Essays on Measuring Climate Change Damages and Adaptation

by

Max Aidas Vilgalys

Submitted to the Institute for Data, Systems, and Society
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Social and Engineering Systems

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author .....
Institute for Data, Systems, and Society
August 26, 2022

Certified by .....
Jing Li
Assistant Professor of Applied Economics
Thesis Supervisor

Certified by .....
Namrata Kala
Assistant Professor of Applied Economics
Doctoral Committee

Certified by .....
Whitney Newey
Professor of Economics
Doctoral Committee

Accepted by .....
Fotini Christia
Program Chair, Social and Engineering Systems

Essays on Measuring Climate Change Damages and Adaptation

by

Max Aidas Vilgalys

Submitted to the Institute for Data, Systems, and Society
on August 26, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Social and Engineering Systems

Abstract

Through changes in average temperature, precipitation patterns, and extreme weather events, climate change is already causing severe ecological and economic damages. Further warming is expected to have a profound effect on the functioning of ecological and human systems worldwide. While it is a top priority to limit carbon emissions and mitigate future climate change, it is also essential to prepare for damages from climate change in the remainder of this century. Research is needed to understand these impacts, and whether it is possible to adapt to these changes.

In this thesis, I measure damages and adaptation to recent climate change in three essays. First, in joint work with Sylvia Klosin, I develop a novel debiased machine learning approach to measure continuous treatment effects in panel settings. We demonstrate benefits of this estimator over standard machine learning or classical statistics approaches. We apply this estimator to measure the degree of damages from climate change in U.S. agriculture, and find that extreme heat is significantly more damaging than linear models suggest. In the second essay, I measure the degree of adaptation to extreme heat in U.S. agriculture using flexible modeling of weather variables and a debiased machine learning estimator. I demonstrate that my double machine learning approach works well in high-dimensional settings. Applying this estimator to the past thirty years of crop yields, I find evidence of considerable adaptation to extreme heat. Finally, I examine the equity of adaptation to increasing wildfire risk in California. I study how electric utilities' power shutoff decisions correlate with community socioeconomic status and health risk factors.

Thesis Supervisor: Jing Li
Title: Assistant Professor of Applied Economics

Acknowledgments

First thanks goes to Jing, for clarity and order in such a turbulent time. Thank you for encouraging me to explore my interests, and helping me find the tools to work on them. Thanks also to Namrata and Whitney, for your insights and time as I developed this thesis. I am grateful to have learned so much from all of you.

Thank you to everyone who made IDSS a welcoming department. To Cate, Amy, Mine, Bora, Saba, Marco, Saeyoung, Bernardo, Aurora, and all the other students who made the place feel like home. To Beth, Barb, and all the staff for making everything work. To Karene for helping me grow as a teacher. I still feel lucky for the chance to study here.

I owe a great deal to my friends who stuck with me through the last five years. Thank you to all of my roommates who kept me sane throughout the pandemic - thank you to Monica, Wen Wen, Alena, Jamie, and George. And of course, a special thank you to Cathy for keeping me insane. Thank you Miguel for making this year so sweet. Thank you Sylvia, Lisa, Kelsey, and all the other Econ kids who let me into their world. Thank you to Megan, Lizzo, and Beyoncé for all the vibes. Thank you to my friends from graduate school who commiserated with me, and my friends from the real world who still put up with me - Emily, Valerie, Mena, Nadia, Gwynn, Deeksha, Emelyn, Sage, Courtney, Jackie, Michelle, Anne, Irene, Alex, Maggie, Mariah, and everyone else.

Thank you to my parents, Pat and Justas, and my brother Tauras - without you I would never have completed this degree. Specifically, without you taking bets on what I would do after I dropped out, I would not have felt the intense need to finish. Also for the years of love and support, and for motivating me to try and make a difference.

Contents

Chapter 1

Introduction

Anthropogenic climate change will lead to over 2 °C of warming by 2100, without considerable global efforts to limit carbon emissions (IPCC, 2022). Through changes in average temperature, precipitation patterns, and extreme weather events, climate change is already causing severe ecological and economic damages. Further warming is expected to have a profound effect on the functioning of ecological and human systems worldwide. While it is a top priority to limit carbon emissions and mitigate future climate change, it is also essential to prepare for damages from climate change in the remainder of this century.

One important element of preparing for these damages is understanding the economic consequences of climate change. Understanding likely climate impacts is necessary in order to motivate climate mitigation policies and to direct interventions towards those most likely to be impacted. Developing this understanding requires a global scientific effort. The climate science community has spent decades developing and improving models to predict climate under carbon emissions scenarios (Eyring et al., 2016). To translate these expected changes into economic implications, the environmental economics community has contributed many empirical approaches to study the impacts of weather and climate change.

One focus among these economic assessments is to estimate the extent of likely adaptation to climate change. By changing behavior or technologies to adapt to a changing climate, it may be possible to offset a great deal of economic damages. However, adaptation requires

knowledge and/or capital that may not be available. IPCC (2022) finds that in many countries, limited access to capital is already impeding efforts to adapt to climate change. A lack of adaptation may exacerbate global inequality if only higher income communities are able to invest in climate adaptation. Adaptation is therefore an important component of the economic consequences of climate change, both to estimate climate damages and to ensure equitable responses to climate change.

In my dissertation, I study climate change damages and adaptation with three essays. Chapter 1 introduces the problem and provides some context for the remainder of the work. In Chapter 2, I (with co-author Sylvia Klosin) develop a novel debiased machine learning estimator and use it to measure expected damages from extreme heat in United States (U.S.) agriculture. In Chapter 3, I estimate the degree of adaptation to extreme heat in U.S. agriculture using a novel statistical approach, also involving debiased machine learning estimators. In Chapter 4, I consider an example of adaptation to wildfire risk among California electric utilities and study the extent that income is correlated with this adaptation response. Chapter 5 concludes.

This dissertation relates to a long history of attempting to project the economic consequences of climate change. Early work on this topic involved assigning a dollar value to physical damage assessments from literature in other sciences (Tol, 2009). Wary about the ability to extrapolate from these studies, economists now typically measure economic impacts of shifts in weather or climate directly. One of the earliest studies was Mendelsohn, Nordhaus, and Shaw (1994), which studied the implications of climate change for U.S. agriculture and concluded that global warming would be slightly beneficial. The authors came to this conclusion by regressing a cross section of property values on average temperatures. This study spurred years of debate among economists as to whether warming would pose a net positive or negative for agriculture in temperate climates such as the United States.

While pursuing this debate, researchers made considerable progress on methodologies to measure economic damages from weather events. Schlenker and Michael J. Roberts (2006) proposed a parsimonious functional form that captures the nonlinear nature of crop growth better than average temperatures, and Deschênes and Greenstone (2007) advocate for using

panel variation to account for unobservable fixed effects and to isolate the impacts of weather shocks. Schlenker and Michael J Roberts (2009) revisited the question of how climate change might impact U.S. agriculture, using panel variation and the functional form from their earlier work. They concluded that climate change would have severe consequences for U.S. agriculture, a conclusion that is now widely accepted. For a history of progress in modeling other relationships between weather and economic outcomes, Dell, Jones, and Olken (2014) and Hsiang (2016) provide two excellent reviews. A consistent theme is that choosing an appropriate functional form and isolating the impacts of weather shocks are essential to correctly measuring damages.

These modeling approaches are now widely used to estimate economic consequences of climate change, including to develop climate damage functions. A climate damage function translates a climate change scenario into economic outcomes, and is an important tool to assess economic consequences of climate scenarios and to inform climate mitigation policies. Dell, Jones, and Olken (2014) discuss the history of these damage functions, and argue that they should be based on credible measurement of damages from weather shocks using panel variation and nonlinear modeling. Recent contributions in the climate damage function literature have followed this advice. Burke, Hsiang, and Miguel (2015) integrate nonlinear panel estimates of weather damages to form projections of global climate damages. Hsiang et al. (2017) use nonlinear panel analyses across a range of sectors to study the likely impacts of climate change within the United States. Rode et al. (2021) use this approach to empirically estimate how the energy demand sector contributes to the social cost of carbon, another important implication of the climate damage function.

These analyses typically rely on binning temperature variables to capture nonlinear impacts, although this approach has a limited ability to incorporate high-dimensional weather variation. Deschênes and Greenstone (2011) is an early paper to use this approach, separating daily records into bins by average temperature to measure nonlinearities in how temperature influences both mortality and energy consumption. Dell, Jones, and Olken (2014) and Hsiang (2016) advocate for binning temperature variables to nonlinearly model the role of weather variables. This approach is now widely used, including in the damage function studies by

Burke, Hsiang, and Miguel (2015), Hsiang et al. (2017), and Rode et al. (2021). However, this approach has limitations when researchers consider higher dimensional weather variation. As the researcher attempts to flexibly model multiple variables and interactions between them, the dimensionality grows at a rapid rate and OLS fails to give consistent estimates.

Machine learning algorithms are well-suited for capturing nonlinear impacts in high-dimensional settings, such as a climate damage function with high-dimensional weather features. Machine learning approaches like neural networks, random forests, or Lasso can estimate regression functions in high-dimensional settings with minimal functional form assumptions (Mullainathan and Spiess, 2017). Such algorithms work best with large datasets, where there is sufficient variation to train an estimator (Varian, 2014). Researchers now have access to such datasets and the computational resources to analyze them. Through administrative data, weather stations, and satellite observations, there are now many panel datasets including high-dimensional records of weather and economic variables. Machine learning techniques could allow researchers to model complex relationships between high-dimensional weather variation and economic outcomes more effectively than classical approaches like binning variables.

A growing number of environmental economists are using machine learning techniques, although few applications focus on the measurement tasks needed to assess climate damages. This is true more broadly in economics, as machine learning techniques become widespread but are largely limited to processing data or solving prediction problems (Mullainathan and Spiess, 2017). In environmental economics, Jean et al. (2016) uses a machine learning approach to predict missing data, Crane-Droesch (2018) constructs a neural network for panel data to predict crop yields under climate change scenarios, and Knittel and Stolper (2019) uses a random forest to predict customers most likely to respond to an energy efficiency program. Econometricians have developed machine learning techniques for a variety of measurement applications, including approaches to remove bias from naive machine learning estimates (CCDDHNR, 2018; Chernozhukov, Newey, and Singh, 2022a). There are few environmental economics applications using machine learning to measure environmental impacts. Deryugina et al. (2019) use a machine learning approach to measure the mortality

costs of air pollution, and Stetter, Mennig, and Sauer (2022) use a debiased machine learning approach to measure effectiveness of an agricultural intervention. There is an opportunity for broader use of machine learning for measurement in environmental economics, especially given methodological progress in debiased machine learning.

Chapter 2 of this thesis introduces a machine learning approach for panel settings that can be used to measure damages in settings with high-dimensional weather variation. This chapter is joint work with Sylvia Klosin. The existing debiased machine learning literature does not account for common features in panel settings, notably the existence of additive fixed effects in short panels (that is, datasets with few time period observations). In short panels, it is not possible to consistently estimate per-unit fixed effects via dummy variables. Because they do not remove these fixed effects, standard machine learning models are not suitable for measuring the impacts of weather shocks. In this chapter, we introduce a debiased machine learning approach that does account for fixed effects terms via the econometric technique of first differencing. This estimator combines benefits of machine learning, specifically consistent estimation in high-dimensional settings, with benefits of classical econometric approaches for removing additive fixed effect terms. We introduce this estimator and demonstrate its effectiveness in a simulation exercise, and then apply the estimator to study the impact of extreme heat on U.S. corn yields.

To improve our understanding of climate impacts and to inform policies to promote adaptation, a recent literature in climate economics has begun to measure the extent of adaptation to climate change. In an early study to argue about the extent of possible adaptation to climate change, Schlenker and Michael J Roberts (2009) compared elasticities of crop yield with respect to extreme heat across regions. They found similar magnitudes between the relatively cool U.S. North and the relatively warm U.S. south, concluding that adaptation to extreme heat was limited because crops were susceptible to the same damages from extreme heat even after generations in a warmer climate. Other approaches have sought to measure adaptation using time series weather variation. Barreca et al. (2016) studies mortality from heat exposure, comparing elasticities throughout the 20^{th} century to conclude that adaptation greatly reduced the mortality effects of extreme heat. To focus on adaptation to recent

anthropogenic climate change, Burke and Emerick (2016) measures the degree of adaptation to extreme heat by comparing locations that experienced more long-run increases in extreme heat since over several decades. Lemoine (2018) proposes a related approach that partially identifies the degree of possible adaptation by considering the role of ex-ante and ex-post adaptation to heat exposure shocks.

Empirical work on adaptation to climate change has examined many other potential data sources, including historical climate change and simulation studies of future climate change. Meyers and Rhode (2019) and Sutch (2008) and Sutch (2011) present suggestive and narrative evidence, respectively, that the 1934 and 1936 Dust Bowl drought events accelerated hybrid corn seed adoption by farmers in subsequent years, especially in areas that already cultivated a small amount of hybrid corn before the onset of the drought. Olmstead and Rhode (2011) describe the dramatic change in the parts of the US that cultivated wheat, corn, and cotton from 1839 - 2002 as farmers migrated westward, moving through undocumented climate regions along the way. Farmers adopted varieties that were better suited for the different growing conditions in those areas. Costinot, Donaldson, and Smith (2016) use simulation studies to examine the potential for such crop switching to offset future damages. They find that if farmers switch to the most-suited crops, up to 25% of climate damages by the year 2100 could be avoided.

Chapter 3 of this thesis contributes to this literature on measuring adaptation to climate change. My project uses recent variation in climate to identify adaptation, and is most closely related to Burke and Emerick (2016). In this project, I propose a flexible modeling framework that extends the basic estimation strategy from Burke and Emerick (2016). Their strategy relies on the well-established functional form from Schlenker and Michael J Roberts (2009) to model crop yields. In settings where no such functional form exists, my flexible method can be implemented to measure the degree of adaptation without making restrictive assumptions. My framework uses debiased machine learning to estimate high-dimensional functions and measure the degree of adaptation. I apply this estimator to study the degree of adaptation to extreme heat in U.S. corn and soy cultivation.

One of the most important limiting factors to climate adaptation is inequity, as lower income

groups may lack the means to prepare for a changing climate. In their report, IPCC (2022) identifies several settings where inequality and poverty have set "soft limits" on the ability of groups to adapt to climate change. Among environmental advocates, there has long been a call to focus on equity in climate change adaptation (Smit and Pilifosova, 2003; Thomas and Twyman, 2005). Coggins et al. (2021) conducted a review of literature on equity in climate change adaptation and highlighted several examples of work assessing the equity of climate adaptation. Sheller and Leon (2016) use interviews to study how historical inequalities between Haiti and the Dominican Republic impacted government responses to similar environmental crises, and Satyal, Byskov, and Hyams (2021) use environmental justice theory to examine how systemic injustices facing an indigenous group in Uganda undermine adaptation planning. However, Coggins et al. (2021) ultimately conclude that more work is needed in this area, especially in empirical assessment of equity and justice.

In chapter 4, I conduct an empirical analysis of equity in adaptation to climate change among California electric utilities. I study how California investor-owned electric utilities have adapted to rising wildfire risk, specifically with regards to their Public Safety Power Shutoff (PSPS) decisions. In a PSPS, the utility removes electricity from communities to avoid sparking catastrophic wildfires. I focus on how income influences these shutoff decisions. Because it is likely that income may be correlated with objective fire risk as well as shutoff decisions, I examine how utilities treat communities that differ in income but share the same weather-based risk. This highlights the role that utilities have contributed to discrepancies in shutoffs, either by investing more in certain communities or targeting certain communities in shutoff decisions.

In the rest of this proposal, I introduce each essay in more detail. I state the goals for each project and summarize the current status of the project.

1.1 Summaries of Dissertation Chapters

1.1.1 Chapter 2: Estimating Continuous Treatment Effects in Panel Data using Machine Learning with an Agricultural Application (joint with Sylvia Klosin)

This paper introduces and proves asymptotic normality for a new semi-parametric estimator of continuous treatment effects in panel data. Specifically, we estimate an average derivative of the regression function. Our estimator uses the panel structure of data to account for unobservable time-invariant heterogeneity and machine learning methods to flexibly estimate functions of high-dimensional inputs. We construct our estimator using tools from double/debiased machine learning (DML) literature. We show the performance of our method in Monte Carlo simulations and also apply our estimator to real-world data and measure the impact of extreme heat in United States (U.S.) agriculture. We use the estimator on a county-level dataset of corn yields and weather variation, measuring the elasticity of yield with respect to a marginal increase in extreme heat exposure. Our DML procedure finds that extreme heat will cause $17.7 billion in annual damages by the year 2050, under median climate scenarios. OLS finds that damages will be $16.2 billion, and this difference is statistically significant. We find little evidence that this elasticity is changing over time.

1.1.2 Chapter 3: A Machine Learning Approach to Measuring Climate Adaptation

I measure adaptation to climate change by comparing elasticities from short-run and long-run changes in damaging weather. I propose a debiased machine learning approach to flexibly measure these elasticities in panel settings. In a simulation exercise, I show that debiased machine learning has considerable benefits relative to standard machine learning or ordinary least squares, particularly in high-dimensional settings. I then measure adaptation to damaging heat exposure in United States corn and soy production. Using rich sets of temperature and precipitation variation, I find evidence that short-run impacts from damaging

heat are significantly offset in the long run. I show that this is because the impacts of long-run changes in heat exposure do not follow the same functional form as short-run shocks to heat exposure.

1.1.3 Chapter 4: Equity and Adaptation to Wildfire Risk: Evidence from California Public Safety Power Shutoffs

In the past decade, California investor-owned electric utilities have begun implementing Public Safety Power Shutoffs (PSPS) as part of their effort to adapt to increasing risk of catastrophic wildfires. I examine the extent that these decisions are correlated with two measures of community vulnerability: health risk factors and socioeconomic status (SES). I first construct a dataset linking weather, vulnerability indices, and PSPS decisions for electric circuits in California's three largest investor-owned utilities. I show that PSPS is used more frequently in circuits with lower average SES among two of California's major utilities, and circuits with higher average health risk in one of the major utilities. To focus on utilities' decisions, rather than other sources of inequality that may place vulnerable communities in areas with higher wildfire risk, I repeat this analysis after controlling for weather variation. The results are qualitatively similar. I then model the utility's decision problem, as reported in regulatory filings, and measure which components of the model may be responsible for the PSPS decisions. After controlling for weather variation, I find that ignitions are more frequent in low-SES circuits and in lower health risk circuits for one utility. I cannot reject that utilities' estimated costs from declaring PSPS shutoffs or expected damages from wildfires are equitably distributed.

Chapter 2

Estimating Continuous Treatment Effects in Panel Data using Machine Learning with an Agricultural Application

This project is joint work with Sylvia Klosin.

2.1 Introduction

Estimating continuous treatment effects in panel data is essential for many applications. For example, environmental researchers study how health or economic outcomes change with exposure to high temperatures, air pollution, or other continuous environmental factors. In economics, applied work commonly studies the impact of a continuous variable such as price (e.g. for demand models in industrial organization) or distance (e.g. for evaluating programs in urban economics). Researchers seldom observe all relevant variables for each unit in a panel, but can still measure treatment effects by using repeated observations to control for unobservable factors. One popular technique is the fixed effects approach, where researchers model time-constant factors as indicator variables for each unit in the data (Wooldridge

2005). Applications using the fixed effects approach typically rely on linear models.This can can cause significant biases if the linear model is not correctly specified, or if there is heterogeneity in effects. With standard statistical approaches, flexible modeling can result in large standard errors or inconsistent estimation.

Double/debiased machine learning (DML) can maintain small standard errors while flexibly estimating functions, but the existing DML literature has not studied continuous treatment effects in typical panel settings. Machine learning (ML) approaches like Lasso allow researchers to estimate functions with many terms while preserving low standard errors, although these approaches can induce biases. There are a growing number of DML approaches to overcome regularization and overfitting biases from standard ML (CCDDHNR, 2018; Chernozhukov, Newey, and Singh, 2022a; Chernozhukov, Newey, and Singh, 2022b; Rothenhäusler and Yu, 2019). These approaches allow the researcher to form unbiased estimates and valid confidence intervals, while preserving the statistical benefits of ML such as greater flexibility and power. However, current DML estimators either focus on discrete treatments, do not consider the panel nature of data, or use stronger assumptions than those common in applied work.

This paper introduces a new DML estimator that uses common assumptions in panel settings to flexibly measure continuous treatment effects. Our estimator measures the average derivative, using the fixed effects approach to address unobservable per-unit factors and DML to form debiased estimates and valid confidence intervals. The estimator allows for high dimensionality and general treatment effect heterogeneity, and works well in short (few time period) panels. Our assumptions on the panel data match those commonly used in applied work, while our DML approach offers greater statistical power and modeling flexibility than standard linear approaches. Chernozhukov, Goldman, et al. (2017) and Chernozhukov, Newey, and Singh (2022a) also apply DML in panel settings with continuous treatment variables. However, both papers use a random correlated effects approach, which imposes a restriction on the relationship between the covariates and the unobserved factors. Our fixed-effects approach does not impose that restriction.

Our estimator uses first differencing to remove additive fixed effects, and then estimates two

machine learning problems to construct a debiased estimate of the average derivative. After taking nonlinear transformations of input variables, we construct a first-differenced dataset by taking the difference between values in each time period and their lagged values. This technique is widely used to account for the role of the persistent unobservable fixed effects (Wooldridge, 2010). For our DML approach, we apply the estimator from Chernozhukov, Newey, and Singh (2022a). We estimate both the regression function, and an orthogonal machine learner to correct for bias in estimating the derivative of a regression function. We introduce a novel estimation procedure for this second machine learning problem, using an optimization package to solve a regularized minimization problem. We prove that this estimator is asymptotically normal, and give standard errors that account for within-panel-unit correlation.

After introducing the estimator, we show that it performs well via Monte Carlo simulation. We simulate a short panel dataset according to a nonlinear function, and demonstrate that our estimator is able to estimate an average derivative with low bias and valid confidence intervals. Standard OLS and Lasso both introduce bias in this setting, while OLS with a flexible set of polynomial basis functions results in much larger standard errors. We also compare the performance of our estimator to the iterative algorithm from Chernozhukov, Newey, and Singh (2022a), and show that our approach results in lower bias and variance.

We then apply our estimator to study the impact of extreme heat on United States (U.S.) corn yields from 1980-2019. Like Schlenker and Michael J Roberts (2009), we measure the elasticity of corn yield with respect to a marginal increase in extreme heat. Our estimator is well suited to this setting because of the importance of fixed effects modeling to isolate the impact of weather shocks (Deschênes and Greenstone 2007) and the nonlinear relationships between crop yields and continuous environmental factors such as temperature and precipitation (Schlenker and Michael J Roberts 2009). Schlenker and Michael J Roberts (2009) introduce a parsimonious model that captures the response of crop yields to temperature and precipitation. In our approach, we also consider polynomial functions of these terms and the interactions between them.

Confirming results from the literature, we find significant damages from annual extreme

heat shocks and little evidence of adaptation to these damages. Our procedure estimates a elasticity is similar to those with a linear model, and within the range Burke and Emerick (2016) find when they examine how this elasticity differs with long-run and short-run weather variation. With the DML procedure, our standard errors are significantly lower to those from using OLS with the same flexible set of terms. Using our approach, we project $17.7 billion (2017 dollars) in annual damages by the year 2050, under median climate scenarios. Using the weather variation and linear model from Schlenker and Michael J Roberts (2009), we find damages of $16.2 billion. These estimates have 95% confidence intervals of $[17.0, 18.4]$ and $[15.2, 17.1]$, respectively. The difference is statistically significant (p value < 0.001), but much smaller than the difference between different climate models. From least to most severe climate model, we find between $9.4 and $33.2 billion dollars in damages.

We then use this estimator to measure how the elasticity shifts over time, which could provide evidence of adaptation to climate change. We estimate the value of this elasticity in shifting 2-year panels over our sample, and cannot reject that the elasticity remains constant over time. This corroborates findings from Schlenker and Michael J Roberts (2009) and Burke and Emerick (2016) who conclude that there has been limited adaptation to extreme heat in U.S. agriculture. We find evidence of substantial interannual variability in this coefficient. Some of this variability can be explained by large historical weather events, but we are unable to fully explain this variability.

Our work relates to the panel, DML, continuous treatment effect, and climate economics literature. Flexible continuous effects in panel data are under-explored in the theoretical literature relative to binary treatment effects. While there are many interesting flexible panel methods for binary treatment effects, including factor models and synthetic controls, there are no such methods for continuous effects. The paper studying binary treatment effects in panel flexibly that is most related to ours is Belloni et al. (2016). This paper uses a fixed-effects approach, as we do, and allows for the outcome model to be a general function of covariates estimated with ML, but does not allow for general treatment effect heterogeneity as we do. Chernozhukov, Newey, and Singh (2022a) and Chernozhukov, Goldman, et al. (2017) allow for general treatment effect heterogeneity and continuous treatment variables.

However, Chernozhukov, Goldman, et al. (2017) considers the dynamic panel and imposes the potentially strong assumption of sequential exogeneity, and Chernozhukov, Newey, and Singh (2022a) does not consider panels with few observations per unit. Both these papers use a random correlated effects approach, which imposes a restriction on the relationship between the covariates and the unobserved fixed effects. Our fixed-effects approach does not impose that restriction.

This paper focuses on the *average derivative*, although there are many other potentially interesting continuous treatment effects. For example, Colangelo and Y.-Y. Lee (2020) and Klosin (2021) study the dose-response curve, and Chernozhukov, Newey, and Singh (2022a) considers average policy effects, average treatment effects, and the average equivalent variation bound. The methods in our paper could be adapted to other objects, such as average policy effects or conditional average derivatives. We focus on the average derivative because it can be estimated at parametric root n rates and because there are many applications of the average derivative in practice. Many papers estimate an elasticity, which is equivalent to estimating an average derivative when the outcome variable is the logarithm of a raw value. The average derivative is also studied by, among others, Imbens and Newey (2009) and Rothenhäusler and Yu (2019).

Within the DML literature, our approach is most related to recent work in automatic debiased machine learning. CCDDHNR (2018) prove that sample splitting and constructing Neyman-orthogonal moment conditions can yield approximately debiased machine learning estimates in certain settings. Semenova and Chernozhukov (2021) extend the Neyman-orthogonal moment condition approach to several other statistical targets, including structural derivatives. In an approach known as Auto-DML, Chernozhukov, Newey, and Singh (2022a) and Chernozhukov, Newey, and Singh (2022b) give an approximately debiased estimator for a more general class of linear functionals based on the Riesz representation theorem. Our approach is most closely related to Chernozhukov, Newey, and Singh (2022a), but we extend their estimator to consider panel settings with additive fixed effect terms.

The application is also related to a literature in environmental economics on studying the economic impacts of climate change. Rode et al. (2021) estimates elasticities of energy use

with respect to temperature, to construct a global estimate of the social cost of carbon from energy use. Burke, Hsiang, and Miguel (2015) and Hsiang et al. (2017) estimate economic damages of climate change globally and within the U.S., respectively, by combining estimates of elasticities of multiple economic sectors with respect to temperature. These approaches model the impacts of temperature by separating heat exposure into bins and estimating a coefficient for each bin, a model advocated by Dell, Jones, and Olken (2014) and Hsiang (2016). These approaches can have limitations for high-dimensional weather variation, as the number of variables grows exponentially if the researcher wishes to include interactions between weather factors. Our project is the first, to our knowledge, to estimate these elasticities using a machine learning approach, allowing us to flexibly model higher-dimensional variation.

Our project is also related to a growing field applying machine learning techniques in environmental economics. Many researchers have used machine learning for predictive properties, such as forecasting crop yields (Crane-Droesch 2018), filling missing data to track global poverty (Jean et al. 2016), or more effectively assign treatment (Knittel and Stolper 2019). Recent work has begun using machine learners for measurement tasks. Deryugina et al. (2019) uses a machine learning approach to measure the costs of air pollution. Stetter, Mennig, and Sauer (2022) use a de-biased machine learning approach to measure effectiveness of an agricultural intervention. Our paper is the first application in environmental economics, to our knowledge, to use DML to estimate continuous treatment effects.

The paper is structured in the following way. Section 2.2 sets up the framework of the paper, introduces the parameter of interest, and presents our estimator. Simulation design and results are given in Section 2.3. Section 2.4 covers our application. Section 2.5 concludes.

2.2 Estimation

2.2.1 Notation and Definitions

We work in a panel data setting with n individuals and T time periods. As is often the case in economic data, we assume that n is large but T is small. We assume we have

independent and identically distributed data $(W_1, \cdots, W_n)$ where the $W_i = \{(X_i, D_i, Y_i)\}_{t=1}^T$ are copies of a random variable W with support $\{\mathcal{W} = \mathcal{X} \times \mathcal{D} \times \mathcal{Y}\}_{t=1}^T$, with a cumulative distribution function (cdf) $F_{YDX}(Y, D, X)$. We use capital letters to denote random variables and lowercase letters to denote their possible values. For each unit in a population $X_{i,t} \in \mathbb{R}^h$ denotes a vector of covariates, with h potentially large, and $D_{i,t} \in \mathbb{R}$ denotes the treatment variable.

For a given variable X, we use the notation $\Delta X_{i,t} := X_{i,t} - X_{i,t-1}$ for the first difference transformation. For the first difference transformation of a function f of a variable X, we apply the function f before taking the difference: $\Delta f(X_{i,t}) := f(X_{i,t}) - f(X_{i,t-1})$.

Define $|\cdot|_1$ as the ℓ_1 norm; that is, $|\beta|_1 = \sum_{j=1}^p |\beta_j|$ where β_j is the j^{th} component of β and p is the length of β.

2.2.2 Parameter of Interest

We are estimating a general additive fixed effects panel model.

$$Y_{i,t} = a_i + \gamma_0(D_{i,t}, X_{i,t}) + \epsilon_{i,t} \qquad E[\epsilon_{i,t} | a_i, X_{i,1}, \cdots, X_{i,T}, D_{i,1}, \cdots, D_{i,T}] = 0 \tag{2.1}$$

Here a_i represents individual fixed effects, and γ_0 is a flexible function of treatment and covariates. We are generic about the form of γ_0, allowing high-dimensional covariates, interactions between terms, and higher order polynomial terms. We assume that γ_0 is constant throughout time and that it can be estimated well with Lasso. This assumption implies that γ_0 is a sparse linear combination of some transformation of the treatment and covariates. However, we do not assume that the fixed effects a_i are sparse.

We estimate Lasso after applying a set of basis functions to transform $\{D_{i,t}, X_{i,t}\}$ into a high-dimensional set of covariates. We define a $p \times 1$ dictionary of basis functions b that transforms our original vector of covariates, so that $b(D_{i,t}, X_{i,t}) \in \mathbb{R}^p$. Basis functions can include any desired transformations of the covariates, such as polynomial terms or interactions between variables. The assumption that Lasso estimates γ implies that there exists a sparse parameter vector $\beta_0 \in \mathbb{R}^p$ such that $\gamma_0(D_{i,t}, X_{i,t}) = b(D_{i,t}, X_{i,t})'\beta_0$.

We choose these modeling assumptions because they match those commonly used in applied work, while relaxing functional form assumptions on γ_0. We assume the additive error term $\epsilon_{i,t}$ is mean zero conditional on the history of covariates, an assumption called strict exogeneity that is frequently used in applied work (Wooldridge 2010).

Our estimation target is the average of a continuous treatment effect, the moment function m given in (2.2). We consider the average derivative:

$$\tau_0 = \mathbb{E}[m(W_{i,t}, \gamma_0)] = \mathbb{E}\left[\frac{\partial \gamma_0(D_{i,t}, X_{i,t})}{\partial D_{i,t}}\right] \tag{2.2}$$

Note here that the expectation is over both $D_{i,t}$ and $X_{i,t}$, and that the target derivative is only a function of $D_{i,t}$ and not its lagged value. The causal interpretation of 2.2 is the average causal effect of a marginal increase in treatment. The average derivative has been studied by, among others, Imbens and Newey (2009) and Rothenhäusler and Yu (2019). When $y_{i,t}$ is in log scale, this parameter captures the elasticity of y with respect to a marginal change in D.

To account for the fixed effect term a_i, we introduce a first-differenced version of 2.1. In short panels, it is not possible to consistently estimate a_i. By taking a first difference, we can remove the time-invariant factor a_i and consistently estimate γ_0 (Wooldridge 2010).

$$Y_{i,t} - Y_{i,t-1} = \gamma_0(D_{i,t}, X_{i,t}) - \gamma_0(D_{i,t-1}, X_{i,t-1}) + \epsilon_{i,t} - \epsilon_{i,t-1} \tag{2.3}$$

$$\Delta Y_{i,t} = \Delta\gamma_0(D_{i,t}, X_{i,t}) + \Delta\epsilon_{i,t} \tag{2.4}$$

Because we apply the first difference transformation after taking the function γ_0 of our data, our linear representations of γ_0 and $\Delta\gamma_0$ share the same parameter vector β_0. That is, if $\gamma_0(D_{i,t}, X_{i,t}) := b(D_{i,t}, X_{i,t})'\beta_0$, then $\Delta\gamma_0(D_{i,t}, X_{i,t}) := \Delta b(D_{i,t}, X_{i,t})'\beta_0$.

Note that we can express the estimation target in terms of the average derivative of $\Delta\gamma_0$:

$$\tau_0 = \mathbb{E}\left[\frac{\partial \gamma_0(D_{i,t}, X_{i,t})}{\partial D_{i,t}}\right] = \mathbb{E}\left[\frac{\partial \Delta\gamma_0(D_{i,t}, X_{i,t})}{\partial D_{i,t}}\right] \tag{2.5}$$

2.2.3 Estimator

We construct a de-biased estimator of the average derivative using two Lasso estimated high dimensional functions, $\Delta\hat{\gamma}$ and $\hat{\alpha}$. The first, $\Delta\hat{\gamma}$, is an estimate of the regression function $\Delta\gamma_0$ in equation 2.4. This estimation procedure is explained in Section 2.2.3. The second, $\hat{\alpha}$ is a de-biasing term. We describe $\hat{\alpha}$ in more detail and explain how we construct it in Section 2.2.3 after introducing the full estimator. In short one can consider $\Delta\hat{\gamma}$ to be the initial ML estimate of the effect, and the $\hat{\alpha}$ to be an add on that de-biases this initial estimate.

We use a cross-folds procedure to reduce the risk of overfitting. First the researcher chooses the number of splits L ($L = 5$ is commonly used). Then each unit's indices are randomly partitioned into the L equally sized groups. We use ℓ to denote these groups, $\ell = 1, \ldots, L$. Denote observations in group ℓ by W_ℓ. All observations from a single panel unit are placed in the same fold. This is important because the data in different folds should be independent, and there is a dependence within the observations of a single unit. Our functions $\hat{\gamma}_\ell$ and $\hat{\alpha}_\ell$ are trained using observations not in group ℓ. The full estimator takes these two functions $\Delta\hat{\gamma}_\ell$ and $\hat{\alpha}_\ell$ from each fold and uses them to construct our estimate $\hat{\tau}$:

$$\begin{aligned} \hat{\tau} &= \frac{1}{n(T-1)} \sum_{\ell=1}^{L} \sum_{i \in \ell} \sum_{t=2}^{T} \hat{\tau}_{\ell;i,t} \\ \hat{\tau}_{\ell;i,t} &= \frac{\partial \Delta\hat{\gamma}_\ell(D_{i,t}, X_{i,t})}{\partial D_{i,t}} + \hat{\alpha}_\ell(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1})(\Delta Y_{i,t} - \Delta\hat{\gamma}_\ell(D_{i,t}, X_{i,t})) \end{aligned} \tag{2.6}$$

To compute the asymptotic variance of our estimator, it is necessary to account for correlation of our target within panel units. We assume that errors have a constant correlation within a panel unit but are uncorrelated between panel units. Let $\hat{\tau}_{\ell;i} = 1/(T-1)\sum_{t=2}^{T} \hat{\tau}_{\ell;i,t}$. Then the asymptotic variance is:

$$\hat{V} = \frac{1}{n(T-1)} \sum_{\ell=1}^{L} \sum_{i \in \ell} \left\{ \sum_{t=2}^{T} (\hat{\tau}_{\ell;i,t} - \hat{\tau})^2 + 2 \sum_{t=2}^{T-1} \sum_{t'=t+1}^{T} (\hat{\tau}_{\ell;i,t} - \hat{\tau}_{\ell;i})(\hat{\tau}_{\ell;it'} - \hat{\tau}_{\ell;i}) \right\} \tag{2.7}$$

Assumptions and the proof for asymptotic normality of our estimator are given in Appendix 2.A. We work with $(T-1)$ time periods rather than T because we removed one time period

by first-differencing the data.

In the following subsections, we describe how we compute the derivative of the regression function and how we estimate the bias correction term α.

Derivative of Regression Function

There are two general steps to find the derivative of the regression function , $\frac{\partial \Delta\hat{\gamma}_\ell(D_{i,t},X_{i,t})}{\partial D_{i,t}}$, in 2.6.

1. Estimate $\Delta\hat{\gamma}_\ell$ for each fold.

   (a) Transform the covariates $\{D_{i,t}, X_{i,t}\}$ using a flexible specification. We do so by using polynomial basis functions of terms and interactions, although other approaches like kernel functions or splines could be used as long as the derivatives are bounded. Let $b(D_{i,t}, X_{i,t})$ denote the resulting $p \times 1$ dictionary of functions [1]. Then let $\Delta b(D_{i,t}, X_{i,t}) := b(D_{i,t}, X_{i,t}) - b(D_{i,t-1}, X_{i,t-1})$. We set each function in the dictionary b to have mean 0 and variance 1; Appendix 2.B.2 discusses this procedure.

   (b) Find a vector of coefficients $\hat{\beta}$ for our dictionary such that $\Delta\hat{\gamma}(D_{i,t}, X_{i,t}) := \Delta b(D_{i,t}, X_{i,t})'\hat{\beta}$ is a sparse linear approximation of $\Delta\gamma_0(D_{i,t}, X_{i,t})$. We do so by solving the following Lasso problem:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T} (\Delta Y_{i,t} - \Delta b(D_{i,t}, X_{i,t})'\beta)^2 + r_L|\beta|_1 \right\} \tag{2.8}$$

   This procedure depends on the regularization weight r_L, which we determine by finding values that minimize test-set error in a cross-folds procedure. This procedure is described in Appendix 2.B.1.

2. Calculate the derivative.

   We calculate the derivative analytically. Many DML approaches use numeric differentiation; we discuss this alternative in Appendix 2.B.4. Our procedure uses the estimate

[1] For example, when we estimate $\Delta\gamma_\ell$ we set $b(D_{i,t}, X_{i,t})$ to be a third order polynomial set of the covariate variables and interactions between D and each covariate in X.

of $\hat{\beta}_\ell$ from the previous step to compute the derivative and derivatives of each function in our dictionary of basis functions.

(a) Construct the dictionary b_D, a $p \times 1$ dictionary of derivatives of each basis function in b. For each basis function b^j for $j = 1, \ldots, p$ in our dictionary of basis functions, define its derivative as follows:

$$b_D^j(D, X) = \frac{\partial b^j(D, X)}{\partial D} \tag{2.9}$$

(b) Estimate the average derivative as:

$$\mathbb{E}\left[\frac{\partial \Delta\hat{\gamma}(D_{i,t}, X_{i,t})}{\partial D_{i,t}}\right] = \mathbb{E}[b_D(D_{i,t}, X_{i,t})'\hat{\beta}] \tag{2.10}$$

Example 1. *Consider a simple setting where $X_{i,t} \in \mathbb{R}$, and where $\gamma_0(D_{i,t}, X_{i,t}) = D_{i,t}^2 X_{i,t}$. Our basis function dictionary is $b(D_{i,t}, X_{i,t}) = \{D_{i,t}, X_{i,t}, D_{i,t}^2 X_{i,t}\}$. In our linear representation, $\beta_0 = \{0, 0, 1\}$.*

In step 1, we obtain an estimate $\hat{\beta}$ using Lasso. In step 2, we first define the derivative of the basis functions. Here, $b_D(D_{i,t}, X_{i,t}) = \{1, 0, 2D_{i,t}X_{i,t}\}$. The estimated average derivative is then: $\hat{\beta}_1 + \mathbb{E}[2D_{i,t}X_{i,t}]\hat{\beta}_3$, where $\hat{\beta}_j$ is the j^{th} component of $\hat{\beta}$.

De-biasing term

Now the next term of equation (2.6) is the de-biasing term, $\hat{\alpha}_\ell(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1})$. Our α is based on the methods of Chernozhukov, Newey, and Singh (2022a), which use the Riesz Representation theorem. This theorem states that for a given linear functional m, there exists a function α such that the following holds for any function b:

$$\mathbb{E}[m(W, b)] = \mathbb{E}[\alpha(W)b(W)] \tag{2.11}$$

In our case using the definition of the moment and our derivative bases, we have $m(W, b) = \Delta b_D(D_{i,t}, X_{i,t})$. These basis functions are the same standardized basis functions as in Section 2.2.3. Plugging this into the above yields:

$$\mathbb{E}[\Delta b_D(D_{i,t}, X_{i,t})] = \mathbb{E}[\alpha(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1})\Delta b(D_{i,t}, X_{i,t})] \tag{2.12}$$

Because this equality holds regardless of the function b, it is possible to estimate α from data independently of estimating the function γ. In order to construct an estimate $\hat{\alpha}$, we first assume that α_0 has a sparse linear form: $\alpha_0(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1}) = \Delta b(D_{i,t}, X_{i,t})'\rho_0$. Our estimate is then $\hat{\alpha}(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1}) = \Delta b(D_{i,t}, X_{i,t})'\hat{\rho}$. We follow Chernozhukov, Newey, and Singh (2022a) and find $\hat{\rho}$ to minimize the squared loss between α_0 and $\hat{\alpha}$:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} \left\{ \mathbb{E}[(\alpha_0(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1}) - \Delta b(D_{i,t}, X_{i,t})'\rho)^2] + r_\alpha |\rho|_1 \right\} \tag{2.13}$$

Additional details about this solution are given in Appendix 2.B.3. Chernozhukov, Newey, and Singh (2022a) provides an iterative procedure to solve this problem; we also introduce an alternate approach using an optimization package to solve the minimization problem.

