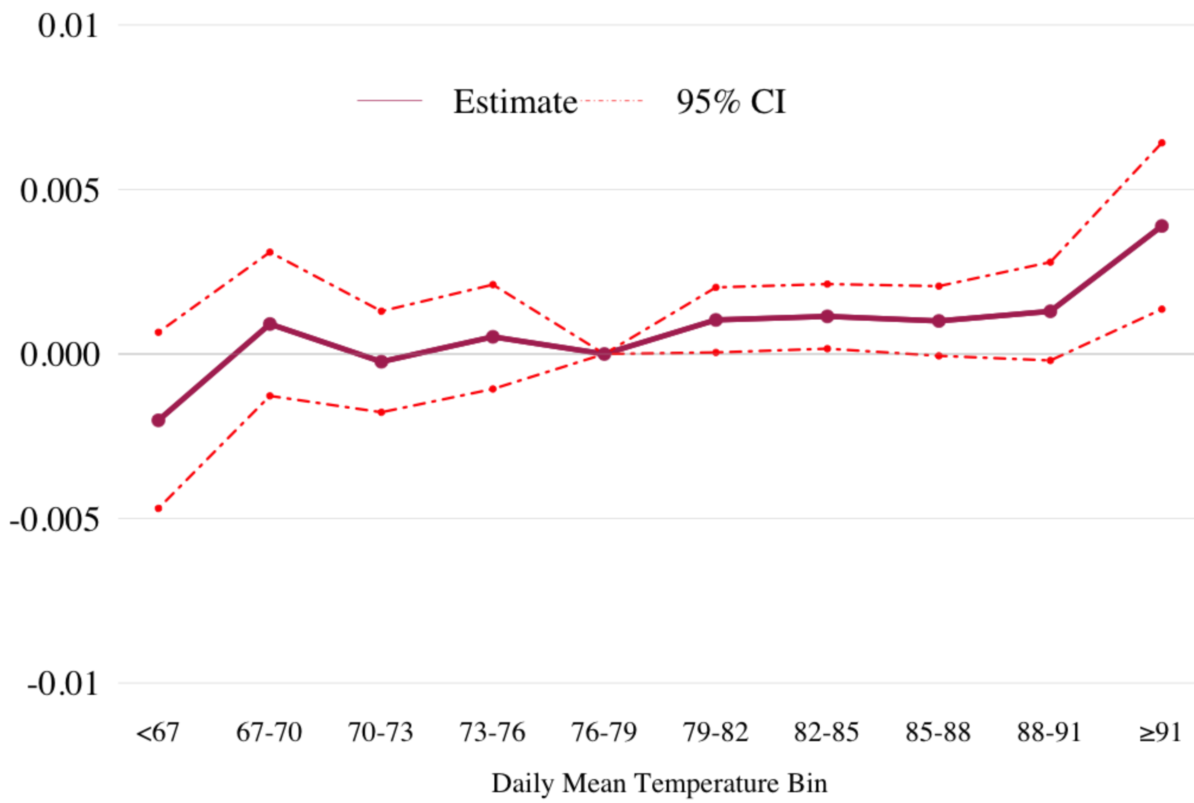
Figure 3.4: The Effect of Daily Average Temperature on Log Mortality




Estimated Impact of a Day in 9 Temperature-Day (current Year) Bins on Log Mortality,
Relative to a Day in the 76-79 Farenheit Bin

Notes: The solid line is the graphical representation of the coefficient estimates in Table 3.3, Column 2. The dashed line is representative of the 95% confidence intervals for the coefficient estimates.

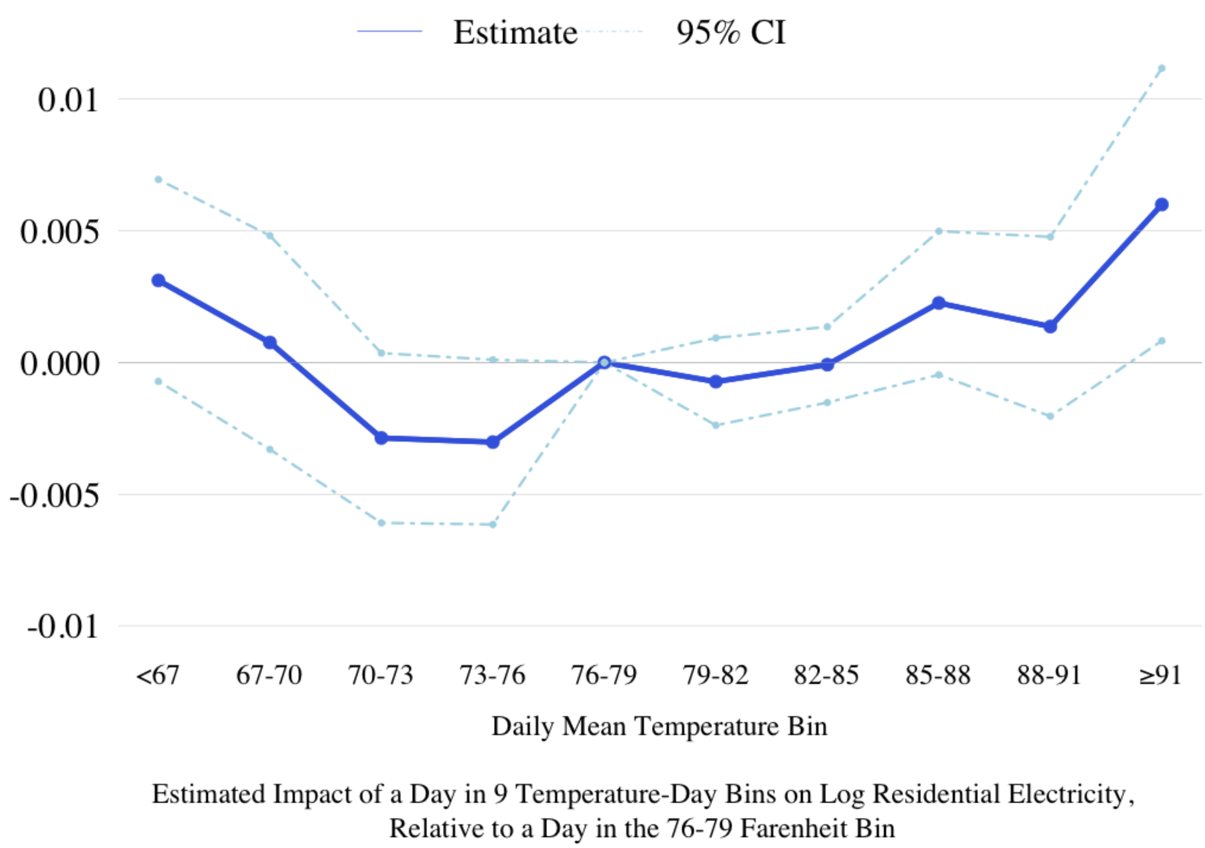
Figure 3.5: The Effect of Daily Average Temperature on Log Mortality




Estimated Impact of a Day in 9 Temperature-Day (next Year) Bins on Log Mortality,
Relative to a Day in the 76-79 Farenheit Bin

Notes: The solid line is the graphical representation of the coefficient estimates in Table 3.3, Column 3. The dashed line is representative of the 95% confidence intervals for the coefficient estimates.

