

© 2022 Chang Cai

ESSAYS ON HETEROGENEITY IN OUTDOOR RECREATION DEMAND AND CLIMATE
CHANGE'S IMPACT ON AGRICULTURE

BY
CHANG CAI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Associate Professor Benjamin Gramig, Chair and Director of Research
Professor Amy W. Ando
Assistant Professor Peter Christensen
Associate Professor Carena van Riper

ABSTRACT

This dissertation presents three chapters related to problems in natural resource management, namely, recreation demand, park management, and climate change impacts on the agriculture sector. The first essay quantifies the response of tourists to increasing wildfire activities in 32 national parks across the western United States. Results from a large-scale analysis show that wildfires will on average reduce annual visitation by about 1.35% each year, which is about a half of the estimates suggested by previous literature. In addition, I find the majority of the visitation losses caused by wildfires are attributed to emergency fire closures, and national park visitors, comprised of both day users and campers, who are surprisingly resilient to smoky conditions. The second essay investigates the possibility of applying Temporal Disaggregation to SafeGraph's daily foot traffic data to downscale the monthly visitation to 31 western national parks to the daily scale. Using a grid search design to achieve the best model performance, my results confirm the applicability of SafeGraph data to downscale monthly visitation for the majority of the parks in my sample. Overall, the approach proposed in this study provides an opportunity for park managers to reconstruct daily visitor use data and be better equipped for adaptive management in national parks. The third essay examines how the choice of grouping can affect expected future agricultural profits when one wants to quantify the heterogeneous impacts of climate change on agriculture. The results indicate that accounting for grouping uncertainty greatly increases the confidence interval around projected climate impacts. In addition, I do not find that one type of grouping is superior to any other. I suggest two potential solutions and emphasize the importance of explicitly controlling for grouping uncertainty in future studies.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deep and sincere gratitude to my advisor, Professor Ben Gramig, for his invaluable insights, consistent support, and unwavering guidance during my Ph.D. study. Professor Gramig has been an encouraging, patient, and genuine mentor, who supported me in pursuing my research interests, encouraged confidence from the initial step, helped me establish my research identity, and read my numerous revisions. This thesis would not have been possible without his guidance and inspiration. I'm grateful for his investment in my success throughout the entire thesis process, and it's my absolute honor to learn from him.

My committee members, Amy Ando, Peter Christensen, and Carena van Riper, have been instrumental in arriving at this point. Amy Ando deserves my gratitude for being a source of inspiration and a role model as I started to dive deeper into this topic. Peter Christensen provided critical advice on my research design and framing, which has pushed me to be a rigorous researcher. I especially want to thank Professor van Riper, who opened the door to the exciting domain of outdoor recreation and encouraged me to share my ongoing work with the greater scientific community. It's her encouragement that makes me confident that the topic of this thesis is worth pursuing. I would also like to thank my coauthor, Sandy Dall'erba, who provided excellent collaboration opportunities in the first three years of my Ph.D. program and helped me with my very first publication.

I want to thank the wonderful support from the staff at Mumford Hall. Pam Splittstoesser, Dung Quach-Wisdom, Melissa Warmbier, Donna Stites, and Susan King have provided outstanding administrative support over the years. A special thanks to Ian Chong for always keeping his door open to my endless IT questions and providing excellent IT support. Thanks to

the University of Illinois Department of Agricultural and Consumer Economics for providing me with financial support.

Lastly, I owe my deepest gratitude to my beloved family for their love and tremendous support: my mother Ming Xu for calming me down and always standing behind me; my father Yuanfa Cai for teaching me to keep an open mind and embrace new ideas, and my grandparents for emotional support. Many thanks to my partner, Zhangliang Chen, for being there for me, for patiently listening to me talk things out, and for your unconditional support (even during the darkest days). I must also thank my friends, whose calls and conversations sheltered me from the pressure as I hurdle through the path of a Ph.D. degree.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
CHAPTER 2: WILDFIRE AND VISITATION IN U.S. NATIONAL PARKS	6
2.1 Introduction.....	6
2.2 Data	12
2.3 Model	15
2.4 Results.....	18
2.5 Conclusion	32
2.6 Tables	34
2.7 Figures.....	38
CHAPTER 3: TEMPORAL DISAGGREGATION OF VISITATION TIME SERIES IN U.S. NATIONAL PARKS USING MOBILE PHONE GPS DATA.....	45
3.1 Introduction.....	45
3.2 Data	50
3.3 Method	54
3.4 Results.....	63
3.5 Discussion	69
3.6 Conclusion	74
3.7 Tables	76
3.8 Figures.....	78
CHAPTER 4: ON THE EVALUATION OF HETEROGENEOUS CLIMATE CHANGE IMPACTS ON US AGRICULTURE: DOES GROUP MEMBERSHIP MATTER?.....	84
4.1 Introduction.....	84
4.2 A review of the conceptual framework.....	88
4.3 Data	93
4.4 Research method	95
4.5 Results.....	101
4.6 The way forward	108
4.7 Conclusion	110
4.8 Tables	113
4.9 Figures.....	116