Our implementation using an optimization package guarantees that we find an optimal solution to this minimization problem. Iterative approaches may fail to converge to the true value of the parameter, with finite training time. Given the convex nature of the minimization problem, we are able to find an exact solution using optimization software. We use the Python package CvxPy (Diamond and Boyd 2016) to set up the problem and the convex optimization software Mosek (ApS 2021) to solve the problem. In simulation trials, we compare the performance of iterative and optimizer approaches for determining $\hat{\alpha}$.

With an estimate $\hat{\alpha}$, we can estimate the average derivative using only the $Y_{i,t}$ observations: $\hat{\tau} = \mathbb{E}[\hat{\alpha}(D_{i,t}, X_{i,t}, D_{i,t-1}, X_{i,t-1})\Delta Y_{i,t}]$. We now provide an example to illustrate the role of the Riesz Representer function in a simple setting.

Example 2. *Consider the average derivative of a function γ of X where $X \sim N(0,1)$. Via integration by parts after expanding the expectation, $\mathbb{E}\left[\dfrac{\partial \gamma(X)}{\partial X}\right] = \mathbb{E}[X\gamma(X)]$.*

That is, the Riesz Representer is $\alpha_0(X) = X$ when $X \sim N(0,1)$ and m is the derivative operator. The Riesz Representer depends on the distribution of data and the operator m, but does not depend on the choice of function γ.

If our basis function is the identity (i.e. $b(X_i) = X_i$), then in our linear representation above, $\rho_0 = \{1\}$. We construct an estimate $\hat{\rho}$ using our Lasso procedure, and then estimate the average derivative as: $\hat{\tau} = \mathbb{E}[X_i \hat{\rho} Y_i]$

2.3 Simulations

We demonstrate the performance of our estimator through a Monte Carlo simulation exercise. We use the data generating process (DGP) defined below to create 1000 different datasets. We report the true value of the derivative, our estimate, the average bias of our estimates, and mean squared error between true and estimated values (MSE τ). We also include mean squared error of prediction for in-sample values (MSE γ In Sample) and out-of-sample values (MSE γ Cross Folds). We compare the performance of our estimator with OLS on the untransformed set of covariates (OLS Linear), OLS on the basis function transformation of covariates (OLS Poly), Lasso without a bias correction term, and DML using an iterative estimation procedure (DML Iterative). We show the results of this simulation trial in Table 2.1 and plot the distribution of the error in Figure 2-1.

DGP : $N = 1000$ number of individuals, $T = 2$ number of time periods, $h = 20$ original number of X covariates. Our basis function transformation takes third order polynomials of each variable, then adds interactions between each $\{D, X^{(j)}\}$ pair and their polynomials. We have a total of $p = 244$ covariates after applying the basis function transformation.

We generate outcome variables according to the following function:

$$Y_{i,t} = a_i + D_{i,t} + D_{i,t}^2 + D_{i,t}^3 + D_{i,t} X_{i,t}^{(1)} + .1\theta \mathbf{X}_{i,t} + \epsilon_{i,t} \tag{2.14}$$

To match real-world panel settings, we impose a correlation between $a_i, D_{i,t}$ and $\mathbf{X}_{i,t}$. We set θ so that the j^{th} element is $\theta_j = 1/j^2$. Fixed effects a_i, covariates $\mathbf{X}_{i,t}$, and random noise $\epsilon_{i,t}$ are drawn from Gaussian distributions: $a_i \sim N(1,1)$, $X_{i,t}^{(j)} \sim N(a_i, 1)$ $j = 1, \ldots, h$, and $\epsilon_{i,t} \sim N(0,1)$, while the treatment is correlated with $\mathbf{X}_{i,t}$ but includes simulation draws from a Beta distribution: $D_{i,t} \sim .1\theta \mathbf{X}_{i,t} + Beta(1,7)$

method	DML	DML Iterative	Lasso	OLS Linear	OLS Poly
True Value	2.96	2.96	2.96	2.96	2.96
Average Derivative	2.958	2.94	2.683	3.246	2.939
Bias	-0.002252	-0.02015	-0.2771	0.2861	-0.0208
Standard Deviation	0.2991	0.341	0.3581	0.3311	0.5573
MSE τ	0.08013	0.1069	0.1957	0.1836	0.3009
Coverage	0.924	0.962	0.224	0.886	0.95
MSE γ In Sample	1.951	1.951	1.951	2.338	1.515
MSE γ Cross Folds	2.048	2.048	2.048	2.454	10.04

Table 2.1: Summary of derivative estimates from 1000 bootstrap trials of our simulation procedure. Bias is the average of the estimated value of the derivative minus the true value of the derivative in each simulation draw. "MSE τ" is the mean squared error between the true average derivative and the estimated average derivative in each simulation draw, while "MSE γ in sample" and "MSE γ cross folds" refer to the mean squared error of regression from own-sample and out-of-sample estimation.

Table 2.1 demonstrates the value of our approach for applied research. If an applied researcher were faced with our simulated data with 20 covariates and they wanted to use classic OLS methods they could run either a classic linear regression with 20 covaraiates ("OLS Linear") or they could run a more flexible linear regression with polynomial terms of these 20 base covaraites leading to 244 covariates ("OLS Poly"). Note that in order to recover the average derivative using OLS Poly, the researcher must use either derivatives of the basis functions, as described in Section 2.2.3, or numerical differentiation.

Researchers may prefer OLS Linear to OLS Poly because the large number of regressors in OLS Poly can can lead to large standard errors, even though OLS Linear may be biased due to misspecification. The bias and variance trade off of OLS Linear and OLS Poly is exactly what we see in the table. OLS Linear has smaller standard errors than OLS Poly, but OLS Linear has a large bias relative to OLS Poly.

Our DML method provides an alternative that preserves the low standard errors of OLS Linear and the low bias of flexible modeling such as OLS Poly. Due to the regularization of Lasso, our estimates have considerably lower standard errors as OLS Poly even though they use the same set of basis functions. Figure 2-1 visualizes this benefit - the error from DML and OLS Poly are both approximately centered around 0, but the distribution of errors using OLS Poly is much wider. In this simulation trial, our DML method has lower bias than OLS

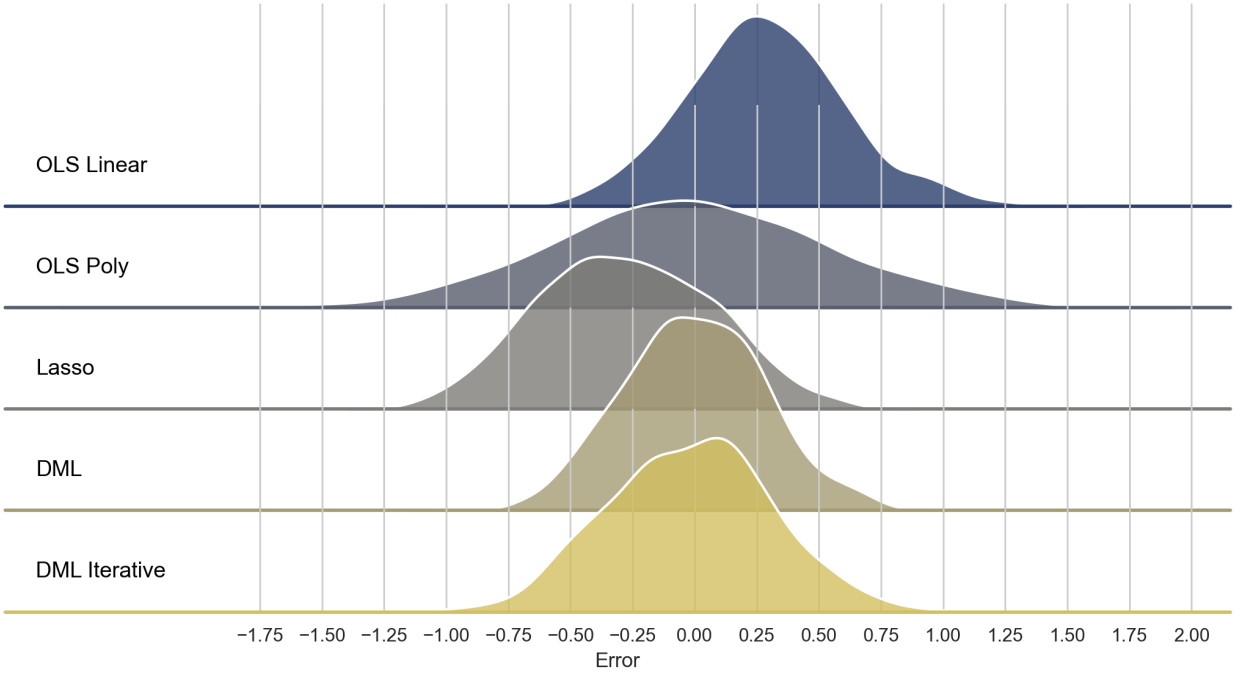


Figure 2-1: Visualization of distributions of error per simulation trial, for the five methods above. Error is defined as estimated value minus true value.

Poly and lower standard errors than OLS Linear.

Reassuringly, our DML procedure results in an approximately de-biased estimate of the average derivative. This is especially clear relative to Lasso without any correction. The bias of base Lasso is 100 times larger than the bias of our proposed method. Our estimates also have substantially lower bias than OLS using the untransformed covariates ("OLS Linear") and lower bias than OLS using the basis function transformation ("OLS Poly"). We expect these results, as misspecification of OLS Linear can induce a bias in estimating the average derivative, and OLS Poly may be overfitting the data (as shown by the high error in MSE cross folds). Our optimization-based DML approach results in lower bias and standard deviation in the result than DML Iterative, although both dramatically improve over Lasso without a correction. This comparison was done by creating our estimator and then using two different algorithms to solve for the de-baising term.

We use the mean squared error of estimating the true parameter (MSE τ) to compare results, incorporating differences in both bias and variance. The MSE τ using our estimation procedure is the lowest among all models considered, roughly half the magnitude from Lasso

or OLS Linear and one quarter the magnitude from OLS Poly. The MSE τ is closest to the result using DML Iterative, and shows that our general optimization solution to de-biasing is leading to lower bias.

Note that simply comparing mean squared error in prediction (MSE γ) does not find the model with the lowest error in estimating τ. OLS Poly has the lowest in-sample MSE γ, but the highest MSE γ in out-of-sample prediction. This indicates that OLS Poly may be overfitting the data, which is unsurprising given the high value of p relative to the number of observations. Lasso has higher in-sample MSE γ than OLS Poly, but the lowest out-of-sample MSE γ. However, Lasso has substantially higher bias than OLS Poly.

In Appendix 2.C, we include results from additional simulation trials where we vary the number of covariates or the number of time periods. These results show that, as expected, the bias generally decreases as the number of samples per covariate increases. Both DML and DML Iterative continues to reduce bias relative to Lasso in all trials, and our DML approach has the lowest MSE τ in all trials. This confirms that our approach has benefits over standard OLS procedures and current DML approaches.

2.4 Application

In our application, we measure the elasticity of corn yields with respect to extreme heat exposure and use this to project the damages from expected changes in the distribution of extreme heat following climate change. We do so by regressing log crop yields on a set of weather variables. The treatment variable $D_{i,t}$ is annual aggregate exposure to temperatures above 29 °C; for simplicity we refer to this variable as extreme heat. This variable roughly captures the amount of heat stress a plant experiences. As Schlenker and Michael J Roberts (2009) demonstrate, crop yields are generally decreasing in extreme heat, while increasing in heat exposure below a crop-specific damaging threshold. Schlenker and Michael J Roberts (2009) find that this threshold is 29°C for corn, and Burke and Emerick (2016) find that the threshold is 29°C for the long-run impacts of heat exposure. In Section 2.D.1, we consider alternate temperature thresholds and find that model fit slightly improves with alternate thresholds; we use 29°C for the closest comparison to the existing literature. The covariates

$X_{i,t}$ include other weather features, such as heat exposure below this temperature threshold and precipitation. Their model is:

$$y_{it} = a_i + \beta_1 D_{it} + beta_2 X_{it} + \varepsilon_{it} \tag{2.15}$$

where y_{it} is log crop yields, D_{it} is the treatment of interest, extreme heat exposure, and $X_{it} \in \mathbb{R}^2$ is the vector of temperature and precipitation. In robustness checks, they demonstrate that this model performs comparably to models including flexible specifications of temperature.

In our estimation, we include models that include interactions between all weather variables. Interaction terms between variables are of particular interest. Schlenker and Michael J Roberts (2009) test some models including interactions between temperature and precipitation, but find that those models have limited improvements for out-of-sample prediction. We examine whether including these terms changes the coefficient estimates.

We also use this estimator to evaluate how the target elasticity changes over time. Barreca et al. (2016) measure the mortality-temperature relationship in the U.S. over time to estimate the degree of adaptation to high temperatures. We conduct a similar exercise to evaluate the extent that this crop yield-extreme heat relationship changes over time.

2.4.1 Data Description

We work with two datasets for the main estimation: corn yield from the USDA's Survey of Agriculture, and weather records from Abatzoglou (2013). To form projections about the impacts under climate change, we use data generously shared from Burke and Emerick.

For corn yields, we use county-level reports from the USDA Survey of Agriculture, supplemented with additional data generously provided by Burke and Emerick (2016). We use the metric bushels per acre, and focus on U.S. counties east of the 100°West meridian. This region is commonly used in studies of U.S. agriculture because the region west of this meridian relies on heavily subsidized irrigation. In our sample period, over 90% of corn in the U.S. was grown in this region. This dataset also includes the planted area of corn per county in

each year.

For weather observations, we use the GridMET weather dataset from Abatzoglou (2013). This dataset contains a rich set of daily weather variables since 1979 at high spatial resolution (4 km) across the United States. In our main results we follow Schlenker and Michael J Roberts (2009) and include only precipitation, beneficial heat exposure, and damaging heat exposure. In Appendix 2.D.2, we include additional results with the dataset's 9 raw variables[2]: specific humidity, precipitation, minimum relative humidity, maximum relative humidity, surface downwelling shortwave flux in air (a measure of solar radiation), minimum air temperature, maximum air temperature, wind speed, and wind direction. We generate a county-level daily dataset by taking the average of daily weather from all grid observations within each U.S. county.

We aggregate daily weather observations to the March-August growing season. We take averages of the variables specific humidity, minimum relative humidity, maximum relative humidity, surface downwelling shortwave flux in air (a measure of solar radiation), wind speed, and wind direction. We take the sum of precipitation. We then construct the variables beneficial and damaging heat exposure, total growing season heat exposure below and above (respectively) 29 °C. Heat exposure is measured in Growing Degree Days (GDD), the amount of time a crop is exposed to temperatures between an upper and lower bound during the March-August growing season. This gives us a county-level dataset of weather variation for our empirical application.

2.4.2 Empirical Results

We compare the estimated elasticity of corn yield with respect to extreme heat using our DML procedure, Lasso, OLS Linear, and OLS Poly. We regress log corn yields on a set of weather covariates, and recover the average derivative of log corn yields with respect to increasing extreme heat exposure. In our primary results the set of weather covariates is the widely-used set of weather variables from Schlenker and Michael J Roberts (2009): total growing season precipitation and heat exposure below and above 29 °C. We use this speci-

[2]GridMET also includes several derived variables, such as Energy Release Component, for specific ecological applications.

fication in the main analysis as it allows for the cleanest comparison between DML and the current literature. In Appendix 2.D.2, we include results using additional GridMET observations at a growing-season level. The central magnitudes using these additional weather data are similar, although the differences between OLS and DML estimates is smaller and not statistically significant.

Our estimation procedure uses the standardization, differencing, cross folds, analytical derivative, and de-biasing procedures described in Section 2.2. Standard errors are clustered at the county level for all analyses. For our basis functions, we use third order polynomial expansions and interactions up to depth 2 of all variables and polynomial expansions. Estimation of Lasso for both the regression problem and the Riesz representer uses the optimization package Mosek (ApS 2021).

We weight each observation by the area of corn planted when computing coefficients or average derivatives. This choice is common in the applied literature, as the object of interest is the elasticity of yield in an acre of corn with a marginal increase in damaging heat exposure. We adjust the Lasso objective function to find the weighted mean squared error, and use the weighted average of observations when estimating the Riesz representer. Appendix 2.D.3 includes results without this weighting choice; in these results, DML estimates a greater degree of damages.

Table 2.3 shows the results from the main analysis. The average derivative can be interpreted as the elasticity of corn yield with respect to additional exposure to extreme heat. That is, each estimate is the percentage by which yields change with an additional growing degree day of heat exposure above 29 °C over the growing season. The magnitude is relatively large - this suggests that an increase of a single growing degree day is associated with corn yields declining by 0.51%-0.58%. This finding is line with findings form other analyses of damages to crop yields from extreme heat.[3]

Mean squared error (MSE) from both in-sample and out-of-sample prediction is slightly higher for OLS Linear than any of the flexible methods. There is not a large difference

[3]Burke and Emerick, 2016 estimate this same elasticity in a range from -0.0036 to -0.0062, depending on the specification.

Table 2.2: Temperature and Precipitation

method	OLS Linear	OLS Poly	Lasso	DML
Average Derivative	-0.005193	-0.005657	-0.005821	-0.005823
	(0.000099)	(0.000135)	(0.000011)	(0.000073)
MSE γ In Sample	0.080929	0.077958	0.077878	0.077878
MSE γ Cross Folds	0.080975	0.078353	0.078079	0.078079
Number of Observations	63662	63662	63662	63662
Number of Covariates	3	36	36	36

Table 2.3: Estimates of elasticity of corn yields with respect to increase in growing season exposure to extreme heat, using two sets of weather covariates. Standard errors are in parentheses. See text for estimation details.

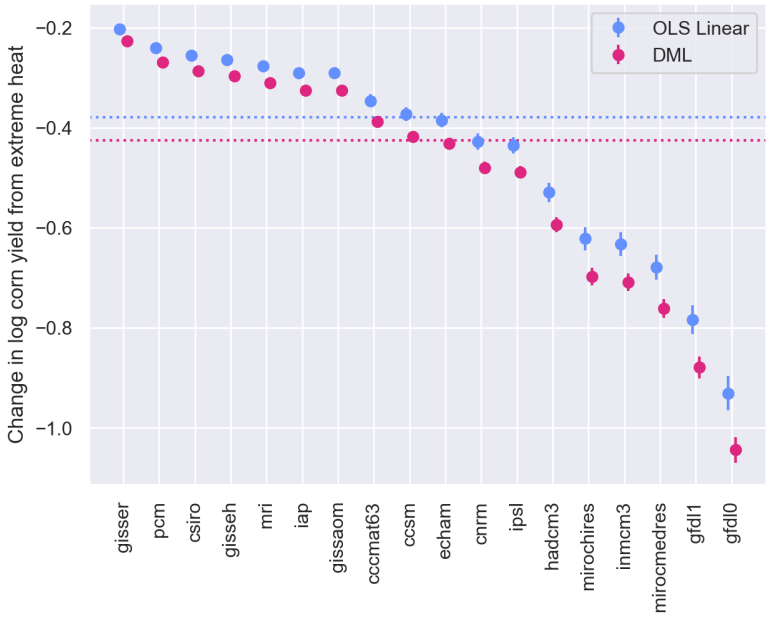


Figure 2-2: Temperature and Precipitation

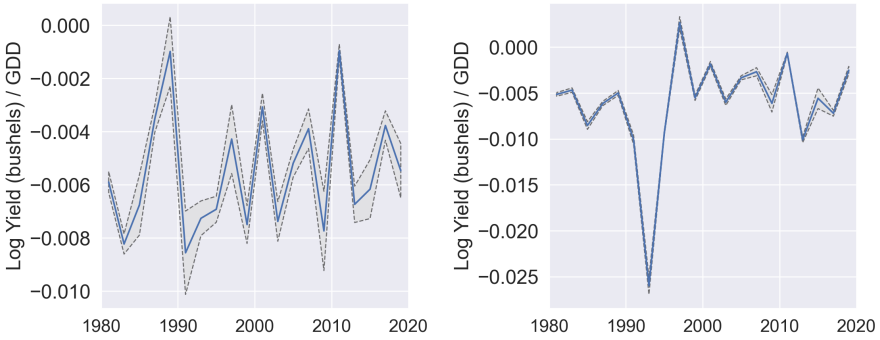
Figure 2-3: Extrapolating impacts of extreme heat to crop yields by the year 2050, using elasticities from OLS Linear and DML. Each dot represents a central estimate from a model, and the error bar represents the 95% confidence interval. Dotted line represents the median value across climate models. See text for estimation details.

between the MSE of the OLS Poly and Lasso models. This indicates that, for these sets of weather variables, OLS Poly does not over-fit the data.

To illustrate the significance of these estimates, Figure 2-3 shows the projected overall change in crop yield by 2050 due solely to an increase in damaging heat exposure. We apply the estimated elasticity to climate projections from a range of scenarios. To generate these projections, we find the weighted average degree of expected warming from 2015 to 2050 under various climate scenarios, where we weigh the degree of warming by the average area of corn cultivated per county. The expected degree of damages is then the degree of expected warming multiplied by the coefficient from Table 2.3. Climate scenarios are derived from 18 global climate models running the A1B emissions scenario. The per-county degree of expected warming under each climate scenario are generously shared by Burke and Emerick (2016). We compare the OLS Linear and DML estimates. In Appendix 2.D.2, we include the same figure using all GridMET weather covariates.

These results illustrate that modeling assumptions significantly impact overall damage projections with the simple set of weather variables. The projected damages are quite significant - in the median emissions scenario, log yields are 0.38-0.43 lower than a world that does not experience climate change. There is also considerable variance in these estimates; from the least to most severe climate scenario, log yields decline from 0.20-1.04. To put this in dollar value, the 2017 Census of Agriculture reported the total value of sales of corn in the United States as $ 51.2 billion . The median range of damage estimates translates to a dollar value of $16.2-$17.7 billion (in 2017 dollars), and the overall range is $9.4-$33.2 billion.

The difference between OLS Linear and DML estimates is statistically significant and economically meaningful. OLS Poly and DML find similar results, although the standard error from OLS Poly is substantially higher than that of DML. OLS Linear, in the median specification, finds a 95% confidence interval of $[15.2, 17.1]$ billion dollars in damages; DML finds $[17.0, 18.4]$. The difference between OLS Linear and DML estimates is significant (p value < 0.001), as is the difference between OLS Linear and OLS Poly estimates (p value 0.0055). Median estimates using DML instead of OLS Linear correspond to an additional $1.55 billion in damages.



(a) OLS Linear
Trend 3.43e-05, p-val 0.433

(b) DML
Trend 9.52e-05, p-val 0.387

Figure 2-4: Estimating elasticity of corn yield with respect to extreme heat over time. We use our estimation procedure on 2-period samples from 1980 through 2019. Line shows central estimate, and grey band shows 95% confidence interval.

These values overstate the degree of damages from climate change for several reasons, but still provide a valuable illustration of the economic significance of using this measurement technique. This is not a complete projection of climate change damages, but only highlights the contribution of the elasticity estimated above using various modeling assumptions. Also note that an elasticity is the marginal impact at the observed distribution of weather characteristics; as the distribution of weather patterns changes with climate change, this elasticity will likely change. Additionally, this estimate does not attempt to account for adaptation to climate change.

We use our estimation procedure to study how the elasticity has changed over time, using 2-period panels from 1980 through 2019. A key advantage of our DML procedure is that it allows de-biased estimation of flexible functions, even in relatively short panels. This allows us to study how the elasticity has changed over time. This research design is similar to Barreca et al. (2016), who study the mortality response from temperature during separate decades across the 20th century to conclude that there has been significant adaptation to extreme heat exposure.

Our estimates suggest that the degree of damages from extreme heat has remained approximately stable from 1980 to 2019. Figure 2-4 shows these estimates, as well as the value and statistical significance of the trend in elasticity. The trend and standard errors are computed using weighted least squares, weighted by the inverse of the sum of the variance of the

elasticity estimate and the residual from the regression of elasticity on the year. Using both methods with both sets of weather variables, we are unable to reject the null hypothesis at the $p = 0.05$ level, suggesting little adaptation to climate change. This finding is in line with the literature on adaptation in U.S. agriculture. Schlenker and Michael J Roberts (2009) and Burke and Emerick (2016) also find limited evidence of adaptation to extreme heat.

While there is no statistically significant trend, the variability of the elasticity is somewhat surprising. In Appendix 2.D.2, we show that this variability persists after incorporating additional weather variation. Appendix 2.D.4 includes some additional regression results, and show that this variation cannot be explained by including state-by-year fixed effects or by removing the interactions or polynomial terms from the basis functions. Significant events in U.S. agriculture help explain some patterns in the data. The large negative elasticity from 1992-1993 in Figure 2-4b is likely due to significant flooding in 1993, leading to a negative correlation of heat and crop yields when heat is interacted with precipitation. Similarly, a 1996 drought in the Midwest may be responsible for the relatively high elasticity estimated from 1996-1997. The remaining variation could be explained by regional differences or extreme heat occurring during different phases of plant growth (Ortiz-Bobea, 2013; Butler and Huybers, 2015; Ortiz-Bobea et al., 2019).

2.5 Conclusion

In this paper, we presented the first DML based estimator for continuous treatment effects in the classic fixed effect panel setting. It is also the first DML estimator in the fixed effects panel setting to allow for general unrestricted heterogeneity in treatment effects. Our estimator is one of the first methods for flexibly estimating continuous treatment effects in high-dimensional panel data, and our estimation approach improves upon the current state-of-the-art methods in simulations. We gave clear guidance on how to implement our methods so they can be of use to an applied audience.

Our estimator is of broad applied interest - it allows for fixed effects in panel data, unrestricted heterogeneity in treatment effects, and ML tools, and keeps standard errors small even with very flexible models. This allows the researcher to flexibly model high-dimensional

data while addressing unobservable fixed effects and machine learning bias.

This is one of the first applications of DML in environmental economics. We applied the estimator to study how extreme heat impacts crop yields. In our application, our DML estimator leads to estimates of treatment effects that are significantly different from the estimates based on OLS. Using a widely used set of weather variation, the difference is significant at the $p = 0.001$ level and economically meaningful. Our method predicts an effect of warmer temperature that is \$1.55 billion larger than current OLS based estimates.

Appendix to Chapter 2

2.A Asymptotic Normality

Proving our estimator is asymptotically normal follows from Theorem 14 of Chernozhukov, Escanciano, et al. (2022). Theorem 14 gives us that if Assumptions (1), (2), and (3) below are satisfied, asymptotic normality follows. In this appendix we explain either why our estimator satisfies these assumptions, or give primitive conditions that justify the assumptions.

We start with notation used in the three assumptions before introducing them. Recall from equation (2.2) from earlier our parameter of interest τ_0 is defined to be the solution to the following moment equation.

$$\tau_0 = \mathbb{E}[m(W, \gamma)] = \mathbb{E}\left[\frac{\partial \gamma_0(D_{i,t}, X_{i,t})}{\partial D_{i,t}}\right]$$

We define

$$g(W, \gamma, \tau) = m(W, \gamma) - \tau \tag{2.16}$$

We define ϕ to be our first step influence function. In our application $\phi = \alpha_0(Y - \gamma_0)$. Adding the first step influence function to our moment function to create our de-biased moment function ψ.

$$\psi(W, \gamma, \alpha_0, \tau_0) = g(W, \gamma, \tau) + \phi(W, \gamma, \alpha_0, \tau_0) \tag{2.17}$$

$$= g(W, \gamma, \tau) + \alpha_0(Y - \gamma_0) \tag{2.18}$$

We also define $\hat{\delta}_\ell$ [4] below. This is a known as the interaction reminder, it shows up in the decomposition of $\hat{\psi} - \psi$, and we have to make an assumption about its behavior for asymptotic linearity.

$$\begin{aligned}\hat{\delta}_\ell &= \hat{\alpha}_\ell(Y - \hat{\gamma}_\ell) - \hat{\alpha}_\ell(Y - \gamma_0) - \alpha_0(Y - \hat{\gamma}_\ell) - \alpha_0(Y - \gamma_0) \\ &= (\hat{\alpha}_\ell - \alpha_0)(\hat{\gamma}_\ell - \gamma_0)\end{aligned} \tag{2.19}$$

We use notation $\|\cdot\|_{F,2}$ to denote the $L_2(F)$-norm and $\|\cdot\|$ for the ℓ^2 vector norm.

Assumption 1. *(mild mean-square consistency) :* $E[\|\psi(W, \tau_0, \gamma_0, \alpha_0)^2\|] < \infty$

i) $\int \|g(w, \hat{\gamma}_\ell, \tau_0) - g(w, \gamma_0, \tau_0)\|^2 F_0(dw) \xrightarrow{p} 0$

ii) $\int \|\alpha_0(Y - \hat{\gamma}_\ell) - \alpha_0(Y - \gamma_0)\|^2 F_0(dw) \xrightarrow{p} 0$

iii) $\int \|\hat{\alpha}_\ell(Y - \gamma_0) - \alpha_0(Y - \gamma_0)\|^2 F_0(dw) \xrightarrow{p} 0$

These are mild mean-square consistency conditions for $\hat{\gamma}_\ell$ and $(\hat{\alpha}_\ell, \tilde{\theta}_\ell)$

Assumption 2. *(rate on interaction remainder) For each* $\ell = 1, \cdots, L$

$$\sqrt{n} \int \hat{\delta}_\ell(w) F_0(d_w) \xrightarrow{p} 0, \int \|\hat{\delta}_\ell\|^2(w) F_0(d_w) \xrightarrow{p} 0 \tag{2.20}$$

Assumption 3. *(double robust) For each* $\ell = 1, \cdots, L$

i) $\int \phi(w, \gamma_0, \hat{\alpha_0}, \theta_0) F_0(dw) = 0$ *with probability approaching one*

ii) $\mathbb{E}[\psi(\gamma, \alpha_0, \theta_0)]$ *is affine in* γ

As noted in Chernozhukov, Escanciano, et al. (2022), in our case because $\mathbb{E}[\psi(\gamma, \alpha_0, \theta_0)]$ is affine in γ, Assumption 3 imposes no conditions additional to Assumption 1 and 2, but we write it for clarity.

[4] note in the original paper by Chernozhukov, Escanciano, et al. (2022) this object was defined to be capital $\hat{\Delta}_\ell$ but given that capital delta is used to describe differences over time we changed notation

2.A.1 Primitive Conditions for Assumptions

Primitive Conditions for Assumption (1) Now we go into detail about the primitive conditions for the assumptions starting with Assumption 1 part i). Plugging in the definition of g into the equation, we get that this assumption is the following. Given that our moment is one dimensional we drop the norm notation for clarity.

$$\int \left((\frac{\partial \hat{\gamma}_\ell}{\partial D_{i,t}} - \theta_0) - (\frac{\partial \gamma_0}{\partial D_{i,t}} - \theta_0) \right)^2 F_0(dw) \xrightarrow{p} 0 \tag{2.21}$$

$$\int \left(\frac{\partial \hat{\gamma}_\ell}{\partial D_{i,t}} - \frac{\partial \gamma_0}{\partial D_{i,t}} \right)^2 F_0(dw) \xrightarrow{p} 0 \tag{2.22}$$

For conciseness this we write this with expectation notation.

$$\mathbb{E}\left[\left(\frac{\partial \hat{\gamma}_\ell}{\partial D_{i,t}} - \frac{\partial \gamma_0}{\partial D_{i,t}} \right)^2 \right] \xrightarrow{p} 0 \tag{2.23}$$

Which is a mean square consistency condition for the average derivative. Mean square consistency for the average derivative follows from two primitive assumptions 1) mean square consistency first step, stated in Assumption (4) and 2) mean square continuity for average derivative, stated in Assumption (5). We start by writing down these two assumptions and show how together they imply mean square consistency for the average derivative.

The first primitive assumption is mean square consistency of the first step estimator $\hat{\gamma}$.

Assumption 4. *(mean square consistency first step)*

$$\|\hat{\gamma}_\ell - \gamma_0\|_{F,2} \xrightarrow{p} 0 \tag{2.24}$$

The second primitive assumption is mean square continuity for average derivative.

Assumption 5. *(mean square continuity for average derivative)*

First we define a function class $\Gamma_{\bar{Q}}$ where $\bar{Q} < \infty$ and such that for all $\gamma \in \Gamma_{\bar{Q}}$

$$\mathbb{E}(\frac{\partial\gamma}{\partial D_{i,t}}^2) \leqslant \bar{Q}[\mathbb{E}(\gamma^2)]^{\frac{1}{2}} \tag{2.25}$$

We assume that $\gamma_0 \in \Gamma_{\bar{Q}}$ *and* $\hat{\gamma} \in \Gamma_{\bar{Q}}$ *with high probability.*

Now to show that Assumption (4) and Assumption (5) give us the mean square consistency for the average derivative.

Proof.

$$\mathbb{E}\left[\left(\frac{\partial\hat{\gamma}_\ell}{\partial D_{i,t}} - \frac{\partial\gamma_0}{\partial D_{i,t}}\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial(\hat{\gamma}_\ell - \gamma_0)}{\partial D_{i,t}}\right)^2\right] \tag{2.26}$$

$$\leqslant \bar{Q}[\mathbb{E}(\hat{\gamma}_\ell - \gamma_0)^2]^{\frac{1}{2}} \tag{2.27}$$

$$= \bar{Q}\|\hat{\gamma}_\ell - \gamma_0\|_{F,2} \xrightarrow{p} 0 \tag{2.28}$$

□

We can bound the left hand side of equation (2.23) using the two assumptions. First the linearity of the average derivative give us the first equality below. Then the inequality follows by mean square continuity since it holds for every realization of $\hat{\gamma}$. Then the last line follows by Assumption (5). Hence we have mean square consistency condition for the average derivative.

Please note that further primitive conditions Assumption (5) are given in Chernozhukov, Newey, and Singh (2021) in Lemma B.4. Though we leave the details to Chernozhukov, Newey, and Singh (2021), in short the conditions are given in Assumption (6)

Assumption 6. *Primitive conditions for Assumption* (5)

i) $f(d|x)$ *vanishes for each* d *in the boundary of the support of* D *given* $X = x$ *almost everywhere*

ii) $-\partial_d \log f(d|x)$ *is bounded above*

iii) Γ *consists of functions* γ *that are twice continuoysly differentiable in the first argument and satisfy* $\mathbb{E}[\{\partial_d \gamma(D,X)\}^2] < \infty$ *and* $\mathbb{E}[\{\partial_d^2 \gamma(D,X)\}^2] < \infty$

Assumption (6) is satisfied if Γ satisfies Sobolev conditions.

Primitive Conditions for Assumption (2) The rate condition on the interaction term of Assumption (2) requires that $\hat{\alpha}_\ell$ and $\hat{\gamma}_\ell$ can be estimated fast enough.

$$\|\hat{\alpha}_\ell - \alpha_0\|_{F,2} \|\hat{\gamma}_\ell - \gamma_0\|_{F,2} = o_p(n^{1/2}) \tag{2.29}$$

Rates and conditions for Lasso estimation of $\hat{\gamma}_\ell$ are given in Bickel, Ritov, Tsybakov, et al. (2009). As for $\hat{\alpha}_\ell$, either an approximate sparsity assumption or sparse eigenvalue assumption can justify rates needed for (2.29) as explained in detail in Chernozhukov, Newey, and Singh (2022a).

2.B Details of Estimation Procedure

2.B.1 Tuning

In this section we explain tuning $\hat{\alpha}$ and $\hat{\gamma}$.

We use a data-driven process to select the hyperparameters for Lasso and the Reisz representer function. We select hyperparameters by minimizing loss on the test set during each fold of our cross validation procedure. Let $\hat{\gamma}_\ell$ and $\hat{\alpha}_\ell$ denote the estimates of γ and α trained using indices not in set ℓ using the above procedures, for the given hyperparameter value. Let $\hat{\beta}_\ell$ and $\hat{\rho}_\ell$ be the corresponding estimates of parameter vectors in our sparse linear models. Recall that W_ℓ denotes all observations in the fold ℓ. Let $\mathcal{L}_\gamma(\gamma, W_\ell; r_L)$ be the sum of squared error of the function γ with the hyperparameter r_L on the data in $\mathcal{I}_\ell$:

$$\mathcal{L}_\gamma(\gamma, W_\ell; r_L) = \sum_i \sum_t (\Delta \hat{\gamma}_\ell(D_{i,t}, X_{i,t}) - \Delta Y_{i,t})^2$$

Let $\mathcal{L}_\alpha(\alpha, \mathcal{I}_\ell; r_\alpha)$ be the loss function of the function α with the hyperparameter r_α on the

data in $\mathcal{I}_\ell$. That loss function is the sum of the distance between α_0 and $\hat{\alpha}$ using our dictionary of basis functions. See Appendix 2.B.3 for an explanation of this loss function. Minimizing this distance is equivalent to minimizing:

$$\mathcal{L}_\alpha(\hat{\alpha}_\ell, W_\ell; r_\alpha) = \sum_i \sum_t -2b_D(D_{i,t}, X_{i,t})'\hat{\rho}_\ell + \hat{\rho}_\ell' \Delta b(D_{i,t}, X_{i,t}) \Delta b(D_{i,t}, X_{i,t})' \hat{\rho}_\ell$$

Then, select hyperparameter r_L that minimize test-set mean squared error of the regression:

$$r_L = \underset{r}{\text{argmin}} \frac{1}{n(T-1)} \sum_{\ell=1}^{k} \mathcal{L}_\gamma(\hat{\gamma}_\ell, W_\ell; r)$$

And select hyparameter r_α that minimize test-set loss of the Riesz representer:

$$r_\alpha = \underset{r}{\text{argmin}} \frac{1}{n(T-1)} \sum_{\ell=1}^{k} \mathcal{L}_\alpha(\hat{\alpha}_\ell, W_\ell; r)$$

2.B.2 Normalization

Many machine learning models perform better when the independent variables are standardized, that is when the data has mean zero and variance 1. In this section, we include some details about how to conduct this standardization so that the researcher is able to recover derivatives after that step.

For each basis function b^j for $j = 1, \ldots, p$ in the dictionary b, define the mean and standard deviation of the transformed data: $\mu^j := \mathbb{E}[b^j(W_i)]$ and $\sigma^j := \sqrt{\mathbb{E}[(b^j(W_i) - \mu^j)^2]}$. Their sample equivalents are: $\hat{\mu}^j := \frac{1}{n}\sum_i b^j(W_i)$ and $\hat{\sigma}^j := \sqrt{\frac{1}{N-1}\sum_i (b^j(W_i) - \hat{\mu}^j)^2}$.

To generate a standardized basis function, we apply the following transformation: $\dot{b}^j := (b^j(W_i) - \hat{\mu}^j)/\hat{\sigma}^j$. We assumed that there was some β_0 such that $\gamma_0(W_i) = \beta_0 b(W_i)$. Then there exists $\tilde{\beta}_0$ such that $\gamma_0(W_i) = \tilde{\beta}_0 \dot{b}(W_i) + C$, where $\tilde{\beta}_0^j \sigma^j = \beta_0^j$ for all j components, $\dot{b}$ is the dictionary of all standardized basis functions $\dot{b}^j$, and C is some generic constant. This is easy to confirm via algebraic manipulation. When we take differences to remove the unobserved individual fixed effect, this C term is also removed.

Recall that the average derivative is $\mathbb{E}[\beta_0 b_D(W_i)]$. We could also write this in terms of $\tilde{\beta}_0$, with the fact that $\beta_0^j = \tilde{\beta}_0^j \sigma^j$. Let $\sigma = \{\sigma^1, \ldots, \sigma^p\}$, and $\sigma^{-1} = \{1/\sigma^1, \ldots, 1/\sigma^p\}$ Then, we write this relationship more compactly as $\mathbb{E}[(\tilde{\beta}_0 \circ \sigma^{-1}) b_D(W_i)$, where $\circ$ is elementwise multiplication or the Hadamard product (i.e. $\tilde{\beta}_0 \circ \sigma^{-1} = \{\tilde{\beta}_0^1/\sigma^1, \ldots, \tilde{\beta}_0^p/\sigma^p\}$).

In our estimation procedure, we found it more convenient to producing scaled derivatives of each basis function and multiplying by the beta estimate from using scaled data. [5] Define the scaled derivative of each basis function, $\dot{b}_D^j := b_D^j/\hat{\sigma}^j$. Then, the average derivative is: $\mathbb{E}[\tilde{\beta}_0 \dot{b}_D^j(W_{i,t})]$. These procedures are equivalent, as can be confirmed through algebraic manipulation.