Figure 3.6: The Effect of Daily Average Temperature on Log Residential Energy Use




Notes: The solid line is the graphical representation of the coefficient estimates in Table 3.5. The dashed line is representative of the 95% confidence intervals for the coefficient estimates.

Table 3.1: The Effect of Daily Average Temperature on Log Mortality

| | (1) Previous Year Weather | (2) Current Year Weather | (3) Next Year Weather |
|---|---|---|---|
| Temperature (Average) | 0.0226 | -0.0378*** | 0.0358*** |
| | (0.0159) | (0.0144) | (0.0122) |
| Indicator for Rainfall Shock in Lowest Tercile | 0.0268** | -0.0130 | 0.00399 |
| | (0.0123) | (0.00908) | (0.0122) |
| Indicator for Rainfall Shock in Highest Tercile | -0.000993 | -0.0266*** | -0.00334 |
| | (0.00859) | (0.00756) | (0.00803) |
| Fixed Effects | Dist, Yr | Dist, Yr | Dist, Yr |
| Province Trend | Prov Lin | Prov Lin | Prov Lin |
| No Obs | 13911 | 13932 | 13932 |
| $R^2$ | 0.585 | 0.586 | 0.585 |

Notes: Regressions are weighted by district population.
Standard errors are in parenthesis and are clustered at the district level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3.2: The Effect of Daily Average Temperature on Log Mortality

| | (1) Previous Year Weather | (2) Current Year Weather | (3) Next Year Weather |
|---|---|---|---|
| Temperature (Degree Days Over 88F) | 0.000483 | -0.000936*** | 0.000594** |
| | (0.000391) | (0.000274) | (0.000231) |
| Temperature (Degree Days Under 65F) | -0.00108** | 0.000451 | -0.000519 |
| | (0.000472) | (0.000872) | (0.000546) |
| Indicator for Rainfall Shock in Lowest Tercile | 0.0274** | -0.0136 | 0.00708 |
| | (0.0120) | (0.00883) | (0.0118) |
| Indicator for Rainfall Shock in Highest Tercile | -0.00234 | -0.0250*** | -0.00655 |
| | (0.00836) | (0.00758) | (0.00814) |
| Fixed Effects | Dist, Yr | Dist, Yr | Dist, Yr |
| Province Trend | Prov Lin | Prov Lin | Prov Lin |
| No Obs | 13911 | 13932 | 13932 |
| $R^2$ | 0.585 | 0.586 | 0.585 |

Notes: Regressions are weighted by district population.
Standard errors are in parenthesis and are clustered at the district level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3.3: The Effect of Daily Average Temperature on Log Mortality

| | (1) Previous Year Weather | (2) Current Year Weather | (3) Next Year Weather |
|---|---|---|---|
| $< 67°F$ | -0.00728*** | -0.00214 | -0.00202 |
| | (0.00198) | (0.00149) | (0.00137) |
| $67° - 70°F$ | -0.00547*** | -0.000902 | 0.000911 |
| | (0.00107) | (0.000986) | (0.00111) |
| $70° - 73°F$ | 0.00100 | 0.00133 | -0.000235 |
| | (0.000826) | (0.000854) | (0.000784) |
| $73° - 76°F$ | -0.00131 | 0.00171** | 0.000522 |
| | (0.000826) | (0.000865) | (0.000809) |
| $79° - 82°F$ | -0.000355 | 0.000832 | 0.00104** |
| | (0.000415) | (0.000800) | (0.000505) |
| $82° - 85°F$ | 0.000165 | -0.000148 | 0.00114** |
| | (0.000449) | (0.000874) | (0.000501) |
| $85° - 88°F$ | -0.00136*** | 0.000217 | 0.00100* |
| | (0.000464) | (0.000830) | (0.000540) |
| $88° - 91°F$ | 0.00141 | -0.00198** | 0.00130* |
| | (0.000949) | (0.000909) | (0.000762) |
| $\geq 91°F$ | -0.000268 | -0.00350** | 0.00389*** |
| | (0.00209) | (0.00177) | (0.00129) |
| Indicator for Rainfall Shock in Lowest Tercile | 0.0286** | -0.0116 | 0.00398 |
| | (0.0122) | (0.00894) | (0.0124) |
| Indicator for Rainfall Shock in Highest Tercile | -0.00634 | -0.0283*** | -0.00275 |
| | (0.00868) | (0.00827) | (0.00831) |
| Fixed Effects | Dist, Yr | Dist, Yr | Dist, Yr |
| Province Trend | Prov Lin | Prov Lin | Prov Lin |
| No Obs | 13911 | 13932 | 13932 |
| $R^2$ | 0.590 | 0.588 | 0.585 |

Notes: Regressions are weighted by district population.
Standard errors are in parenthesis and are clustered at the district level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3.4: The Effect of Daily Average Temperature on Log Residential Electricity Use

| | (1) Electricity | (2) Electricity | (3) Electricity |
|---|---|---|---|
| Temperature (Average) | 0.0772*** | | |
| | (0.0215) | | |
| Temperature (Degree Days Over 88F) | | 0.000941*** | |
| | | (0.000315) | |
| Temperature (Degree Days Under 65F) | | 0.000400 | |
| | | (0.000793) | |
| Indicator for Rainfall Shock in Lowest Tercile | 0.000511 | 0.00823 | 0.00398 |
| | (0.0199) | (0.0202) | (0.0124) |
| Indicator for Rainfall Shock in Highest Tercile | 0.0164 | 0.00748 | -0.00275 |
| | (0.0137) | (0.0133) | (0.00831) |
| Fixed Effects | Dist, Yr | Dist, Yr | Dist, Yr |
| Province Trend | Prov Lin | Prov Lin | Prov Lin |
| No Obs | 12661 | 12661 | 13932 |
| $R^2$ | 0.850 | 0.850 | 0.585 |

Notes: Regressions are weighted by number of electricty users per district.
Standard errors are in parenthesis and are clustered at the district level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3.5: The Effect of Daily Average Temperature on Log Residential Electricity Use

| | (1)<br>Electricity |
|---|---|
| $< 67°F$ | 0.00312 |
| | (0.00196) |
| $67° - 70°F$ | 0.000760 |
| | (0.00207) |
| $70° - 73°F$ | -0.00287* |
| | (0.00165) |
| $73° - 76°F$ | -0.00302* |
| | (0.00160) |
| $79° - 82°F$ | -0.000727 |
| | (0.000846) |
| $82° - 85°F$ | -0.0000819 |
| | (0.000734) |
| $85° - 88°F$ | 0.00226 |
| | (0.00139) |
| $88° - 91°F$ | 0.00137 |
| | (0.00174) |
| $\geq 91°F$ | 0.00600** |
| | (0.00264) |
| Indicator for Rainfall Shock in Lowest Tercile | 0.000455 |
| | (0.0201) |
| Indicator for Rainfall Shock in Highest Tercile | 0.0197 |
| | (0.0144) |
| Fixed Effects | Dist, Yr |
| Province Trend | Prov Lin |
| No Obs | 12661 |
| $R^2$ | 0.851 |

Notes: Regressions are weighted by number of electricty users per district. Standard errors are in parenthesis and are clustered at the district level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# Appendix A

# Appendix

## A.1 Chapter 1 Sensitivity Analysis

To explore the robustness of the main results in Chapter 1, I present regression results including a one-month lag for each specification discussed in Section 1.3. These results are presented below. These results demonstrate that the inclusion of lags appears to have no effect on the contemporaneous effect of weather, and the lagged variables are primarily statistically insignificant. This implies that negative serial correlation is not a concern for homicide, and therefore the primary specifications are an appropriate modeling approach.