CHAPTER 5: CONCLUSION.....	120
APPENDIX A: SUPPLEMENTAL MATERIALS FOR CHAPTER 2.....	123
APPENDIX B: SUPPLEMENTAL MATERIALS FOR CHAPTER 3	131
APPENDIX C: SUPPLEMENTAL MATERIALS FOR CHAPTER 4.....	134
REFERENCES:	147

CHAPTER 1:

INTRODUCTION

This dissertation presents three essays on the spatial and temporal heterogeneity of socio-economic response to environmental change, with a particular focus on two aspects: demand for outdoor recreation and agricultural production. Every year, millions of people flock to the vast public lands in the United States (US) to pursue a variety of recreational experiences. Nature-based recreation and tourism are becoming increasingly important to local and regional economies. The rapidly growing visitation along with the warming-mediated expansion of visiting season have raised considerable concerns over the appropriate use levels in protected areas. Too many visitors could result in overcrowding, additional demands on infrastructure, and conflicts among visitors, all of which would diminish the quality of visitor experiences. Visitors' footprints will also threaten natural and cultural resources through trail erosion, water pollution, and disturbance of wildlife. While the expanding recreation in recent years also provides opportunities, many local park-oriented communities are looking toward recreation and tourism to boost local economic development. To address the increasing visitation, park managers are working hard to meet multiple visitor use management objectives, which are to improve the visitor experience, encourage access and promote recreation, while also conserving the natural and cultural resources. Many of the management strategies, such as the permit system, are implemented through managing access based on capacity-related visitation limits. Visitation may not only serve as a crowding indicator but be also a key input for calibrating economic benefits to local and national economies from tourist expenditures. Therefore, it's important to understand how visitation responds to natural shocks such as wildfires and smoke and access restrictions such as permit programs and

fire-related closures. Although the spatial heterogeneity of the impact of climate change on agriculture, i.e., the fact that climate sensitivity is space-dependent, is well acknowledged, there have been considerable inconsistencies in the definition of spatial heterogeneity as represented by various grouping strategies.

This dissertation presents three chapters that use large-scale analyses to study factors that may influence recreation demand and provide evidence on the temporal heterogeneity of visitor use as well as insights on the spatial heterogeneity of climate change's impacts. In the second chapter, I study the responses of tourists to increased wildfire risk in national parks (NP). Previous studies of how wildfires affect visitation to public lands predominantly rely on survey data collected from special population subgroups in a specific region and/or for a specific type of recreation activities such as camping, fishing, or hiking. Yet we know little about the aggregate effect of wildfires on recreation demand. Because recreational users in different regions may not behave similarly, and as the majority of a site's visitors are more likely engaged in sightseeing along scenic roads and short-distance day hiking than backcountry trips, the policy implications of studies based upon narrow and likely unrepresentative samples cannot be broadly applied. To fill these gaps, the first chapter investigates the causal effect of wildfires and wildfire smoke on recreation visits to 32 national parks across the western United States. I link monthly visitation data from the National Park Service (NPS) to an extensive dataset on wildfire occurrences, fire-related park road closures, and wildfire-caused smoke for the last three decades. I construct this novel dataset by combining ground-based historical fire records, social media, and daily satellite-based remote sensing fire detection data. Using fixed effect panel regressions at the park level, I leverage the annual variations in fire events nearby each park to identify the causal relationships between wildfire activities and park visitation. My analysis addresses the identification challenge

of reverse causality by exploiting the plausibly random variation in lightning-caused fires and accounts for potentially confounding spillover effects from neighboring recreation areas. I also investigate the underlying mechanisms that drive these effects, including fire emergency closures. My study yields more applicable policy implications than many prior works did.