This suggests our procedure to standardize data and recover the derivative:

1. For each basis function b^j for $j = 1, \ldots, p$, find $\hat{\mu}^j$ and $\hat{\sigma}^j$. Store these estimates.
2. Create the sample standardized basis function and its derivative, $\dot{b}^j := (b^j(W_i) - \hat{\mu}^j)/\hat{\sigma}^j$ and $\dot{b}_D^j := b_D^j/\hat{\sigma}^j$.
3. Find an estimate $\hat{\tilde{\beta}}$ that satisfies the regression $\mathbb{E}[\Delta y_{i,t}] = \hat{\tilde{\beta}} \Delta \dot{b}(W_{i,t})$.

   Where $\Delta \dot{b}(W_{i,t}) := \dot{b}(W_{i,t}) - \dot{b}(W_{i,t-1})$. This could be via OLS or a cross-folds Lasso procedure.
4. Estimate the average derivative as $\mathbb{E}[\hat{\tilde{\beta}} \dot{b}_D(W_{i,t})]$.

Standardization of our basis functions is also relevant for estimating the Riesz representer. After the standardization, our Riesz representer now takes the form $\hat{\alpha}(W_{i,t}) = \dot{b}(W_{i,t})\hat{\rho}$. Before standardizing the data, we had an estimator of the form:

$$\hat{\rho}_{\text{original}} = \underset{\rho}{\operatorname{argmin}} -2\hat{M}\rho + \rho' \hat{Q} \rho + \lambda |\rho|_1$$

where $\hat{M} = \mathbb{E}[b_D(W_{i,t})]$ and $\hat{Q} = \mathbb{E}[\Delta b(W_{i,t})' \Delta b(W_{i,t})]$. After applying the standardization, we are taking the derivative of the standardized basis functions, so this estimator now takes

[5] This form can also be motivated by the chain rule, taking the derivative of the standardized data. In this case, an additional bias correction would be necessary because we are estimating $\hat{\mu}^j$ and $\hat{\sigma}^j$. See Appendix 2.B.2 for more details.

the form:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} -2\hat{\dot{M}}\rho + \rho'\hat{\dot{Q}}\rho + \lambda|\rho|_1 \tag{2.30}$$

where $\hat{\dot{M}} = \mathbb{E}[\dot{b}_D(W_{i,t})]$ and $\hat{\dot{Q}} = \mathbb{E}[\Delta\dot{b}(W_{i,t})'\Delta\dot{b}(W_{i,t})]$. We use $\dot{b}^j_D(W_{i,t}) = b^j_D(W_{i,t})/\hat{\sigma}^j$, although this neglects an additional correction because standardization involves estimating the mean and variance of the dataset. Below, we include a derivation of a full bias correction term to account for this estimation, and note this bias correction term is equivalent to the above expression for large n. In our estimation, we therefore use the values $\dot{b}^j_D(W_{i,t}) = b^j_D(W_{i,t})/\hat{\sigma}^j$ to construct $\hat{\dot{M}}$ and estimate $\hat{\alpha}$ using 2.30.

In the remainder of this section, we proceed with the full derivation of the term $\dfrac{\partial \dot{b}^j(W_{i,t})}{\partial D_{i,t}}$. Recall that we use the notation ξ_D to denote the partial derivative of ξ with respect to D.

$$\frac{\partial}{\partial D_i}\frac{b^j(W_i) - \hat{\mu}^j}{\hat{\sigma}^j} = \frac{(b^j_D(W_i) - \hat{\mu}^j_D)\hat{\sigma}^j - (b^j(W_i) - \mu^j)\hat{\sigma}^j_D}{(\hat{\sigma}^j)^2}$$

$$(b^j_D(W_i) - \hat{\mu}^j_D) = b^j_D(W_i) - 1/N b^j_D(W_i) = \frac{N-1}{N} b^j_D(W_i)$$

$$\hat{\sigma}^j_D = \frac{1}{2}\left(\frac{\sum_i (b^j(W_i) - \hat{\mu}^j)^2}{N-1}\right)^{-1/2} \frac{1}{N-1}\sum_k 2(b^j(W_i) - \hat{\mu}^j)(b^j_D(W_i) - \hat{\mu_D}^j)$$

$$\sum_k (b^j(W_i) - \hat{\mu}^j)(b^j_D(W_i) - \hat{\mu_D}^j) = (b^j(W_i) - \hat{\mu}^j)b^j_D(W_i) - \sum_k (b^j(W_i) - \hat{\mu}^j)\hat{\mu_D}^j$$

$$\sum_k (b^j(W_i) - \hat{\mu}^j)\hat{\mu}^j)\hat{\mu_D}^j = 0$$

$$\sum_k (b^j(W_i) - \hat{\mu}^j)(b^j_D(W_i) - \hat{\mu_D}^j) = (b^j(W_i) - \hat{\mu}^j)b^j_D(W_i)$$

$$\frac{\partial}{\partial D_i}\frac{b^j(W_i) - \hat{\mu}^j}{\hat{\sigma}^j} = \frac{\frac{N-1}{N} b^j_D(W_i)\hat{\sigma}^j - \frac{(b^j(W_i) - \mu^j)^2}{(N-1)\hat{\sigma}^j} b^j_D(W_i)}{(\hat{\sigma}^j)^2}$$

$$\frac{\partial}{\partial D_i}\frac{b^j(W_i) - \hat{\mu}^j}{\hat{\sigma}^j} = \frac{b^j_D(W_i)}{\hat{\sigma}^j}\left(\frac{N-1}{N} - \frac{(b^j(W_i) - \mu^j)^2}{(N-1)(\hat{\sigma}^j)^2}\right)$$

Note that $\dfrac{b^j_D(W_i)}{\hat{\sigma}^j} - \dfrac{b^j_D(W_i)}{\hat{\sigma}^j}\left(\dfrac{N-1}{N} - \dfrac{(b^j(W_i) - \mu^j)^2}{(N-1)(\hat{\sigma}^j)^2}\right) = O(1/N)$. This term is therefore negligible, as it converges to 0 faster than the $\sqrt{N}$ convergence rate.

2.B.3 Riesz representer Details

Our goal is to find the estimator $\hat{\alpha}(W_{i,t})$ that minimizes the mean squared error (MSE) between $\hat{\alpha}$ and α_0:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}}\, \mathbb{E}[(\alpha_0(W_{i,t}) - \alpha(W_{i,t}))^2]$$

Plugging in our guess at the functional form, $\hat{\alpha} = \Delta b(W_{i,t})\hat{\rho}$, we form the following regularized problem:

$$\begin{aligned}
\hat{\rho} &= \underset{\rho}{\operatorname{argmin}}\, \mathbb{E}\left[(\alpha_0(W_{i,t}) - \Delta b(W_{i,t})\rho)^2\right] + \lambda|\rho|_1 \\
&= \underset{\rho}{\operatorname{argmin}}\, \mathbb{E}[\alpha_0(W_{i,t})^2 - 2\Delta b(W_{i,t})\alpha_0(W_{i,t})\rho + \rho'\Delta b(W_{i,t})'\Delta b(W_{i,t})\rho] + \lambda|\rho|_1 \\
&= \underset{\rho}{\operatorname{argmin}}\, -2\mathbb{E}[b_D(W_{i,t})]\rho + \rho'\mathbb{E}[\Delta b(W_{i,t})'\Delta b(W_{i,t})]\rho + \lambda|\rho|_1 \\
&= \underset{\rho}{\operatorname{argmin}}\, -2\hat{M}\rho + \rho'\hat{Q}\rho + \lambda|\rho|_1
\end{aligned}$$

Where the 3rd equality comes from applying the Riesz Representation theorem, and the 4th equality comes from the definition $\hat{M} = \mathbb{E}[b_D(W_{i,t})]$ and $\hat{Q} = \mathbb{E}[\Delta b(W_{i,t})'\Delta b(W_{i,t})]$.

We use this estimator to find $\hat{\rho}$; we use an optimization package to find the optimal value of $\hat{\rho}$. Chernozhukov, Newey, and Singh (2022a) provides an iterative approach.

More generally, for any linear functional m and any functional form for $\hat{\alpha}$:

$$\begin{aligned}
\hat{\alpha} &= \underset{\alpha}{\operatorname{argmin}}\, \mathbb{E}[(\alpha_0(W_{i,t}) - \alpha(W_{i,t}))^2] \\
&= \underset{\alpha}{\operatorname{argmin}}\, \mathbb{E}[\alpha_0(W_{i,t})^2] - 2\mathbb{E}[\alpha_0(W_{i,t})\alpha(W_{i,t})] + \mathbb{E}[\alpha(W_{i,t})^2] \\
&= \underset{\alpha}{\operatorname{argmin}}\, -2\mathbb{E}[m(W_{i,t}, \alpha)] + \mathbb{E}[\alpha(W_{i,t})^2]
\end{aligned}$$

Plugging in our estimator form, we get $\mathbb{E}[m(W_{i,t}, \alpha)] = \mathbb{E}[b_D(W_{i,t})\rho]$ and $\mathbb{E}[\alpha(W_{i,t})^2] = \rho'\mathbb{E}[\Delta b(W_{i,t})'\Delta b(W_{i,t})]\rho$, confirming the above result.

2.B.4 Analytical vs. Numerical Derivative

We propose calculating the derivative analytically, as opposed to the numerical methods. In current DML and Auto DML papers, derivatives are computed using numerical differentiation. To explain what we mean by analytical vs numerical let's consider a function $\gamma(D, X)$, and let's say we have an estimate of this function $\hat{\gamma}(D, X)$ and we want to estimate the derivative at point $D = D_0$ and $X = X_0$. To estimate derivative numerically we could use Newton's difference quotient (also known as a first-order divided difference) and pick some small h.

$$\widehat{\frac{\partial \gamma(D_0, X_0)}{\partial D}} \sim \frac{\hat{\gamma}(D_0 + h, X_0) - \hat{\gamma}(D_0, X_0)}{\delta} = \frac{b(D_0 + h, X_0)\hat{\beta} - b(D_0, X_0)\hat{\beta}}{h} \tag{2.31}$$

It is well known that this procedure can introduce biases, either through formula error or rounding error. Formula error is introduced because, for most cases, the difference between the true derivative and the numerical approximation is decreasing in h. Formula error is most relevant when h is large. Rounding error is introduced during computation with small h, as computers must round floating point numbers in order to carry out computation.

There are approaches for reducing the error of numerical differentiation, although these approaches are still under development especially for noisy data. Simple modifications include taking a symmetric difference instead of a one-sided difference. The problem of numerical differentiation with noisy data is more challenging and the subject of ongoing research (Mboup, Join, and Fliess 2009; Chartrand 2011; Chartrand 2017; Van Breugel, Kutz, and Brunton 2020). By taking analytical derivatives of our basis function, we avoid these numerical challenges.

2.C Additional Simulation Results

Here, we include simulation results from different parameters of the data generating process described in Section 2.3. Each table summarizes 1000 bootstrap trials, for the specified data generating process. We vary the number of covariates and the number of time periods in

comparison to our main simulation results in the paper which have T = 2 and 20 covariates. These results show that, as expected, the bias generally decreases as the number of samples per covariate increases. This is particularly true of OLS Poly. OLS Poly has the lowest bias in these additional trials, but has higher mean squared error in estimating the true derivative (MSE τ) than DML, DML Iterative, or Lasso in all trials. DML has lower bias than DML Iterative in Table 2.C.1 and Table 2.C.2, and lower MSE τ in all trials.

method	DML	DML Iterative	Lasso	OLS Linear	OLS Poly
True Value	2.937	2.937	2.937	2.937	2.937
Average Derivative	2.929	2.909	2.819	3.245	2.932
Bias	-0.007124	-0.02712	-0.1173	0.3082	-0.004251
Standard Deviation	0.3187	0.3498	0.311	0.3553	0.5017
MSE τ	0.09302	0.1141	0.1021	0.2152	0.2436
Coverage	0.906	0.942	0.287	0.845	0.942
MSE γ In Sample	1.945	1.945	1.945	2.349	1.752
MSE γ Cross Folds	2.033	2.033	2.033	2.411	3.59

Table 2.C.1: Summary of derivative estimates from 1000 bootstrap trials of our simulation procedure. Estimates use $N = 1000$, $T = 2$, and 10 covariates. Flexible basis functions include 3^{rd} order polynomial functions of all terms and all interactions of D and X terms. After applying the basis function transformation, $p = 124$. Bias is the average of the estimated value of the derivative minus the true value of the derivative in each simulation draw. "MSE τ" is the mean squared error between the true average derivative and the estimated average derivative in each simulation draw, while "MSE γ in sample" and "MSE γ cross folds" refer to the mean squared error of regression from own-sample and out-of-sample estimation.

method	DML	DML Iterative	Lasso	OLS Linear	OLS Poly
True Value	2.936	2.936	2.936	2.936	2.936
Average Derivative	2.935	2.922	2.864	3.239	2.936
Bias	-0.0006326	-0.01354	-0.07144	0.3038	0.0003044
Standard Deviation	0.2258	0.2425	0.2187	0.2272	0.2925
MSE τ	0.03628	0.04374	0.03838	0.1293	0.06965
Coverage	0.9	0.923	0.289	0.678	0.937
MSE γ In Sample	1.988	1.988	1.988	2.366	1.918
MSE γ Cross Folds	2.01	2.01	2.01	2.388	2.207

Table 2.C.2: Summary of derivative estimates from 1000 bootstrap trials of our simulation procedure. Estimates use $N = 1000$, $T = 5$, and 10 covariates. Flexible basis functions include 3^{rd} order polynomial functions of all terms and all interactions of D and X terms. After applying the basis function transformation, $p = 124$. Rows are as described in Table 2.C.1.

method	DML	DML Iterative	Lasso	OLS Linear	OLS Poly
True Value	2.964	2.964	2.964	2.964	2.964
Average Derivative	2.977	2.96	2.902	3.258	2.973
Bias	0.01301	-0.003874	-0.06199	0.2943	0.008642
Standard Deviation	0.2292	0.25	0.2222	0.2351	0.2878
MSE τ	0.03879	0.04942	0.0391	0.1245	0.07208
Coverage	0.8747	0.9114	0.27	0.6976	0.9287
MSE γ In Sample	1.98	1.98	1.98	2.369	1.838
MSE γ Cross Folds	2.013	2.013	2.013	2.409	2.577

Table 2.C.3: Summary of derivative estimates from 1000 bootstrap trials of our simulation procedure. Estimates use $N = 1000$, $T = 5$, and 20 covariates. Flexible basis functions include 3^{rd} order polynomial functions of all terms and all interactions of D and X terms. After applying the basis function transformation, $p = 244$. Rows are as described in Table 2.C.1.

2.D Additional Applied Results

In our main estimation, we make some empirical choices to match those in the literature. In this section, we include results with some alternate specifications, such as alternate temperature thresholds for damaging heat exposure, estimates without weighting the county-year observations, and using additional weather variables beyond the commonly used specification involving only temperature and precipitation.

2.D.1 Testing Damaging Weather Threshold

In the main analysis, we define damaging heat as any temperature exposure above 29°C, as in Burke and Emerick (2016). This threshold was selected by iterating over a potential range and finding the value that minimizes the sum of squared errors from a linear model. We inspect whether the same threshold also minimizes the sum of squared errors from a more flexible model. Table 2.D.1 shows the results of this analysis. We find that 28°C minimizes the mean squared error for all methods. We use the threshold from Burke and Emerick (2016) to have a more straightforward comparison with the existing literature, and as there is minimal model fit improvement from an alternate specification.

	MSE in sample			MSE cross folds		
method	Lasso	OLS Linear	OLS Poly	Lasso	OLS Linear	OLS Poly
Threshold						
25 °C	0.07960	0.08191	0.07952	0.07992	0.08194	0.07994
26 °C	0.07863	0.08077	0.07863	0.07893	0.08080	0.07906
27 °C	0.07804	0.08007	0.07802	0.07832	0.08011	0.07844
28 °C	0.07783	0.08005	0.07779	0.07808	0.08010	0.07820
29 °C	0.07799	0.08093	0.07796	0.07823	0.08097	0.07835
30 °C	0.07841	0.08279	0.07840	0.07865	0.08284	0.07879
31 °C	0.07903	0.08557	0.07911	0.07925	0.08562	0.07951
32 °C	0.08024	0.08915	0.08021	0.08056	0.08920	0.08067
33 °C	0.08210	0.09326	0.08200	0.08254	0.09331	0.08253
34 °C	0.08486	0.09740	0.08470	0.08534	0.09745	0.08528

Table 2.D.1: Mean squared error (both own-sample and cross-folds) from the regression of corn yields with alternate temperature thresholds.

2.D.2 Including New Weather Variables

The results are relatively similar between the two sets of weather variation. The results from the main estimation are presented again in Table 2.D.2a next to the results including more covariates in Table 2.D.2b. OLS Linear finds a lower magnitude of the average derivative than the flexible models (OLS Poly, Lasso, and DML), all of which find relatively consistent estimates of the average derivative. The estimates from OLS Poly have the highest variance among any model, with roughly twice the standard deviation as OLS Linear. DML has a slightly lower variance than OLS Linear, and is significantly higher than Lasso. Based on the simulation results, we expect that the standard errors with Lasso do not accurately reflect the uncertainty around the estimate.

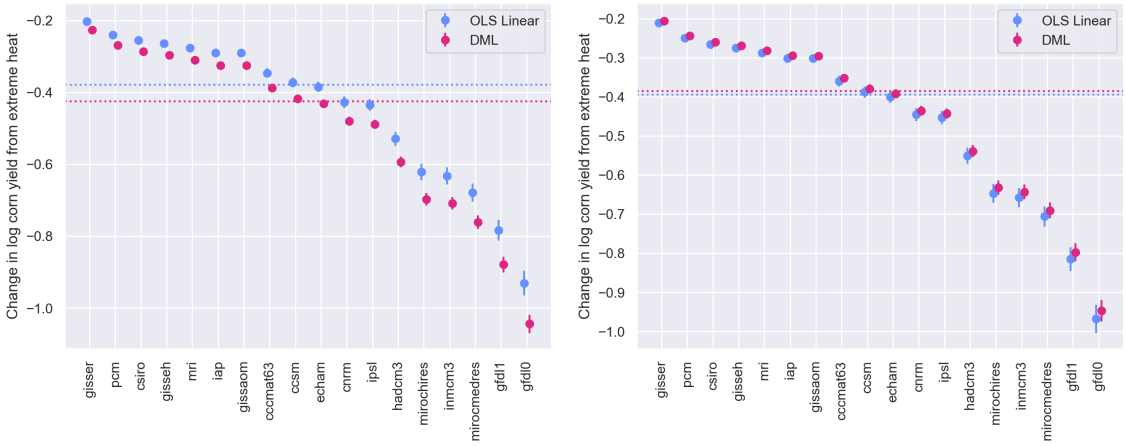
(a) Temperature and Precipitation

method	OLS Linear	OLS Poly	Lasso	DML
Average Derivative	-0.005193	-0.005657	-0.005821	-0.005823
	(0.000099)	(0.000135)	(0.000011)	(0.000073)
MSE γ In Sample	0.080929	0.077958	0.077878	0.077878
MSE γ Cross Folds	0.080975	0.078353	0.078079	0.078079
Number of Observations	63662	63662	63662	63662
Number of Covariates	3	36	36	36

(b) All Weather Covariates

method	OLS Linear	OLS Poly	Lasso	DML
Average Derivative	-0.005402	-0.005247	-0.005298	-0.005283
	(0.000103)	(0.000865)	(0.000019)	(0.000079)
MSE γ In Sample	0.079875	0.065521	0.065379	0.065379
MSE γ Cross Folds	0.079936	0.067955	0.067837	0.067837
Number of Observations	63662	63662	63662	63662
Number of Covariates	9	351	351	351

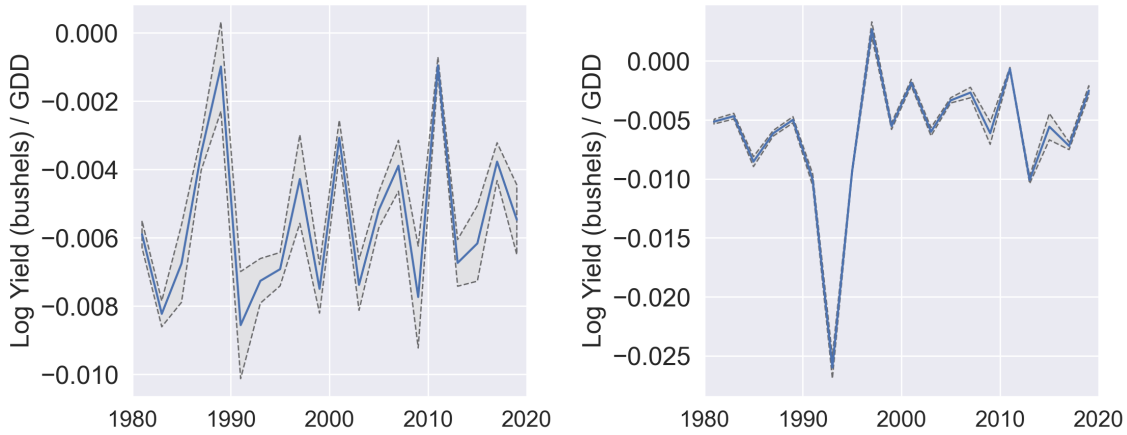
Table 2.D.2: Using the weighted estimates. Estimates of elasticity of corn yields with respect to increase in growing season exposure to extreme heat, using two sets of weather covariates. Standard errors (in parentheses) are clustered at the county level. See text for estimation details.



(a) Temperature and Precipitation

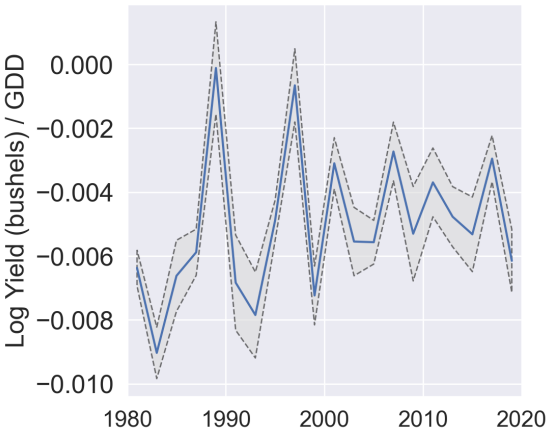
(b) All Covariates

Figure 2.D.1: Extrapolating impacts of extreme heat to crop yields by the year 2050, using elasticities from OLS Linear and DML. This figure shows the weighted estimates. Each dot represents a central estimate from a model, and the error bar represents the 95% confidence interval. Dotted line represents the median value across climate models. See text for estimation details.

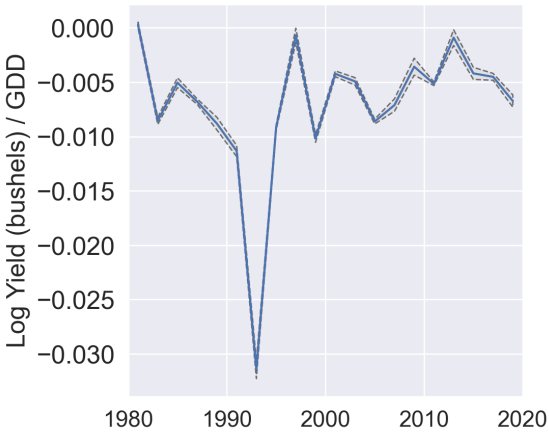


(a) OLS, Simple Covariates
Trend 3.43e-05, p-val 0.433

(b) DML, Simple Covariates
Trend 9.52e-05, p-val 0.387



(c) OLS, All Covariates
Trend 5.34e-05, p-val 0.213



(d) DML, All Covariates
Trend 0.000113, p-val 0.375

Figure 2.D.2: Estimating elasticity of corn yield over time with OLS Linear and DML. This figure shows the weighted estimates. We use our estimation procedure on 2-period samples from 1980 through 2019. Line shows central estimate, and grey band shows 95% confidence interval. The trend and standard errors are computed using weighted least squares, weighted by the inverse of the sum of the variance of the elasticity estimate and the residual from the regression of elasticity on the year.

2.D.3 Results Without Weighting

In our main estimates, we weight the results by the area of corn planted per county; the average derivative we estimate in that panel is the effect averaged over all acres of corn. This estimate is commonly used in the literature, including by Schlenker and Michael J Roberts (2009) and Burke and Emerick (2016). It may also be of interest to consider the effect without weighting, which gives the effect averaged over all U.S. counties. This appendix includes those results, as well as results without weighting and with additional weather variables.

Table 2.D.2 summarizes the estimation results, using each estimation method and the two sets of covariates. Figure 2.D.3 shows the extrapolated damages from extreme heat by the year 2050, using the same weighted procedure as in Burke and Emerick (2016) to generate these extrapolations. That range of damage estimates translates to a dollar value of $17.1 -$21.7 billion (in 2017 dollars). Our preferred estimate, DML with all weather covariates, finds an estimated damage of $21.7 billion. The difference between OLS Linear and DML estimates is statistically significant and economically meaningful. Median damage estimates using DML instead of OLS correspond to an additional $4.23 and $3.84 billion using the simple set of covariates and the set of all weather variables, respectively. This difference in parameter estimates is significant at the $p = 0.001$ level.

Figure 2.D.2 shows the estimated elasticity over time, without weighting to adjust for the crop level. Using both methods for the simple set of covariates, the mean trend in the elasticity is positive and we reject the null hypothesis at the 0.05 level. Using all covariates, we fail to reject the null hypothesis at the 0.05 level for either estimation method.

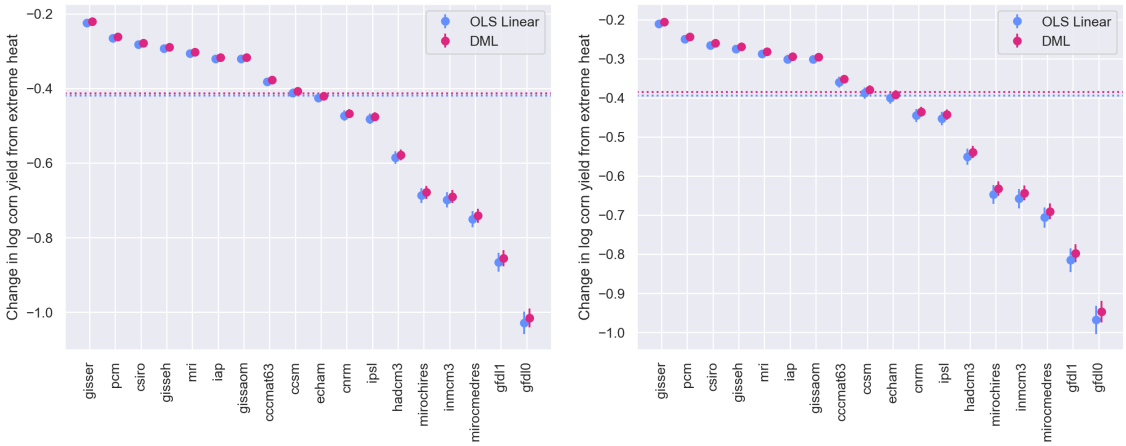
(a) Temperature and Precipitation

method	OLS Linear	OLS Poly	Lasso	DML
Average Derivative	-0.005738	-0.007558	-0.007578	-0.007577
	(8.484e-05)	(0.0001199)	(1.182e-05)	(6.431e-05)
MSE γ In Sample	0.07901	0.07125	0.07124	0.07124
MSE γ Cross Folds	0.07911	0.07166	0.07161	0.07161
Number of Observations	63662	63662	63662	63662
Number of Covariates	3	36	36	36

(b) All Covariates

method	OLS Linear	OLS Poly	Lasso	DML
Average Derivative	-0.005586	-0.007237	-0.007229	-0.007226
	(9.5e-05)	(0.0001926)	(1.708e-05)	(6.417e-05)
MSE γ In Sample	0.07785	0.05822	0.05802	0.05802
MSE γ Cross Folds	0.07797	0.06054	0.06054	0.06054
Number of Observations	63662	63662	63662	63662
Number of Covariates	9	351	351	351

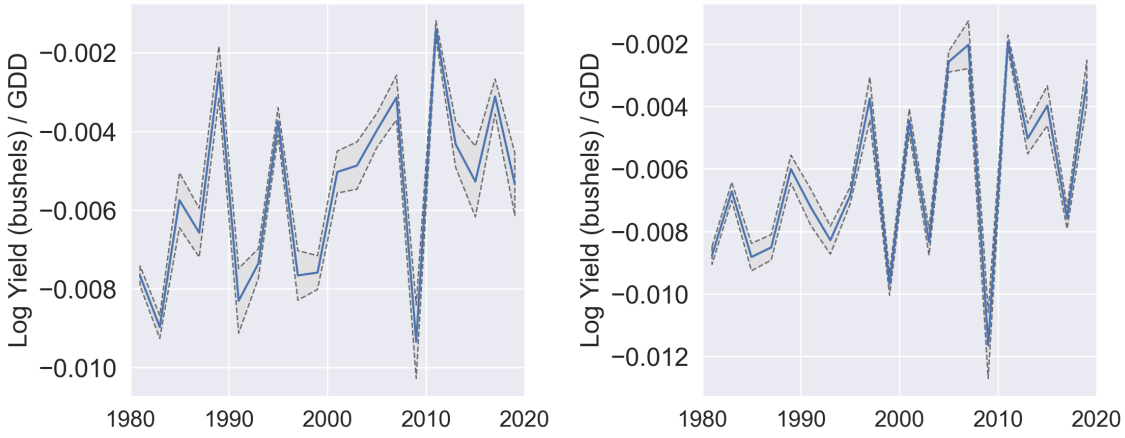
Table 2.D.3: Same as 2.D.2 using the non-weighted estimates. Estimates of elasticity of corn yields with respect to increase in growing season exposure to extreme heat, using two sets of weather covariates. Standard errors (in parentheses) are clustered at the county level. See text for estimation details.



(a) Only Temperature and Precipitation

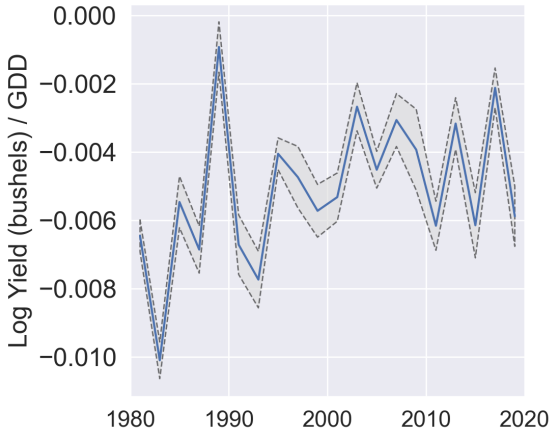
(b) All Covariates

Figure 2.D.3: Same as Figure 2.D.1 using the non-weighted estimates.

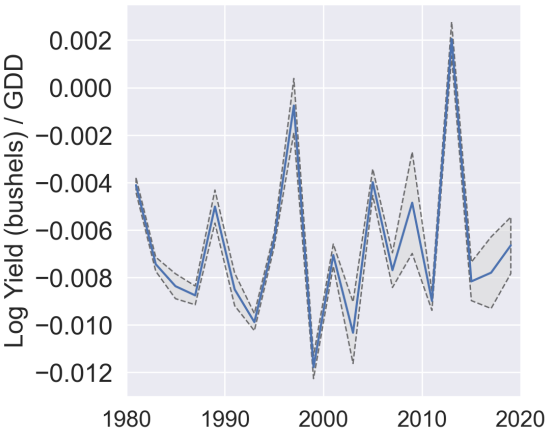


(a) OLS, Simple Covariates
Trend 8.72e-05, p val 0.0286

(b) DML, Simple Covariates
Trend 0.000102, p val 0.0390



(c) OLS, All Covariates
Trend 7.18e-05, p-val 0.065



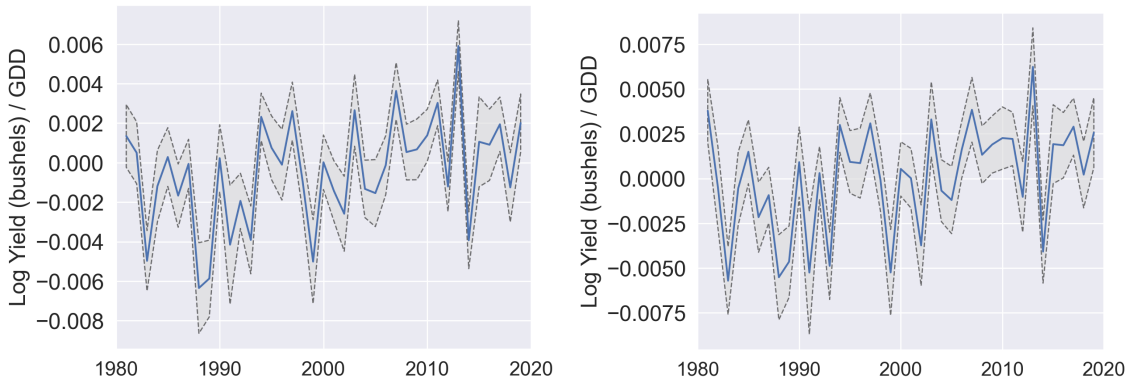
(d) DML, All Covariates
Trend 3.1e-05, p-val 0.631

Figure 2.D.4: Same figure as 2.D.2, using the non-weighted estimates.

2.D.4 Variation in Elasticities Over Time

In this section, I include additional estimation results on the variability in elasticities over time. I run a linear regression of crop yields and weather over the full sample, including yearly fixed effects terms and interactions between each yearly fixed effect and all weather covariates. Here, I use the within transformation to remove per-county unobservable heterogeneity and isolate the effects of weather shocks from climate or other persistent factors. This regression weights each observation by the area of corn planted in each county, and clusters standard errors at the county level. Figure 2.D.5 shows the coefficient of the interaction between the yearly fixed effect term and damaging heat exposure.

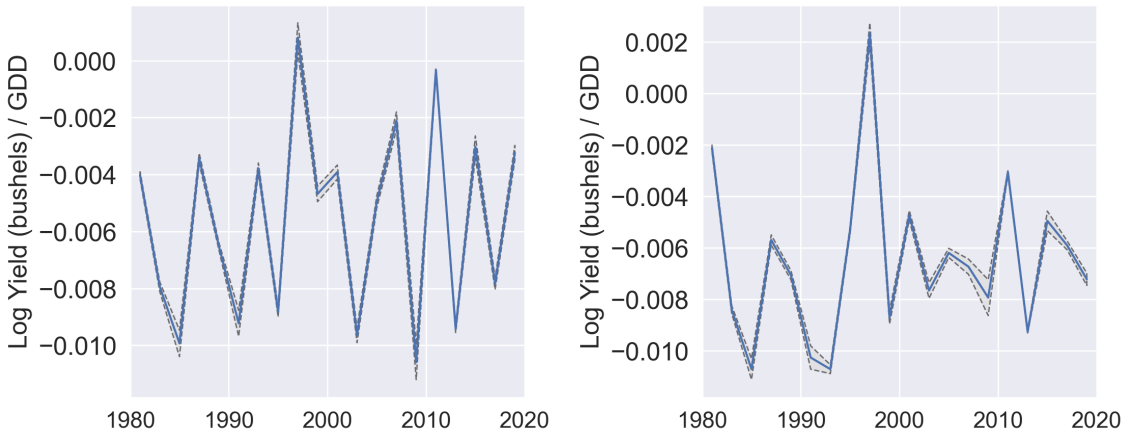
To isolate the effects of the polynomial expansion terms or the interaction terms from results in Section 2.4, I include estimates of the elasticity over time using alternate basis functions. In Figure 2.D.6, I use basis functions that include third-order polynomial expansions of each variable, but do not include interactions. In Figure 2.D.7, I use basis functions that include interactions but do not include polynomial expansions of each term.



(a) Simple Covariates

(b) All Covariates

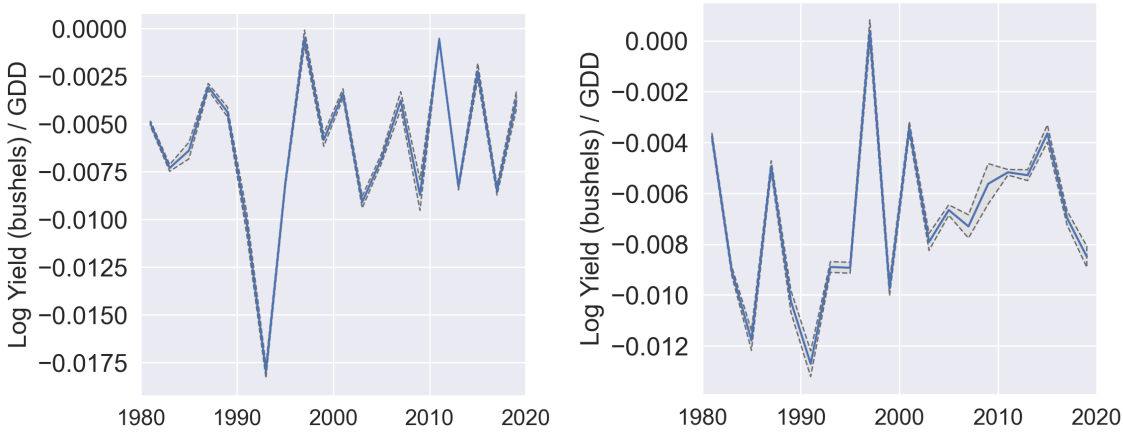
Figure 2.D.5: The coefficient of the interaction between the yearly fixed effect term and damaging heat exposure. This analysis is run using weighted least squares, where weights are the area of corn planted per county, after applying the within transformation to isolate the impact of weather shocks.



(a) DML, Simple Covariates
No interactions
Trend 3.7e-05, p-val 0.579

(b) DML, All Covariates
No interactions
Trend 1.81e-05, p-val 0.772

Figure 2.D.6: Estimating elasticity of corn yield over time with DML. Basis functions used to create this figure include polynomial terms but not interactions. This figure shows the weighted estimates. We use our estimation procedure on 2-period samples from 1980 through 2019. Line shows central estimate, and grey band shows 95% confidence interval. The trend and standard errors are computed using weighted least squares, weighted by the inverse of the sum of the variance of the elasticity estimate and the residual from the regression of elasticity on the year.



(a) DML, Simple Covariates
No polynomials
Trend 5.05e-05, p-val 0.517

(b) DML, All Covariates
No polynomials
Trend 7.01e-05, p-val 0.248

Figure 2.D.7: Same figure as 2.D.6, using the weighted estimates and basis functions that include interactions but do not include polynomial expansions of each term.

Chapter 3

A Machine Learning Approach to Measuring Climate Adaptation

3.1 Introduction

Measuring adaptation to recent climate change is important for informing climate policy and projecting damages of future climate change. The extent of recent adaptation can signal when policy interventions are needed and give a sense of how much adaptation will be possible to projected changes. Researchers can study this by modeling how weather shocks impact economic outcomes, and examining how impacts change with more exposure to the weather shock (e.g. Burke and Emerick 2016) or over time (e.g. Barreca et al. 2016). These studies depend on accurate models of the relationship between weather and economic outcomes.

Machine learning (ML) can help model these relationships when researchers have rich data, but lack expert guidance to suggest a functional form. When domain experts provide a model of how weather shocks impact economic outcomes, researchers can use this to accurately measure and compare impacts. Otherwise the researcher must learn this relationship from data. Economists typically use classic statistical tools to model weather-economic relationships without imposing strong functional form assumptions (Hsiang, 2016). Such tools work well with low-dimensional weather variation, but can lead to high variance or

inconsistent estimates as the dimensionality of covariates increases. Even a model with only temperature and precipitation can become high-dimensional if a researcher flexibly models each variable and the interactions between them. ML is well suited for flexible modeling in such settings (Mullainathan and Spiess, 2017), and can help researchers take full advantage of high-dimensional weather variation in modern panel datasets.

I introduce an ML approach to study adaptation to damaging heat exposure in United States (U.S.) corn and soy production. High temperatures are generally damaging for crop growth, although adaptation may offset some of these damages in the long run. Schlenker and Michael J. Roberts (2006) introduce a parsimonious model of how an annual shock of heat exposure impacts crop yield. They model crop yields as a piecewise-linear function of heat exposure below and above a crop-specific temperature threshold, where heat exposure below the threshold is beneficial and heat exposure above the threshold is damaging. By estimating the parameter on damaging heat using different sources of variation, Schlenker and Michael J Roberts (2009), Burke and Emerick (2016), and Lemoine (2018) argue that the observed degree of adaptation to damaging heat exposure will not be sufficient to offset projected climate damages.

I measure the degree of adaptation after learning the crop yield-weather relationship from data. I compare impacts from short-run and long-run changes in damaging heat exposure. With a low-dimensional linear model, this is equivalent to the method from Burke and Emerick (2016). The approach allows me to incorporate more flexible models and high-dimensional weather variation, making the method suitable for applications with rich weather variation and little expert guidance on how that variation impacts economic outcomes. I use ML to model the crop yield-weather relationship, specifically Least Absolute Shrinkage and Selection Operator (Lasso) and a neural network (NNet). My ML models account for additive fixed effects while flexibly modeling temperature, precipitation, and interactions between them.

To estimate the degree of adaptation, I estimate the elasticity of crop yields with respect to damaging heat exposure. This elasticity summarizes how damaging heat exposure impacts crop yields, and can be computed for each model. There is no single parameter I can compare

across models, as in Burke and Emerick (2016). Instead, I find the elasticity by fitting a regression function of log crop yields on temperature and precipitation and computing the average directional derivative in the direction of a marginal increase in damaging heat exposure. I estimate these regression functions via ordinary least squares (OLS) and ML.