Table A.1: Log Homicide Rate by Race

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temp : $< 10°$ F | 0.0007 | 0.0002 | 0.0007 | 0.0029 |
| | (0.0012) | (0.0011) | (0.0023) | (0.0027) |
| Temp: $10- < 20°$ F | -0.0032 | -0.0019 | -0.0060* | -0.0033* |
| | (0.0021) | (0.0015) | (0.0034) | (0.0018) |
| Temp: $20- < 30°$ F | -0.0015* | -0.0006 | -0.0037** | 0.0021 |
| | (0.0008) | (0.0008) | (0.0016) | (0.0021) |
| Temp: $30- < 40°$ F | -0.0020* | -0.0010 | -0.0035* | 0.0002 |
| | (0.0010) | (0.0008) | (0.0021) | (0.0013) |
| Temp: $40- < 50°$ F | -0.0008 | -0.0004 | -0.0013 | -0.0001 |
| | (0.0005) | (0.0005) | (0.0011) | (0.0008) |
| Temp: $50- < 60°$ F | -0.0009** | -0.0004 | -0.0019** | 0.0003 |
| | (0.0004) | (0.0004) | (0.0009) | (0.0006) |
| Temp: $70- < 80°$ F | 0.0009** | 0.0012*** | 0.0003 | -0.0004 |
| | (0.0003) | (0.0003) | (0.0007) | (0.0004) |
| Temp: $80- < 90°$ F | 0.0019*** | 0.0020*** | 0.0018* | 0.0015* |
| | (0.0005) | (0.0005) | (0.0010) | (0.0008) |
| Temp: $\geq 90°$ F | 0.0011 | 0.0009 | 0.0034 | 0.0002 |
| | (0.0007) | (0.0007) | (0.0030) | (0.0025) |
| Lag T : $< 10°$ F | 0.0071*** | 0.0025 | 0.0138*** | 0.0024 |
| | (0.0016) | (0.0018) | (0.0030) | (0.0036) |
| Lag T: $10- < 20°$ F | -0.0001 | -0.0001 | -0.0002 | -0.0016 |
| | (0.0014) | (0.0010) | (0.0025) | (0.0026) |
| Lag T: $20- < 30°$ F | 0.0002 | -0.0006 | 0.0018 | -0.0000 |
| | (0.0008) | (0.0010) | (0.0014) | (0.0012) |
| Lag T: $30- < 40°$ F | 0.0008 | 0.0005 | 0.0014 | -0.0005 |
| | (0.0008) | (0.0007) | (0.0015) | (0.0013) |
| Lag T: $40- < 50°$ F | 0.0008* | 0.0004 | 0.0017 | 0.0006 |
| | (0.0005) | (0.0005) | (0.0011) | (0.0009) |
| Lag T: $50- < 60°$ F | 0.0006 | 0.0001 | 0.0015 | 0.0004 |
| | (0.0004) | (0.0004) | (0.0011) | (0.0005) |
| Lag T: $70- < 80°$ F | 0.0002 | -0.0001 | 0.0009 | 0.0004 |
| | (0.0004) | (0.0004) | (0.0008) | (0.0004) |
| Lag T: $80- < 90°$ F | -0.0000 | -0.0002 | 0.0004 | -0.0004 |
| | (0.0005) | (0.0006) | (0.0010) | (0.0010) |
| Lag T: $\geq 90°$ F | 0.0005 | 0.0002 | 0.0009 | -0.0020 |
| | (0.0011) | (0.0011) | (0.0034) | (0.0022) |
| Precip: $> 0- < 10$ mm | 0.0003 | -0.0003 | 0.0016*** | 0.0005 |
| | (0.0003) | (0.0004) | (0.0006) | (0.0008) |
| Precip: $10- < 20$ mm | 0.0002 | 0.0006 | 0.0009 | 0.0017 |
| | (0.0009) | (0.0009) | (0.0017) | (0.0019) |
| Precip: $20- < 30$ mm | 0.0003 | 0.0016 | -0.0004 | -0.0020 |
| | (0.0014) | (0.0015) | (0.0027) | (0.0027) |
| Precip: $\geq 30$ mm | -0.0012 | -0.0020 | -0.0008 | 0.0030 |
| | (0.0014) | (0.0015) | (0.0028) | (0.0031) |
| Lag P: $> 0- < 10$ mm | -0.0003 | -0.0005 | 0.0000 | 0.0002 |
| | (0.0003) | (0.0004) | (0.0006) | (0.0005) |
| Lag P: $10- < 20$ mm | 0.0006 | 0.0000 | 0.0021 | -0.0015 |
| | (0.0009) | (0.0010) | (0.0019) | (0.0019) |
| Lag P: $20- < 30$ mm | 0.0002 | 0.0002 | 0.0004 | -0.0004 |
| | (0.0015) | (0.0016) | (0.0029) | (0.0042) |
| Lag P: $\geq 30$ mm | 0.0003 | -0.0001 | 0.0020 | -0.0021 |
| | (0.0015) | (0.0020) | (0.0022) | (0.0030) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.796 | 0.612 | 0.610 | 0.253 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.2: Log Homicide Rate by Race