I implement a debiasing procedure to address bias from standard ML models. ML approaches can induce bias from overfitting or regularization, but there are approaches to reduce this bias when estimating causal parameters or statistics based on regression functions (CCDDHNR, 2018; Chernozhukov, Newey, and Singh, 2022a). I adapt the estimator from Chernozhukov, Newey, and Singh (2022a). This approach uses double machine learning (DML), where standard ML estimates are corrected using a second ML algorithm. I adjust the second ML algorithm to suit the panel setting. I then apply this DML procedure to debias the elasticity estimates.

I compare elasticities to estimate the degree that short-run impacts from damaging heat exposure are offset in the longer term. I construct panel datasets of long-run and short-run variation in crop yield and weather variables from 1990-2019. I then estimate the elasticity of crop yield with respect to damaging heat exposure in each dataset. I compute this elasticity using OLS, DML, and ML without bias correction. Comparing these estimates, I find the extent that short-run impacts from damaging heat exposure are offset in the longer term.

Before taking this approach to the data, I conduct a simulation exercise to evaluate DML estimates of an elasticity relative to OLS and standard ML approaches. Each simulation trial uses the empirical distribution of temperature and precipitation from U.S. counties, but simulates outcome variables. I examine the performance of my DML procedure relative to OLS and ML without bias correction, comparing the bias and variance of recovering the true elasticity. I use three potential sets of weather variables: a simple, commonly used set of annual temperature and precipitation variables, a set that includes richer variation in annual temperature, and a set that includes monthly observations of precipitation and the rich variation in temperature.

This simulation exercise clearly highlights the benefits of using debiased machine learning for high-dimensional settings. Both NNet and Lasso have advantages over OLS in high-

dimensional cases, although NNet has lower bias. With the rich set of annual temperature variation, DML estimates have significantly lower variance than OLS and lower bias than standard ML estimates. With the set of monthly temperature and precipitation, DML estimates have significantly less bias and variance than OLS, and lower bias than standard ML.

I then apply the approach to a dataset of U.S. crop yields, and find evidence that adaptation is offsetting impacts from damaging heat exposure. I implement Lasso, NNet, and OLS estimators on a county-level dataset of corn and soy yields from 1990 to 2019. I compare estimates of the elasticity using short-run and long-run variation. I take 500 bootstrap trials of the estimated elasticity, using the three sets of weather variation as in the simulation exercise. Using the simple annual set of weather variables as in Schlenker and Michael J Roberts (2009) and Burke and Emerick (2016), I find that there has been little to no significant adaptation to climate change in corn or soy production. This confirms the results from Burke and Emerick (2016).

Using more flexible sets of weather variation, I find that a large share of short-run impacts from damaging heat exposure are offset in the long run. With short-run variation, I find statistically and economically significant declines in yield from a marginal increase in damaging heat exposure. However, I do not find evidence of such declines when using long-run variation. These results hold for both corn and soy. The primary difference comes from using a richer set of weather variables. I make this same conclusion using OLS with the more flexible set of weather variation, although the DML approach results in smaller confidence intervals. This shows that a substantial degree of the short-run impacts from damaging heat exposure are offset in the long run, suggesting substantial adaptation to this heat exposure.

This result differs dramatically from the conclusions by Burke and Emerick (2016) and other analyses (Schlenker and Michael J Roberts, 2009; Lemoine, 2018). This is likely explained by model misspecification for the impact of a long-run shift in heat exposure on crop yields. I show that in the panel with long-run variation, the simple model from Schlenker and Michael J. Roberts (2006) does not adequately summarize the flexible role of temperature variables. While damaging heat exposure is correlated with declines in crop yield, other

temperature variation is better able to explain these declines. This suggests that there is limited adaptation to some damaging feature of climate change, but not to the specific feature of marginal increase in damaging heat exposure.

This paper is related to several literatures. First is a literature on estimating the degree of adaptation to climate change. Measuring adaptation to climate change requires understanding how weather influences economic outcomes. For a review of economics literature on measuring the economic impacts of the weather, see Dell, Jones, and Olken (2014). Hsiang (2016) provides an overview of econometric approaches to measuring these impacts. Much of this literature focuses on agriculture, as this sector is directly exposed to weather and hence is particularly vulnerable to climate change (Shukla et al., 2019). The first approach to studying impacts of climate change used the Ricardian approach, where researchers compare the value of agricultural land in cross sections. Mendelsohn, Nordhaus, and Shaw (1994) forecast the impacts of climate change by regressing average temperature and agricultural property value in a cross section of U.S. counties. This approach is susceptible to omitted variable bias, and subsequent work has focused on addressing specific omitted variables such as endogenous changes in farmer technology (Kurukulasuriya, Kala, and Mendelsohn, 2011) or nonfarm income (Ortiz-Bobea, 2020). Other approaches to estimate the potential for adaptation in agriculture involve economy-wide simulations (Costinot, Donaldson, and Smith, 2016), production changes in historical migrations (Sutch, 2011; Olmstead and Rhode, 2011), natural experiments (Hornbeck, 2012; Hagerty, 2021), or panel approaches.

Panel approaches address omitted variable bias by identifying adaptation from annual or long-term variation within a panel dataset. Schlenker and Michael J Roberts (2009) uses panel data to estimate the elasticity of crop yields with respect to extreme heat exposure, and conclude that there is limited potential to adapt to climate change because these damages are similar in the southern and northern U.S. despite climatic differences. Barreca et al. (2016) use a flexible model of temperature exposure to document how the mortality consequences of extreme heat declined over the 20^{th} century. Burke and Emerick (2016) show that panel variation can be used to estimate adaptation to recent climate change by using separate sources of variation to identify the impacts of weather shocks and shifts in average

temperature. Lemoine (2018) provides an alternate approach that partially identifies the degree of possible adaptation by considering the role of ex-ante and ex-post adaptation to heat exposure shocks.

My paper is most closely related to Burke and Emerick (2016). Like their paper, I estimate the degree that damages to corn and soy yields from short-run changes in weather are offset over longer exposures. I also use crop and weather data from U.S. agriculture. My approach differs because I consider richer sets of weather variables, and use DML to model learn the relationship between these data and crop yields. I conclude that there has been a higher degree of adaptation to damaging heat exposure.

Second is a growing literature on applying ML methods in economics. Kleinberg et al. (2015) discuss applications of predictive machine learning in economics, and Varian (2014) and Mullainathan and Spiess (2017) provide a practical guide to algorithms. Several recent papers have used ML to measure important outcomes in environmental economics. Crane-Droesch (2018) proposes a semi-parametric NNet and uses it to study the impact of climate change on corn yields. Deryugina et al. (2019) uses a ML approach to measure the costs of air pollution. Burlig et al. (2020) use ML to refine estimates of energy efficiency improvements. Stetter, Mennig, and Sauer (2022) use a DML approach to measure effectiveness of an agricultural intervention, and Klosin and Vilgalys (2022) introduce a DML approach to measure elasticities in a panel setting. There are also numerous applications within agriculture; for a review, Liakos et al. (2018).

My paper is most related to Crane-Droesch (2018) and Klosin and Vilgalys (2022). Crane-Droesch (2018) estimates a NNet that accounts for unobservable county-level fixed effects, and uses the model to predict yield under counterfactual climate change scenarios. I use the same technique to address county-level fixed effects, although I modify the algorithm in order to recover derivatives from the network and to ensure that the network is differentiable. Crane-Droesch studies the impacts of future climate change, while I use this tool to understand adaptation to recent climate change. Like Klosin and Vilgalys (2022), I estimate the elasticity of crop yield with respect to an increase in damaging heat exposure. I consider more flexible representations of temperature variation, and I apply the estimator to measure

adaptation to recent climate change.

Within the literature on machine learning, this paper applies results from the emerging field of DML. CCDDHNR (2018) prove that sample splitting and constructing Neyman-orthogonal moment conditions can yield approximately debiased machine learning estimates in certain settings. Semenova and Chernozhukov (2021) extend the Neyman-orthogonal moment condition approach to several other statistical targets, including structural derivatives. In an alternate approach, Chernozhukov, Newey, and Singh (2022a) and Chernozhukov, Newey, and Singh (2022b) give an approximately debiased estimator for a more general class of linear functionals based on the Riesz representation theorem. Klosin and Vilgalys (2022) adopt this approach in panel settings and prove asymptotic normality of the estimator for the average derivative. Like Klosin and Vilgalys (2022), my approach applies the result from Chernozhukov, Newey, and Singh (2022a) in panel settings; I use a different approach to address fixed effects and consider a more flexible representation of temperature variation.

The rest of this paper proceeds as follows. In Section 3.2, I describe the data used for this project. I illustrate the degree of climate change and describe the transformations required to generate the growing degree days. In Section 3.3, I describe the methods used; this includes details on how to compute average derivatives and an explanation of the debiased machine learning estimation approach. In Section 3.4, I give details and results of the simulation exercise. In Section 3.5, I present and discuss results from using the estimation procedure to measure the degree of adaptation to climate change. Section 3.6 concludes.

3.2 Data

For the empirical application, I use weather and crop data from U.S. corn and soy production from 1990-2019. I consider counties east of the 100°West meridian, which defines an agricultural region of significant corn and soy cultivation. From 1990-2019, this region produced over 93% of the nation's corn and over 99% of the nation's soy. Crop data are annual yield (bushels per acre) of corn and soy, from the U.S. Department of Agriculture's Survey of Agriculture. These data also include the area planted (acres) in each county.

Weather data are generously shared by Schlenker and Michael J Roberts, who provide a grid-

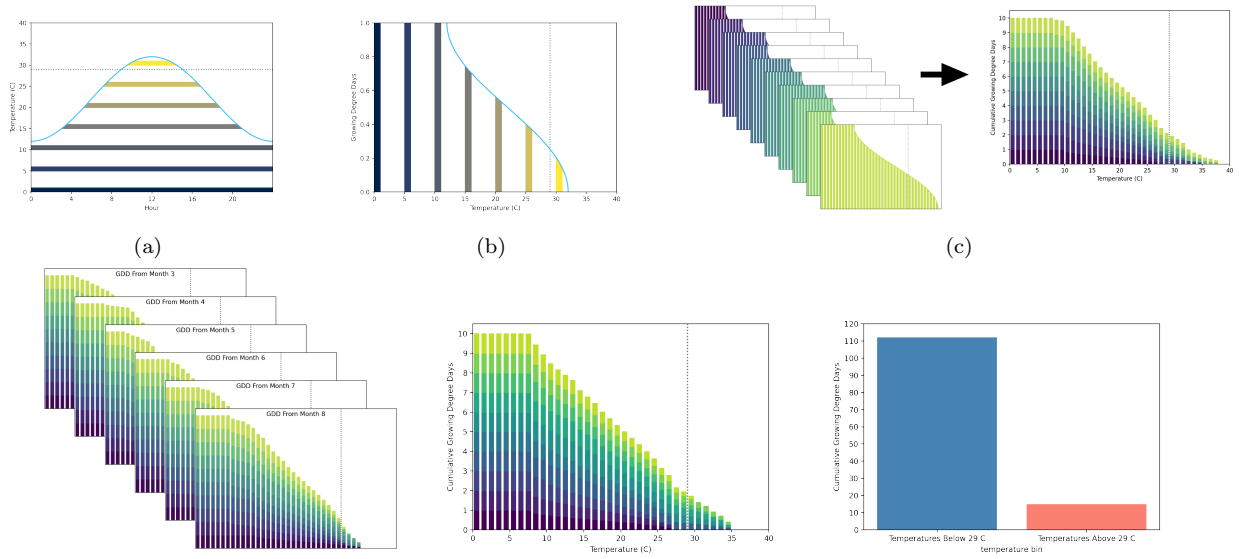
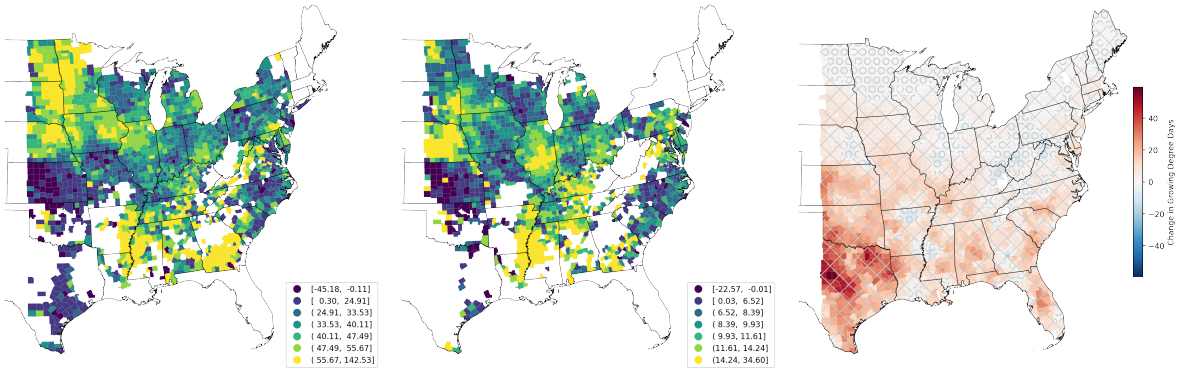


Figure 3.2.1: Transformations from daily minimum and maximum temperature records into the weather variables used in the analysis. (a) shows daily temperature within one day. In (b), this temperature is translated into growing degree days (GDD). (c) illustrates how daily GDD observations are aggregated into cumulative GDD exposure. (d) is the Monthly Flexible variable, where daily temperature is aggregated into GDD in each temperature bin for each month. In (e), this is aggregated into total growing season heat exposure in each temperature bin. In (f), this is further aggregated into total growing season heat exposure above and below the dotted line, 29°C.

ded dataset of daily temperature and precipitation from a network of consistently reporting weather stations. As in Schlenker and Michael J Roberts (2009) and Ortiz-Bobea (2013), I consider weather during the March-August growing season. I aggregate the gridded dataset to a county-level dataset of daily maximum temperature, minimum temperature, and precipitation. I then transform daily temperature exposure into growing degree days (GDD) at a monthly level and for the growing season. Figure 3.2.1 illustrates the transformation from daily temperature observations to GDD.

Throughout this paper, I consider three sets of weather variables per county: total growing season heat exposure above and below 29°C plus total growing season precipitation (Yearly Linear), total growing season heat exposure in each 1°C temperature bin plus total growing season precipitation (Yearly Flexible), and monthly heat exposure in each 1°C bin and monthly precipitation for each month of the growing season (Monthly Flexible). Figure 3.2.1 illustrates the transformations of temperature variables. The Yearly Linear transformation,



(a) Corn yields (b) Soy yields (c) GDD above 29°C

Figure 3.2.2: Differences in county-level values of corn yields, soy yields, and growing-season GDD above 29°C, between 1990-1999 averages and 2010-2019 averages. Maps include counties east of the 100 °W meridian. In (a) and (b), the first color bin includes all counties with yield declines (roughly 4% of counties) and the remaining distribution is divided evenly among the remaining bins. In (c), 'X' ('O') hatching indicates counties with an increase (decrease) in heat exposure above 29°C.

shown in Figure 3.2.1f, is widely used in economic analysis. Schlenker and Michael J Roberts (2009) demonstrate that, in a panel setting, regression based on heat exposure above and below a crop-specific damaging threshold explains as much variation in yields as more flexible models. Burke and Emerick (2016) identify 29°C as the threshold for damaging heat exposure, for both corn and soy. I use the 29°C threshold throughout the analysis, and refer to heat exposure above 29°C as damaging heat exposure.

Variation in the degree of climate change between counties is used to estimate the long-run elasticities in our sample. Figure 3.2.2c shows illustrates the difference in GDD exposure above 29°C per US county, from the period 1990-1999 to the period from 2010-2019. Heat exposure increases in 80% of counties in the sample, but there is substantial variation in the degree of warming even within states. There are many pairs of neighboring counties where one experienced an increase in damaging heat exposure and the other experienced a decrease. This variation is plausibly uncorrelated with unobservable factors such as soil quality or per-state fixed effects. See Burke and Emerick (2016) for a detailed argument that this type of variation can identify the effects of climate change.

While many counties experienced damaging heat exposure increases and decreases, only a handful had average yields decline between these time periods. Figure 3.2.2a and Fig-

ure 3.2.2b show changes in average yields of corn and soy over this time period. Only 4.03% of counties saw corn yields decline and only 3.96% saw soy yields decline, among counties that grew corn or soy in both 1990-1999 and 2010-2019. This suggests that improved farming technology increased yields throughout the sample, and that these improvements exceeded damages from increased damaging heat exposure.

3.3 Methods

In this section, I discuss the methods used to estimate the degree of adaptation. I first define adaptation in terms of average directional derivatives, and describe how to measure these average directional derivatives. I then introduce the ordinary least squares (OLS), Lasso and neural network (NNet) procedures, including details on how to account for fixed effect terms in each model and the cross-folds training procedures. I also describe the double machine learning (DML) approach used to adjust for bias in the standard machine learning (ML) estimates.

3.3.1 Adaptation

I define adaptation as the amount of short-term impact to yield from damaging heat exposure that is offset in the longer term. This definition encompasses all adaptation behaviors an agent makes to their production technology, management practices, or variety choice within each crop as they are exposed to climate change. It does not capture some other important margins of adaptation such as crop switching or exit from agriculture. Burke and Emerick (2016) provide evidence that these margins of adaptation are limited. This definition also does not capture changes that alter the impact of damaging heat independently from an individual's exposure to climate change, such as the decline in heat-related mortality as studied by Barreca et al. (2016). First I define the estimation target, as introduced by Burke and Emerick (2016). I generalize this definition in terms of elasticities, and discuss how to estimate those elasticities using flexible functional forms and/or richer sets of weather variation.

Following Burke and Emerick (2016), I measure adaptation by comparing the elasticity of

crop yields with respect to extreme heat using short-run and long-run variation. Short-run variation comes from year-to-year changes in an annual panel of weather observations and log crop yields. To capture long-run variation, I first average weather and crop yield data over a long period (I use 10-year periods) and then construct a two-period panel. The elasticity computed using short run variation captures the extent that a weather shock of extreme heat in a single year impacts yields, while the elasticity computed using long run variation captures the extent that exposure to a long period of increased heat exposure will impact average crop yields. Estimates using long-run variation arguably capture the extent of damages from climate change, because there is a change in average heat exposure over a relatively long period where farmers have time to adjust to those changes. Figure 3.2.2 illustrates long-run variation in the sample, comparing the differences in long-run average crop yields and damaging temperature exposure over the decades I study. Comparing impacts using these sources of variation, I can conclude whether the short-term damages are offset in the long term.

It is straightforward to recover these elasticities in a model where log yield is linear in total growing season heat exposure. Burke and Emerick (2016) use such a model, which I adapt below:

$$y_{it} = a_i + \beta_1 lower_{it} + \beta_2 higher_{it} + g(prec_{it}) + \varepsilon_{it} \tag{3.1}$$

where $higher_{it}$ ($lower_{it}$) is total growing season heat exposure above (below) the damaging temperature threshold, g is some function of precipitation, a_i is an additive per-county fixed effect term, and ε_{it} is an additive error term. Burke and Emerick (2016) estimate this equation with OLS, after using the within transformation to remove the fixed effect term[1]. The key parameter here is β_2, which captures the extent that log crop yield changes with marginal increase in damaging heat exposure. This is equivalent to the elasticity of crop yield with respect to damaging heat exposure. Note that β_2 is expected to be negative in the short run, as temperature shocks in this range are damaging to crop growth (Schlenker and Michael J. Roberts, 2006; Schlenker and Michael J Roberts, 2009). The estimate of the share of short-

[1]Burke and Emerick (2016) advocate using first differences to remove the fixed effect in the long-run variation panel; I use a within transformation approach as this is numerically equivalent to taking first differences in a two-period panel.

run damages that are offset in the longer term is therefore $(\hat{\beta}_2^{SR} - \hat{\beta}_2^{LR})/\hat{\beta}_2^{SR} = 1 - \hat{\beta}_2^{LR}/\hat{\beta}_2^{SR}$, where $\hat{\beta}_2^{SR}$ ($\hat{\beta}_2^{LR}$) is the estimate using the short-run (long-run) variation. Burke and Emerick (2016) take bootstrap samples of this ratio and fail to reject the null hypothesis that the ratio is different from 0.

To recreate this ratio with a more flexible functional form, I replace the parameter estimates with average directional derivatives. Consider a more general functional form:

$$y_{it} = a_i + \gamma(X_{it}) + \varepsilon_{it} \tag{3.2}$$

Here, X_{it} is a collection of weather variables, γ is a general function, a_i is an additive fixed effect term, and ε_{it} is an additive error term. I use three different collections of weather variables for X_{it}, as described in Figure 3.2.1. As y_{it} is the log of crop yields, the elasticity of crop yield with respect to some variable is equivalent to the average directional derivative of γ with respect to that variable.

The analogue to β_2 from Equation (3.1) is therefore the average directional derivative of γ with respect to $higher_{it}$. Let θ^{SR} (θ^{LR}) be this average directional derivative using short-run (long-run) variation. I then take bootstrap samples of the ratio $1 - \hat{\theta}_2^{LR}/\hat{\theta}_2^{SR}$ to test the hypothesis that damages from short-run heat exposure are offset in the longer run.

This average directional derivative is equivalent to an average partial derivative when using the Yearly Linear set of weather variables. Specifically, take $\hat{\theta} = \mathbb{E}[\partial\hat{\gamma}(X_{it})/\partial higher_{it}]$. When using a linear specification, this average derivative is equivalent to the parameter estimate from OLS. When using NNet, this can be recovered after training the network (see Section 3.3.3). When the specification involves basis functions (such as OLS with polynomial basis functions or Lasso), I recover this derivative by taking the average of the dot product of the gradient of the basis functions and the estimated coefficients. I describe this procedure in Section 3.3.2.

When I use another set of weather variation, it is also necessary to account for the extent that each weather variable contributes to the total growing season heat exposure. I find this

using the chain rule:

$$\hat{\theta} = \sum_{X_{it} \in X_{it}^{higher}} \frac{\partial \hat{\gamma}(X_{it})}{\partial X_{it}} \frac{\partial X_{it}}{\partial higher_{it}} \tag{3.3}$$

Where X_{it}^{higher} is the set of weather variables that are summed to reach total growing season heat exposure above the temperature threshold. I set $\partial X_{it}/\partial higher_{it} = X_{it}/higher_{it}$; this captures the assumption that additional marginal heat will be distributed proportionally to the empirical heat exposure distribution.

This allows me to measure adaptation to climate change from recent panel variation while using a flexible model of high-dimensional weather variation. In the following sections, I describe how I estimate the regression function and average directional derivative using various estimation techniques.

3.3.2 Ordinary Least Squares methods

Ordinary Least Squares (OLS) is a classical statistics approach to estimating this elasticity. I use methods with a linear functional form (OLS Linear), and after applying a basis function transformation of polynomial functions and interactions (OLS Poly). In both cases, I use the within transformation to remove a county-level fixed effect term, and then include yearly fixed effect terms via dummy variables.

I use basis functions that include interactions and flexible functional forms of the original data. I specify polynomial basis functions of all terms, as well as interactions between these polynomial expansions. To produce a tractable model, I limit the space of potential interactions to interactions between heat exposure and precipitation variables within the same time period. For example, in the Monthly Flexible specification, I consider cumulative GDD in July between 28 and 29 C interacted with precipitation in July, as well as squared values of each term and the interactions between those polynomial expansions, but do not consider interactions of that variable with cumulative GDD in any other temperature bin, or precipitation in any other month. I use 3rd order polynomials for the Yearly Linear and Yearly Flexible variable sets, and 2nd order polynomials for the Monthly Flexible variable set. I then scale each flexible basis function so that it has mean zero and variance 1.

For OLS Linear, I use the identity set of basis functions; that is, $b(X_{it}) = X_{it}$. For Yearly Linear, X_{it} has 3 covariates; for Yearly Flexible X_{it} has 41; and for Monthly Flexible X_{it} has 246. For OLS Poly, I use the set of basis functions described above. For Yearly Linear, $b(X_{it})$ has 30 covariates; for Yearly Flexible $b(X_{it})$ has 486; and for Monthly Flexible $b(X_{it})$ has 1464.

OLS assumes that the following is a true model of the relationship:

$$y_{it} = a_i + b(X_{it})\beta_0 + \varepsilon_{it} \tag{3.4}$$

As is common in economics, I assume that we have relatively short panels where it is not possible to consistently estimate a_i by including dummy variables. I therefore use the within transformation to remove county-level fixed effect terms:

$$\ddot{y}_{it} = \ddot{b}(X_{it})\beta_0 + \varepsilon_{it} \tag{3.5}$$

where the double dot denotes the within transformation, i.e. $\ddot{y}_{it} := y_{it} - \text{mean}(\{y_{it} \forall t\})$ and $\ddot{b}(X_{it}) := \text{mean}(\{b(X_{it} \forall t\})$. Note that to construct $\ddot{b}(X_{it})$, the mean of all observations in that panel unit is subtracted after applying the basis function transformation. This ensures that β_0 is the same parameter vector between models. I use OLS to estimate $\hat{\beta}$.

I then compute the average directional derivative by projecting my estimate of β_0 on partial derivatives of the basis functions. Define the gradient of the basis function as b_{higher}; this is a 1 by p dictionary of the derivative of each basis function with respect to $higher_{it}$. The true average directional derivative is then $\theta_0 = \mathbb{E}[b_{higher}(X_{it})\beta_0]$ and its estimate is $\hat{\theta} = \mathbb{E}[b_{higher}(X_{it})\hat{\beta}]$.

In specifications in my main analysis, I include per-year fixed effects as dummy variables. I assume that there are enough observations per time period to consistently estimate these variables separately. The derivative of each per-year fixed effect is zero, so including these terms does not change how I estimate the average directional derivative.

3.3.3 Machine Learning Methods

Machine Learning (ML) methods include a range of estimation techniques that allow consistent function approximation in high-dimensional settings, when the number of covariates is large relative to the number of observations. Such settings poses challenges for classical statistical methods such as OLS, binning, or kernel regression. I use Lasso and Neural Networks (NNets), although the same procedure could be used for another algorithm such as random forests, support vector machines, or other methods. I focus on these two machine learning algorithms because each enables the researcher to incorporate linear fixed effects and to evaluate derivatives without numerical differentiation. Standard ML methods can induce bias in regression analysis; I overcome this bias by using a procedure from Chernozhukov, Newey, and Singh (2022a).

The average derivative is computed from a ML regression of the output variable on weather inputs. Write $\gamma(\cdot;\lambda)$ to denote the flexible machine learning function, emphasizing the dependence on the hyperparameter λ. The hyperparameter is a researcher-specified value that influences the behavior of the model, such as the regularization penalty in Lasso or the network width in NNet. The machine learner is estimated as a regression function: $\mathbb{E}[\ddot{y}_{it}|\ddot{X}_{it}] = \hat{\ddot{\gamma}}(X_{it};\lambda)$. Section 3.3.3 and Section 3.3.3 give details on each estimation procedure. Let $m(\gamma, X_{it};\lambda)$ denote the directional derivative of $\gamma(\cdot;\lambda)$ evaluated on observation X_{it}.

I use the double machine learning (DML) procedure from Chernozhukov, Newey, and Singh (2022a) to find an approximately debiased estimate of the average directional derivative. I discuss this procedure in more detail in Section 3.3.3. Briefly, I estimate a second machine learner and use this to construct an estimate of the average directional derivative that is robust to errors in estimating either the first or second machine learner. Let $\alpha(X;\kappa)$ denote this second machine learner, emphasizing the dependence on hyperparameter κ. Chernozhukov, Newey, and Singh (2022a) show that the expression $\mathbb{E}[m(\hat{\gamma}, X_{it};\lambda) + \hat{\alpha}(X_{it};\kappa)\hat{\varepsilon}_{it}]$ is an approximately unbiased estimate of the true average directional derivative, where $\hat{\varepsilon}_{it}$ is the residual from estimating y_{it}. I modify the form of the doubly robust estimator in Cher-

nozhukov, Newey, and Singh (2022a) to account for the panel structure of the data; details are in Section 3.3.3.

I use a data-driven process to determine the value of the hyperparameters. First split each panel unit (a U.S. county) into one of the k folds for cross validation. Grouping observations from each panel unit into the same fold reduces correlation between training and test data, as counties share unobservable characteristics that likely influence the distribution of weather and crop yields. Let $\mathcal{I}_\ell$ denote the set of indices in fold ℓ, for $\ell \in \{1, 2, \ldots, k\}$. For each of the k folds, I train the ML and DML algorithm on data not in fold ℓ, and evaluate the algorithm only on indices in fold ℓ. Let $\hat{\gamma}_\ell$ and $\hat{\alpha}_\ell$ denote the ML and DML estimators trained on the set of indices not in fold ℓ. Then select a hyperparameter value by searching over a grid of potential values, and selecting the value that minimizes a loss function. Let $\mathcal{L}_\gamma(\gamma_\ell, \mathcal{I}_\ell; \lambda)$ be the mean squared error of the function γ_ℓ with the hyperparameter λ on the data in $\mathcal{I}_\ell$. Let $\mathcal{L}_\alpha(\alpha_\ell, \mathcal{I}_\ell; \kappa)$ be the loss function of the function α_ℓ with the hyperparameter κ on the data in $\mathcal{I}_\ell$; I describe this loss function in Appendix 3.A.3.

Once the hyperparameter is selected, I evaluate the (debiased) score on the test sets using the same folds defined above. The full estimation procedure for the debiased score is below. To use this procedure without the debiasing correction, omit the bias correction term $\hat{\alpha}_\ell(X_{it}; \hat{\kappa})(\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}; \hat{\lambda}))$ from each step.

1. Select hyperparameter $\hat{\lambda}$ that minimize test-set mean squared error of the regression:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \sum_{\ell=1}^{k} \mathcal{L}_\gamma(\hat{\gamma}_\ell, \mathcal{I}_\ell; \lambda)$$

2. Select hyperparameter $\hat{\kappa}$ that minimize test-set loss of the double machine learner:

$$\hat{\kappa} = \underset{\kappa}{\operatorname{argmin}} \sum_{\ell=1}^{k} \mathcal{L}_\alpha(\hat{\alpha}_\ell, \mathcal{I}_\ell; \kappa)$$

3. Evaluate the debiased score using these hyperparameters:

$$\hat{\theta} = \frac{1}{N} \sum_{\ell=1}^{k} \sum_{it \in \mathcal{I}_\ell} m(\hat{\gamma}_\ell, X_{it}; \hat{\lambda}) + \hat{\alpha}_\ell(X_{it}; \hat{\kappa})(\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}; \hat{\lambda}))$$

4. Find the asymptotic variance of the estimator, after adjusting for within-panel-unit correlations. Let $\hat{\theta}_{\ell;it} := m(\hat{\gamma}_\ell, X_{it}; \hat{\lambda}) + \hat{\alpha}_\ell(X_{it}; \hat{\kappa})(\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}; \hat{\lambda}))$ and $\hat{\theta}_{\ell;i} := 1/T \sum_{t=1}^{T} \hat{\theta}_{\ell;it}$. Then the asymptotic variance is:

$$\hat{V} = \frac{1}{N} \sum_{\ell=1}^{k} \sum_{i \in \mathcal{I}_\ell} \left\{ \sum_{t=1}^{T} (\hat{\theta}_{\ell;it} - \hat{\theta})^2 + 2 \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} (\hat{\theta}_{\ell;it} - \hat{\theta}_{\ell;i})(\hat{\theta}_{\ell;it'} - \hat{\theta}_{\ell;i}) \right\}$$

In the following subsections, I describe how to train and evaluate the Lasso and NNet estimators, and introduce the DML procedure.

Lasso

Least absolute shrinkage and selection operator (Lasso) is a regression procedure that selects a sparse linear model from a researcher-specified set of basis functions. The procedure finds this sparse linear combination by minimizing squared error of estimation, while penalizing more complex models via regularization. The procedure is similar to OLS Poly, but estimates can differ greatly because of this penalization. The hyperparameter λ is the magnitude of the regularization term, which I determine via the cross fitting procedure outlined above.

As with OLS, I assume that there is a true linear model of the form Equation (3.4) and take within transformations to remove county-level fixed effects to result in Equation (3.5). I use the same flexible set of basis functions used in OLS Poly, as defined in Section 3.3.2. Unlike in OLS Poly, I assume that the true parameter vector is sparse and find an estimate by solving a regularized optimization problem.

For each cross-validation fold, I find the estimate $\hat{\beta}_\ell$ via the following minimization problem:

$$\hat{\beta}_\ell = \underset{\beta}{\operatorname{argmin}} \sum_{i \in \mathcal{I}_\ell} \sum_{t=1}^{T} (\ddot{y}_{it} - \ddot{b}(X_{it})\beta)^2 + \lambda |\beta|_1 \tag{3.6}$$

where $|\beta|_1$ is $\ell 1$ penalty or the sum of the absolute value of each component of β. To incorporate yearly fixed effects, I include dummy variables and do not apply the regularization penalty to the coefficients on those dummy variables. This reflects the assumption that while coefficients in γ are sparse, coefficients in the fixed effects terms are not (Belloni et al., 2016). Additional details on the Lasso procedure are included in Appendix 3.A.1.

Once I have estimated $\hat{\beta}_\ell$, I compute the directional derivative by projecting my estimate of β_0 on partial derivatives of the basis functions. This is the same procedure to compute the derivative using OLS, as described in Section 3.3.2.

Neural Network

A neural network (NNet) learns a relationship between inputs and outputs through an iterative process, where the best fitting model is selected from a large space of flexible transformations of all potential interactions of input features. NNets have several advantages for my application. First, it is straightforward to compute a gradient of the output of the entire network with respect to each input feature. Second, NNets can incorporate fixed effects in panel models, as demonstrated by Crane-Droesch (2018). Third, estimating NNets does not require the researcher to specify basis functions.

I compute derivatives of the NNet by extracting gradients computed during training the network. A NNet is a weighted composition of activation functions (user-specified transformations) applied to an input vector. After a random initialization of internal parameters, the model goes through many iterations of predicting the output variable, calculating a loss from training data, and updating the parameter values. I use the mean squared error for this loss function. While training or evaluating the network, the algorithm computes the derivative of the output with respect to each input. This is commonly used to adjust parameter values during the training procedure, in a process known as back propagation. I also use these automatic derivatives to compute the partial derivative of the prediction with respect to each weather input.

I follow Crane-Droesch (2018) to estimate the model with nonparametric treatment of weather inputs and additive linear fixed effects. Training this network involves an itera-

tive procedure that can be interpreted as selecting basis functions from a large space of candidate functions. After this training procedure is completed, I use standard econometric techniques to account for linear terms, treating the nonlinear transformation of inputs as a feature in a linear model. Specifically, I use the within transformation to remove individual fixed effects and use OLS to find other linear terms such as yearly fixed effects. The iterative procedure is performed on the set of training data, and the OLS step is performed on the test set. This requires an architecture with a top layer that is linear in all fixed effect components and the output of the nonlinear network transformation. See Crane-Droesch (2018) for more details on such networks and their performance relative to linear models or fully nonparametric models in estimating agricultural yields. Additional details of the NNet architecture and training procedure are included in Appendix 3.A.2.

Double Machine Learning

Double machine learning (DML) is an approach to remove bias from standard machine learning algorithms. I use an approach introduced by Chernozhukov, Newey, and Singh (2022a), relying on the statistical theory of the Riesz representer.

There are two related, yet orthogonal, problems in estimating the average derivative. First is the regression function, and second is the derivative operator on a regression function. Chernozhukov, Newey, and Singh (2022a) show how to conduct this second step without estimating a regression function, and that these two methods can be combined to construct an approximately debiased estimate of the original target derivative.

This second function can be estimated from data using the Riesz representation theorem. I denote this second function α_0, and its estimate $\hat{\alpha}$. The Riesz representation implies that $\mathbb{E}[\alpha_0(X_{i,t})\ddot{\gamma}_0(X_{i,t})] = m(\ddot{\gamma}_0, X_{i,t})$. Chernozhukov, Newey, and Singh (2022a) show how to use this fact to estimate α_0 from data. In a panel setting, I use the within-transformed function $\ddot{\gamma}$; I use $m(\ddot{\gamma}_0, X_{i,t}) = m(\gamma_0, X_{i,t})$ because the functional m is a directional derivative and the derivative of the mean value is 0. I assume that α_0 is linear in the set of basis functions b. The true Riesz representer is then $\alpha_0(X_{it}) := \ddot{b}(X_{it})\rho_0$ and its estimate is $\hat{\alpha}(X_{it}) := \ddot{b}(X_{it})\hat{\rho}$. I estimate $\hat{\rho}$ using an optimization package, based on the moment conditions identified by

Chernozhukov, Newey, and Singh (2022a). Klosin and Vilgalys, 2022 demonstrate that this procedure is effective in panel settings. Appendix 3.A.3 includes more details on the motivation of this Riesz representer and the estimation procedure.

This estimate is then used to construct the following doubly robust score:

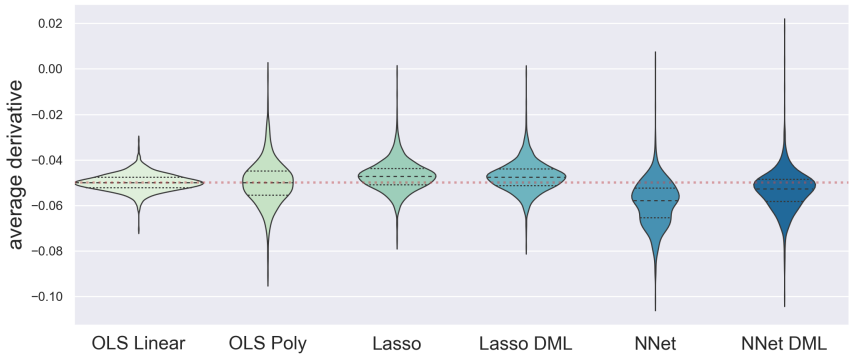
$$\hat{\theta} = \frac{1}{N} \sum_{\ell=1}^{k} \sum_{it \in \mathcal{I}_\ell} m(\hat{\gamma}_\ell, X_{it}; \hat{\lambda}) + (\ddot{y}_{it} - \hat{\ddot{\gamma}}_\ell(X_{it}))\hat{\alpha}_\ell(X_{it}; \hat{\kappa})$$

3.4 Simulation Exercise

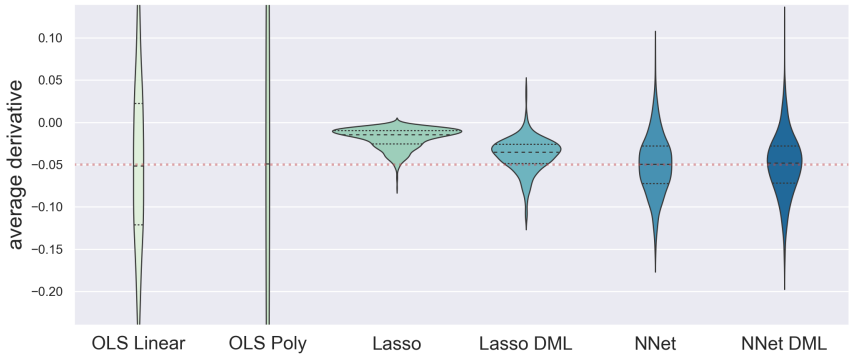
I conduct a simulation exercise to compare the performance of the estimation procedure to OLS while varying the set of weather variables used. To match the setting as closely as possible, I use the empirical distribution of weather covariates in all trials. I then apply a simulated production function of the the piecewise-linear functional form from Schlenker and Michael J Roberts (2009).

The simulation exercise focuses on comparing ML, DML, and OLS in a regression setting with high-dimensional variation. OLS is correctly specified in all trials. In cases where OLS is not correctly specified, ML and DML would likely perform better because they are able to represent a richer set of flexible functional forms. Simulation exercises by CCDDHNR (2018) demonstrate this property.

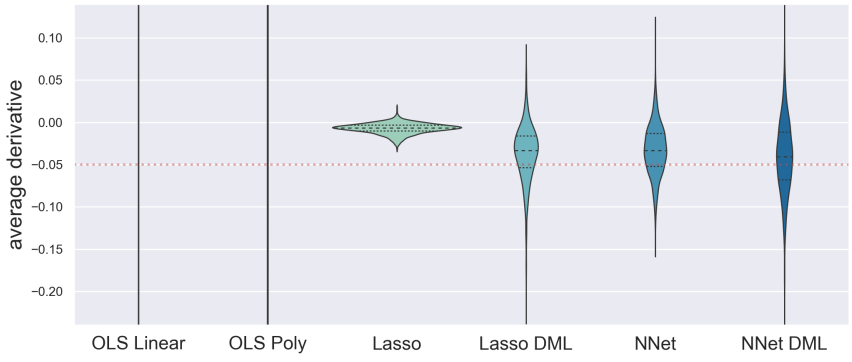
I conduct 1,000 Monte Carlo simulation trials of this estimation procedure. In each trial, I first randomly draw a 1,000 county sample, and randomly select a 2-year panel from these counties. I use the weather observations from this sample, guaranteeing that there are realistic correlations between weather variables. Then, I generate a y variable according to the following function: $y_{it} = a_i + \beta_1 lower_{it} + \beta_2 higher_{it} + \beta_3 prec_{it} + \varepsilon_{it}$, where $lower$ ($higher$) is the total growing season accumulated GDD below (above) 29°C and $prec$ is the total growing season precipitation. This functional form closely matches the parsimonious form suggested by Schlenker and Michael J Roberts (2009). Note that in this functional form, OLS Linear is correctly specified for all sets of weather variables I consider. I set $\beta_1 = 0.02; \beta_2 = -0.05; \beta_3 = 0.001$, and take $a_i \sim N(1, 1)$ and $\varepsilon_{it} \sim N(0, 1)$.