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temperature (Degree Days Over 70F) | 0.000110*** | 0.000106*** | 0.000144** | 0.0000637 |
| | (0.0000308) | (0.0000340) | (0.0000566) | (0.0000557) |
| Temperature (Degree Days Under 50F) | -0.0000353 | -0.0000173 | -0.0000894** | 0.0000192 |
| | (0.0000223) | (0.0000204) | (0.0000368) | (0.0000417) |
| Lag T (Degree Days Over 70F) | -0.00000589 | -0.00000382 | -0.0000194 | -0.0000241 |
| | (0.0000311) | (0.0000369) | (0.0000626) | (0.0000546) |
| Lag T (Degree Days Under 50F) | 0.0000290 | 0.000000305 | 0.0000663** | -0.0000360 |
| | (0.0000186) | (0.0000234) | (0.0000313) | (0.0000311) |
| Precip: $> 0- < 10$ mm | 0.000166 | -0.000360 | 0.00148** | 0.000564 |
| | (0.000301) | (0.000397) | (0.000571) | (0.000804) |
| Precip: $10- < 20$ mm | 0.0000521 | 0.000542 | 0.000582 | 0.00166 |
| | (0.000849) | (0.000875) | (0.00170) | (0.00179) |
| Precip: $20- < 30$ mm | 0.000214 | 0.00155 | -0.000492 | -0.00194 |
| | (0.00140) | (0.00149) | (0.00269) | (0.00273) |
| Precip: $\geq 30$ mm | -0.00136 | -0.00213 | -0.000884 | 0.00290 |
| | (0.00143) | (0.00145) | (0.00283) | (0.00295) |
| Lag P: $> 0- < 10$ mm | -0.000172 | -0.000439 | 0.000210 | 0.000330 |
| | (0.000337) | (0.000356) | (0.000611) | (0.000445) |
| Lag P: $10- < 20$ mm | 0.000737 | 0.000138 | 0.00218 | -0.00137 |
| | (0.000833) | (0.00104) | (0.00179) | (0.00181) |
| Lag P: $20- < 30$ mm | 0.000555 | 0.000429 | 0.000870 | -0.0000912 |
| | (0.00149) | (0.00161) | (0.00289) | (0.00424) |
| Lag P: $\geq 30$ mm | 0.000545 | -0.0000421 | 0.00226 | -0.00202 |
| | (0.00147) | (0.00203) | (0.00221) | (0.00303) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.796 | 0.612 | 0.609 | 0.253 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.3: Log Homicide Rate for Young Males (15-34) by Race

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temp : $< 10°$ F | -0.0017 | -0.0016 | -0.0039 | 0.0029 |
| | (0.0027) | (0.0033) | (0.0044) | (0.0027) |
| Temp: $10- < 20°$ F | -0.0064 | -0.0030 | -0.0127** | -0.0033* |
| | (0.0040) | (0.0031) | (0.0055) | (0.0018) |
| Temp: $20- < 30°$ F | -0.0046*** | -0.0008 | -0.0112*** | 0.0021 |
| | (0.0015) | (0.0020) | (0.0024) | (0.0021) |
| Temp: $30- < 40°$ F | -0.0036** | -0.0020 | -0.0069** | 0.0002 |
| | (0.0018) | (0.0019) | (0.0028) | (0.0013) |
| Temp: $40- < 50°$ F | -0.0029*** | -0.0010 | -0.0059*** | -0.0001 |
| | (0.0008) | (0.0010) | (0.0018) | (0.0008) |
| Temp: $50- < 60°$ F | -0.0024*** | -0.0010 | -0.0058*** | 0.0003 |
| | (0.0007) | (0.0008) | (0.0014) | (0.0006) |
| Temp: $70- < 80°$ F | 0.0011* | 0.0024*** | -0.0013 | -0.0004 |
| | (0.0007) | (0.0007) | (0.0014) | (0.0004) |
| Temp: $80- < 90°$ F | 0.0026*** | 0.0040*** | 0.0004 | 0.0015* |
| | (0.0010) | (0.0012) | (0.0020) | (0.0008) |
| Temp: $\geq 90°$ F | 0.0063** | 0.0095*** | 0.0010 | 0.0002 |
| | (0.0026) | (0.0031) | (0.0056) | (0.0025) |
| Lag T : $< 10°$ F | 0.0150*** | 0.0055 | 0.0219*** | 0.0024 |
| | (0.0034) | (0.0034) | (0.0054) | (0.0036) |
| Lag T: $10- < 20°$ F | -0.0018 | 0.0003 | -0.0019 | -0.0016 |
| | (0.0034) | (0.0030) | (0.0049) | (0.0026) |
| Lag T: $20- < 30°$ F | -0.0004 | -0.0016 | 0.0017 | -0.0000 |
| | (0.0018) | (0.0022) | (0.0029) | (0.0012) |
| Lag T: $30- < 40°$ F | 0.0031* | 0.0027 | 0.0025 | -0.0005 |
| | (0.0018) | (0.0017) | (0.0031) | (0.0013) |
| Lag T: $40- < 50°$ F | 0.0016 | 0.0007 | 0.0023 | 0.0006 |
| | (0.0011) | (0.0011) | (0.0023) | (0.0009) |
| Lag T: $50- < 60°$ F | 0.0025*** | 0.0018** | 0.0029 | 0.0004 |
| | (0.0008) | (0.0009) | (0.0019) | (0.0005) |
| Lag T: $70- < 80°$ F | 0.0004 | -0.0007 | 0.0024* | 0.0004 |
| | (0.0007) | (0.0009) | (0.0013) | (0.0004) |
| Lag T: $80- < 90°$ F | -0.0004 | -0.0011 | 0.0025 | -0.0004 |
| | (0.0011) | (0.0013) | (0.0021) | (0.0010) |
| Lag T: $\geq 90°$ F | 0.0009 | -0.0018 | 0.0084 | -0.0020 |
| | (0.0016) | (0.0019) | (0.0084) | (0.0022) |
| Precip: $> 0- < 10$ mm | 0.0002 | -0.0011 | 0.0027** | 0.0005 |
| | (0.0006) | (0.0008) | (0.0011) | (0.0008) |
| Precip: $10- < 20$ mm | 0.0001 | -0.0001 | 0.0040 | 0.0017 |
| | (0.0021) | (0.0025) | (0.0034) | (0.0019) |
| Precip: $20- < 30$ mm | 0.0024 | 0.0000 | 0.0049 | -0.0020 |
| | (0.0029) | (0.0031) | (0.0049) | (0.0027) |
| Precip: $\geq 30$ mm | -0.0037 | -0.0029 | -0.0058 | 0.0030 |
| | (0.0031) | (0.0038) | (0.0050) | (0.0031) |
| Lag P: $> 0- < 10$ mm | 0.0003 | 0.0003 | 0.0005 | 0.0002 |
| | (0.0008) | (0.0008) | (0.0014) | (0.0005) |
| Lag P: $10- < 20$ mm | -0.0011 | -0.0009 | -0.0007 | -0.0015 |
| | (0.0020) | (0.0022) | (0.0038) | (0.0019) |
| Lag P: $20- < 30$ mm | 0.0021 | -0.0023 | 0.0087 | -0.0004 |
| | (0.0033) | (0.0038) | (0.0063) | (0.0042) |
| Lag P: $\geq 30$ mm | 0.0009 | -0.0019 | 0.0019 | -0.0021 |
| | (0.0039) | (0.0046) | (0.0069) | (0.0030) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.678 | 0.534 | 0.514 | 0.253 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.4: Log Homicide Rate for Young Males (15-34) by Race

| | (1) Homicide Rate: All Victims | (2) Homicide Rate: White Victims | (3) Homicide Rate: Black Victims | (4) Homicide Rate: Other Race |
|---|---|---|---|---|
| Temperature (Degree Days Over 70F) | 0.000220*** | 0.000302*** | 0.000107 | 0.0000637 |
| | (0.0000573) | (0.0000739) | (0.000111) | (0.0000557) |
| Temperature (Degree Days Under 50F) | -0.0000879* | -0.0000263 | -0.000199*** | 0.0000192 |
| | (0.0000488) | (0.0000511) | (0.0000697) | (0.0000417) |
| Lag T (Degree Days Over 70F) | -0.0000497 | -0.0000918 | 0.0000958 | -0.0000241 |
| | (0.0000570) | (0.0000694) | (0.000125) | (0.0000546) |
| Lag T (Degree Days Under 50F) | 0.00000612 | -0.0000269 | 0.0000611 | -0.0000360 |
| | (0.0000439) | (0.0000471) | (0.0000698) | (0.0000311) |
| Precip: $> 0- < 10$ mm | -0.000245 | -0.00142* | 0.00223* | 0.000564 |
| | (0.000653) | (0.000823) | (0.00114) | (0.000804) |
| Precip: $10- < 20$ mm | -0.000509 | -0.000236 | 0.00310 | 0.00166 |
| | (0.00216) | (0.00248) | (0.00340) | (0.00179) |
| Precip: $20- < 30$ mm | 0.00202 | -0.0000375 | 0.00403 | -0.00194 |
| | (0.00293) | (0.00312) | (0.00494) | (0.00273) |
| Precip: $\geq 30$ mm | -0.00376 | -0.00277 | -0.00621 | 0.00290 |
| | (0.00312) | (0.00376) | (0.00496) | (0.00295) |
| Lag P: $> 0- < 10$ mm | 0.000524 | 0.000460 | 0.000783 | 0.000330 |
| | (0.000769) | (0.000794) | (0.00133) | (0.000445) |
| Lag P: $10- < 20$ mm | -0.000865 | -0.000661 | -0.000808 | -0.00137 |
| | (0.00189) | (0.00218) | (0.00354) | (0.00181) |
| Lag P: $20- < 30$ mm | 0.00283 | -0.00171 | 0.00932 | -0.0000912 |
| | (0.00328) | (0.00383) | (0.00622) | (0.00424) |
| Lag P: $\geq 30$ mm | 0.00168 | -0.00147 | 0.00242 | -0.00202 |
| | (0.00392) | (0.00466) | (0.00690) | (0.00303) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.678 | 0.534 | 0.513 | 0.253 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.5: Log Homicide Rate by Victim and Offender Race

| | (1) Homicide Rate: White-White | (2) Homicide Rate: Black-Black | (3) Homicide Rate: White-Black | (4) Homicide Rate: Black-White |
|---|---|---|---|---|
| Temp : $< 10°$ F | -0.0003 | -0.0040 | -0.0003 | 0.0001 |
| | (0.0009) | (0.0029) | (0.0003) | (0.0006) |
| Temp: $10- < 20°$ F | -0.0012 | -0.0059** | -0.0002 | -0.0005 |
| | (0.0011) | (0.0030) | (0.0003) | (0.0004) |
| Temp: $20- < 30°$ F | 0.0004 | -0.0038** | -0.0003 | -0.0001 |
| | (0.0006) | (0.0018) | (0.0002) | (0.0002) |
| Temp: $30- < 40°$ F | -0.0001 | -0.0040** | -0.0001 | -0.0001 |
| | (0.0005) | (0.0018) | (0.0001) | (0.0002) |
| Temp: $40- < 50°$ F | 0.0000 | -0.0004 | -0.0001 | 0.0000 |
| | (0.0003) | (0.0009) | (0.0001) | (0.0001) |
| Temp: $50- < 60°$ F | -0.0004 | -0.0025*** | -0.0000 | -0.0001 |
| | (0.0002) | (0.0008) | (0.0001) | (0.0001) |
| Temp: $70- < 80°$ F | 0.0004* | 0.0004 | 0.0001 | 0.0001 |
| | (0.0002) | (0.0006) | (0.0001) | (0.0001) |
| Temp: $80- < 90°$ F | 0.0010*** | 0.0016* | 0.0001 | 0.0001 |
| | (0.0003) | (0.0010) | (0.0001) | (0.0001) |
| Temp: $\geq 90°$ F | -0.0001 | 0.0008 | 0.0000 | 0.0003 |
| | (0.0005) | (0.0021) | (0.0002) | (0.0002) |
| Lag T : $< 10°$ F | 0.0004 | 0.0055** | 0.0003 | 0.0003 |
| | (0.0013) | (0.0027) | (0.0004) | (0.0004) |
| Lag T: $10- < 20°$ F | 0.0003 | 0.0010 | -0.0001 | 0.0002 |
| | (0.0006) | (0.0019) | (0.0004) | (0.0004) |
| Lag T: $20- < 30°$ F | -0.0000 | 0.0006 | 0.0001 | -0.0002 |
| | (0.0005) | (0.0017) | (0.0002) | (0.0003) |
| Lag T: $30- < 40°$ F | -0.0000 | -0.0009 | 0.0002 | 0.0001 |
| | (0.0005) | (0.0015) | (0.0001) | (0.0002) |
| Lag T: $40- < 50°$ F | 0.0002 | 0.0008 | 0.0000 | 0.0000 |
| | (0.0002) | (0.0010) | (0.0001) | (0.0001) |
| Lag T: $50- < 60°$ F | 0.0003 | -0.0001 | 0.0001* | 0.0001 |
| | (0.0003) | (0.0014) | (0.0001) | (0.0001) |
| Lag T: $70- < 80°$ F | 0.0001 | -0.0006 | 0.0001** | 0.0001 |
| | (0.0002) | (0.0009) | (0.0001) | (0.0001) |
| Lag T: $80- < 90°$ F | 0.0002 | -0.0013 | 0.0001 | 0.0001 |
| | (0.0003) | (0.0009) | (0.0001) | (0.0001) |
| Lag T: $\geq 90°$ F | -0.0002 | -0.0011 | 0.0001 | 0.0001 |
| | (0.0009) | (0.0016) | (0.0002) | (0.0003) |
| Precip: $> 0- < 10$ mm | -0.0001 | 0.0011* | -0.0000 | -0.0000 |
| | (0.0003) | (0.0007) | (0.0001) | (0.0001) |
| Precip: $10- < 20$ mm | 0.0003 | 0.0000 | 0.0000 | -0.0003 |
| | (0.0007) | (0.0016) | (0.0002) | (0.0003) |
| Precip: $20- < 30$ mm | 0.0005 | -0.0014 | -0.0002 | -0.0003 |
| | (0.0012) | (0.0031) | (0.0005) | (0.0005) |
| Precip: $\geq 30$ mm | -0.0008 | 0.0010 | -0.0001 | -0.0002 |
| | (0.0010) | (0.0027) | (0.0004) | (0.0004) |
| Lag P: $> 0- < 10$ mm | -0.0004 | -0.0002 | -0.0000 | 0.0002* |
| | (0.0003) | (0.0006) | (0.0001) | (0.0001) |
| Lag P: $10- < 20$ mm | -0.0006 | 0.0018 | 0.0002 | -0.0004 |
| | (0.0008) | (0.0020) | (0.0003) | (0.0004) |
| Lag P: $20- < 30$ mm | 0.0002 | -0.0031 | 0.0002 | 0.0000 |
| | (0.0015) | (0.0030) | (0.0005) | (0.0005) |
| Lag P: $\geq 30$ mm | 0.0016 | 0.0035 | 0.0002 | -0.0005 |
| | (0.0011) | (0.0032) | (0.0005) | (0.0005) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.445 | 0.519 | 0.166 | 0.183 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.6: Log Homicide Rate by Victim and Offender Race