(a) Yearly Linear



(b) Yearly Flexible



(c) Monthly Flexible

Figure 3.4.1: Violin plots of distribution of parameter estimates from 1,000 Monte Carlo simulations. The subfigures indicate which set of weather variables are used, using the notation from Figure 3.2.1. Each subfigure has a separate violin plot for each method used. The plot shows the density of the parameter at the value on the y axis, based on the simulation trials. The dotted lines within each plot show the median and upper/lower quartiles of the distribution. The dotted horizontal line through the plots shows the true value of the parameter, -0.05.

Weather	method	OLS Linear	OLS Poly	Lasso	Lasso DML	NNet	NNet DML
Yearly Linear	θ	-0.0498	-0.0499	-0.047	-0.0473	-0.0588	-0.0533
		(0.003756)	(0.009589)	(0.006619)	(0.006594)	(0.009762)	(0.008802)
	MSE	0.995	0.968	1.01	1.01	1.08	1.08
Yearly Flexible	θ	-0.0507	-0.0469	-0.0184	-0.0383	-0.0498	-0.0498
		(0.108)	(0.386)	(0.01192)	(0.02)	(0.03537)	(0.03649)
	MSE	0.954	0.511	1.03	1.03	1.1	1.1
Monthly Flexible	θ	2.36	-0.0317	-0.00716	-0.0367	-0.0327	-0.0413
		(58.22)	(0.9779)	(0.006234)	(0.03336)	(0.02993)	(0.04387)
	MSE	0.752	7.32e-09	1.08	1.08	1.13	1.13

Table 3.4.1: Summary of results from 1000 Monte Carlo simulation draws of the estimation procedure. The true value of the average derivative in all trials is -0.05; θ is the average of the estimated average derivative from all trials. Each column represents a separate method used for the regression function. Standard errors are in parentheses, and are computed from the distribution of bootstrap values.

The results of this simulation exercise are summarized visually in Figure 3.4.1 and numerically in Table 3.4.1. I estimate β_2, the parameter of interest, using OLS, ML, and DML for weather variables in the Yearly Linear, Yearly Flexible, and Monthly Flexible sets (as illustrated in Figure 3.2.1). For OLS, I OLS Linear and OLS Poly as described in Section 3.3.2. For ML, I consider Lasso and NNet estimators as described in Section 3.3.3 and Section 3.3.3. For DML, I adjust each ML result with the DML procedure in Section 3.3.3. I use Lasso DML and NNet DML to describe the Lasso and NNet estimates with the DML correction. I use the cross-folds and sample splitting procedure described in Section 3.3.

With Yearly Linear weather inputs, OLS performs best among all models. OLS Poly and OLS Linear have the lowest bias, and OLS Linear has the lowest variance. The machine learning models performed reasonably well, especially with the DML correction, although they have greater variance and bias than the OLS results. The central estimate from both DML approaches lie within 7% of the true elasticity, and a 95% confidence interval contains the true elasticity.

With Yearly Flexible weather variables, ML estimates have considerably lower variance than OLS. The estimates using OLS Linear have a standard deviation approximately three times as large as the DML estimates, while estimates from OLS Poly have a standard deviation approximately ten times as large. There is substantial bias from the Lasso estimates, although the DML correction greatly reduces this bias. The NNet and NNet DML estimates

have very little bias; central estimates from both are within 0.4% of the true value and have less bias than OLS Linear. The 95% confidence interval from both DML procedures contains the true elasticity, although the 95% confidence interval from Lasso without DML does not.

With Monthly Flexible variables, the benefits of using DML are even more dramatic. OLS Linear and OLS Poly have extremely high variance, and both have greater bias than the DML methods. The high variance is especially limiting, as confidence intervals using these approaches are uninformative. As Figure 3.4.1c shows, the interquartile range of the bootstrap estimates is not visible on a plot whose range is 8 times the magnitude of the true target elasticity. OLS Linear estimates a mean value of 2.359, relative to the true value of -0.05. OLS Poly has a bias of .018, which is greater than bias from Lasso DML (0.012) or NNet DML (0.0087).

Note that while mean squared error (MSE) can help suggest a preferred model, a straightforward comparison of MSE does not select the model with lowest bias. MSE appears higher among ML models than OLS because the MSE reported by ML is the MSE of the model evaluated on a test set. When MSE from an ML model is low, this indicates that ML is providing a better fit because the model extrapolates well to unseen data. When MSE from an OLS approach is low, this could indicate overfitting. OLS Poly has the lowest MSE in all trials, but is overfitting the data in the Yearly Flexible and Monthly Flexible cases. In the data generating process (DGP), the additive error term has variance 1. The MSE from OLS Poly is much lower than the true squared error from the DGP, indicating that OLS Poly is fitting the noise instead of the desired pattern in the data. The MSE from OLS Linear with Monthly Flexible weather terms is significantly lower than the true squared error, indicating that OLS Linear may be overfitting the data in that setting. This suggests that a comparison of MSE alone should not be used to select the preferred model, but the researcher can consider test-set MSE and performance in simulation trials.

This simulation exercise shows that when a researcher wishes to measure a relationship with a high-dimensional set of weather variables, DML can provide results with lower bias and less variance than OLS. With Yearly Flexible or Monthly Flexible weather, NNet DML has the lowest bias among all models, and the DML methods have significantly lower variance.

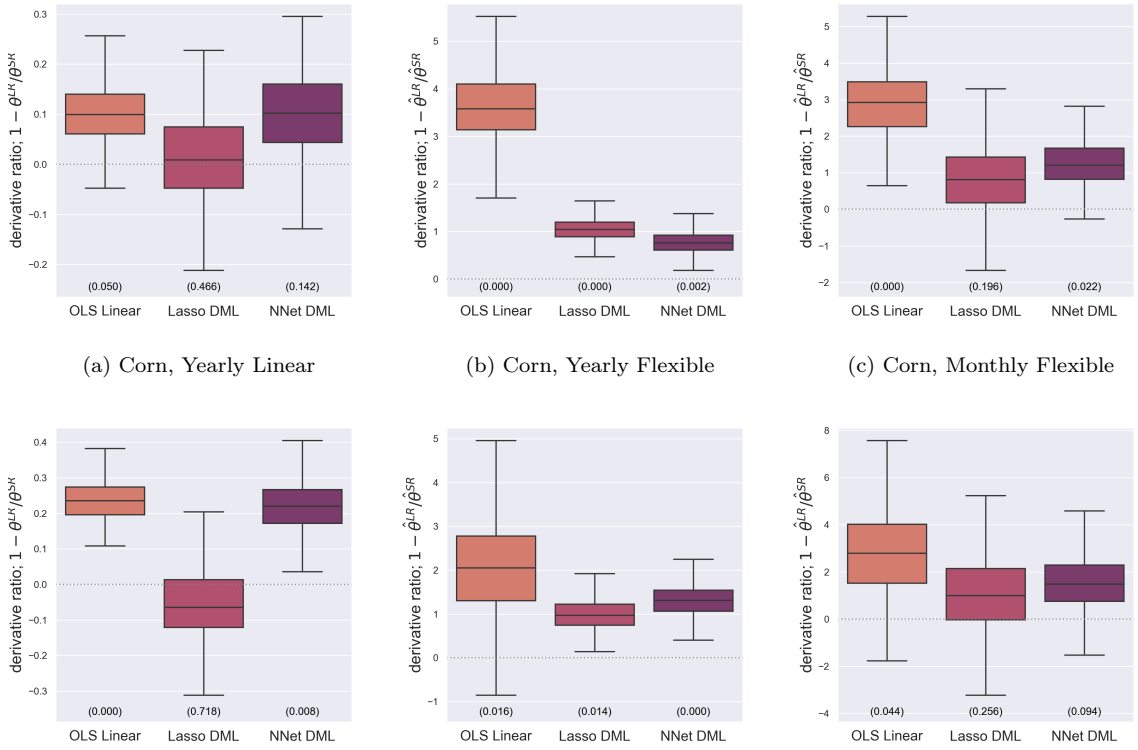
OLS performs best in the Yearly Linear case, which is expected because OLS is correctly specified and the weather variation is low-dimensional.

3.5 Results

I implement the above estimation procedure to study adaptation to damaging heat exposure in U.S. corn and soy production. When using the Yearly Linear set of weather variables (as in Burke and Emerick 2016), I find little to no evidence of adaptation. However, when using a richer set of weather variation, I find evidence that a considerable share of the short-run damages from extreme heat are offset with greater exposure to those temperatures. To help explain this discrepancy, I visualize OLS results with Yearly Flexible temperature variation to show that the simple linear model may not accurately describe the long-run role of extreme heat.

I run 500 bootstrap trials of the procedure described in Section 3.3 to measure the elasticity of crop yields with respect to extreme heat. As described in Section 3.2, I study corn and soy production from 1990-2019 in counties east of the 100 °W meridian. I use a panel of the full data to capture short-run variation, and capture long-run variation by comparing average yield and weather from 1990-1999 to 2010-2019. In each bootstrap trial, I draw a random subsamples of 80% of counties. This resampling scheme addresses the intertemporal correlation as suggested by Kapetanios (2008), while avoiding the risk of own-sample bias from machine learning methods by including the same observations in train and test data.

I estimate the elasticity using the three sets of weather variation described in Figure 3.2.1 and all estimation approaches described in Section 3.3. Each method uses the within transformation to remove individual fixed effects, and includes additive yearly fixed effects. For estimation methods that use a set of basis functions, I use polynomial expansions of all temperature and precipitation variables, and interactions of the polynomial expansions of each precipitation variable with the polynomial expansions of each temperature variable. I use third-order polynomial expansions for Yearly Linear and Yearly Flexible weather variables, and second-order polynomial expansions for Monthly Flexible. The machine learning models use a 5-fold cross validation procedure.



(a) Corn, Yearly Linear (b) Corn, Yearly Flexible (c) Corn, Monthly Flexible

(d) Soy, Yearly Linear (e) Soy, Yearly Flexible (f) Soy, Monthly Flexible

Figure 3.5.1: Box plots of 500 bootstrap samples of ratio $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$, from both corn and soy production. The subfigures indicate which set of weather variables are used, using the notation from Figure 3.2.1. The number in parentheses at the bottom of each box plot is the p value for the one-sided test that the ratio is greater than 0. Each subfigure has a separate box plot for each method used. The line inside each box is the median, the edges of the box are the upper and lower quartile of the distribution, and the whiskers extending from each box are 1.5 times the interquartile range.

I only report results from OLS Linear, Lasso DML, and NNet DML in this section. The simulation results showed that the debiasing procedures are effective at reducing bias from naive machine learning methods, and that estimates using OLS Poly can have extremely high variance. Results from the machine learning models without the debiasing procedure and from OLS with polynomial basis functions are included in Appendix 3.B. Generally, estimates using machine learning without bias correction are similar to their debiased counterparts, and estimates using OLS Poly have unacceptably high variance when using richer sets of weather variation.

Reassuringly, the results using Yearly Linear weather variation are similar to those of Burke and Emerick (2016). Table 3.5.1 summarizes the results from the bootstrap trials. Long-run

Crop	Weather	method	Long-Run OLS Linear	Long-Run Lasso DML	Long-Run NNet DML	Short-Run OLS Linear	Short-Run Lasso DML	Short-Run NNet DML
Corn	Yearly Linear	θ	-0.00481***	-0.007***	-0.00597***	-0.00535***	-0.00709***	-0.00663***
			(0.000326)	(0.000568)	(0.000464)	(5.68e-05)	(8.05e-05)	(0.000337)
		MSE	0.00652	0.00648	0.00673	0.0387	0.0363	0.0362
		N	2850	2850	2850	38633	38633	38633
	Yearly Flexible	θ	0.0322***	0.000429	-0.00249	-0.0125***	-0.0106***	-0.0104***
			(0.00826)	(0.00234)	(0.00247)	(0.00106)	(0.00115)	(0.000603)
		MSE	0.00531	0.00582	0.00603	0.0376	0.0343	0.0352
		N	2850	2850	2850	38633	38633	38633
	Monthly Flexible	θ	0.0213*	-0.000793	0.00148	-0.0114***	-0.00505***	-0.00674***
			(0.00874)	(0.00533)	(0.00404)	(0.00101)	(0.000612)	(0.000718)
		MSE	0.00263	0.00544	0.00461	0.0334	0.0303	0.0306
		N	2850	2850	2850	38633	38633	38633
Soy	Yearly Linear	θ	-0.00409***	-0.0055***	-0.00409***	-0.00537***	-0.00525***	-0.00525***
			(0.000307)	(0.000574)	(0.000331)	(5.2e-05)	(8.07e-05)	(0.000252)
		MSE	0.00579	0.00562	0.00578	0.0308	0.029	0.0291
		N	2422	2422	2422	33799	33799	33799
	Yearly Flexible	θ	0.00805	-7.48e-05	0.00239	-0.00788***	-0.00698***	-0.00764***
			(0.0079)	(0.00276)	(0.0026)	(0.00109)	(0.000843)	(0.000648)
		MSE	0.00432	0.0056	0.00519	0.03	0.0276	0.0282
		N	2422	2422	2422	33799	33799	33799
	Monthly Flexible	θ	0.00955	0.000192	0.00247	-0.00565***	-0.00362***	-0.00465***
			(0.00914)	(0.0069)	(0.00572)	(0.00104)	(0.000717)	(0.000611)
		MSE	0.00204	0.00451	0.00421	0.0253	0.0236	0.0234
		N	2422	2422	2422	33799	33799	33799

Table 3.5.1: Summary of results from 500 bootstrap draws of the estimation procedure. θ is the mean of the average derivative over bootstrap trials. Standard errors are in parentheses, and are computed from the distribution of bootstrap values. N is the average number of samples per simulation trial. Stars indicate significance at the $p = 0.05(*), 0.01(**)$, and $0.001(***)$ levels, based on a Z-score from the mean and standard error of bootstrap trials.

and short-run datasets both find that additional extreme heat decreases crop yields, with comparable magnitudes for both corn and soy cultivation. The magnitude of these elasticities are economically significant – an estimate of -0.005 implies that crop yields decline by 0.5% for each additional day crops are exposed to temperatures above 29°C. The findings are similar to the results reported by Burke and Emerick (2016). They find that the elasticity of corn yield ranges from -0.0037 to -0.0062.

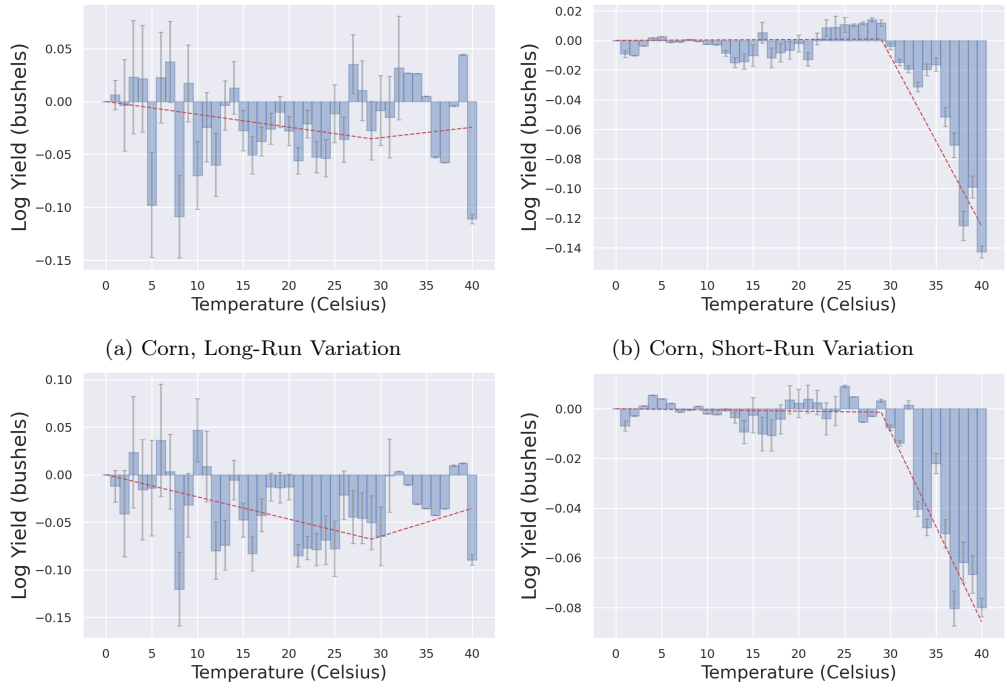
As in Burke and Emerick (2016), with Yearly Linear weather variation I find little to no evidence that damages from the short run are offset in the long run. Figure 3.5.1 shows the results of taking bootstrap samples of the ratio $1 - \hat{\theta}^{LR}/\hat{\theta}^{SR}$. As described in Section 3.3.1, $\hat{\theta}^{SR}$ ($\hat{\theta}^{LR}$) is the estimated elasticity with short-run (long-run) variation. For corn cultivation, I fail to reject the null hypothesis that this ratio is different from 0 for any estimation method.

I test at the $p = 0.05$ level, with a Bonferroni correction to account for taking three hypothesis tests. For soy cultivation, I find mixed results: for both OLS Linear and NNet DML I reject the null hypothesis at this level, while with Lasso DML I fail to reject the null hypothesis. For these soy estimates, the 95% confidence interval is $[0.1278, 0.3488]$ using OLS Linear and $[0.07153, 0.3649]$ using NNet DML. Confidence intervals are computed using the sample mean and standard deviation among bootstrap trials.

Using Yearly Flexible weather variables, I find significant evidence of adaptation. The panel with short-run variation finds statistically and economically significant damages from a marginal increase in extreme heat exposure. Central estimates range from -0.01057 to -0.01248 for corn, and -0.006983 to -0.00788 for soy. However, I fail to reject the null hypothesis that the long-run elasticity is different from zero, for both crops using all specifications. Figure 3.5.1 confirms this finding. For both crops and using all estimation methods, I reject the null hypothesis that the degree of short-run damages that are offset in the long run is equal to zero.

Using this set of weather variables, my preferred estimator is NNet DML. Simulation results showed the this estimator performed best with Yearly Flexible weather variables. The MSE also suggests that the NNet Double is performing well in this case. The MSE of NNet DML decreases for long-run regressions for both corn and soy. As the MSE of NNet DML is the test-set MSE, an improvement relative to the MSE from the Yearly Linear weather variables reflects am improvement in modeling the true functional form instead of overfitting. Using this preferred estimate, I find a 95% confidence interval that the ratio is within $[0.3023, 1.222]$ for corn and $[0.6331, 2.001]$ for soy.

The results from using the Monthly Flexible set of weather variables are similar to those using Yearly Flexible weather variables, although with greater variance. This greater variance is not surprising, as the simulation results demonstrated that DML and OLS have higher bias and variance using this set of weather variables. With short-run variation using this set of weather variables, there is an economically and statistically significant decline in yields associated with a marginal increase in heat exposure. With long-run variation, there is no evidence of significant declines in yields from a marginal increase in heat exposure. Reas-



(a) Corn, Long-Run Variation

(b) Corn, Short-Run Variation

(c) Soy, Long-Run Variation

(d) Soy, Short-Run Variation

Figure 3.5.2: Comparison of relationship between temperature and yields with short-run and long-run variation. Each plot summarizes the coefficient estimates from results with the Yearly Flexible set of weather variables, using NNet DML. The bar plot shows the expected change in yields from exposure to a single degree day at the given temperature level. Error bars show one standard deviation above/below each estimated coefficient. The line shows the best piecewise linear fit to the bar plot, with one piece from 0-29°C and another from 29-40 °C.

suringly, this evidence is consistent with the result using Yearly Flexible weather variation. Given the greater variance of these estimates, I fail to reject the null hypothesis that the ratio of these elasticities is different from zero except using DML methods. Using OLS Linear, I find that a marginal increase in heat exposure significantly increases crop yields.

These results show that, when the temperature-crop yield relationship is modeled flexibly, there are not significant average declines in yields associated with a marginal change in long-run damaging heat exposure. I do not estimate a significant decline from long-run exposure to temperatures above 29°C because other variation is better able to explain differences in average long-run yields. This is not the case when using short-run variation, where flexible modeling confirms that a marginal increase in exposure to temperatures above 29°C is significantly associated with a decline in crop yields. This finding is novel, and suggests significant adaptation to damaging heat exposure.

To better explain this result, I examine the estimates of temperature coefficients from each weather bin of the Yearly Flexible weather variables. I use my NNet DML procedure to estimate these plots. Section 3.C includes results using alternate estimation procedures. Due to the computational cost of estimating each average derivative using the DML procedure, I do not replicate the DML procedure for each of these temperature coefficients; the results are similar. Figure 3.5.2 visualizes the results of these regressions. Figure 3.5.2b and Figure 3.5.2d show the results of this estimate using short-run variation. These results are similar to the results from Schlenker and Michael J Roberts (2009), who noted that the piecewise linear fit (the dotted line in these figures) explains most of the variation from the flexible modeling approach (the bar chart in these figures). Here, the piecewise linear model seems to be a good fit: before around 29°C, additional heat exposure is generally associated with an increase in yields. After this point, additional heat exposure is associated with a decline in yields. Like Burke and Emerick (2016), I find that 29°C appears to be threshold for damaging heat exposure.

The results using long-run variation do not share this pattern. Figure 3.5.2a and Figure 3.5.2c show the results of this estimate using long-run variation. The coefficients do not demonstrate a clear piecewise linear pattern. This suggests that the piecewise linear model relating temperature exposure and crop yield may not be appropriate in long-run models. In Appendix 3.C, I replicate this figure for different time periods, starting from 1950-1979. These figures confirm that the piecewise linear model is a better fit when using short-run variation than when using long run variation. While the results from the Yearly Linear analysis show that long-run damaging heat exposure is correlated with a decline in crop yields, these results show that other temperature variation is better able to explain the long-run changes in crop yields. This shows the importance of flexible modeling to understanding the role of a marginal increase in damaging heat exposure.

3.6 Discussion

I introduced a DML procedure that can estimate average directional derivatives in high-dimensional settings, and demonstrated the benefits of this procedure over OLS estimation in

a simulation exercise. The results show that this procedure can estimate average directional derivatives with lower bias and variance than OLS. The simulation results also show that my method is less reliable than having access to a true, parsimonious model of the underlying function. In settings where the researcher has high-dimensional weather variation and does not have a strong prior about the true functional form, this DML approach can be used to estimate elasticities and the degree of adaptation to a weather feature.

Applying this estimator to panel of U.S. corn and soy yields with a rich set of temperature variation, I conclude that there has been significant adaptation to damaging heat exposure. Using this flexible method, a panel of short-run damages finds evidence that a marginal increase in heat exposure above 29°C is damaging for both corn and soy yields. However, by constructing a dataset of average changes to capture long-run variation, I cannot reject the null hypothesis that a marginal increase in long-run damaging heat exposure is unrelated to long-run crop yields. OLS estimates using a Yearly Flexible set of weather variables support this finding. I demonstrate that long-run exposure to temperatures above 29°C is not clearly associated with yield declines, as is the case with short-run exposures. This implies that there has been significant adaptation to climate change in this setting, as the short-run damages are significantly offset in the long run.

This approach does not offer insight into what form this adaptation may take. Farmers have many possible adaptation mechanisms, such as investing in irrigation, purchasing improved seeds, or adjusting planting times. It is important to learn which mechanisms may be responsible for offsetting damages in the long run, in order to study the long-term effectiveness of these mechanisms or the ability to use them in other agricultural contexts. More research is needed to understand how these damages are offset.

The apparent contradiction between this result and prior literature suggests that while there is adaptation to a marginal increase in damaging heat exposure, there is limited adaptation to some other damaging feature of climate change. Prior literature, such as Burke and Emerick (2016), did not find evidence of substantial adaptation to damaging heat exposure. My analysis shows that while a long-run increase in damaging heat exposure is correlated with a decline in crop yields, variation in heat exposure at other temperature levels is better able

to explain this pattern. This is evidence of omitted variable bias in long-run estimates using only beneficial and damaging heat exposure. While prior results indicate limited adaptation to some change in the temperature distribution, my detailed analysis finds that there has been substantial adaptation to damaging heat exposure.

Appendix to Chapter 3

3.A Machine Learning Estimation Details

This appendix additional details on the ML estimation procedures.

3.A.1 Lasso

I vary the Lasso hyperparameter (the regularization penalty) on a grid of 15 values, evenly distributed (in log space) from 10^{-10} to 10^{0}. The cross folds validation procedure generally selects a hyperparameter from the interior of this grid, suggesting that the range is appropriate.

Recall the form of the optimization problem:

$$\hat{\beta}_\ell = \underset{\beta}{\operatorname{argmin}} \sum_{i \in \mathcal{I}_\ell} \sum_{t=1}^{T} w_i (\ddot{y}_{it} - \ddot{b}(X_{it})\beta)^2 + \lambda |\beta|_1 \tag{3.7}$$

As this is a (weakly) convex optimization problem, I solve for $\hat{\beta}_\ell$ using the optimization package CVXPY (Diamond and Boyd, 2016) with the optimizer Mosek (ApS, 2021). In the optimization package implementation, the optimizer can occasionally fail to find a unique optimal solution. In these cases, I introduce an $\ell 2$ of 10^{-20} to find a unique solution. Using the optimization package ensures that the Lasso procedure converges, although this procedure is more memory intensive than coordinate descent and will raise an error if it fails to converge. I find that the model fails to converge in roughly 5% of simulation trials.

In the main specification, I include yearly fixed effects terms but do not apply a regularization

to these terms. I accomplish this by selectively applying the regularization factor to terms in β not in those yearly fixed effects terms. The additional fixed effects are retuned before evaluating the estimator on the test set; this does not impact the estimate of the gradient, but does improve the mean squared error of the estimator.

I use the optimization package for two reasons. When the solver does not return an error, the solution is guaranteed to be optimal, unlike in iterative methods such as stochastic gradient descent or coordinate descent. Some packages contain this functionality for iterative methods, but I am not aware of an implementation in Python.

3.A.2 Neural Network

I use a relatively simple network configuration, and vary the width of the neural network (NNet) in each simulation trial. Three key researcher degrees of freedom when using a NNet are the depth (the number of layers), the width (the number of nodes per hidden layer), and the activation function (the nonlinear function applied to the outputs of each layer). I use a network with one hidden layer, the minimal depth for which a NNet can approximate an arbitrary function (Hornik, Stinchcombe, and White, 1990; Park and Sandberg, 1991). In each trial run, I use cross-folds validation to select the width of the network. To ensure that the network is differentiable, I use the Continuously Differentiable Exponential Linear Unit (Barron, 2017) as the activation function in the network.

I select the hyperparameter from a grid space of 2 to 256, evenly spaced in log terms. The cross folds procedure generally selects a width from the interior of this set, suggesting that the range is appropriate. To train the network, I use batch normalization, the Adam optimizer (Kingma and Ba, 2015), a learning rate of 0.01, and 1000 epochs of training. I retune the parametric components (the county and yearly fixed effects terms) before evaluating on the test set. This does not impact the estimate of the gradient, but does improve the mean squared error of the estimator.

Simulation results show that this simple configuration performs well, although it is possible that a more complex network configuration would perform better. Given the computational cost of exploring a wide range of potential network configurations, I leave such exploration

for future work.

3.A.3 Automatic Double Machine Learning

The theoretical basis for the double machine learning (DML) procedure comes from the Riesz representation theorem. This theorem states that, for the linear functional[2] m, there exists a function α_0 such that for any function h:

$$\mathbb{E}[m(h, X_i)] = \mathbb{E}[h(X_i)\alpha_0(X_i)] \tag{3.8}$$

Because this relationship holds regardless of function h, one can set known functions for which the target $m(h, X_i)$ is known and estimate $\hat{\alpha}$ from the empirical distribution of X. Alternate procedures involve the researcher solving for the functional form of α given the functional m, and often estimating densities or derivatives of densities. I use this procedure because it does not require the researcher specifying the form of α, but instead learns $\hat{\alpha}$ from data.

I follow Chernozhukov, Newey, and Singh (2022a) to estimate $\hat{\alpha}$ using Lasso and the basis functions used to estimate the regression function. The goal is to find the estimator that minimizes the mean squared error between $\hat{\alpha}$ and α_0, where α_0 is the true Riesz representer:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}}\, \mathbb{E}[(\alpha_0(X_{i,t}) - \alpha(X_{i,t}))^2]$$

Using the Lasso functional form, I have $\hat{\alpha} = \ddot{b}(X_{i,t})\hat{\rho}$. I find $\hat{\rho}$ by solving the following regularized problem:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}}\, \mathbb{E}\left[(\alpha_0(X_{i,t}) - \ddot{b}(X_{i,t})\rho)^2\right] + \kappa|\rho|_1$$

Expanding the polynomial and applying the definition of the Riesz representer, the expression simplifies:

$$\hat{\rho} = \underset{\rho}{\operatorname{argmin}} -2\mathbb{E}[m(\ddot{b}, X_{i,t})]\rho + \rho'\mathbb{E}[\ddot{b}(X_{i,t})'\ddot{b}(X_{i,t})]\rho + \kappa|\rho|_1$$

[2] A functional is a scalar summary of a function. The functional in this setting is the directional derivative.

Chernozhukov, Newey, and Singh (2022a) include an additional iterative procedure to scale each component's regularization term by the inverse of its variance; I implement their suggested procedure without modification.

As this is a (weakly) convex optimization problem, I use the optimization package Mosek (ApS, 2021) to find the optimal value of parameter vector $\hat{\rho}$ for a given value of regularization κ. Klosin and Vilgalys (2022) present simulation results comparing this optimization method to the iterative approach from Chernozhukov, Newey, and Singh (2022a); they find that the optimization approach results in lower mean squared error in estimating the true average derivative.

I select the value of the hyperparameter κ by minimizing the above loss function through the cross-folds procedure described in Section 3.3.3. That is:

$$\mathcal{L}_\alpha(\hat{\alpha}_\ell, X_{i,t}; \kappa) = -2m(\ddot{b}, X_{i,t})\hat{\rho}_\ell + \hat{\rho}_\ell' \ddot{b}(X_{i,t})' \ddot{b}(X_{i,t})\hat{\rho}_\ell$$

I search the hyperparameter over a grid of values suggested by Chernozhukov, Newey, and Singh (2022a): $\kappa = c(N - N_\ell)^{-1/2}\Phi^{-1}(1 - .05/p)$, for $c \in \{5/4, 1, 3/4, 5/8, 9/16, 1/2\}$, where Φ^{-1} is the inverse of the standard normal density function and N_ℓ is the size of the set of observations in fold ℓ.

3.B Additional Results

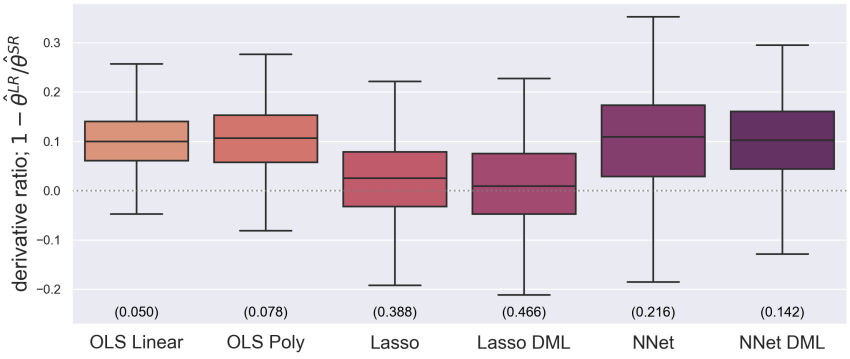
This appendix includes additional results from the bootstrap trials described in Section 3.5. Table 3.B.1 shows the full results from the estimation of corn yields, and Table 3.B.2 shows the full results from the estimation of soy yields. As expected, the OLS Poly results have higher variance than other methods, particularly in higher-dimensional settings. After applying the bias correction to Lasso, the variance increases. Figure 3.B.1 show the bootstrapped values of the ratio $1 - \hat{\theta}^{LR}/\hat{\theta}^{SR}$ from all models for corn yields, and Figure 3.B.2 show the bootstrapped values of this ratio from soy yields. Note that the variance from the OLS Poly estimates dominates. These standard errors are large and in charge, particularly in Figure 3.B.1c and Figure 3.B.2c.

Variation	Weather	method	OLS Linear	OLS Poly	Lasso	Lasso DML	NNet	NNet DML
Long-Run	Yearly Linear	θ	-0.00481***	-0.0063***	-0.00692***	-0.007***	-0.00583***	-0.00597***
			(0.000326)	(0.00052)	(0.000547)	(0.000568)	(0.000485)	(0.000464)
		MSE	0.00652	0.00584	0.00648	0.00648	0.00673	0.00673
		N	2850	2850	2850	2850	2850	2850
	Yearly Flexible	θ	0.0322***	0.014	-0.00135	0.000429	0.00147	-0.00249
			(0.00826)	(0.0294)	(0.000882)	(0.00234)	(0.00155)	(0.00247)
		MSE	0.00531	0.00159	0.00582	0.00582	0.00603	0.00603
		N	2850	2850	2850	2850	2850	2850
	Monthly Flexible	θ	0.0213*	-5.83e+04	-0.00184***	-0.000793	0.00356**	0.00148
			(0.00874)	(1.25e+06)	(0.00042)	(0.00533)	(0.00126)	(0.00404)
		MSE	0.00263	3.72e-06	0.00544	0.00544	0.00461	0.00461
		N	2850	2850	2850	2850	2850	2850
Short-Run	Yearly Linear	θ	-0.00535***	-0.00704***	-0.00708***	-0.00709***	-0.00645***	-0.00663***
			(5.68e-05)	(7.64e-05)	(7.96e-05)	(8.05e-05)	(0.000595)	(0.000337)
		MSE	0.0387	0.0363	0.0363	0.0363	0.0362	0.0362
		N	38633	38633	38633	38633	38633	38633
	Yearly Flexible	θ	-0.0125***	-0.0115***	-0.0106***	-0.0106***	-0.00845***	-0.0104***
			(0.00106)	(0.000988)	(0.00112)	(0.00115)	(0.00055)	(0.000603)
		MSE	0.0376	0.0316	0.0343	0.0343	0.0352	0.0352
		N	38633	38633	38633	38633	38633	38633
	Monthly Flexible	θ	-0.0114***	-14	-0.0038***	-0.00505***	-0.00562***	-0.00674***
			(0.00101)	(491)	(0.000589)	(0.000612)	(0.000677)	(0.000718)
		MSE	0.0334	0.0258	0.0303	0.0303	0.0306	0.0306
		N	38633	38633	38633	38633	38633	38633

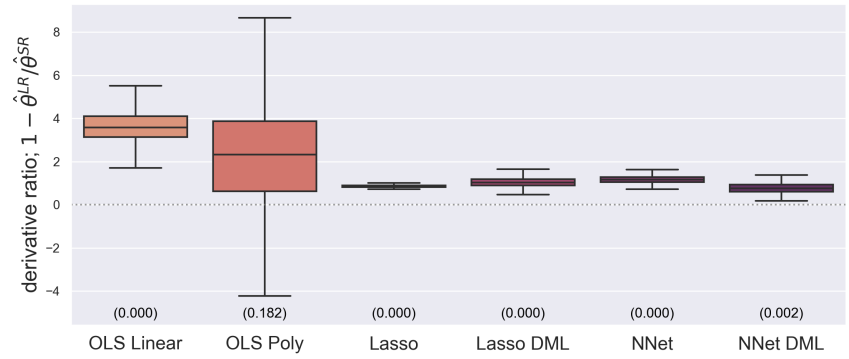
Table 3.B.1: Summary of results from 500 bootstrap draws of the estimation procedure, for corn production. θ is the mean of the average derivative over bootstrap trials. Standard errors are in parentheses, and are computed from the distribution of bootstrap values. N is the average number of samples per simulation trial. Stars indicate significance at the $p = 0.05(*), 0.01(**)$, and $0.001(***)$ levels, based on a Z-score from the mean and standard error of bootstrap trials.

Variation	Weather	method	OLS Linear	OLS Poly	Lasso	Lasso DML	NNet	NNet DML
Long-Run	Yearly Linear	θ	-0.00409***	-0.0042***	-0.00538***	-0.0055***	-0.00404***	-0.00409***
			(0.000307)	(0.000485)	(0.000585)	(0.000574)	(0.000374)	(0.000331)
		MSE	0.00579	0.00479	0.00562	0.00562	0.00578	0.00578
		N	2422	2422	2422	2422	2422	2422
	Yearly Flexible	θ	0.00805	-0.0554	-0.00169*	-7.48e-05	0.00559**	0.00239
			(0.0079)	(0.0328)	(0.000677)	(0.00276)	(0.00204)	(0.0026)
		MSE	0.00432	0.000905	0.0056	0.0056	0.00519	0.00519
		N	2422	2422	2422	2422	2422	2422
	Monthly Flexible	θ	0.00955	0.00374	-0.00182***	0.000192	0.00518***	0.00247
			(0.00914)	(0.0779)	(0.000299)	(0.0069)	(0.00128)	(0.00572)
		MSE	0.00204	2.74e-23	0.00451	0.00451	0.00421	0.00421
		N	2422	2422	2422	2422	2422	2422
Short-Run	Yearly Linear	θ	-0.00537***	-0.00522***	-0.00525***	-0.00525***	-0.00509***	-0.00525***
			(5.2e-05)	(7.43e-05)	(8.01e-05)	(8.07e-05)	(0.000321)	(0.000252)
		MSE	0.0308	0.029	0.029	0.029	0.0291	0.0291
		N	33799	33799	33799	33799	33799	33799
	Yearly Flexible	θ	-0.00788***	-0.00651***	-0.00686***	-0.00698***	-0.00649***	-0.00764***
			(0.00109)	(0.00102)	(0.000884)	(0.000843)	(0.000877)	(0.000648)
		MSE	0.03	0.0258	0.0276	0.0276	0.0282	0.0282
		N	33799	33799	33799	33799	33799	33799
	Monthly Flexible	θ	-0.00565***	-38.3	-0.00396***	-0.00362***	-0.00444***	-0.00465***
			(0.00104)	(3.4e+03)	(0.000839)	(0.000717)	(0.000544)	(0.000611)
		MSE	0.0253	0.0193	0.0236	0.0236	0.0234	0.0234
		N	33799	33799	33799	33799	33799	33799

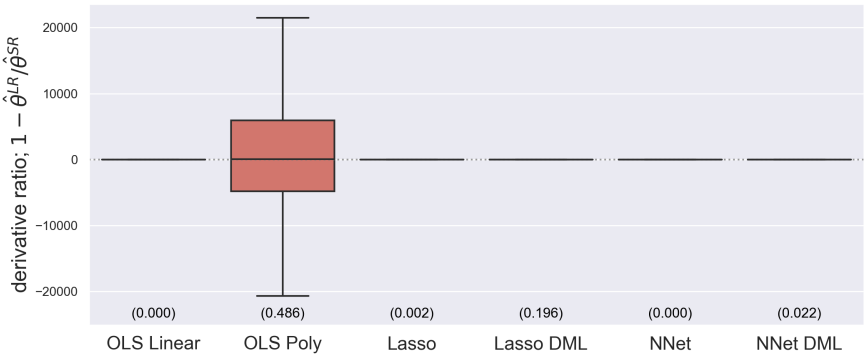
Table 3.B.2: Summary of results from 500 bootstrap draws of the estimation procedure, for soy production. θ is the mean of the average derivative over bootstrap trials. Standard errors are in parentheses, and are computed from the distribution of bootstrap values. N is the average number of samples per simulation trial. Stars indicate significance at the $p = 0.05(*), 0.01(**)$, and $0.001(***)$ levels, based on a Z-score from the mean and standard error of bootstrap trials.