| | (1)<br>Homicide Rate:<br>White-White | (2)<br>Homicide Rate:<br>Black-Black | (3)<br>Homicide Rate:<br>White-Black | (4)<br>Homicide Rate:<br>Black-White |
|---|---|---|---|---|
| Temperature<br>(Degree Days Over 70F) | 0.0000456$^{*}$ | 0.0000836 | 0.00000191 | 0.00000213 |
| | (0.0000247) | (0.0000564) | (0.00000670) | (0.00000684) |
| Temperature<br>(Degree Days Under 50F) | 0.00000565 | -0.000103$^{**}$ | -0.00000627$^{*}$ | -0.00000534 |
| | (0.0000147) | (0.0000411) | (0.00000372) | (0.00000443) |
| Lag T<br>(Degree Days Over 70F) | -0.00000186 | -0.000103$^{*}$ | 0.00000534 | 0.00000688 |
| | (0.0000204) | (0.0000540) | (0.00000546) | (0.00000675) |
| Lag T<br>(Degree Days Under 50F) | -0.00000359 | 0.0000256 | 0.00000131 | 0.000000282 |
| | (0.0000131) | (0.0000358) | (0.00000332) | (0.00000452) |
| Precip: $> 0- < 10$ mm | -0.000237 | 0.000913 | -0.0000387 | -0.0000260 |
| | (0.000275) | (0.000690) | (0.0000650) | (0.0000786) |
| Precip: $10- < 20$ mm | 0.000519 | -0.000269 | 0.0000418 | -0.000256 |
| | (0.000638) | (0.00159) | (0.000223) | (0.000302) |
| Precip: $20- < 30$ mm | 0.000605 | -0.00151 | -0.000138 | -0.000301 |
| | (0.00121) | (0.00308) | (0.000463) | (0.000485) |
| Precip: $\geq 30$ mm | -0.00101 | 0.00102 | -0.0000986 | -0.000209 |
| | (0.00105) | (0.00268) | (0.000406) | (0.000442) |
| Lag P: $> 0- < 10$ mm | -0.000271 | -0.0000234 | -0.0000220 | 0.000148$^{**}$ |
| | (0.000264) | (0.000575) | (0.0000756) | (0.0000742) |
| Lag P: $10- < 20$ mm | -0.000270 | 0.00177 | 0.000198 | -0.000325 |
| | (0.000778) | (0.00201) | (0.000264) | (0.000395) |
| Lag P: $20- < 30$ mm | 0.000374 | -0.00305 | 0.000215 | 0.0000716 |
| | (0.00148) | (0.00299) | (0.000457) | (0.000528) |
| Lag P: $\geq 30$ mm | 0.00156 | 0.00341 | 0.000213 | -0.000485 |
| | (0.00109) | (0.00316) | (0.000446) | (0.000457) |
| No Obs | 30356 | 30356 | 30356 | 30356 |
| $R^2$ | 0.445 | 0.518 | 0.166 | 0.183 |

Notes:
Regressions weighted by relevant populations.
Standard errors are in parenthesis and are clustered at the county level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# Appendix B

# Appendix

## B.1 Chapter 2 Sensitivity Analysis

Here we present additional results to explore the robustness of our weather-yield model. These tables reflect that temperature drives much of the effect seen on yields, and across higher order specifications the returns to temperature are increasing. When we consider soy grown outside the amazon we find decreasing returns, which is consistent with the literature.

Figure B.1: Temperature Yield Relationships in the Amazon

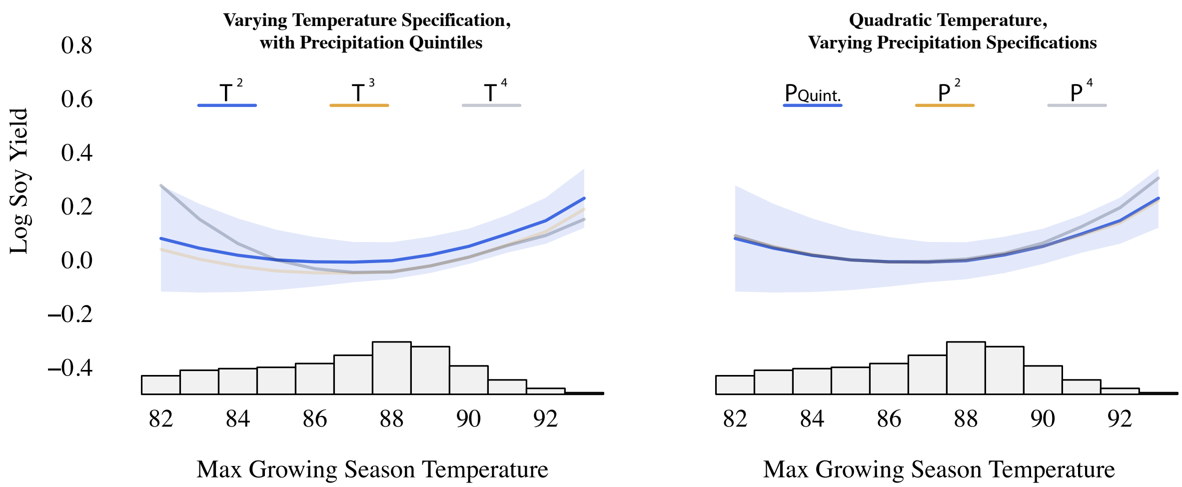



Figure 4 shows the estimated temperature-yield relationships for different models. The left figure allows for higher order temperature specification, keeping precipitation controls fixed. The right more flexible precipitation controls for a quadratic temperature specification. The 95% confidence intervals shown correspond to our primary specification in column (2) from Table 2.2 with quadratic temperature and precipitation quintiles. All regressions have municipal and year fixed effects and a linear state time trend.