(a) Corn, Yearly Linear

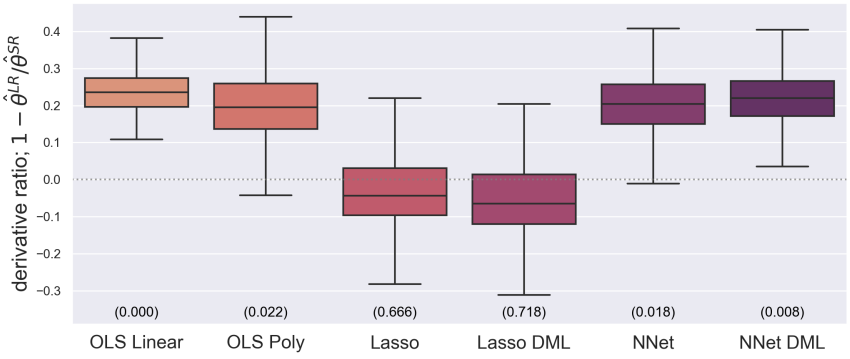


(b) Corn, Yearly Flexible

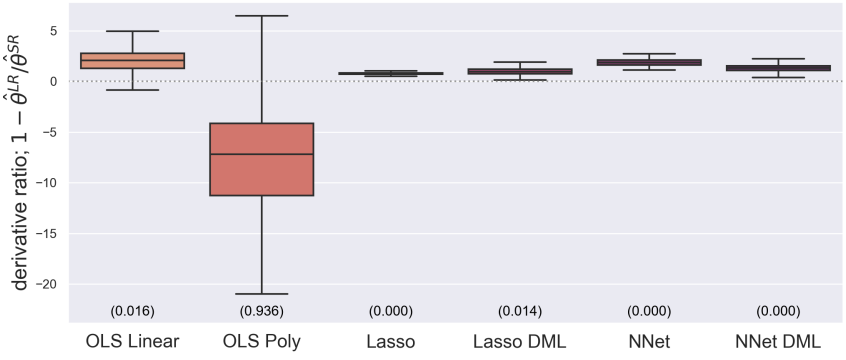


(c) Corn, Monthly Flexible

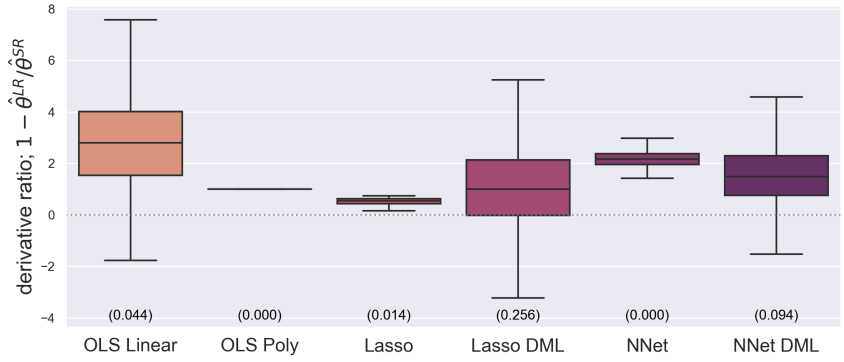
Figure 3.B.1: Box plots of 500 bootstrap samples of ratio $1 - \hat{\theta}^{LR}/\hat{\theta}^{SR}$, from corn production. The subfigures indicate which set of weather variables are used, using the notation from Figure 3.2.1. The number in parentheses at the bottom is the p value for the one-sided test that the ratio is greater than 0. Each subfigure has a separate box plot for each method used. The line inside each box is the median, the edges of the box are the upper and lower quartile of the distribution, and the whiskers extending from each box are 1.5 times the interquartile range.



(a) Soy, Yearly Linear



(b) Soy, Yearly Flexible



(c) Soy, Monthly Flexible

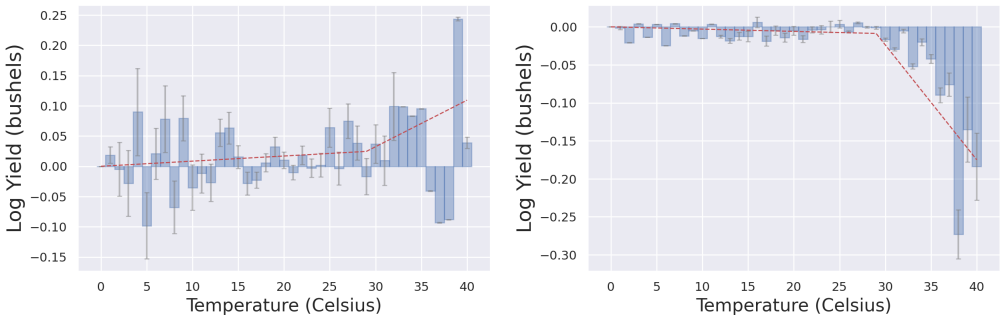
Figure 3.B.2: Box plots of 500 bootstrap samples of ratio $1-\hat{\theta}^{LR}/\hat{\theta}^{SR}$, from soy production. The subfigures indicate which set of weather variables are used, using the notation from Figure 3.2.1. The number in parentheses at the bottom is the p value for the one-sided test that the ratio is greater than 0. Each subfigure has a separate box plot for each method used. The line inside each box is the median, the edges of the box are the upper and lower quartile of the distribution, and the whiskers extending from each box are 1.5 times the interquartile range.

3.C Short-Run and Long-Run Estimates of Yearly Coefficients

I provide additional results to show the difference between short-run and long-run estimates of yearly coefficient bins. These results confirm that the functional form of long-run damages does not follow the well-established pattern of short-run damages. I demonstrate that this conclusions is robust to different modeling choices and time periods.

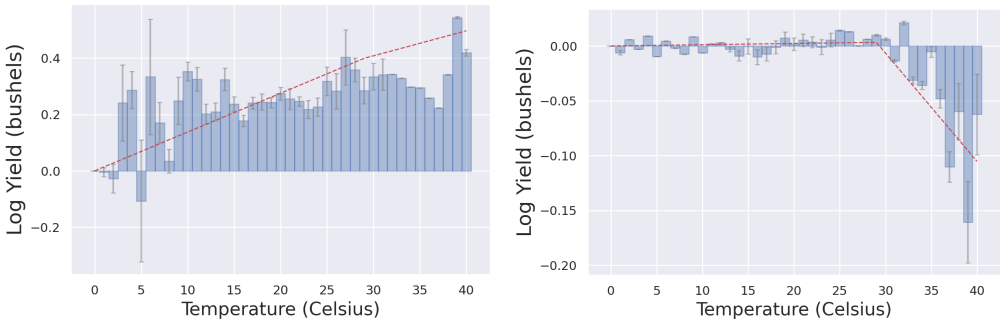
3.C.1 Comparisons using alternate estimators

The following plots show results from the comparison between short-run and long-run estimates for each yearly coefficient bin. All plots use the sample from the main estimation, corn and soy yields from 1990-2019. Each plot summarizes the coefficient estimates from results with the Yearly Flexible set of weather variables, using either Lasso DML, OLS Poly, or OLS Linear. The bar plot shows the expected change in yields from exposure to a single degree day at the given temperature level. Error bars show one standard deviation above/below each estimated coefficient. The line shows the best piecewise linear fit to the bar plot, with one piece from 0-29°C and another from 29-40 °C.



(a) Corn, Long-Run Variation, Lasso DML

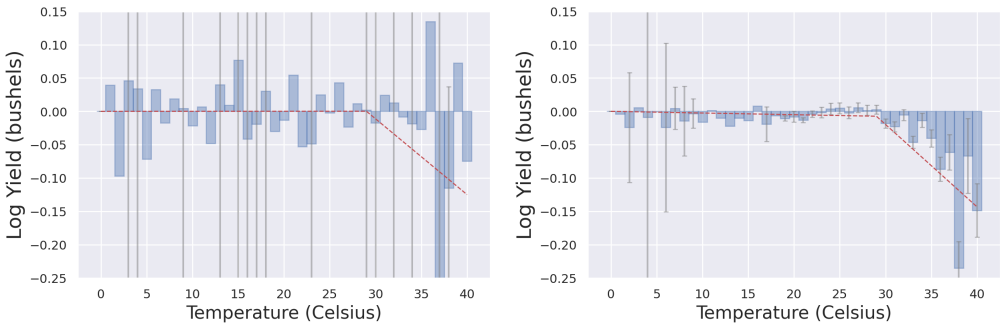
(b) Corn, Short-Run Variation, Lasso DML



(c) Soy, Long-Run Variation, Lasso DML

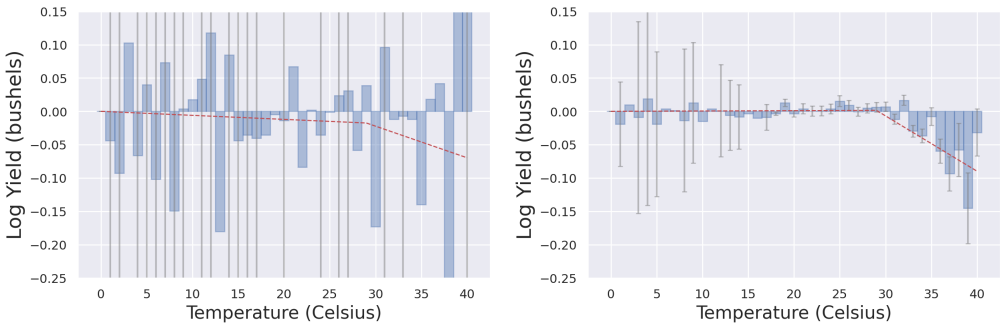
(d) Soy, Short-Run Variation, Lasso DML

Figure 3.C.1: Comparison of relationship between temperature and yields with short-run and long-run variation, using Lasso DML.



(a) Corn, Long-Run Variation, OLS Poly

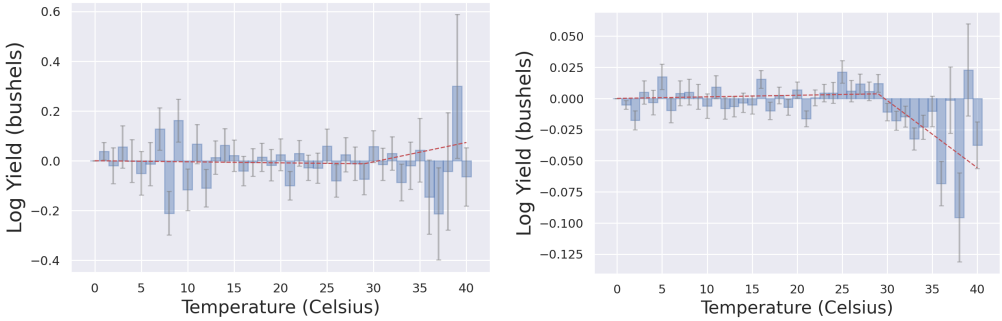
(b) Corn, Short-Run Variation, Lasso DML



(c) Soy, Long-Run Variation, OLS Poly

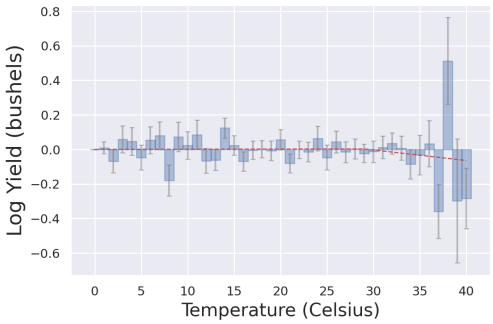
(d) Soy, Short-Run Variation, OLS Poly

Figure 3.C.2: Comparison of relationship between temperature and yields with short-run and long-run variation, using OLS Poly.

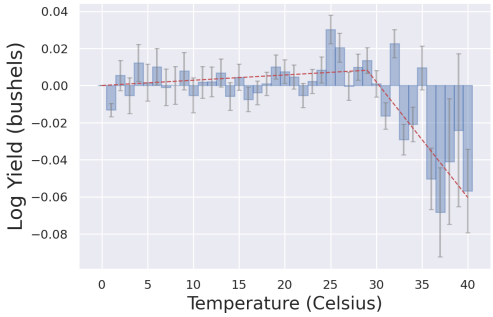


(a) Corn, Long-Run Variation, OLS Linear

(b) Corn, Short-Run Variation, Lasso DML



(c) Soy, Long-Run Variation, OLS Linear



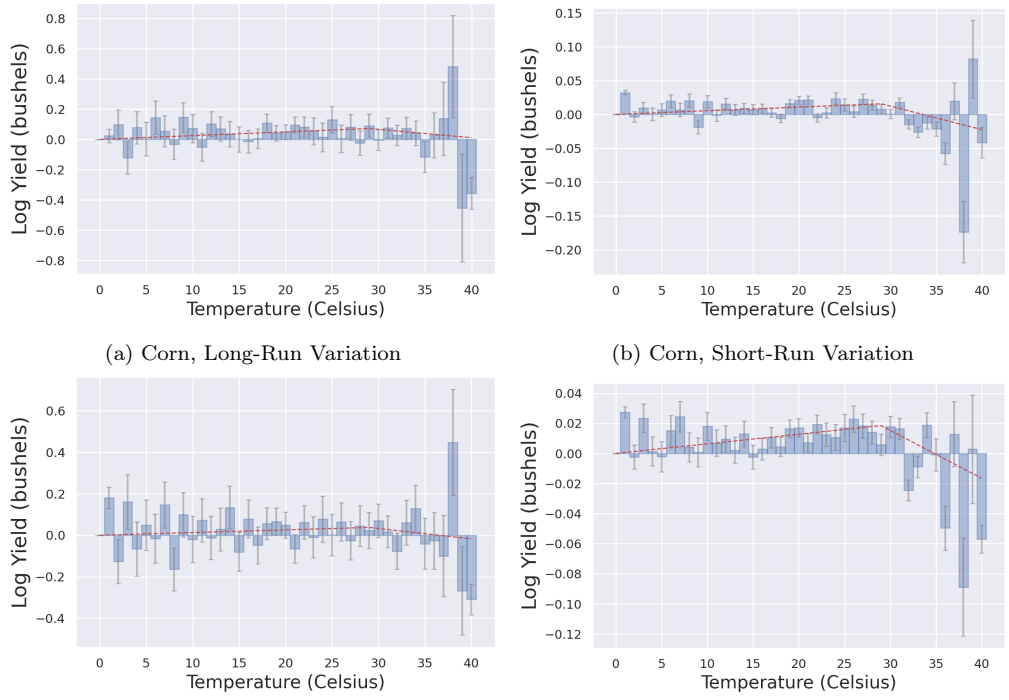
(d) Soy, Short-Run Variation, OLS Linear

Figure 3.C.3: Comparison of relationship between temperature and yields with short-run and long-run variation, using OLS Linear.

3.C.2 Comparisons Over Time

In this section, I provide OLS estimates of Yearly Flexible weather variables using short-run and long-run variation from different historical periods. In each figure, I compare short-run estimates using a panel dataset all years in the 30-year range, and long-run estimates comparing average weather and crop yield values from the first and last decades in that range.

Each figure contains four separate plots. Each plot summarizes the coefficient estimates from OLS results with the Yearly Flexible set of weather variables. The bar plot shows the expected change in yields from exposure to a single degree day at the given temperature level. Error bars show one standard deviation above/below each estimated coefficient. The line shows the best piecewise linear fit to the bar plot, with one piece from 0-29°C and another from 29-40 °C. Regression is taken after the within transformation, and also includes yearly fixed effects terms and precipitation; these terms are omitted from the figures.



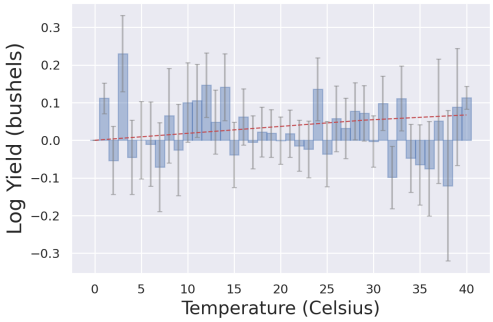
(a) Corn, Long-Run Variation

(b) Corn, Short-Run Variation

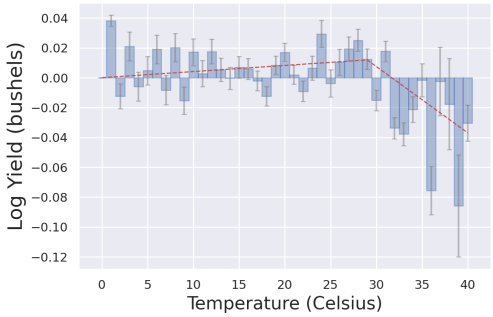
(c) Soy, Long-Run Variation

(d) Soy, Short-Run Variation

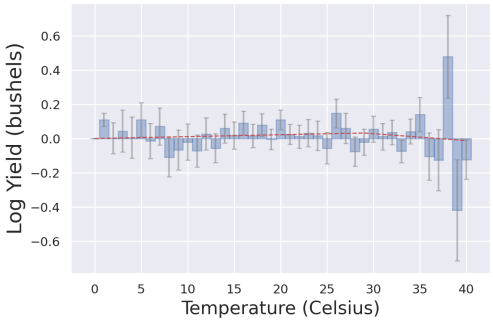
Figure 3.C.4: Comparison of relationship between temperature and yields with short-run and long-run variation, from 1950-1979.



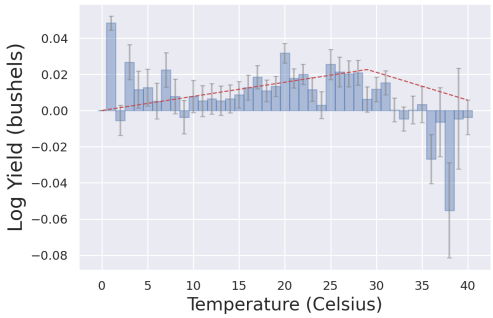
(a) Corn, Long-Run Variation



(b) Corn, Short-Run Variation

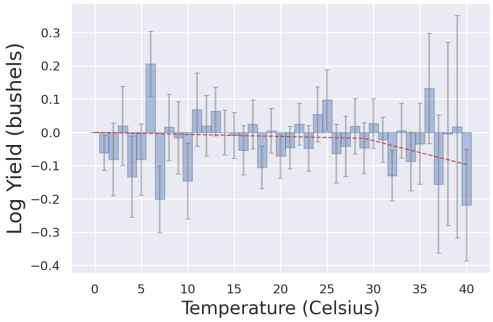


(c) Soy, Long-Run Variation

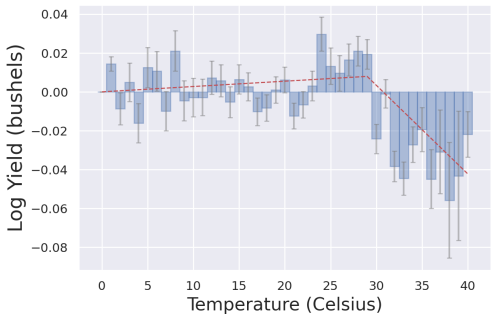


(d) Soy, Short-Run Variation

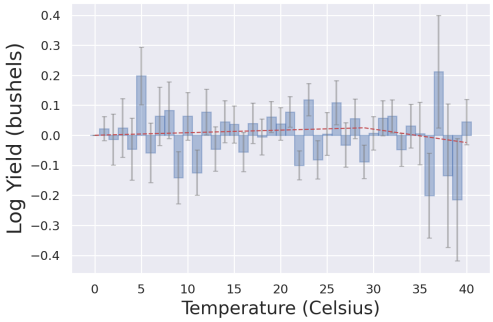
Figure 3.C.5: Comparison of relationship between temperature and yields with short-run and long-run variation, from 1960-1989.



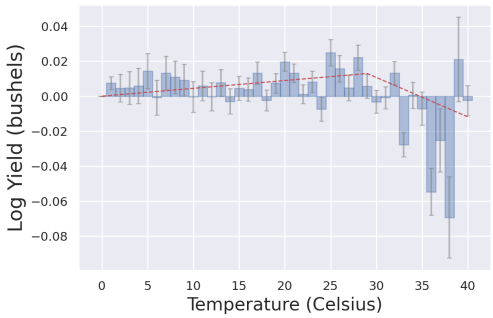
(a) Corn, Long-Run Variation



(b) Corn, Short-Run Variation

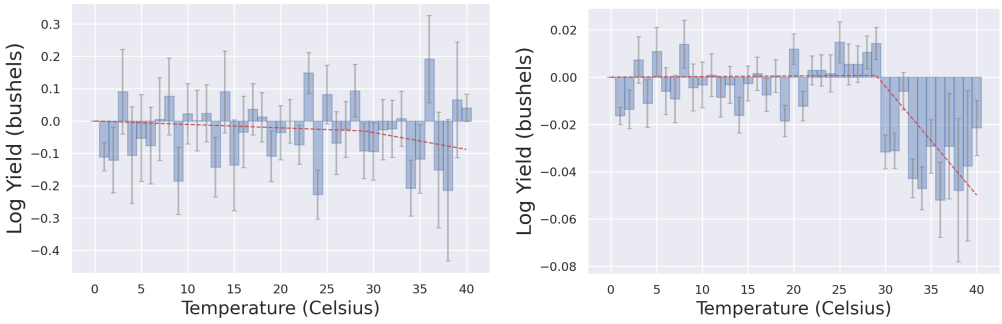


(c) Soy, Long-Run Variation



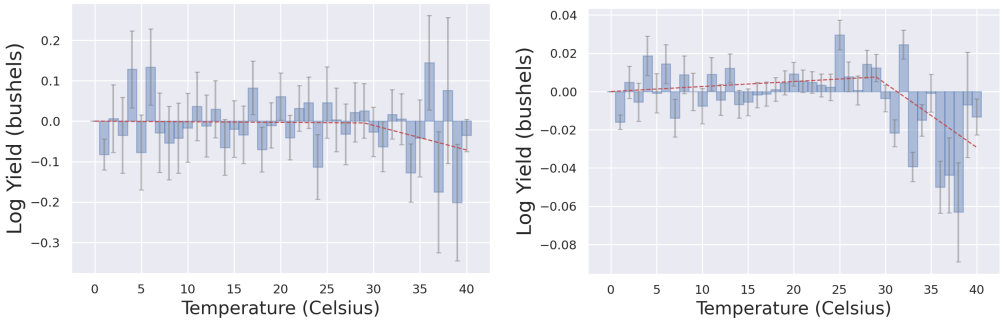
(d) Soy, Short-Run Variation

Figure 3.C.6: Comparison of relationship between temperature and yields with short-run and long-run variation, from 1970-1999.



(a) Corn, Long-Run Variation

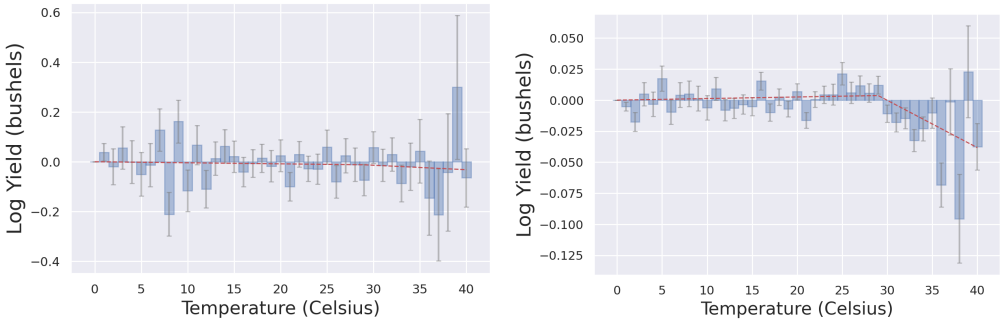
(b) Corn, Short-Run Variation



(c) Soy, Long-Run Variation

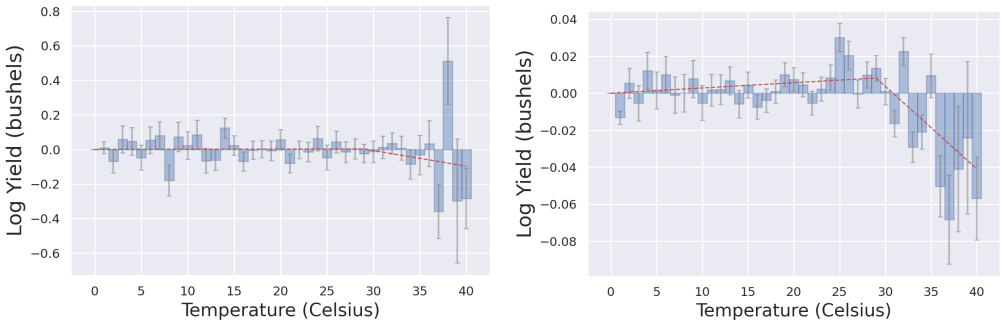
(d) Soy, Short-Run Variation

Figure 3.C.7: Comparison of relationship between temperature and yields with short-run and long-run variation, from 1980-2009.



(a) Corn, Long-Run Variation

(b) Corn, Short-Run Variation



(c) Soy, Long-Run Variation

(d) Soy, Short-Run Variation

Figure 3.C.8: Comparison of relationship between temperature and yields with short-run and long-run variation, from 1990-2019.

Chapter 4

Equity and Adaptation to Wildfire Risk: Evidence from California Public Safety Power Shutoffs

4.1 Introduction

In the last decade, electric utilities in California have been forced to adapt to increasing risk of catastrophic wildfire. Climate change, forest management practices, and shifting wildland-urban interface have contributed to the most severe wildfire seasons in California's history. Some of the costliest wildfires have been sparked by California electric utilities, and utilities are responsible for those financial damages under California law. Utilities have already faced billions in dollars in fines, notably driving Pacific Gas and Electric (PG&E) to declare bankruptcy in 2019. To make their electric lines safer, utilities invest in managing vegetation, upgrading infrastructure, and moving lines underground. However, these improvements are relatively slow and wildfire risk can require utilities to respond quickly.

This paper focuses on utilities' last-ditch effort to prevent wildfire, the Public Safety Power Shutoff (PSPS). In a PSPS, a utility preemptively de-energizes lines that are likely to spark large wildfires. PSPS imposes concentrated costs to impacted communities, and diffuse

benefits to the utility and the overall public if utilities are successfully preventing wildfire. These shutoffs are subject to strict regulation, and utilities must demonstrate that they carefully weigh the costs and benefits of each de-energization event (CPUC, 2019). The first PSPS was in 2013, and since then over 5,000 circuits (small segments of the electric grid) have been de-energized and over 1 million customers impacted (Hill et al., 2020). I focus on shutoffs by the three largest investor-owned utilities in California: Southern California Edison (SCE), Pacific Gas and Electric (PG&E), and San Diego Gas and Electric (SDG&E). Utilities must disclose PSPS events, unlike other investments for wildfire resilience that are protected as information for critical national infrastructure. This makes PSPS decisions a useful data source to examine the equitability of utilities' adaptation to wildfire risk.

While PSPS is necessary in the short run to adapt to rising wildfire risk, shutoffs could exacerbate inequalities if disadvantaged communities receive fewer benefits or bear more costs. This paper focuses on evaluating the distribution of costs of PSPS. It is challenging to evaluate the benefit of PSPS because this depends on predicting the damages from wildfires that might occur without these shutoffs. Predicting wildfire size and damages are notoriously challenging problems, even given modern machine learning techniques (Taylor et al., 2013; Xi et al., 2019; Jain et al., 2020). Utilities use proprietary models and datasets to predict wildfire size, and only declare a PSPS event when they believe that there is a high probability of catastrophic wildfire (SCE, 2021; PG&E, 2021; SDG&E, 2021a). Many Californians benefit from reduced wildfire risk, and rural populations or those with health risk factors may benefit most from reducing wildfire smoke (D'Evelyn et al., 2022). I do not attempt to model the benefits of PSPS, but note that these shutoffs are likely preventing catastrophic wildfires and their associated health risks.

I examine how the costs of PSPS are distributed by measuring the extent that PSPS decisions have been equitably targeted. Equitable targeting means that communities with the same observable risk factors (e.g. weather variation), but differ in health risk factors or socioeconomic status (SES), experience the same rate of shutoffs. I focus on health risk and SES as measures of vulnerability because low-SES communities may have limited resources to adapt to electricity failures and those with health risks may experience health complica-

tions from wildfire smoke or electricity outages. I use definitions of health risk and SES from CalEnviroScreen (August et al., 2021). Correlation between PSPS decisions and vulnerability indices could be explained by differences in treatment by the utilities or by various weather, vegetation, or development conditions that impact wildfire risk. If this correlation persists after accounting for all factors outside the utility's control, actions by the utility (such as unequal infrastructure investments or biased PSPS decision rules) are responsible. I compare circuits that differ in average SES or health risk, with and without population and weather controls. This provides evidence of whether vulnerable populations experience different rates of shutoffs and whether that is explained by observable factors outside the utility's control.

To conduct this analysis, I first construct an extensive dataset linking vulnerability indices, weather, and PSPS records from 2014-2021. For vulnerability, I use Census tract-level data on health risk factors and SES indices from August et al. (2021). The health risk index includes rates of asthma, cardiovascular disease, and low birth weights. The SES index includes unemployment, rates of high school completion, and linguistic isolation. My weather observations come from GridMET, a dataset of daily observations of 13 weather variables used to predict wildfire size (Abatzoglou, 2013). I use records of PSPS outages and ignitions along power lines from filings to the California Public Utilities Commission (CPUC). I merge all these datasets with geospatial records of circuits from utilities' Integrated Capacity Analysis maps.

I find that there are significant associations between vulnerability scores and PSPS decisions, although these vary by utility. Without controlling for weather factors, circuits with lower SES are significantly more likely to have a shutoff in PG&E and SDG&E (at the $p = 0.001$ level), and less likely in SCE (p value 0.005). Populations with higher health risk are significantly more likely to have a shutoff in SDG&E (at the $p = 0.001$ level) and less likely (at the $p = 0.001$ level) in PG&E. After controlling for population and weather variation, model fit improves but these patterns remain largely consistent. This shows that the difference in rates of PSPS by vulnerability indices is largely unexplained by population or weather differences.

To better understand what factors lead to these decisions, I develop a model of PSPS decisions based on guidelines in utilities' published Wildfire Mitigation Plans (SCE, 2021; PG&E, 2021; SDG&E, 2021a). When weather conditions suggest that large fires are likely, utilities identify circuits that could spark a large wildfire. Teams of meteorologists, fire scientists, and data scientists predict regions where ignitions are likely to spread to large fires. They use data of line conditions from public weather reports, service crews, and private weather stations to identify lines that could spark wildfires. These experts form predictions using these rich sources of public and private information, relying on machine learning models and extensive simulations of wildfire behavior. If their predictions find that the likely costs of wildfire exceed the costs of shutting off power, they notify residents and de-energize the circuit. Power remains off until weather conditions are less severe and the utility inspects affected circuits for any debris or damage.

I examine how health risk and SES indices correlate with the probability of an ignition and the firm's cost of declaring PSPS. To find the probability of ignition, I use logistic regression with records of ignitions along circuits from 2014-2021. I find that ignitions are significantly (at the $p = 0.001$ level) more frequent in low-SES circuits and in lower health risk circuits in PG&E, after controlling for weather variation. To find the cost firms incur from declaring PSPS, I examine Wildfire Mitigation Plans and post-event reports. In their regulatory filings, utilities state that the cost of declaring PSPS is linear in the expected size of interruption (SCE, 2021; PG&E, 2021; SDG&E, 2021a). However, the calculations I find from post-event reports indicate that cost is linear in number of customers (Valdberg, Tozer, and Kilberg, 2021). The results using either metric as a proxy are fairly noisy, and I do not observe a strong association between vulnerability indices and the utility's cost of declaring PSPS.

I can infer how vulnerability indices correlate with the firm's estimated cost of wildfire by imposing a model of the utility's decision problem. With my model, the estimated coefficients relating probability of PSPS, probability of ignition, and utility's cost of declaring PSPS imply the coefficients on the missing component: the utility's estimated damage from a wildfire. I do not access this data directly, because utilities use complex, proprietary software to project wildfire damages. After controlling for weather and population, my

results are fairly noisy. The only significant finding (after controlling for population and weather variation, at the $p = 0.001$ level) is that PG&E estimates higher wildfire damages in lower-SES circuits. Overall, I cannot reject that utilities are equitably estimating damages.

This project relates to several literatures. First is a literature studying the environmental justice of wildfire risk. In early work to explore this topic, Niemi and K. S. Lee (2001) describe how poverty can increase wildfire incidence and damages and Ojerio (2008) shows that federal wildfire preparedness grants are concentrated in higher-SES communities. One strand of this literature focuses on comparing populations that live in high wildfire risk regions. Wigtil et al. (2016) document that places with higher wildfire potential generally have lower social vulnerability to wildfire risk. Wibbenmeyer and Robertson (2022) find higher average property value, older residents, and more white residents in places with high wildfire potential. Another strand focuses on the impacts and responses of wildfires. D'Evelyn et al. (2022) argue that the health effects of wildfire smoke disproportionately impact populations with limited adaptive capacity. Anderson, A. Plantinga, Wibbenmeyer, et al. (2020) study inequality in firefighting responses, and document preferential treatment to higher SES communities following salient wildfire events. A. J. Plantinga, Walsh, and Wibbenmeyer (2022) study the historical spread of fires and find that firefighting efforts prioritize high-value properties.

Within this literature, several recent studies have examined PSPS as a tool to combat wildfire risk. Guliasi (2021) gives an analysis of the political economy and history of the PSPS. Hill et al. (2020) examines potential health costs from PSPS, and Wong-Parodi (2020) surveys impacted California residents about attitudes towards PSPS events. Rhodes, Ntaimo, and Roald (2020) studies the PSPS as an optimization problem, and suggests improvements to current decision processes using a test case. My paper is the first, to my knowledge, to empirically study the equity of these shutoff decisions.

This project is also related to a literature on identifying bias in decision making, specifically in cases where agents make decisions relying on complex algorithms. There is a broad literature on studying discrimination in decision-making, dating back to at least Becker (1957). Lang and Kahn-Lang Spitzer (2020) and Mehrabi et al. (2021) provide reviews of economics and

machine learning literature, respectively, on identifying bias in decision making. Recent examples examining bias in human decisions include an analysis of racial bias in healthcare decision rules (Obermeyer et al., 2019) and in pretrial appearance risk (Rambachan, 2021). Examples examining bias in algorithms include facial recognition software (Buolamwini and Gebru, 2018) and predicting risks from medical records data (Gianfrancesco et al., 2018; Parikh, Teeple, and Navathe, 2019). Like these studies, I examine decisions and look for evidence of unequal treatment after controlling for relevant, exogenous variation. I focus on a setting that is less well-studied in the literature, where agents make algorithm-supported decisions.

My project is also related to literature on measuring equity in adaptation to climate change. Among environmental advocates, there has long been a call to focus on equity in climate change adaptation (Smit and Pilifosova, 2003; Thomas and Twyman, 2005). In their report, IPCC (2022) identifies several settings where inequality and poverty have set "soft limits" on the ability of groups to adapt to climate change. Coggins et al. (2021) conducted a review of literature on equity in climate change adaptation and highlighted several examples of work assessing the equity of climate adaptation. Sheller and Leon (2016) use interviews to study how historical inequalities between Haiti and the Dominican Republic impacted government responses to similar environmental crises, and Satyal, Byskov, and Hyams (2021) use environmental justice theory to examine how systemic injustices facing an indigenous group in Uganda undermine adaptation planning. However, Coggins et al. (2021) ultimately conclude that more work is needed in this area, especially in empirical assessment of equity and justice. This paper addresses this by providing more work on empirical assessment of equity and justice in these shutoff decisions.

The remainder of this paper is structured as follows. Section 4.2 describes the various datasets used in the analysis, and provides summary statistics. Section 4.3 describes my modeling approach including references to utilities' filings that justify my modeling decisions. Section 4.4 gives the results of my analysis, and discusses their interpretation. Section 4.5 concludes.

4.2 Data

I use a variety of sources to construct a dataset with circuit-level records of weather variation, vulnerability indices, and shutoff decisions from 2014-2021. The unit of observation is a circuit-day. An electric circuit is a small unit of the electricity distribution network, and generally the level at which PSPS decisions are recorded. I treat vulnerability as fixed over the sample period.

4.2.1 PSPS Events

For PSPS events, I use filings from firms to the CPUC. Firms are required to report statistics after each shutoff, so this dataset represents the universe of shutoffs between October 2013 and December 2021. The CPUC summarizes these reports and publishes a record of each shutoff. Each record includes the circuit targeted, the date and time of the shutoff, the duration of the outage, the number of customers impacted, and information on what types of customers are impacted. Table 4.2.1 summarizes these filings by year and firm.

PSPS events are generally reported at the circuit level. In some cases, a firm reports a sub-circuit level outage. I sum these outages to the circuit level to match the weather records in my data.

In order to link these with other geospatial records, I use integrated capacity analysis (ICA) maps from each electric utility. ICA maps are circuit-level maps of the distribution infrastructure, although some circuit segments are not published due to privacy concerns. I am able to match over 98% of PSPS records to their corresponding geographic file. The ICA maps include 5,411 circuits; there are PSPS events recorded on 20.3% of these circuits. Figure 4.2.1a shows the location of these circuits.

4.2.2 Fire Data

My main analysis uses data on fire ignitions along utility lines from filings to the CPUC from 2014 through 2021. Per CPUC guidelines, firms must report all fires to their knowledge larger than one meter (CPUC, 2014). This dataset includes 4,550 ignitions from the three firms I

Utility	year	2013	2014	2017	2018	2019	2020	2021
SCE	Customers	–	–	–	–	196,879	235,879	117,690
	Million CMI	–	–	–	–	353	280	372
	# PSPS Events	–	–	–	–	246	1,501	122
PG&E	Customers	–	–	–	47,324	1,987,783	645,859	79,630
	Million CMI	–	–	–	89.8	6,670	1,560	174
	# PSPS Events	–	–	–	32	1,458	670	219
SDG&E	Customers	179	884	17,111	21,036	45,337	93,058	–
	Million CMI	0.0797	0.665	40.5	65.4	78.2	165	–
	# PSPS Events	3	6	51	38	218	110	–

Table 4.2.1: Number of PSPS events by firm, by year, and the number of customers impacted. CMI is Customer Minutes Impacted, the product of the minutes of shutoff and number of customers per circuit. Note that number of customers impacted is the sum of customer shutoffs experiences, but not the unique number of customers impacted.

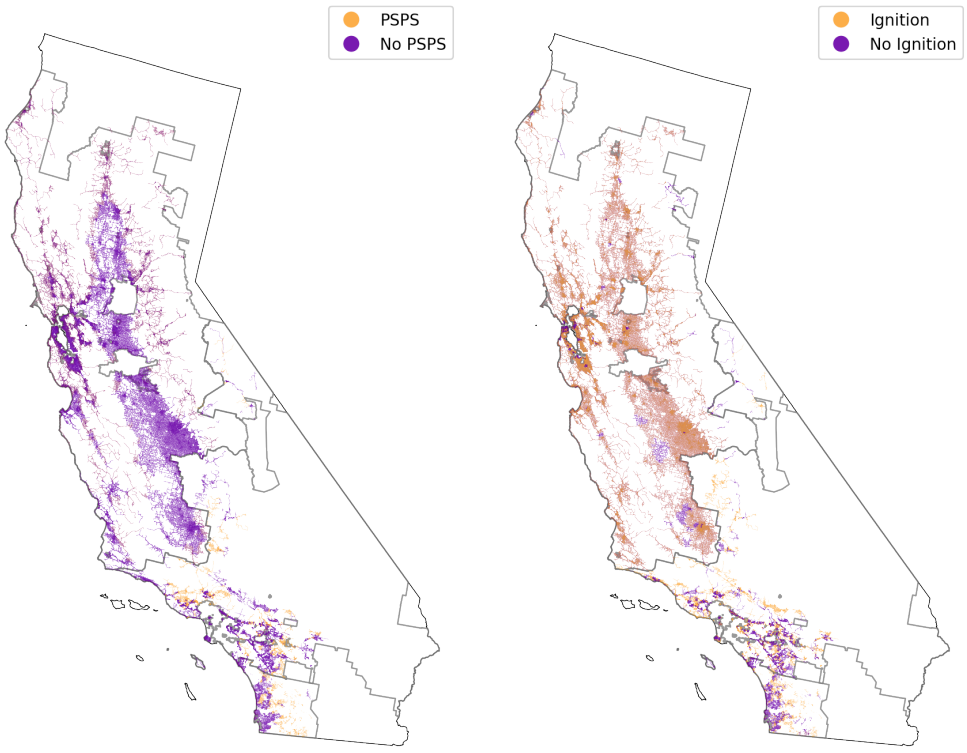
study. These filings are required to include the ignition location, but not the corresponding circuit segment. To match these to the circuit records, I find the closest circuit segment from the ICA maps to the ignition location. Figure 4.2.1b shows the location of these circuits.

4.2.3 Vulnerability

I use indices from CalEnviroScreen to measure population vulnerability (August et al., 2021). The authors construct a Census tract-level database of health risk factors and socioeconomic status (SES) indicators. I use these indexes, as well as the tract-level population, in my analysis. This database is primarily intended to assess environmental and energy justice in the state of California.

In the main analysis, I summarize these data with an SES index and a health risk index. Each index ranges from 0-100, with 100 being the most vulnerable and 0 being the least. The indices are constructed as the average of ranks of several factors, as in August et al. (2021). For socioeconomic vulnerability, this includes rate of high school non-attainment, rent-burdened low-income households, limited English proficiency, living below twice the federal poverty line, and share unemployed. For the health risk index, this includes asthma incidence, cardiovascular disease incidence, and rate of low birth weight infants.

To match these records to circuits, I take the average of values from each census tract that



(a) PSPS Events

(b) Ignitions

Figure 4.2.1: Map of which circuits had PSPS events (a) and ignitions (b) from my data. State borders and boundaries between electric utilities are shown.

contains a given circuit segment. I weight these averages by the length of the circuit in each census tract. I am able to match records for 5,000 out of 5,411 circuits, and for 1,071 of the 1,103 circuits with a PSPS event.

Figure 4.2.2 plots these scores per circuit against the total number of PSPS events (among circuits with at least one event), the total number of recorded ignitions (among circuits with at least one ignition), and the total customer minutes interrupted (among circuits with at least one event). Each plot also includes the best-fitting line to these observations, to help summarize the trend among these scatter plots.

4.2.4 Weather data

For weather observations, I use the GridMET weather dataset from Abatzoglou (2013) and an archive of areas with a red flag warning. This dataset was designed to support applications

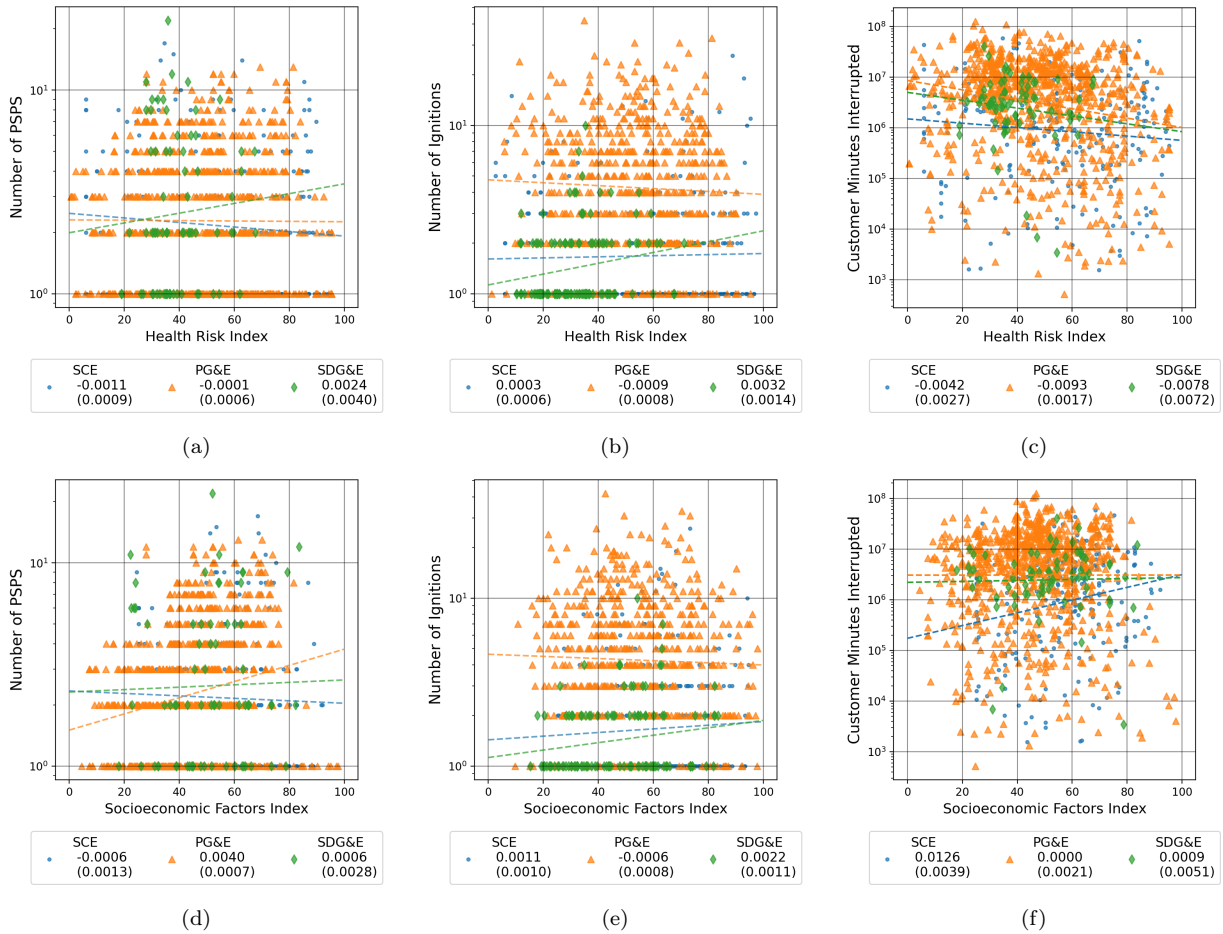


Figure 4.2.2: Scatter plots showing vulnerability indices and various outcomes, for circuits with nonzero values of PSPS events or ignitions. Each plot includes a best-fitting line for the observations. The coefficient on the index and and the standard error (in parentheses) of each line are reported in the legend.

in modeling wildfire risk, and includes a rich set of relevant weather variables. Each variable is reported daily at a high spatial resolution (4 km) across the United States; I include observations from California.

GridMET includes primary variables, constructed via satellite- and geography-guided interpolation from weather stations, and variables derived from these primary observations. To merge GridMET records with my dataset, I find all grid points within $2\sqrt{2}$ km of a circuit and take the simple average of weather records at each observation. Primary variables are specific humidity, precipitation, minimum relative humidity, maximum relative humidity, surface downwelling shortwave flux in air (a measure of solar radiation), minimum air temperature, maximum air temperature, wind speed, and wind direction. Derived variables are

	(1)	(2)	(3)
Max Air Temperature (C)	23.80	28.17	23.31
	(8.048)	(6.364)	(6.390)
Min Air Temperature (C)	10.01	11.91	9.897
	(5.564)	(5.073)	(4.850)
Precipitation Amount (daily mm)	1.265	0.0595	0.00312
	(5.598)	(0.775)	(0.0579)
Specific Humidity (kg/kg)	0.00639	0.00521	0.00394
	(0.00216)	(0.00251)	(0.00167)
Wind Velocity at 10 m (m/s)	3.318	3.670	5.325
	(1.581)	(1.655)	(2.102)
Wind From Direction (Degrees past North)	233.6	228.1	224.2
	(82.76)	(90.89)	(111.6)
Mean Vapor Pressure Deficit (kPa)	1.274	1.945	1.565
	(0.946)	(0.836)	(0.694)
Max Relatively Humidity (%)	77.82	57.08	50.94
	(19.37)	(21.26)	(19.13)
Min Relatively Humidity (%)	33.80	16.62	15.40
	(18.47)	(11.80)	(9.803)
Surface Downwelling Shortwave Flux (W/m^2)	223.8	232.1	190.1
	(96.88)	(74.95)	(47.95)
Burning Index (Derived)	36.15	54.96	68.55
	(20.91)	(16.57)	(19.45)
Energy Release Component (Derived)	46.07	67.08	70.08
	(23.77)	(14.89)	(14.03)
Potential Evapotranspiration (Derived, mm)	4.204	5.440	5.092
	(2.409)	(2.138)	(1.783)
Reference Evapotranspiration (Derived, mm)	5.769	7.994	7.978
	(3.342)	(3.076)	(2.787)
Dead Fuel Moisture 100 hr (Derived, %)	12.91	8.538	7.821
	(5.100)	(3.055)	(2.591)
Dead Fuel Moisture 1000 hr (Derived, %)	14.03	9.771	9.301
	(5.494)	(2.631)	(2.689)
Elevation	241.0	314.8	493.6
	(325.2)	(382.9)	(400.1)
Observations	16303343	652310	3333

Table 4.2.2: Summary statistics of GridMET data from October 2013 through 2021. Columns separate the full sample, the sample during a red flag warning, and the sample during a PSPS event. Observations are weighted by the length of each circuit segment. Standard errors of the mean for each column are in parentheses.

expected to be relevant for predicting wildfire risk: burning index, energy release component, potential evapotranspiration, reference evapotranspiration, dead fuel moisture at 100 hours, and dead fuel moisture at 1000 hours. See Abatzoglou (2013) for more details on the development of this dataset.

In addition to weather variation, firms use Red Flag Warnings from the National Weather Service to make PSPS decisions. Over 98% of shutoffs occur during a Red Flag Warning, a period when the National Weather Service has identified weather conditions that could sustain catastrophic fires. I include an indicator of whether a Red Flag Warning was in effect in any part of a circuit by merging a historical archive of Red Flag Warning shapefiles.[1]

Summary statistics of each weather variable are given in Table 4.2.2. Relative to the full sample, Red Flag Warnings are drier, hotter, and more elevated, and PSPS events occur in windier and drier conditions and in higher locations.

4.3 Model

My empirical model comes from firms' descriptions of their shutoff decision process. Per their filings to the CPUC, firms initiate a PSPS if the expected degree of damages (that is, the product of expected damages conditional on ignition and the probability of ignition) exceeds the cost of failing to provide power.[2] I am interested in the degree that health risk factors and SES are correlated with PSPS decisions, and whether this is due to the channels of expected damages from wildfire, probability of ignition, or the firm's cost of declaring a PSPS event. I am unable to directly estimate the degree that these vulnerability indices are associated with expected damages, but I can infer this parameter from my model and estimates of the two other channels.

The shutoff decision is a binary choice model, where the firm weighs the expected damages from a wildfire ("Wildfire Risk") against the firm's cost of failing to provide power ("PSPS Risk"). Per filings to the state regulator, firms use separate prediction problems for probabil-

[1] From `https://mesonet.agron.iastate.edu/info/datasets/vtec.html`, accessed 9 December 2021.

[2] Legislation requires that firms making shutoff decisions must quantify benefits and risks of de-energization events, and document "how the power disruptions to customers, residents, and the general public is weighed against the benefits of a proactive de-energization."

ity of an ignition and size of fire conditional on ignition (PG&E, 2021; SCE, 2021; SDG&E, 2021a). This approach is common in both classical statistical (Xi et al., 2019) and machine learning (Jain et al., 2020) approaches to predicting wildfire size. This implies that fire size and ignition probability are conditionally independent, given an appropriate set of controls. Conditional independence allows environmental factors and income to influence both fire damages and ignition probability, but assumes that other shocks to fire ignition are unrelated to fire damages. This allows firms to have private information of shocks that influence either ignition probability or fire size (e.g. having a line crew detect a fallen tree along a power line), as long as those shocks provide no additional information on the other outcome variable conditional on our control variables.

I therefore write the firm's problem as:

$$\text{PSPS}_i = \mathbf{1}\left\{\text{Prob.}(\text{ignition}_i|X_i, Z_i)\mathbb{E}[C(\text{damages}_i)|X_i, Z_i] \geqslant \mathbb{E}[\{\text{PSPS Risk}\}_i]\right\} \tag{4.1}$$

where X_i is the set of vulnerability indices of interest, and Z_i are additional controls including weather variation, elevation, and population.

I use a generic function C for the expected degree of damages because firms' profit function is likely not linear in damages. If firms face nonlinearly increasing consequences from large fires (e.g. bankruptcy, as PG&E experienced after the 2019 fire season) or are risk averse.[3]

Let $\pi(X_i, Z_i) := \mathbb{E}[\mathbf{1}\{\text{ignition}_i\}|X_i, Z_i]$ be the conditional probability of ignition, $\phi(X_i, Z_i) := \mathbb{E}[U(\text{damages}_i)|X_i, Z_i]$ be the firm's expected cost from damages conditional on ignition, and $\mathbf{V}(X_i, Z_i) := \mathbb{E}[\{\text{PSPS Risk}\}_i|X_i, Z_i]$ be the firm's expected cost from declaring PSPS. The firm faces uncertainty over the cost of PSPS because this depends on the duration of an outage, which is determined by how long weather factors remain in effect. Replacing each term by its expectation and taking logs, I can rewrite Equation (4.1) as a sum of these expectations plus an expectational error term:

$$\text{PSPS}_i = \mathbf{1}\left\{\log(\pi(X_i, Z_i)) + \log(\phi(X_i, Z_i)) - \log(\mathbf{V}(X_i, Z_i)) + \epsilon_i \geqslant 0\right\} \tag{4.2}$$

[3]PG&E explicitly includes nonlinear risk weighting in their decision function; see PG&E (2021), section 4.2.a, for a description, and justifies this behavior as risk aversion in PG&E (2020).

where ϵ_i is an expectational error term.

To summarize how factors X_i influence decisions, I introduce a partially linear approximation to each function $\pi, \phi, \mathbf{V}$.

$$\log \pi(X_i, Z_i) = \sigma_1 X_i + \tilde{\pi}(Z_i); \quad \log \phi(X_i, Z_i) = \sigma_2 X_i + \tilde{\phi}(Z_i); \quad \log \mathbf{V}(X_i, Z_i) = \sigma_3 X_i + \tilde{\mathbf{V}}(Z_i) \tag{4.3}$$

Let γ be the overall nuisance function: $\gamma(Z_i) := \log(\tilde{\pi}(Z_i)) + \log(\tilde{\phi}(Z_i)) - \log(\tilde{\mathbf{V}}(Z_i))$. Then, I can estimate the following equation:

$$\text{PSPS}_i = \mathbf{1}\left\{(\sigma_1 + \sigma_2 - \sigma_3)X_i + \gamma(Z_i) + \epsilon_i \geqslant 0\right\} \tag{4.4}$$

I make the standard assumption that ϵ_i is a type-I extreme value random variable, and estimate this decision as a logistic model. The coefficient on X_i from this logistic regression is the overall degree that vulnerability indices influence PSPS decisions, $(\sigma_1 + \sigma_2 - \sigma_3)$. I assume that γ is a linear function of log population and weather variables, although it is also possible to use a more flexible approach. By taking hypothesis tests on whether this overall parameter is different from zero, I evaluate the research question of whether circuits with different vulnerability indices experience different rates of shutoffs.

I also wish to find which parts of the firm's decision problem explain any differences in the rates of shutoffs between circuits with different vulnerability indices. In terms of the model, I wish to test whether each parameter $\sigma_1, \sigma_2,$ and σ_3 is significantly different from zero. I am able to estimate σ_2 (the contribution to the overall coefficient from ignition probabilities) and σ_3 (the contribution to the overall coefficient from PSPS cost) through separate regression problems, but am not able to estimate σ_1 directly.

The value of σ_1 is implied given estimates of σ_2, σ_3, and $(\sigma_1 + \sigma_2 - \sigma_3)$. I cannot estimate σ_1 directly, as I am not able to estimate the damages from a fire or the firms' cost functions based on those damages. Damages from a wildfire are a function of wildfire size and the features of land damaged by the wildfire. Predicting fire size is a notoriously challenging problem, even given modern machine learning techniques (Taylor et al., 2013; Xi et al., 2019; Jain

et al., 2020). Firms use proprietary software to make these fire size predictions, and achieve relatively high levels of accuracy. In Section 4.A, I document an attempt to predict fire size using linear regression and random forest regressions. I am unable to provide informative bounds on the degree of fire size, and am therefore unable to estimate the expected damages or the firm's expected damages (after applying the cost function).

To estimate σ_2, I use a two-stage procedure with data of ignitions along utility lines. To estimate σ_3, I use a regression of proxies to the firm's cost, according to their filings. I discuss these estimation problems in subsections below.

4.3.1 Ignition Probability

I estimate σ_2 (the contribution to the overall coefficient from ignition probabilities) by first finding the probability of ignition, and then finding the coefficient on vulnerability indices X_i from a partially linear model to those predicted probabilities. To estimate the probability of ignition, I use logistic regression of ignitions along power lines with control variables $\{X_i, Z_i\}$. I adjust the standard errors from the second stage estimates, as the second stage estimates depend on the results from the first stage.

To construct the probability, I model ignition probability as a binomial logit model. I assume that there exists some latent model of fire ignitions, for some function g of vulnerability indices and environmental factors and a type-I extreme value distributed error term ε_i:

$$\text{ignition}_i = \mathbf{1}\{g(X_i, Z_i) + \varepsilon_i > 0\} \tag{4.5}$$

With this model, I can calculate the ignition probability $\hat{\pi}(X_i, Z_i)$ given an estimate of g:

$$\hat{\pi}(X_i, Z_i) = \frac{\exp \hat{g}(X_i, Z_i)}{1 + \exp \hat{g}(X_i, Z_i)} \tag{4.6}$$

I estimate g using a subset of data from years where firms do not use PSPS. Table 4.2.1 shows the years with PSPS observations. I use data from all firms in 2015 and 2016 and from a subset of firms in 2014, 2017, 2018, and 2021. When firms use PSPS, data is censored: the

researcher does not observe whether an ignition would have occurred without PSPS. With endogenously censored outcome variables, it is generally only possible to partially identify regression functions (Khan and Tamer, 2009). Subsets from years when firms do not use PSPS are not subject to this censoring concern. I assume that the relationship between weather variation and ignition probability is consistent between years when utilities do and do not use PSPS; this could be violated if dry vegetation accumulates and fire risk increases over time, or utilities choose to use other wildfire management strategies in years without PSPS. In Appendix 4.C, I conduct a robustness exercise and estimate ignition probabilities using the full sample. While I do not fully characterize the partially identified set, these exercises support my main findings.

To estimate σ_2, I regress the log probability computed from the first stage on X_i and Z_i. Per Equation (4.3), I assume that the log probability is additively separable in X_i and Z_i. The parameter σ_2 can then be estimated via partially linear regression.

$$\log \hat{\pi}(X_i, Z_i) = \sigma_2 X_i + \tilde{\pi}(Z_i) + \varepsilon_i \tag{4.7}$$

As this second stage depends on the first-stage estimation, I adjust standard errors for σ_2 to incorporate uncertainty in my estimate of g. I do so by finding the influence function of the first-stage problem and incorporating this into standard errors of the second-stage problem, as in Newey and Daniel McFadden (1994). Section 4.B shows the derivation.

I implement these stages to estimate σ_2. I use linear models for the functions g and $\tilde{\pi}$. The procedure is as follows:

1. Estimate $\hat{g}$ via logistic regression of ignitions on X_i and Z_i. With a linear model, estimating $\hat{g}$ means finding a parameter vector $\beta = \{\beta_0, \beta_1, \beta_2\}$ that maximizes the likelihood of the model $\text{ignition}_i = \mathbf{1}\{\beta_0 + \beta_1 X_i + \beta_2 Z_i + \varepsilon_i > 0\}$.

2. Find $\hat{\sigma}_2$ using Equation (4.7) and the results from the first stage. For the log probabilities, use the estimated first-stage parameter vector to compute the probabilities $\hat{\pi}(X_i, Z_i)$. Then find σ_2 via linear regression of these log probabilities on X_i and Z_i. To find standard errors for σ_2, I adjust the standard errors from linear regression to

incorporate error in estimating $\hat{g}$.

4.3.2 Cost of PSPS

I estimate σ_3 (the contribution to the overall coefficient from PSPS cost) through regression of proxies to the firm's cost of PSPS, based on their filings to regulators. Due to ambiguity between various documents, I use both the number of customers impacted and the customer minutes interrupted (CMI) as proxies to the firm's cost of PSPS. These proxies capture major sources of variance in the firm's expected costs of PSPS. While neither is a perfect approximation, they provide a reasonable estimate of how vulnerability indices influence the firm's cost of PSPS.

In firms' Wildfire Mitigation Plans, they state formulas to calculate cost of a PSPS that depend on customer minutes interrupted (CMI) and the total number of customers interrupted (SCE 2021, p. 61; PG&E 2021, p. 52; SDG&E 2021a, p. 26). Firms also incorporate the safety cost and financial cost of PSPS, as well as a reliability score. This safety cost is calculated as a constant factor multiplied by CMI, and the financial cost scales with the cost of shutoff (SCE, 2022; PG&E, 2020; SDG&E, 2021b). PG&E incorporates a scaling function if the safety, reliability, or financial costs of PSPS in a circuit exceed 10% of the largest recorded wildfire damages; I assume that the damage at any circuit never exceeds this threshold. SDG&E plans to incorporate the health sensitivity of subpopulations, but I do not observe decisions made with these rules (SDG&E, 2021a, p. 30). In 2021, SCE began weighting some components of its cost function by the number of vulnerable customers per line; I do not have access to their conversion formula and do not attempt to model this improvement. While the conversion factors are not published, this is absorbed into the constant if I take a regression of log CMI or log number of customers.

SCE is the only firm to specify how they form ex-ante predictions of the CMI. In their post-event reports, SCE calculates their CMI as a constant number of minutes multiplied by number of customers impacted, effectively making the cost of a shutoff a function of function solely of the number of customers (Valdberg, Tozer, and Kilberg, 2021, p. 16). No other firms publish their ex-ante PSPS cost calculations. I assume that they either use a constant

	(1)	(2)	(3)	(4)	(5)	(6)
Health Risk Index	53.28	50.01	50.16	48.21	39.65	39.19
	(1.610)	(1.160)	(1.010)	(0.439)	(1.130)	(0.768)
SES Index	54.18	57.34	49.01	48.65	50.35	51.91
	(1.149)	(0.712)	(0.758)	(0.349)	(2.018)	(1.193)
Observations	223	469	451	1885	81	181
Utility	SCE	SCE	PGE	PGE	SDGE	SDGE
⩾24 Hours		X		X		X

mean coefficients; se in parentheses

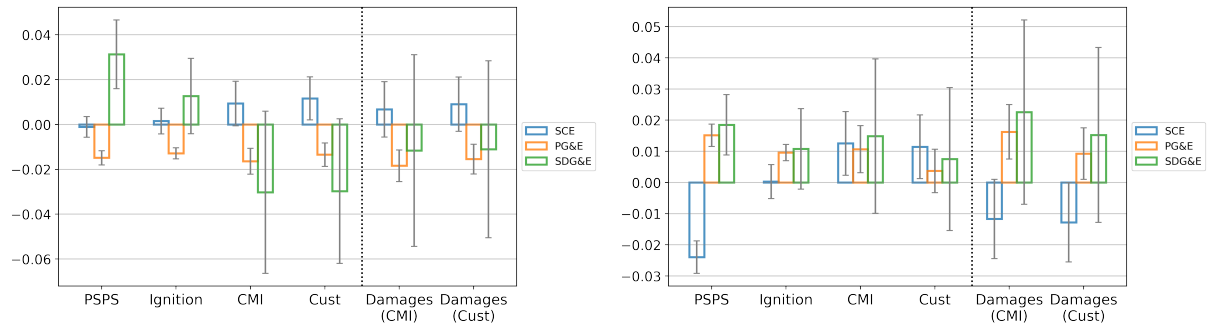
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.3.1: Summary of vulnerability indices (SES index and health risk index), by whether the observed outage exceeds 24 hours. Standard error of the mean is in parentheses.

factor, or the expected CMI per outage based on the empirical duration of PSPS outages.

With these proxies selected, I estimate σ_3 using linear regression on the log of the proxy. The number of customers impacted and CMI for a given outage are stochastic; depending on weather conditions, firms may be able to de-energize a smaller section of the circuit or be forced to prolong the outage. I estimate this relationship using outage duration and number of customers impacted during each reported outage, with vulnerability indices, weather controls, and log population as additional controls. I assume that for circuits with zero reported shutoffs, there is the same relationship between average vulnerability indices and CMI.

If firms use a constant outage duration to estimate costs, this approximation may systematically undervalue the cost to low SES or high health risk communities. To inspect this, I compare the average health risk and SES indices for circuits with PSPS outages above and below 24 hours. Table 4.3.1 shows these summary statistics. For SCE, outages over 24 hours occur in circuits with significantly higher SES index (indicating lower-SES circuits), and significantly lower health risk index. This shows that SCE's stated decision systematically undervalues the cost of an outage to low-SES populations.



(a) Health Risk Index

(b) SES Index

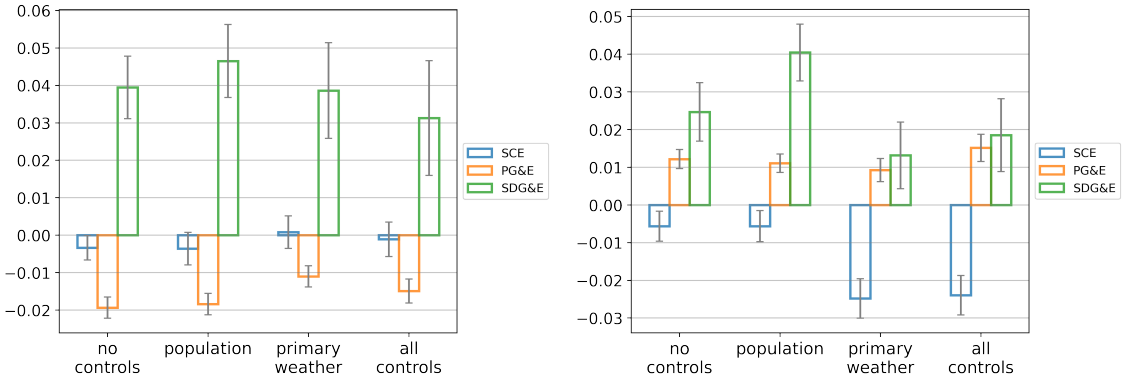
Figure 4.4.1: Collected coefficient estimates of health risk or SES index on various outcomes, using all weather and population controls. Each group of plots on the x axis show coefficient estimates for a given outcome variable. Damages (to the right of dotted line) are implied from estimates to the left of dotted line and my model. Error bars show ± 1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.

4.4 Results

As each firm has unique decision rules, I report separate coefficients for each firm in all trials. All regressions include utility-by-year fixed effects, and all population or weather control variables are interacted with these fixed effects. In each subsection, I use four specifications: no controls (beyond fixed effects), only population as a control variable, primary weather variables plus population, and all weather variables plus population.

The dependent variables of interest are the health risk index and socioeconomic factor index from CalEnviroScreen. Recall that in these indices, 0 is the least vulnerable and 100 is the most vulnerable. Increasing the socioeconomic or health risk index by 1 is equivalent to an average increase of 1 percentage point across the ranks of the sub-indices. A positive coefficient indicates that the event in a given logistic regression is more likely, or the expected outcome in a linear regression is larger, when the population on the circuit has higher average health risk or lower average SES. I refer to circuits where the population has a lower (higher) average health risk index as lower (higher) health risk circuits, and circuits where the population has a lower (higher) average SES index as lower (higher) SES circuits.

Figure 4.4.1 summarizes the overall coefficient estimates, for regressions including population and all weather controls. This figure shows that, after controlling for my set of covariates, there is a significant association between PSPS decisions and my vulnerability characteristics,



(a) Health Risk Index

(b) SES Index

Figure 4.4.2: Coefficient estimates of vulnerability indices for logistic regression of PSPS events. Groups on the x axis collect results from regression with a given set of controls. Error bars show ± 1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.

in some utilities. The estimates of CMI and number of customers impacted (and the damage estimates that depend on these values) are especially noisy and I am unable to reject that these metrics are equitably distributed. Additional summary figures are shown in Section 4.D using different sets of weather variation.

The following subsections describe the results from each separate regression.

4.4.1 PSPS Decisions

To study PSPS decisions, I use a subset of data during red flag warnings, from October 2013 (the month of the first PSPS event) onward. I limit the scope to red flag events to recreate the firm's problem, as over 90% of PSPS events are declared during a red flag warning. Coefficient estimates from the logistic regression are reported in Table 4.4.1, and visualized in Figure 4.4.2.

A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of PSPS shutoffs. These coefficients represent the amount that log-odds change with an increase of 1 unit of the index. They can be approximately interpreted as the percentage change in PSPS likelihood given an increase in 1 unit of the index, as the coefficients are fairly close to 0. For example, an estimated coefficient of 0.01 indicates that PSPS events are 1% more likely in circuits with 1 higher index.

	(1) psps	(2) psps	(3) psps	(4) psps
SCE x Health	-0.00335*	-0.00355	0.000820	-0.00109
	(0.00167)	(0.00222)	(0.00222)	(0.00236)
SCE x SES	-0.00565**	-0.00562**	-0.0248***	-0.0240***
	(0.00202)	(0.00211)	(0.00268)	(0.00267)
PG&E x Health	-0.0193***	-0.0184***	-0.0110***	-0.0149***
	(0.00145)	(0.00147)	(0.00143)	(0.00163)
PG&E x SES	0.0122***	0.0111***	0.00926***	0.0151***
	(0.00129)	(0.00124)	(0.00157)	(0.00183)
SDG&E x Health	0.0394***	0.0465***	0.0386***	0.0313***
	(0.00425)	(0.00498)	(0.00652)	(0.00781)
SDG&E x SES	0.0247***	0.0404***	0.0131**	0.0185***
	(0.00397)	(0.00385)	(0.00451)	(0.00494)
Observations	375064	375064	370263	370263
Pseudo R^2	0.081	0.090	0.297	0.378
Population		X	X	X
Primary			X	X
Derived				X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4.1: Results from logistic regression of PSPS events. Perfectly predicted failures are omitted. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of PSPS shutoffs.

Without controlling for weather factors (column 1), higher health risk circuits are significantly less likely to have a PSPS in both SCE and PG&E, and more likely in SDG&E. This finding is significant at the $p = 0.001$ level for PG&E and SDG&E, and at the $p = 0.05$ level for SCE. Lower SES circuits are more likely to have a shutoff in PG&E and SDG&E, and less likely in SCE; this finding is significant at the $p = 0.001$ level for PG&E and SDG&E, and at the $p = 0.01$ level for SCE. These magnitudes are on the order of 0.01, so a 1 point increase in the index corresponds to roughly one percent difference in the likelihood of PSPS.

After controlling for weather variation (columns 2-3), model fit improves but these patterns remain largely consistent. D. McFadden (1973) suggests that a pseudo-R squared of 0.2-0.4 suggests good model fit for logistic regression, indicating that our model acceptably fits the PSPS decisions after controlling for weather variation. The exception is that the coefficient on the health risk index for SCE is no longer significant, but the coefficient on socioeconomic

factors for SCE is larger in magnitude and is statistically significant at the $p = 0.001$ level.

I do not observe the full set of relevant variation that firms have while making these decisions. In linear models, Oster (2019) gives an approach to quantify the degree of omitted variable bias by comparing the stability of coefficients as the model fit improves. I am not aware of an analogous approach for logistic regression. Informally, the sign and magnitude of coefficients remain relatively stable as the model fit improves from the null model to the model including population and all weather controls, indicating that these conclusions may be robust to incorporating additional variables.

4.4.2 Ignition Probability

I measure how vulnerability indices influence ignition probability via a two-stage procedure, as described in Section 4.3.1. In the first stage, I use logistic regression of ignitions versus vulnerability indices, log population, and weather covariates. In the second stage, I find the best fitting linear approximation to the log probability given the first-stage results. Standard errors from the second stage are adjusted to account for uncertainty from the first stage estimation, as described in Section 4.B.

As described in Section 4.3.1, I use data from years where utilities did not conduct PSPS. This avoids the identification concern that when a firm conducts a PSPS, I do not observe whether an ignition would have occurred without that intervention. In Section 4.C, I conduct a robustness exercise using the full set of data. I evaluate the coefficients assuming that each PSPS event would be an ignition, or that no PSPS event would be an ignition. The conclusions below still hold in both alternate specifications, suggesting that my findings hold regardless of any changes in the relationship between weather and ignition probability over time.

Table 4.4.2 shows the results from the first stage, logistic regression of ignitions on the vulnerability indices and additional controls. The pseudo-R squared value is relatively low, even for the model with both primary and derived weather covariates. Many of the coefficient estimates are statistically indistinguishable from 0. In PG&E circuits, higher health risk circuits are significantly (at $p = 0.001$ level) less likely to have an ignition and lower SES

	(1) ignition	(2) ignition	(3) ignition	(4) ignition	(5) ignition	(6) ignition	(7) ignition	(8) ignition	(9) ignition
Health Risk Index	-0.00314	-0.0113***	0.0213**	-0.00178	-0.0111***	0.0162*	0.00652*	-0.0121***	0.0113
	(0.00209)	(0.00114)	(0.00665)	(0.00264)	(0.00115)	(0.00723)	(0.00305)	(0.00127)	(0.00859)
SES Index	0.00308	0.0148***	0.0174**	0.00244	0.0146***	0.0196**	-0.00404	0.00916***	0.00955
	(0.00258)	(0.00110)	(0.00664)	(0.00262)	(0.00110)	(0.00625)	(0.00278)	(0.00133)	(0.00666)
Observations	1950168	4939641	700344	1950168	4939641	700344	1950168	4939641	700344
Pseudo R^2	0.006	0.007	0.013	0.007	0.007	0.017	0.047	0.060	0.080
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4.2: Results from first stage of ignition estimation, logistic regression of PSPS events. Health risk and SES are both measured as indices, with 0 being least vulnerable and 100 being most vulnerable. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of ignitions. Robust standard errors are reported.

circuits are more likely to have an ignition. This finding is robust to including population and weather variables. At the $p = 0.01$ level, ignitions in SDG&E lines are positively correlated with higher vulnerability indices, although these relationships are not significant after controlling for weather factors.

Table 4.4.3 shows the results from the second stage, fitting a linear model to the log predicted probabilities from the first stage. These coefficients show the degree that vulnerability indices are associated with the log probability of ignition. I suppress the R^2 value from the second stage procedure, as this statistic does not incorporate the uncertainty from the first stage and may be misleading. Magnitudes and significance of second stage estimates are generally quite similar to the first stage results.

Some patterns from the coefficient estimates in Table 4.4.3 are similar to those of the PSPS decisions, although less precisely estimated. Without controlling for weather variation, I find that lower SES circuits have higher rates of ignition in PG&E and SDG&E, and lower rates of ignition in SCE. I find that higher health risk circuits have higher rates of ignition in SDG&E, and lower rates in PG&E. Controlling for population and weather variation, the only significant associations that remain are that lower SES circuits in PG&E have higher rates of ignition and that higher health risk circuits in PG&E have lower rates of ignition. This is similar to the findings from Table 4.4.1, although there is greater uncertainty. This suggests that population and weather variation are able to explain much of the observed

	(1)	(2)	(3)	(4)
SCE x Health	-0.00015	-0.00178	-0.00141	0.00151
	(0.00085)	(0.00264)	(0.00284)	(0.00291)
SCE x SES	-0.00299**	0.00244	0.00074	0.00028
	(0.00094)	(0.00262)	(0.00278)	(0.00279)
PG&E x Health	-0.01231***	-0.01110***	-0.01294***	-0.01290***
	(0.00045)	(0.00115)	(0.00125)	(0.00126)
PG&E x SES	0.01340***	0.01463***	0.01025***	0.00959***
	(0.00046)	(0.00110)	(0.00130)	(0.00133)
SDG&E x Health	0.03221***	0.01620*	0.01388	0.01264
	(0.00182)	(0.00723)	(0.00840)	(0.00857)
SDG&E x SES	0.01972***	0.01955**	0.01409*	0.01079
	(0.00181)	(0.00625)	(0.00624)	(0.00659)
Observations	14925285	7590153	7590153	7590153
Popuation		X	X	X
Primary			X	X
Derived				X

Table 4.4.3: Results from second stage of ignition probability regression. A positive coefficient indicates that higher health risk (lower SES) circuits have a greater rate of ignitions. Standard errors are computed using Appendix 4.B. Stars indicate significance at the $p = 0.05(*), 0.01(**),$ and $0.001(***)$ levels.

differences in ignitions between more and less vulnerable communities.

4.4.3 PSPS Costs

I use two proxies to find how vulnerability indices correlate with the utility's computed cost of PSPS. This cost is the value the utility uses when weighing the costs and benefits of a shutoff; it reflects the estimated size of the disruption from declaring a PSPS event. As discussed in Section 4.3.2, I use customer minutes interrupted (CMI) and number of customers impacted. Results using log of CMI are shown in columns 1-4 of Table 4.4.4, and log of the number of customers impacted are in columns 5-8.

The patterns are generally similar between regressions, although both are relatively noisy. Estimates with log CMI are generally larger in magnitude and more precisely estimated than those using log number of customers. Without controlling for weather variation, there

	(1) Log CMI	(2) Log CMI	(3) Log CMI	(4) Log CMI	(5) Log Cust	(6) Log Cust	(7) Log Cust	(8) Log Cust
SCE x Health	-0.00808*	-0.00780*	0.00871	0.00933	-0.00633*	-0.00606	0.0106*	0.0116*
	(0.00327)	(0.00384)	(0.00489)	(0.00506)	(0.00309)	(0.00364)	(0.00479)	(0.00491)
SCE x SES	0.0303***	0.0298***	0.0148**	0.0125*	0.0253***	0.0246***	0.0121*	0.0115*
	(0.00497)	(0.00502)	(0.00512)	(0.00521)	(0.00486)	(0.00493)	(0.00507)	(0.00521)
PG&E x Health	-0.0208***	-0.0207***	-0.0157***	-0.0164***	-0.0167***	-0.0170***	-0.0140***	-0.0135***
	(0.00275)	(0.00280)	(0.00286)	(0.00294)	(0.00255)	(0.00258)	(0.00261)	(0.00267)
PG&E x SES	0.0119**	0.0118**	0.00779*	0.0107**	0.00731*	0.00783*	0.00297	0.00372
	(0.00363)	(0.00369)	(0.00385)	(0.00384)	(0.00335)	(0.00340)	(0.00355)	(0.00355)
SDG&E x Health	-0.0108	-0.0315**	-0.0389*	-0.0303	-0.00472	-0.0274**	-0.0342*	-0.0297
	(0.00803)	(0.0104)	(0.0167)	(0.0185)	(0.00654)	(0.00856)	(0.0140)	(0.0164)
SDG&E x SES	0.00697	-0.00131	0.0132	0.0149	0.00375	-0.00582	0.00683	0.00754
	(0.00545)	(0.00620)	(0.0110)	(0.0126)	(0.00476)	(0.00531)	(0.00955)	(0.0117)
Observations	3270	3270	3270	3270	3270	3270	3270	3270
R^2	0.172	0.176	0.309	0.351	0.106	0.112	0.237	0.268
Population		X	X	X		X	X	X
Primary			X	X			X	X
Derived				X				X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.4.4: Results from linear regression of proxies to PSPS cost, CMI (columns 1-4) and number of customers (columns 5-8). A positive coefficient indicates that higher health risk (lower SES) circuits have higher average CMI or number of customers impacted. Robust standard errors are reported. Outcome variable values of 0 are omitted.

is a significant positive correlation between low SES and the cost proxy for SCE (p value < 0.001) and PG&E (p value 0.001 for CMI, 0.029 for customers). There is a significant negative correlation for the health risk index for PG&E (p value < 0.001). After controlling for weather variation, these observations are largely similar. The difference between estimates with no controls and all controls are not significant at the $p = 0.05$ except the "SCE x SES" row, where the association is no longer significant after including all controls.

Rules that determine the cost of PSPS may disadvantage low-SES or high-health risk populations if the number of customers is negatively correlated with these indices. This occurs regardless of whether the rules intend to discriminate based on these characteristics; that is, it is an example of statistical rather than taste-based discrimination (Guryan and Charles, 2013). The utility's decision rule places more weight on circuits with higher historical customer outages. If circuits with a higher share of vulnerable individuals are less impacted by shutoffs, the utility's rule calculates a lower cost from shutoffs in those circuits. These findings indicate that PG&E's decision rules may disadvantage high health risk circuits.

Utilities can adjust their rules to avoid this potential discrimination, and some already are. SCE already scales part of their PSPS risk score by the size of populations with medical needs (Valdberg, Tozer, and Kilberg, 2021, p. 16), and SDG&E has plans to implement a similar program (SDG&E, 2021a, p. 30).

4.4.4 Implied Coefficient of Expected Damages

While I am not able to estimate the firm's expected damages from a wildfire, the results of the other estimation steps and the firm's decision-making rule allow me to infer how vulnerability indices influence the firm's expected damages from a wildfire. In Equation (4.2), I write a model of firms' decision making involving three parameters: σ_1 (the extent that vulnerability indices influence the firms' expected damages), σ_2 (the extent that vulnerability indices influence log probability of ignition), and σ_3 (the extent that vulnerability indices influence log cost of declaring PSPS). I am not able to estimate σ_1 directly, but estimate $\sigma_1 + \sigma_2 - \sigma_3$ in Section 4.4.1. With separate estimates of σ_2 and σ_3, I can infer a value of σ_1 as $\hat{\sigma}_1 = \{\sigma_1 + \sigma_2 - \sigma_3\} - \hat{\sigma}_2 + \hat{\sigma_3}$.

Table 4.4.5 shows the implied estimate of σ_1 given this model and my estimates of σ_2 and σ_3 from previous sections. I find standard errors by assuming that these problems are uncorrelated; I take the square root of the sum of squared standard deviations from other estimates. Overall, these estimates are quite noisy. I fail to reject that expected damages from fires are equitably estimated by utilities.

These implied values assume that utilities have no control over the duration of an outage or the number of customers impacted. This is supported by regulatory filings, which state that the size and duration of an outage are as small as permitted by weather and infrastructure conditions. This may not be accurate in practice. For example, I conclude that PG&E's expected damage is smaller for populations with higher health risk because there fewer customers are impacted in circuits with higher health risk. This observation would also be consistent with a model where firms face different costs for different populations, and set the length of an outage or the number of customers impacted to prioritize populations with higher health risk. I do not have the data necessary to falsify this model, although it would

	from CMI	from CMI	from CMI	from CMI	from Cust	from Cust	from Cust	from Cust
SCE x Health	-0.0113**	-0.00957	0.0109	0.00673	-0.00953**	-0.00784	0.0129*	0.00902
	(0.00377)	(0.00516)	(0.00607)	(0.0063)	(0.00361)	(0.00501)	(0.00599)	(0.00617)
SCE x SES	0.0277***	0.0217***	-0.0107	-0.0117	0.0226***	0.0166**	-0.0134*	-0.0128*
	(0.00544)	(0.00604)	(0.00641)	(0.00649)	(0.00535)	(0.00596)	(0.00637)	(0.00648)
PG&E x Health	-0.0278***	-0.0279***	-0.0137***	-0.0184***	-0.0237***	-0.0243***	-0.0121***	-0.0154***
	(0.00315)	(0.00336)	(0.00343)	(0.00359)	(0.00297)	(0.00318)	(0.00323)	(0.00338)
PG&E x SES	0.0106**	0.00824*	0.00679	0.0162***	0.00606	0.00428	0.00197	0.00927*
	(0.00388)	(0.00405)	(0.00436)	(0.00446)	(0.00362)	(0.00379)	(0.00409)	(0.00421)
SDG&E x Health	-0.0036	-0.00118	-0.0142	-0.0116	0.00252	0.00295	-0.00949	-0.0111
	(0.00927)	(0.0136)	(0.0198)	(0.0218)	(0.00801)	(0.0123)	(0.0176)	(0.0201)
SDG&E x SES	0.0119	0.0196*	0.0122	0.0226	0.00869	0.0151	0.00589	0.0153
	(0.00698)	(0.00961)	(0.0134)	(0.0151)	(0.00645)	(0.00906)	(0.0123)	(0.0143)
Population		X	X	X		X	X	X
Primary			X	X			X	X
Derived				X				X

Table 4.4.5: Inferred value of σ_1, or the coefficient of vulnerability indices on log of expected utility loss from a fire. Computed using coefficient estimates from previous regressions. Standard errors are the square root of sum of squares from previous regressions. Stars indicate significance at the $p = 0.05(*), 0.01(**),$ and $0.001(***)$ levels.

be of interest to distinguish between these potential models of firm conduct.