Table B.1: Contemporaneous Effects of Growing Season (Nov - April) Weather on Soy Yields in Brazil (Non-Amazon)

| | (1) Log Yield | (2) Log Yield |
|---|---|---|
| Average Maximum Temperature | 0.529*** | 0.409** |
| | (0.173) | (0.167) |
| Average Max Temperature Squared | -0.00304*** | -0.00241** |
| | (0.00101) | (0.000986) |
| Indicator for Rainfall Shock in Lowest Tercile | -0.0796*** | -0.0772*** |
| | (0.0106) | (0.0103) |
| Indicator for Rainfall Shock in Highest Tercile | 0.0508*** | 0.0609*** |
| | (0.00916) | (0.0105) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 22130 | 22130 |
| $R^2$ | 0.544 | 0.551 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table B.2: Contemporaneous Effects of Growing Season (Nov - April) Weather on Soy Yields in Brazilian Amazon

| | (1)<br>Log Yield | (2)<br>Log Yield |
|---|---|---|
| Average Maximum Temperature | -0.695$^{*}$ | -0.963$^{**}$ |
| | (0.416) | (0.376) |
| Average Max Temperature Squared | 0.00411$^{*}$ | 0.00559$^{***}$ |
| | (0.00235) | (0.00213) |
| Total Precipitation | 0.0000224 | 0.0000222 |
| | (0.000104) | (0.000103) |
| Total Precipitation Squared | 1.14e-08 | 1.04e-08 |
| | (3.68e-08) | (3.73e-08) |
| Fixed Effects | Munic, Year | Munic, Year |
| Trend | No | State Linear |
| Weight | Soy-Area | Soy-Area |
| No Obs | 2791 | 2791 |
| $R^2$ | 0.522 | 0.557 |

Notes: Standard errors are in parenthesis and are clustered at the Municipal level.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# Bibliography


Abman, R. and C. Carney (2016). Agriculture productivity and deforestation: Evidence from input subsidies and ethnic favoritism in malawi. *Working Paper*.

Anderson, C. A. (1987). Temperature and aggression: effects on quarterly, yearly, and city rates of violent and nonviolent crime. *Journal of personality and social psychology 52*(6), 1161.

Anderson, C. A. (1989). Temperature and aggression: ubiquitous effects of heat on occurrence of human violence. *Psychological bulletin 106*(1), 74.

Anderson, C. A. (2001). Heat and violence. *Current Directions in Psychological Science 10*(1), 33–38.

Anderson, C. A., B. J. Bushman, and R. W. Groom (1997). Hot years and serious and deadly assault: empirical tests of the heat hypothesis. *Journal of personality and social psychology 73*(6), 1213.

Angelsen, A. (1999). Agricultural expansion and deforestation: Modelling the impact of population, market forces and property rights. *Journal of Development Economics 58*(1), 185–218.

Angelsen, A. and D. Kaimowitz (2001). *Agricultural technologies and tropical deforestation.* CABi.

Arima, E. Y., P. Richards, R. Walker, and M. M. Caldas (2011). Statistical confirmation of indirect land use change in the brazilian amazon. *Environmental Research Letters 6*(2), 024010.

Assunção, J., M. Lipscomb, A. Mushfiq Mobarak, and D. Szerman (2016). Agricultural productivity and deforestation in brazil. *Working Paper*.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013a). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, ret016.

Auffhammer, M., S. M. Hsiang, W. Schlenker, and A. Sobel (2013b). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*.

Barbier, E. B. and J. C. Burgess (2001). The economics of tropical deforestation. *Journal of Economic Surveys 15*(3), 413–433.

Baron, R. A. (1972). Aggression as a function of ambient temperature and prior anger arousal. *Journal of Personality and Social Psychology 21*(2), 183.

Baron, R. A. and P. A. Bell (1976). Aggression and heat: The influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of Personality and Social Psychology 33*(3), 245.

Baron, R. A. and S. F. Lawton (1972). Environmental influences on aggression: The facilitation of modeling effects by high ambient temperatures. *Psychonomic Science 26*(2), 80–81.

Baron, R. A. and V. M. Ransberger (1978). Ambient temperature and the occurrence of collective violence: the" long, hot summer" revisited. *Journal of personality and social psychology 36*(4), 351.

Barona, E., N. Ramankutty, G. Hyman, and O. T. Coomes (2010). The role of pasture and soybean in deforestation of the brazilian amazon. *Environmental Research Letters 5*(2), 024002.

Barreca, A., K. Clay, O. Deschênes, M. Greenstone, and J. S. Shapiro (2015). Convergence in adaptation to climate change: Evidence from high temperatures and mortality, 1900–2004. *The American Economic Review 105*(5), 247–251.

Barreca, A. I. (2012). Climate change, humidity, and mortality in the united states. *Journal of Environmental Economics and Management 63*(1), 19–34.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Brown, J. C., M. Koeppe, B. Coles, and K. P. Price (2005, 2016/05/16). Soybean production and conversion of tropical forest in the brazilian amazon: The case of vilhena, rondônia. *AMBIO: A Journal of the Human Environment 34*(6), 462–469.

Bureau of Labor Statistics (2017). BLS: Labor Force Statistics from the Current Population Survey. https://www.bls.gov/data/.

Burgess, R., O. Deschenes, D. Donaldson, and M. Greenstone (2017). Weather, climate change and death in india. *Working Paper*.

Burke, M., S. M. Hsiang, and E. Miguel (2015). Climate and conflict. *Annual Review of Economics 7*(1), 577–617.

Bushman, B. J., M. C. Wang, and C. A. Anderson (2005). Is the curve relating temperature to aggression linear or curvilinear? assaults and temperature in minneapolis reexamined. *Journal of Personality and Social Psychology 89*(1), 62–66.

Carlsmith, J. M. and C. A. Anderson (1979). Ambient temperature and the occurrence of collective violence: a new analysis. *Journal of personality and social psychology 37*(3), 337.

Chassang, S. and G. Padro-i Miquel (2009). Economic shocks and civil war. *Quarterly Journal of Political Science 4*(3), 211–228.

Cohen, F. and A. Dechezlepretre (2017). Mortality inequality, temperature, and public health provision: Evidence from mexico. *Working Paper*.

Cohen, L. E. and M. Felson (1979). Social change and crime rate trends: A routine activity approach. *American sociological review*, 588–608.

Cohn, E. G. (1990). Weather and crime. *British Journal of Criminology 30*(1), 51–64.

Cohn, E. G. and J. Rotton (1997). Assault as a function of time and temperature: A moderator-variable time-series analysis. *Journal of Personality and Social Psychology 72*(6), 1322.

Compeán, R. G. (2013). Weather and welfare: Health and agricultural impacts of climate extremes, evidence from mexico. *Inter-American Development Bank: Working Paper Series* (391).

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states. *International journal of climatology 28*(15), 2031–2064.

Davis, L. W. and P. J. Gertler (2015). Contribution of air conditioning adoption to future energy use under global warming. *Proceedings of the National Academy of Sciences 112*(19), 5962–5967.

Dell, M., B. F. Jones, and B. A. Olken (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics 4*(3), 66–95.

Dell, M., B. F. Jones, and B. A. Olken (2014). What do we learn from the weather? the new climate–economy literature. *Journal of Economic Literature 52*(3), 740–798.

Deryugina, T. and S. M. Hsiang (2014). Does the environment still matter? daily temperature and income in the united states. Technical report, National Bureau of Economic Research.

Deschenes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics 46*, 606 – 619.

Deschênes, O. (2014). Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics 46*, 606–619.

Deschenes, O. and M. Greenstone (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *The American Economic Review 97*(1), 354–385.

Deschênes, O. and M. Greenstone (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us. *American Economic Journal: Applied Economics 3*(4), 152–85.

Deschênes, O. and E. Moretti (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics 91*(4), 659–681.

Economist (2015a). Midsummer murder: Lack of trust in police forces is contributing to a spike in murder rates. http://www.economist.com/news/united-states/21661020-lack-trust-police-forces-contributing-spike-murder-rates-midsummer-murder.

Economist (2015b). Summer in chicago: Packing heat. http://www.economist.com/blogs/democracyinamerica/2015/07/summer-chicago.

Ewers, R. M., J. P. Scharlemann, A. Balmford, and R. E. Green (2009). Do increases in agricultural yield spare land for nature? *Global Change Biology 15*(7), 1716–1726.

Federal Bureau of Investigation (2016). Crime in the united states, 2015: About the uniform crime reporting program.

Field, S. (1992). The effect of temperature on crime. *Brit. J. Criminology 32*, 340.

Garrett, R. D., E. F. Lambin, and R. L. Naylor (2013). Land institutions and supply chain configurations as determinants of soybean planted area and yields in brazil. *Land Use Policy 31*, 385 – 396. Themed Issue 1-Guest Editor Romy GreinerThemed Issue 2- Guest Editor Davide Viaggi.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 507–548.

Gorner, J. (2016). After 90 killed in august, chicago may soon pass lat year's homicide total. *The Chicago Tribune*.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. Goetz, T. Loveland, et al. (2013). High-resolution global maps of 21st-century forest cover change. *science 342*(6160), 850–853.

Harris, I., P. Jones, T. Osborn, and D. Lister (2014). Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International Journal of Climatology 34*(3), 623–642.

IBGE (2014). Sistema ibge de recuperacao automatica-sidra. http://www.sidra.ibge.gov.br.

Jacob, B., L. Lefgren, and E. Moretti (2007). The dynamics of criminal behavior: Evidence from weather shocks. *The Journal of Human Resources 42*(3), 489–527.

Kaimowitz, D. and A. Angelsen (1998). *Economic models of tropical deforestation: a review*. Cifor.

Kenrick, D. T. and S. W. MacFarlane (1986). Ambient temperature and horn honking a field study of the heat/aggression relationship. *Environment and behavior 18*(2), 179–191.

Kirkos, B. (2016). Fears of a 'long, hot summer' as chicago racks up a deadly record. *CNN*.

Kotlowitz, A. (2016, September). Solving chicago's murders could prevent more. *The New Yorker*.

Krueger, P., S. B. Huie, R. Rogers, and R. A. Hummer (2004). Neighbourhoods and homicide mortality: an analysis of race/ethnic differences. *Journal of Epidemiology and Community Health 58*(3), 223–230.

Larrick, R. P., T. A. Timmerman, A. M. Carton, and J. Abrevaya (2011). Temper, temperature, and temptation heat-related retaliation in baseball. *Psychological Science*.

Lehren, A. and A. Baker (2009). In new york, number of killings rises with heat. *New York Times*.

Leovy, J. (2015). *Ghettoside*. Susan Mallery.

Lobell, D. B., M. Banziger, C. Magorokosho, and B. Vivek (2011, 04). Nonlinear heat effects on african maize as evidenced by historical yield trials. *Nature Clim. Change 1*(1), 42–45.

Macedo, M. N., R. S. DeFries, D. C. Morton, C. M. Stickler, G. L. Galford, and Y. E. Shimabukuro (2012, 01). Decoupling of deforestation and soy production in the southern amazon during the late 2000s. *Proceedings of the National Academy of Sciences 109*(4), 1341–1346.

Montesquieu, C. d. S. B. D. (1748). *The Spirit of The Laws*.

Morton, D. C., R. S. DeFries, Y. E. Shimabukuro, L. O. Anderson, E. Arai, F. del Bon Espirito-Santo, R. Freitas, and J. Morisette (2006, 09). Cropland expansion changes deforestation dynamics in the southern brazilian amazon. *Proceedings of the National Academy of Sciences 103*(39), 14637–14641.

National Statistics Office of Thailand (2006). Survey of population change.

NPR (2016, November). National public radio. http://www.npr.org/sections/thetwo-way/2016/11/30/503867628/deforestation-of-the-amazon-up-29-percent-from-last-year-study-finds.

PRISM Climate Group (2016). Oregon State Universty.

Ranson, M. (2014). Crime, weather, and climate change. *Journal of Environmental Economics and Management 67*(3), 274 – 302.

Regoeczi, W. C. and D. Banks (2014). The nation's two measures of homicide. *U.S. Deparment of Justice, Office of Justice Programs, Bureau of Justice Statistics*.

Rotton, J. and E. G. Cohn (2000). Violence is a curvilinear function of temperature in dallas: a replication. *Journal of personality and social psychology 78*(6), 1074.

Rotton, J. and E. G. Cohn (2003). Global warming and us crime rates an application of routine activity theory. *Environment and Behavior 35*(6), 802–825.

Rudel, T. K., L. Schneider, M. Uriarte, B. L. Turner, R. DeFries, D. Lawrence, J. Geoghegan, S. Hecht, A. Ickowitz, E. F. Lambin, et al. (2009). Agricultural intensification and changes in cultivated areas, 1970–2005. *Proceedings of the National Academy of Sciences 106*(49), 20675–20680.

Sampson, R. J., S. W. Raudenbush, and F. Earls (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science 277*(5328), 918–924.

Schlenker, W. and M. J. Roberts (2009). Nonlinear temperature effects indicate severe damages to u.s. crop yields under climate change. *Proceedings of the National Academy of Sciences 106*(37), 15594–15598.

Smith, K. R., A. Woodward, D. Campbell-Lendrum, D. D. Chadee, Y. Honda, Q. Liu, J. M. Olwoch, B. Revich, and R. Sauerborn (2014). Human health: impacts, adaptation, and co-benefits. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change*, 709–754.

The Chicago Tribune (2017). Crime in chicagoland: Chicago homicides. http://crime.chicagotribune.com/chicago/homicides.

Tropek, R., O. Sedláček, J. Beck, P. Keil, Z. Musilová, I. Šímová, and D. Storch (2014). Comment on äúhigh-resolution global maps of 21st-century forest cover changeäù. *Science 344*(6187), 981–981.

United Nations Office on Drugs and Crime (2016). UNODC Statistics: Homicide Counts and Rates (2000 - 2014). https://data.unodc.org/.

United States Riot Commission (1968). Report of the national advisory commission on civil disorders.

US Census Bureau (2012). Intercensal county estimtates by age, sex, race: 1980 - 1989, 1990 - 1999, 2000 - 2010.

US Census Bureau (2016). Annual county resident population estimates by age, sex, race, and hispanic origin: April 1, 2010 - july 1, 2015.

Villoria, N. B., D. Byerlee, and J. Stevenson (2014). The effects of agricultural technological progress on deforestation: What do we really know? *Applied Economic Perspectives and Policy 36*(2), 211–237.

Vrij, A., J. Van der Steen, and L. Koppelaar (1994). Aggression of police officers as a function of temperature: An experiment with the fire arms training system. *Journal of community & applied social psychology*.

WHO (2014). World malaria report. Technical report, Wolrd Health Organization.