4.5 Conclusions

I find that PSPS is used more frequently in low-SES circuits among two of California's major utilities, and among higher health risk circuits in one of the major utilities. This finding is robust to controlling for weather variation. After controlling for weather variation, I find that ignitions are more frequent in low-SES circuits and in lower health risk circuits in PG&E, but otherwise do not find significant evidence. With my proxies for the firm's cost of declaring a PSPS event, I am unable to make strong claims about the degree that these vulnerability indices influence the utilities' estimated costs from declaring a PSPS shutoff or their expected damages from wildfires. I cannot reject that these costs are equitably distributed.

This work starts to explore a gap in the literature on empirically assessing the equity of adaptation mechanisms. More research is needed in this area more broadly, as well as to better understand the impacts of electric utilities' response to wildfire risk. This research agenda is challenging without better data about the firm's problem, particularly how the firm computes costs and benefits of PSPS. These data would allow researchers to explore a broader range of research questions, such as the explorations of systematic bias in Obermeyer et al. (2019) or Rambachan (2021).

Appendix to Chapter 4

4.A Predicting Fire Size

I use public data to attempt to predict fire size given weather covariates. From 1992-2018, comprehensive records of fire size are available from the US Forest Service (K C Short, 2014; Karen C Short, 2021). From 2019-2021, I include records from the National Interagency Fire Service.[4] Records include the date, fire size, and latitude and longitude of ignition. The final database includes 240,239 records within California. I then merge these data with my weather observations from GridMET.

I model the problem of predicting catastrophic fires both as a regression and classification problem. To predict fire size, I regress the log of fire size against the full set of weather variables from GridMET, as well as yearly fixed effects and fixed effect terms per utility's service area. For classification trials, I use three definitions of "large fire": top 0.02 quantile (my definition), larger than 300 acres[5], and larger than 500 acres.[6]. In each classification trial, I weight each observation by the inverse frequency of its class to predict the relatively rare event of a large fire. I consider a linear set of weather variables, linear regression with interactions between weather variables, and random forests with 5-folds cross validation.

Table 4.A.1 summarizes the results of these trials. For classification trials, I report the specificity (share of negative outcomes that are correctly predicted) and sensitivity (share

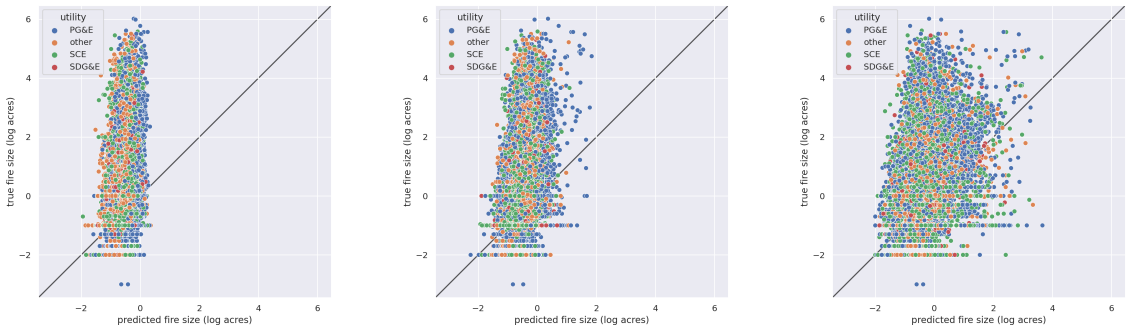
[4]From `https://data-nifc.opendata.arcgis.com/datasets/nifc::wfigs-current-wildland-fire-perimeters/about`, accessed 11 January 2022.

[5]Definition from `https://www.nps.gov/olym/learn/management/upload/fire-wildfire-definitions-2.pdf`, accessed 1 April 2022.

[6]Definition from Holmes, Huggett, and Westerling (2008).

	Regression R squared	Fire size ⩾ 300 acres Specificity	Fire size ⩾ 300 acres Sensitivity	Fire size ⩾ 500 acres Specificity	Fire size ⩾ 500 acres Sensitivity	Fire size ⩾ top 2% Specificity	Fire size ⩾ top 2% Sensitivity
Linear	0.06311	0.6907	0.6455	0.7017	0.6624	0.6691	0.6115
Linear Interacted	0.08927	0.723	0.6612	0.7302	0.6893	0.6888	0.6271
Random Forest	0.08446	0.02007	0.9994	0.01766	0.9994	0.02622	0.9991

Table 4.A.1: Results from random forest and linear regression at predicting large fires.



(a) Linear (b) Linear interacted (c) Random forest

Figure 4.A.1: Predicted vs actual fire size, using various regression methods.

of positive outcomes that are correctly predicted) of each prediction method, for each "large fire" definition. For regression, I report the R^2 value. Figure 4.A.1 shows the scatter plots of predicted fire size vs. actual fire size.

Overall, these results indicate poor performance at predicting fire size. I have limited ability to extrapolate fire size, meaning I cannot construct informative bounds on the missing data as required to identify counterfactuals in Rambachan (2021).

4.B Standard Errors for Ignition Probability

In this section, I derive the asymptotic variance of the two-step estimation procedure from Section 4.3.1. I write the problem as a two-step M estimation, and apply the method from Newey and Daniel McFadden (1994) to find asymptotic variance after accounting for error in estimating g.

I rewrite Equation (4.7) as a least squares problem. Let $m(\theta, X_i, Z_i, g) := -(\log \hat{\pi}(X_i, Z_i; g) - \sigma_2 X_i - \tilde{\pi}(Z_i))^2$, emphasizing the dependence of predicted probabilities $\hat{\pi}$ on the first-step function g. The set of parameters θ includes σ_2 and the parameters defining the function $\tilde{\pi}$.

Then:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^{N} m(\theta, X_i, Z_i, g) \tag{4.8}$$

I can then express the asymptotic variance of estimating θ, as long as $\hat{g}$ is asymptotically linear. The form comes from applying the Mean Value Theorem to the differences $\sqrt{N}(\hat{\theta}-\theta_0)$ and $\sqrt{N}(\hat{g} - g_0)$.

$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$, where

$$V = H^{-1}\mathbb{E}\left[\left\{\frac{\partial m(\theta, X_i, Z_i, g)}{\partial \theta} + M\psi(X_i, Z_i)\right\}\left\{\frac{\partial m(\theta, X_i, Z_i, g)}{\partial \theta} + M\psi(X_i, Z_i)\right\}'\right] H^{-1} \tag{4.9}$$

where H is the Hessian matrix of m: $H = \mathbb{E}\left[\frac{\partial^2 m(\theta, X_i, Z_i, g)}{\partial \theta \partial \theta'}\right]$, M is the matrix of cross derivatives of m: $M = \mathbb{E}\left[\frac{\partial^2 m(\theta, X_i, Z_i, g)}{\partial \theta \partial g'}\right]$, and $\psi(X_i, Z_i)$ is the influence function of g for an observation $\{X_i, Z_i\}$.

With linear approximations to the functions θ and g, it is straightforward to derive these expressions. Let $W_i = \{X_i, Z_i\}$ be the column vector of all covariates for unit i. Then let $\beta = \theta$ so that $\sigma_2 X_i + \tilde{\pi}(Z_i) = W_i'\beta$, and let ρ be a parameter vector so that $g(X_i, Z_i) = W_i'\rho$. Recall that with the logistic assumption, $\hat{\pi}(W_i; \rho) = (1 + \exp(-W_i'\rho))^{-1}$. I write $\hat{\pi}_i := \hat{\pi}(W_i; \rho)$ Then:

$$m(\beta, W_i, \rho) = -(\log \hat{\pi}_i - W_i'\beta)'(\log \hat{\pi}_i - W_i'\beta) \tag{4.10}$$

$$\frac{\partial m(\beta, W_i, \rho)}{\partial \beta} = 2W_i(\log \hat{\pi}_i - W_i'\beta) \tag{4.11}$$

$$H = -2\mathbb{E}[W_i W_i'] \tag{4.12}$$

$$M = 2\mathbb{E}\left[W_i \frac{\exp(-W_i'\rho)}{1 + \exp(-W_i'\rho)} W_i'\right] = 2\mathbb{E}[W_i(1 - \hat{\pi}_i)W_i'] \tag{4.13}$$

Following Jann (2020), I derive the influence function for logistic regression after writing the problem as a Generalized Method of Moments estimator:

$$\psi(W_i) = \left(\mathbb{E}[W_i \hat{\pi}_i (1 - \hat{\pi}_i) W_i']\right)^{-1} W_i(y_i - \hat{\pi}_i) \tag{4.14}$$

Plugging these into Equation (4.9), I can determine the asymptotic variance of my estimator. Note that this framework allows for more general semiparametric estimation, as long as $\hat{g}$ is asymptotically linear. This allows estimators like random forests or general double machine learning techniques; see Athey, Tibshirani, and Wager (2019) for advice on deriving the influence function of a random forest and Ichimura and Newey (2022) for influence functions of general semiparametric functions.

4.C Ignitions regression with alternate sample

In the main text, I estimate ignition probability using data from years where firms do not declare PSPS, to avoid a potential data censoring problem. As discussed in Section 4.3.1, this choice if the relationship between weather variables and ignition probability is not consistent between years when utilities do and do not use PSPS. As a robustness exercise, I include results using the full time period, with two different assumptions on the observed ignitions and shutoffs: that, absent a shutoff, each circuit with a PSPS event would either have an ignition (Table 4.C.1 and Table 4.C.2) or no ignition (Table 4.C.3 and Table 4.C.4). Overall, these results suggest that my conclusions in the main analysis are robust to including ignition data from the full sample.

This provides suggestive evidence about the partially identified set that contains the true parameter. To fully characterize that set, I could enumerate all possible potential realizations of the missing data and repeat the estimation procedure for each potential outcome. Due to the immense computational cost of such a procedure, I only repeat the two-stage estimation procedure for these two scenarios.

Table 4.C.1 and Table 4.C.2 show the results from the first- and second-stage estimation (respectively), using the full sample and treating each missing value as a true positive. The findings that are statistically significant (at the $p = 0.001$ level) from Table 4.4.3 in the main analysis are also significant in these regressions: low-SES circuits in PG&E have higher rates of ignition, as do lower health risk circuits. This regression finds additional significant associations, although many of these are not robust to an alternate assumption on the missing data.

Table 4.C.3 and Table 4.C.4 show the results from the first- and second-stage estimation (respectively), using the full sample and treating each missing value as a true negative. Again, the statistically significant findings from the main analysis are confirmed in these regressions. Many of the additional significant associations from treating each missing value as a true positive are not significant in this exercise, although both find that lines in low-SES circuits in SDG&E are significantly more likely to have an ignition.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	ignition	ignition	ignition	ignition	ignition	ignition	ignition	ignition	ignition
Health Risk Index	-0.000153	-0.0123***	0.0322***	0.00128	-0.0117***	0.0243***	0.00238	-0.0115***	0.0189***
	(0.00113)	(0.000615)	(0.00272)	(0.00141)	(0.000617)	(0.00303)	(0.00151)	(0.000691)	(0.00374)
SES Index	-0.00299*	0.0134***	0.0197***	-0.00359**	0.0129***	0.0253***	-0.0148***	0.00918***	0.0100***
	(0.00134)	(0.000583)	(0.00292)	(0.00138)	(0.000575)	(0.00255)	(0.00163)	(0.000746)	(0.00290)
Observations	3120696	9879282	1925307	3120696	9879282	1925307	3120696	9879282	1895814
Pseudo R^2	0.034	0.029	0.045	0.035	0.030	0.069	0.198	0.168	0.334
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.C.1: First stage ignition results, using the full sample, with PSPS results counted as ignitions. This is the assumption that all censored results would have been true positives.

	(1)	(2)	(3)	(4)
SCE x Health	-0.00015	0.00128	-0.00204	-0.00224
	(0.00113)	(0.00141)	(0.00131)	(0.00143)
SCE x SES	-0.00299*	-0.00359**	-0.01008***	-0.01056***
	(0.00134)	(0.00138)	(0.00161)	(0.00162)
PG&E x Health	-0.01231***	-0.01170***	-0.01136***	-0.01241***
	(0.00062)	(0.00062)	(0.00067)	(0.00069)
PG&E x SES	0.01340***	0.01285***	0.00807***	0.00965***
	(0.00058)	(0.00057)	(0.00071)	(0.00074)
SDG&E x Health	0.03221***	0.02432***	0.02337***	0.02043***
	(0.00272)	(0.00303)	(0.00372)	(0.00374)
SDG&E x SES	0.01972***	0.02532***	0.01335***	0.01204***
	(0.00292)	(0.00255)	(0.00277)	(0.00290)
Observations	14925285	14925285	14850777	14889258
Popuation		X	X	X
Primary			X	X
Derived				X

Table 4.C.2: Results from second stage of ignition probability regression, using the full sample, with PSPS results counted as ignitions. Standard errors are adjusted to account for error in estimating the probabilities; derivation is given in Section 4.B. Stars indicate significance at the $p = 0.05(*), 0.01(**),$ and $0.001(***)$ levels.

	(1) ignition	(2) ignition	(3) ignition	(4) ignition	(5) ignition	(6) ignition	(7) ignition	(8) ignition	(9) ignition
Health Risk Index	0.00144	-0.0119***	0.0179***	0.00460*	-0.0117***	0.0133**	0.00849***	-0.0134***	0.00875
	(0.00155)	(0.000769)	(0.00430)	(0.00187)	(0.000776)	(0.00461)	(0.00207)	(0.000866)	(0.00547)
SES Index	0.00298	0.0167***	0.0205***	0.00178	0.0166***	0.0224***	-0.00347	0.0123***	0.0132**
	(0.00180)	(0.000759)	(0.00404)	(0.00182)	(0.000761)	(0.00376)	(0.00199)	(0.000925)	(0.00406)
Observations	3120696	9879282	1867158	3120696	9879282	1867158	3120696	9879282	1845942
Pseudo R^2	0.007	0.007	0.013	0.008	0.007	0.019	0.052	0.061	0.088
Utility	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E	SCE	PG&E	SDG&E
Population				X	X	X	X	X	X
Primary							X	X	X
Derived							X	X	X

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

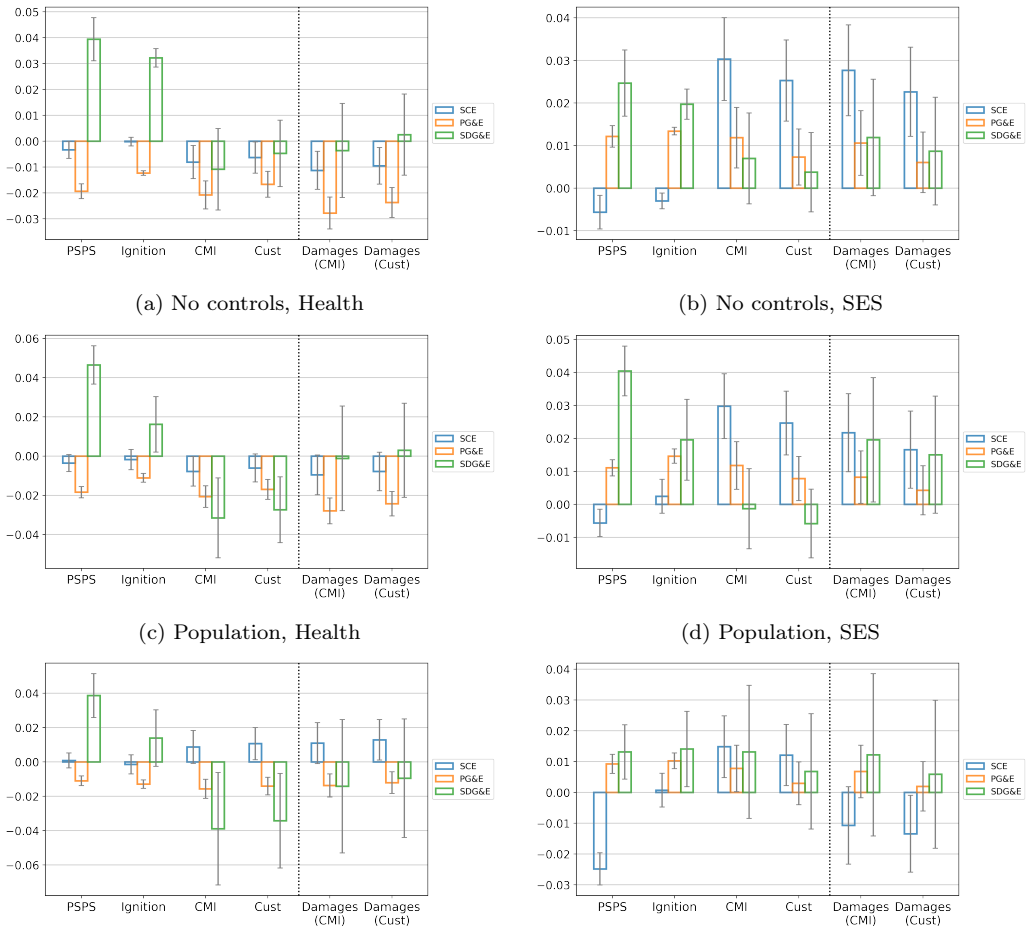
Table 4.C.3: First stage ignition results, using the full sample, with no PSPS results counted as ignitions. This is the assumption that all censored results would have been true negatives.

	(1)	(2)	(3)	(4)
SCE x Health	0.00144	0.00460*	0.00177	0.00400*
	(0.00155)	(0.00187)	(0.00198)	(0.00202)
SCE x SES	0.00298	0.00178	0.00073	0.00051
	(0.00180)	(0.00182)	(0.00196)	(0.00197)
PG&E x Health	-0.01185***	-0.01175***	-0.01385***	-0.01391***
	(0.00077)	(0.00078)	(0.00086)	(0.00086)
PG&E x SES	0.01669***	0.01660***	0.01268***	0.01256***
	(0.00076)	(0.00076)	(0.00090)	(0.00092)
SDG&E x Health	0.01787***	0.01328**	0.01025	0.01004
	(0.00430)	(0.00461)	(0.00542)	(0.00546)
SDG&E x SES	0.02049***	0.02238***	0.01569***	0.01445***
	(0.00404)	(0.00376)	(0.00375)	(0.00404)
Observations	14867136	14867136	14844334	14844604
Popuation		X	X	X
Primary			X	X
Derived				X

Table 4.C.4: Results from second stage of ignition probability regression, using the full sample, with no PSPS results counted as ignitions. Standard errors are adjusted to account for error in estimating the probabilities; derivation is given in Section 4.B. Stars indicate significance at the $p = 0.05(*), 0.01(**),$ and $0.001(***)$ levels.

4.D Additional Coefficient Plots

In the main text, I include a visualization of coefficient estimates using all controls (weather and population). In this section, I include additional visualizations for the other specifications.



(a) No controls, Health

(b) No controls, SES

(c) Population, Health

(d) Population, SES

(e) Primary weather, Health

(f) Primary weather, SES

Figure 4.D.1: Collected coefficient estimates of health risk or SES index on various outcomes, using varying sets of controls. Each group of plots on the x axis show coefficient estimates for a given outcome variable. Damages (to the right of dotted line) are implied from estimates to the left of dotted line and my model. Error bars show ± 1.96 times the standard error. Each group of plots is ordered SCE, PG&E, SDG&E.

Chapter 5

Conclusion

It is becoming increasingly urgent to understand and prepare for the damages from anthropogenic climate change. This thesis explored several important measurement questions for evaluating these damages. I used tools from econometrics, statistics, machine learning, and economics to measure damages from and adaptation to extreme heat in U.S. agriculture, and the equitability of adaptation to wildfire risk by California electric utilities.

There is need for further research to understand the impacts of climate change and how adaptation can offset those impacts. In Chapter 3, I find that the damages from extreme heat in agriculture are significantly offset in the longer run. However, further work is needed to measure what farming practices contribute to this adaptation. Without this knowledge, we are unable to assess how sustainable these farming practices are, and whether they should be applied to other settings. Datasets with records of farmers' practices over time are needed to address the role of unobservable characteristics and evaluate the effectiveness of farming practices.

Should such data become more widely available, panel DML approaches (like those in Chapters 2 and 3) could be instrumental in measuring the impacts of these practices. New datasets could pose significant challenges for standard statistical approaches. There is no guarantee that widely used linear approaches will continue to be good models for underlying physical processes once the researcher starts to incorporate additional variables. Without

expert guidance, it is desirable to flexibly model the role of these additional variables. Panel DML approaches allow researchers to flexibly model these relationships with potentially high-dimensional variation. In Chapter 2, we demonstrate the effectiveness of DML for estimating unknown functions with higher power than conventional flexible modeling approaches. In Chapter 3, I show that these approaches are also effective at recovering true functions from high dimensional variation.

By augmenting DML approaches to account for unobservable variation, these estimators can be useful for a variety of applications beyond the agricultural settings in this thesis. Researchers have access to large datasets of weather variables and economic factors, but largely rely on tools from classical statistics that struggle to represent the full space of interactions and flexible terms in these variables. One benefit of classical approaches is in addressing unobservable variation, such as persistent fixed effects. Literature in environmental economics has demonstrated the importance of addressing these factors to identify the impacts of weather shocks and other variables of interest. Panel DML methods allow researchers to estimate functions with the benefits of modern machine learning, while addressing these unobservable factors. This could be useful for measuring the impact of weather shocks or climate change as inputs to a climate damage function, or for other applications studying the economic consequences of environmental variation.

In addition to better understanding the damages from and adaptation to climate change, more research is needed to study how the consequences of adaptation to climate change for equity. Equity of climate change adaptation has long been a concern, as some of the most impacted nations are those with the least resources to adapt. Even within countries, adaptation has the potential to exacerbate or address equality. In Chapter 4, I show that an adaptation response to wildfire risk is used more frequently in low-socioeconomic status communities among two of California's major utilities, and among sensitive populations in one of the major utilities. This shows that this adaptation mechanism is not being applied equally, although I do not have the data to identify what is responsible for this inequality. Again, better data availability would help researchers identify these channels.

This thesis contributes several relevant methods and analyses to the literature on the eco-

nomic impacts of climate change. More research is needed to understand the consequences of climate change, but there is clearly an urgent need for political action to mitigate future climate change and address the resulting inequality.

Bibliography

Abatzoglou, John (2013). "Development of gridded surface meteorological data for ecological applications and modelling". In: *International Journal of Climatology* 33.1, pp. 121–131.

Anderson, Sarah, Andrew Plantinga, Matthew Wibbenmeyer, et al. (2020). "Inequality in agency responsiveness: evidence from salient wildfire events". In: *Resources for the Future: Washington, DC, USA*, p. 36.

ApS, MOSEK (2021). *MOSEK Optimizer API for Python 9.3.6*.

Athey, Susan, Julie Tibshirani, and Stefan Wager (2019). "Generalized random forests". In: *The Annals of Statistics* 47.2, pp. 1148–1178.

August, Laura, Komal Bangia, Laurel Plummer, Shankar Prasa, Kelsey Ranjbar, Andrew Slocombe, and Walker Wieland (2021). *CalEnviroScreen 4.0*. California Environmental Protection Agency, Office of Environmental Health Hazard Assessment.

Barreca, Alan, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro (2016). "Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the Twentieth Century". In: *Journal of Political Economy* 124.1, pp. 105–159.

Barron, Jonathan T. (2017). "Continuously Differentiable Exponential Linear Units". In: *arXiv preprint*. arXiv: 1704.07483.

Becker, Gary S (1957). *The economics of discrimination*. University of Chicago press.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur (2016). "Inference in high-dimensional panel models with an application to gun control". In: *Journal of Business & Economic Statistics* 34.4, pp. 590–605.

Bickel, Peter J, Ya'acov Ritov, Alexandre B Tsybakov, et al. (2009). "Simultaneous analysis of Lasso and Dantzig selector". In: *The Annals of statistics* 37.4, pp. 1705–1732.

Buolamwini, Joy and Timnit Gebru (2018). "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, accountability and transparency*. PMLR, pp. 77–91.

Burke, Marshall and Kyle Emerick (2016). "Adaptation to climate change: Evidence from US agriculture". In: *American Economic Journal: Economic Policy* 8.3, pp. 106–140.

Burke, Marshall, Solomon Hsiang, and Edward Miguel (2015). "Global non-linear effect of temperature on economic production". en. In: *Nature* 527.7577, pp. 235–239.

Burlig, Fiona, Christopher Knittel, David Rapson, Mar Reguant, and Catherine Wolfram (2020). "Machine learning from schools about energy efficiency". In: *Journal of the Association of Environmental and Resource Economists* 7.6, pp. 1181–1217.

Butler, Ethan E and Peter Huybers (2015). "Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase". In: *Environmental Research Letters* 10.3, p. 034009.

California Public Utilities Commission (2014). *Order instituting rulemaking to revise and clarify commission regulations relating to the safety of electric utility and communications infrastructure provider facilities. Decision 19-05-042.*

– (2019). *Decision adopting de-energization (Public Safety Power Shut-off) guidelines (Phase 1 Guidelines). Decision 19-05-042.*

Chartrand, Rick (2011). "Numerical differentiation of noisy, nonsmooth data". In: *International Scholarly Research Notices* 2011.

– (2017). "Numerical differentiation of noisy, nonsmooth, multidimensional data". In: *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, pp. 244–248.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney K Newey, and James Robins (2018). "Double/debiased machine learning for treatment and structural parameters". In: *Econometrics Journal* 21.1, pp. C1–C68.

Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins (2022). "Locally robust semiparametric estimation". In: *Econometrica* 90.4, pp. 1501–1535.

Chernozhukov, Victor, Matt Goldman, Vira Semenova, and Matt Taddy (2017). "Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels". In: *arXiv*, arXiv–1712.

Chernozhukov, Victor, Whitney K Newey, and Rahul Singh (2021). "A simple and general debiased machine learning theorem with finite sample guarantees". In: *arXiv preprint arXiv:2105.15197*.

– (2022a). "Automatic debiased machine learning of causal and structural effects". In: *Econometrica* 90.3, pp. 967–1027.

– (2022b). "Debiased machine learning of global and local parameters using regularized Riesz representers". In: *The Econometrics Journal*.

Coggins, Shaugn, Lea Berrang-Ford, Keith Hyams, Poshendra Satyal, James Ford, Jouni Paavola, Ingrid Arotoma-Rojas, and Sherilee Harper (2021). "Empirical assessment of equity and justice in climate adaptation literature: A systematic map". In: *Environmental Research Letters* 16.7, p. 073003.

Colangelo, Kyle and Ying-Ying Lee (2020). "Double debiased machine learning nonparametric inference with continuous treatments". In: *arXiv preprint arXiv:2004.03036*.

Costinot, Arnaud, Dave Donaldson, and Cory Smith (2016). "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World". In: *Journal of Political Economy* 124.1.

Crane-Droesch, Andrew (2018). "Machine learning methods for crop yield prediction and climate change impact assessment in agriculture". In: *Environ. Res. Lett* 13, p. 114003.

D'Evelyn, Savannah M et al. (2022). "Wildfire, Smoke Exposure, Human Health, and Environmental Justice Need to be Integrated into Forest Restoration and Management". In: *Current environmental health reports*, pp. 1–20.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken (2014). "What Do We Learn from the Weather? The New Climate-Economy Literature". In: *Journal of Economic Literature* 52.3, pp. 740–798.

Deryugina, Tatyana, Garth Heutel, Nolan H. Miller, David Molitor, and Julian Reif (2019). "The mortality and medical costs of air pollution: Evidence from changes in wind direction". In: *American Economic Review* 109.12, pp. 4178–4219.

Deschênes, Olivier and Michael Greenstone (2007). "The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather". In: *American Economic Review* 97.1, pp. 354–385.

– (2011). "Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the US". In: *American Economic Journal: Applied Economics* 3.4, pp. 152–185.

Diamond, Steven and Stephen Boyd (2016). "CVXPY: A Python-embedded modeling language for convex optimization". In: *Journal of Machine Learning Research* 17.83, pp. 1–5.

Eyring, Veronika, Sandrine Bony, Gerald A Meehl, Catherine A Senior, Bjorn Stevens, Ronald J Stouffer, and Karl E Taylor (2016). "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization". In: *Geoscientific Model Development* 9.5, pp. 1937–1958.

Gianfrancesco, Milena A, Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk (2018). "Potential biases in machine learning algorithms using electronic health record data". In: *JAMA internal medicine* 178.11, pp. 1544–1547.

Guliasi, Leslie (2021). "Toward a political economy of public safety power shutoff: Politics, ideology, and the limits of regulatory choice in California". In: *Energy Research & Social Science* 71, p. 101842.

Guryan, Jonathan and Kerwin Kofi Charles (2013). "Taste-based or statistical discrimination: The economics of discrimination returns to its roots". In: *The Economic Journal* 123.572, F417–F432.

Hagerty, Nick (2021). "Adaptation to surface water scarcity in irrigated agriculture". In: *Working paper*.

Hill, L L, R Blythe, E M Krieger, A Smith, A McPhail, and S B C Shonkoff (2020). "The public health dimensions of California wildfire and Wildfire Prevention, Mitigation, and Suppression". In: *Technical Report; Physicians and Scientists for Healthy Energy*.

Holmes, Thomas P, Robert J Huggett, and Anthony L Westerling (2008). "Statistical analysis of large wildfires". In: *The economics of forest disturbances*. Springer, pp. 59–77.

Hornbeck, Richard (2012). "The enduring impact of the American Dust Bowl: Short- and long-run adjustments to environmental catastrophe". In: *American Economic Review* 102.4, pp. 1477–1507.

Hornik, Kurt, Maxwell Stinchcombe, and Halbert White (1990). "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". In: *Neural Networks* 3.5, pp. 551–560.

Hsiang, Solomon (2016). "Climate econometrics". In: *Annual Review of Resource Economics* 8.1, pp. 43–75.

Hsiang, Solomon et al. (2017). "Estimating economic damage from climate change in the United States". In: *Science* 356.6345, pp. 1362–1369.

Ichimura, Hidehiko and Whitney K Newey (2022). "The influence function of semiparametric estimators". In: *Quantitative Economics* 13.1, pp. 29–61.

Imbens, Guido W and Whitney K Newey (2009). "Identification and estimation of triangular simultaneous equations models without additivity". In: *Econometrica* 77.5, pp. 1481–1512.

IPCC (2022). *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the IPCC Sixth Assessment Report*. Ed. by H.-O. Pörtner et al. Cambridge University Press.

Jain, Piyush, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan (2020). "A review of machine learning applications in wildfire science and management". In: *Environmental Reviews* 28.4, pp. 478–505.

Jann, Ben (2020). "Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment". In.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon (2016). "Combining satellite imagery and machine learning to predict poverty". In: *Science* 353.6301, pp. 790–794.

Kapetanios, George (2008). "A bootstrap procedure for panel data sets with many cross-sectional units". In: *The Econometrics Journal* 11.2, pp. 377–395.

Khan, Shakeeb and Elie Tamer (2009). "Inference on endogenously censored regression models using conditional moment inequalities". In: *Journal of Econometrics* 152.2, pp. 104–119.

Kingma, Diederik P. and Jimmy Lei Ba (2015). "Adam: A method for stochastic optimization". In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR. arXiv: 1412.6980.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015). "Prediction policy problems". In: *American Economic Review*. Vol. 105. 5, pp. 491–495.

Klosin, Sylvia (2021). "Automatic Double Machine Learning for Continuous Treatment Effects". In: *arXiv preprint arXiv:2104.10334*.

Klosin, Sylvia and Max Vilgalys (2022). "Estimating Continuous Treatment Effects in Panel Data using Machine Learning with an Agricultural Application". In: *arXiv preprint arXiv:2207.08789*.

Knittel, Christopher R and Samuel Stolper (2019). *Using machine learning to target treatment: The case of household energy use*. Tech. rep. National Bureau of Economic Research.

Kurukulasuriya, Pradeep, Namrata Kala, and Robert Mendelsohn (2011). "Adaptation and climate change impacts: a structural Ricardian model of irrigation and farm income in Africa". In: *Climate Change Economics* 2.02, pp. 149–174.

Lang, Kevin and Ariella Kahn-Lang Spitzer (2020). "Race discrimination: An economic perspective". In: *Journal of Economic Perspectives* 34.2, pp. 68–89.

Lemoine, Derek (2018). "Estimating the Consequences of Climate Change from Variation in Weather". In: *Manuscript. National Bureau of Economic Research* w25008.

Liakos, Konstantinos G., Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis (2018). "Machine learning in agriculture: A review". In: 18.8, p. 2674.

Mboup, Mamadou, Cédric Join, and Michel Fliess (2009). "Numerical differentiation with annihilators in noisy environment". In: *Numerical algorithms* 50.4, pp. 439–467.

McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behaviour". In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York, NY, USA: Academic Press New York, pp. 105–142.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan (2021). "A survey on bias and fairness in machine learning". In: *ACM Computing Surveys (CSUR)* 54.6, pp. 1–35.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw (1994). "The Impact of Global Warming on Agriculture: A Ricardian Analysis". In: *The American Economic Review* 84, pp. 753–771.

Meyers, Keith and Paul W Rhode (2019). "Exploring the Causes Driving Hybrid Corn Adoption from 1933 to 1955". In: *Economics of Research and Innovation in Agriculture*. University of Chicago Press.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine learning: An applied econometric approach". In: *Journal of Economic Perspectives*. Vol. 31. 2, pp. 87–106.

Newey, Whitney K and Daniel McFadden (1994). "Large sample estimation and hypothesis testing". In: *Handbook of econometrics* 4, pp. 2111–2245.

Niemi, Ernest G. and Kristin S Lee (2001). "Wildfire and poverty: an overview of the interactions among wildfires, fire-related programs, and poverty in the Western States". In: The Center for Watershed and Community Health, Mark O. Hatfield School of Government, Portland State University.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan (2019). "Dissecting racial bias in an algorithm used to manage the health of populations". In: *Science* 366.6464, pp. 447–453.

Ojerio, Ryan S (2008). "Equity in Wildfire Risk Management: Does Socioeconomic Status Predict Involvement in Federal Programs to Mitigate Wildfire Risk?" MA thesis. Department of Planning, Public Policy & Management, University of Oregon.

Olmstead, Alan L and Paul W Rhode (2011). "Adapting North American wheat production to climatic challenges, 1839-2009." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.2, pp. 480–485.

Ortiz-Bobea, Ariel (2013). "Is Weather Really Additive in Agricultural Production? Implications for Climate Change Impacts". In: *SSRN Electronic Journal* 1, pp. 3–41.

– (2020). "The role of nonfarm influences in Ricardian estimates of climate change impacts on US agriculture". In: *American Journal of Agricultural Economics* 102.3, pp. 934–959.

Ortiz-Bobea, Ariel, Haoying Wang, Carlos M Carrillo, and Toby R Ault (2019). "Unpacking the climatic drivers of US agricultural yields". In: *Environmental Research Letters* 14.6, p. 064003.

Oster, Emily (2019). "Unobservable selection and coefficient stability: Theory and evidence". In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Pacific Gas and Electric Company (2020). *2020 risk assessment and mitigation phase report*. Tech. rep.

– (June 2021). *2021 Wildfire mitigation plan – revised*. Tech. rep.

Parikh, Ravi B, Stephanie Teeple, and Amol S Navathe (2019). "Addressing bias in artificial intelligence in health care". In: *Jama* 322.24, pp. 2377–2378.

Park, J. and I. W. Sandberg (1991). "Universal Approximation Using Radial-Basis-Function Networks". In: *Neural Computation* 3.2, pp. 246–257.

Plantinga, Andrew J, Randall Walsh, and Matthew Wibbenmeyer (2022). "Priorities and effectiveness in wildfire management: evidence from fire spread in the western United States". In: *Journal of the Association of Environmental and Resource Economists* 9.4, pp. 603–639.

Rambachan, Ashesh (2021). "Identifying prediction mistakes in observational data". In: *Working Paper, Harvard University Department of Economics*.

Rhodes, Noah, Lewis Ntaimo, and Line Roald (2020). "Balancing wildfire risk and power outages through optimized power shut-offs". In: *IEEE Transactions on Power Systems* 36.4, pp. 3118–3128.

Rode, Ashwin et al. (2021). "Estimating a social cost of carbon for global energy consumption". In: *Nature* 598.7880, pp. 308–314.

Rothenhäusler, Dominik and Bin Yu (2019). "Incremental causal effects". In: *arXiv preprint arXiv:1907.13258*.

San Diego Gas and Electric Company (Feb. 2021a). *2020-2022 Wildfire mitigation plan update*. Tech. rep.

– (2021b). *San Diego Gas and Electric Company: risk assessment mitigation phase*. Tech. rep.

Satyal, Poshendra, Morten Fibieger Byskov, and Keith Hyams (2021). "Addressing multidimensional injustice in indigenous adaptation: the case of Uganda's Batwa community". In: *Climate and Development* 13.6, pp. 529–542.

Schlenker, Wolfram and Michael J Roberts (2009). "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.37, pp. 15594–8.

– (2006). "Nonlinear Effects of Weather on Corn Yields*". In: *Review of Agricultural Economics* 28.3, pp. 391–398.

Semenova, Vira and Victor Chernozhukov (2021). "Debiased machine learning of conditional average treatment effects and other causal functions". In: *Econometrics Journal* 24.2, pp. 264–289. arXiv: 1702.06240.

Sheller, Mimi and Yolanda M Leon (2016). "Uneven socio-ecologies of Hispaniola: Asymmetric capabilities for climate adaptation in Haiti and the Dominican Republic". In: *Geoforum* 73, pp. 32–46.

Short, K C (2014). *A spatial database of wildfires in the United States, 1992-2011*.

Short, Karen C (2021). *Spatial wildfire occurrence data for the United States, 1992-2018*.

Shukla, P R et al. (2019). "IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems". In.

Smit, Barry and Olga Pilifosova (2003). "Adaptation to climate change in the context of sustainable development and equity". In: *Sustainable Development* 8.9, p. 9.

Southern California Edison Company (June 2021). *2021 Wildfire mitigation plan update (revision)*. Tech. rep.

– (2022). *Southern California Edison Company: risk assessment mitigation phase*. Tech. rep.

Stetter, Christian, Philipp Mennig, and Johannes Sauer (2022). "Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study". In: *European Review of Agricultural Economics*.

Sutch, Richard (2008). *Henry Agard Wallace, the Iowa Corn Yield Tests, and the Adoption of Hybrid Corn*. Tech. rep. Cambridge, MA: National Bureau of Economic Research.
– (2011). "The Impact of the 1936 Corn Belt Drought on American Farmers' Adoption of Hybrid Corn". In: *Economics of Research and Innovation in Agriculture*. University of Chicago Press, pp. 195–223.
Taylor, Steve W, Douglas G Woolford, CB Dean, and David L Martell (2013). "Wildfire prediction to inform fire management: statistical science challenges". In: *Statistical Science* 28.4, pp. 586–615.
Thomas, David SG and Chasca Twyman (2005). "Equity and justice in climate change adaptation amongst natural-resource-dependent societies". In: *Global environmental change* 15.2, pp. 115–124.
Tol, Richard SJ (2009). "The economic effects of climate change". In: *Journal of economic perspectives* 23.2, pp. 29–51.
Valdberg, Anna, Andrea Tozer, and Elena Kilberg (2021). *Southern California Edison Company's (U 338-E) public safety power shutoff post-event report for November 21, 2021 de-energization event*. Tech. rep.
Van Breugel, Floris, J Nathan Kutz, and Bingni W Brunton (2020). "Numerical differentiation of noisy data: A unifying multi-objective optimization framework". In: *IEEE Access* 8, pp. 196865–196877.
Varian, Hal R (2014). "Big data: New tricks for econometrics". In: *Journal of Economic Perspectives* 28.2, pp. 3–28.
Wibbenmeyer, Matthew and Molly Robertson (2022). "The distributional incidence of wildfire hazard in the western United States". In: *Environmental Research Letters* 17.6, p. 064031.
Wigtil, Gabriel, Roger B Hammer, Jeffrey D Kline, Miranda H Mockrin, Susan I Stewart, Daniel Roper, and Volker C Radeloff (2016). "Places where wildfire potential and social vulnerability coincide in the coterminous United States". In: *International Journal of Wildland Fire* 25.8, pp. 896–908.
Wong-Parodi, Gabrielle (2020). "When climate change adaptation becomes a "looming threat" to society: Exploring views and responses to California wildfires and public safety power shutoffs". In: *Energy Research and Social Science* 70.
Wooldridge, Jeffrey M (2005). "Violating ignorability of treatment by controlling for too many factors". In: *Econometric Theory* 21.5, pp. 1026–1028.
– (2010). *Econometric analysis of cross section and panel data*. MIT press.
Xi, DD, Stephen W Taylor, Douglas G Woolford, and CB Dean (2019). "Statistical models of key components of wildfire risk". In: *Annual Review of Statistics and Its Application* 6.1, pp. 197–222.