In the third chapter, I explore the possibility of improving the temporal resolution of official visitor counts reported by National Park Service (NPS). While NPS monitors visitor use and publishes visitation statistics on a consistent basis, such statistics are generally a number aggregated on the monthly level. Monthly visitation numbers provide relatively very little information on the temporal distribution of when individual visits are made within a month, making it challenging to model time-varying determinants of visitation. As the newly proposed management plans are considering timed entry permit systems that would manage the number of people in the park during peak visitation seasons. The time-entry system is designed to spread visitation throughout the week or the day, and some of the day-use reservation systems impose varying daily limits on weekdays versus weekends/holidays. More granular daily visitation data is always preferred when evaluating a time-varying management program and estimating the effect of visitor spending on the local economy over a short time period. In this chapter, I combine newly available mobile phone location data with Temporal Disaggregation (TD) approach from the time series literature to downscale the monthly visitation to 31 national parks in the western US to the daily level. To improve the model performance, I develop a grid search design that simultaneously refines the cellphone location records and achieves the best TD model performance, allowing me to move beyond park-based case studies to a large-scale investigation. The findings of this chapter, which includes historical daily visitation, would allow scholars and park managers to investigate

if visitation shifts are a direct result of management policy changes (e.g. the implementation of day use permits) or acute events (e.g. wildfires).

The fourth chapter diverges from previous chapters' topics of tourism and protected areas to examine the effect of climate change on agriculture. Despite the fact that this chapter has a different topic, it has a common theme with the others in that it seeks to understand how incorporating more granular data in the presence of spatial heterogeneity might influence the estimated impacts of climate change and natural disasters. This chapter examines how the choice of grouping can affect group-specific estimates of climate change's impact on agricultural revenues. I survey all existing economic literature studying the spatial heterogeneity of the relationship between climate and agriculture and find that most studies tend to group geographic units based on one single pre-determined criterion such as the differences in irrigation, and then test the relevance of these groups. Additionally, there have been considerable variations in the way of grouping counties. Because various forms of grouping are available in the literature, the question of whether the grouping that is ultimately chosen may affect the value and the precision of the estimated impact of climate change has not been investigated so far. This inconsistency in the grouping choices can also pose an empirical challenge to the practitioners in terms of which grouping should be preferred. In this chapter, I address the uncertainty in groupings by examining, first, what the effect of various groupings is on the point estimates of group-specific impact projection and, second, whether one can evaluate the relative effectiveness of each grouping and thus select the preferred one. Using panel data of climate and agricultural outcomes for US agriculture from 1997 to 2012, I estimate the county-level changes in agricultural revenues due to climate change for each of the preexisting groupings on the same sample and same time period. Furthermore, I conduct cross-validation exercises and multiple statistical tests to evaluate whether

one grouping is a better choice than the others. If the uncertainty associated with grouping choice is not addressed, the actual confidence intervals associated with climate change impact estimates may be much larger than that reported in previous studies.

CHAPTER 2:

WILDFIRE AND VISITATION IN U.S. NATIONAL PARKS

2.1 Introduction

Millions of people flock to the vast expanse of public lands in the United States each year. In 2019 alone, these outdoor recreation activities contributed \$459.8 billion in economic activity (BEA, 2020). In recent years, US national parks have attracted an increasing number of visitors, as reflected by three new visitation records since 2014 (Ziesler, 2020). Most parks' peak visitation seasons coincide with fire seasons that are projected to last longer with more severe occurrences in the foreseeable future (Abatzoglou and Williams, 2016). Catastrophic fires can cause large-scale disturbance to ecosystems and result in emergency closures and smoke-related health risks, which may trigger shifts in visitation patterns (Englin et al., 1996; Starbuck et al., 2006; Sánchez et al., 2016). Intensified fire events pose unprecedented challenges both to park managers working to mitigate fire risk using limited resources during extended peak seasons (Fisichelli et al., 2015; Rothman, 2005), and to gateway communities that have grown heavily reliant on the tourism industry stemming from the outdoor recreation (White et al., 2016). While the effects of wildfires on outdoor recreation demand are well studied (e.g., see Bawa, 2017), the previous literature has predominantly focused on specific regions (Duffield et al., 2013; Kim and Jakus, 2019; Loomis et al., 2001) and/or a specific type of recreation use, such as camping or biking (Hesseln et al., 2003; Gellman et al., 2022), leaving the aggregated impacts of wildfires largely unmeasured. However, recreation demand in different regions may not respond to wildfires similarly and most park visitors are often day users, who would rather spend a day driving along scenic roads and hiking short trails than venturing into the backcountry (Vaske and Lyon, 2014). As a result, policy

implications drawn from existing studies are likely based upon narrow and likely unrepresentative samples that cannot be broadly applied to the entire population of park visitors.

This paper provides a large-scale evaluation of the effects of wildland fires on recreational visits to 32 national parks in the western United States. My starting point is to develop an extensive dataset on wildfire and smoke that integrates ground-based historical fire records obtained from several wildland fire management agencies, fire closure history recovered from social media, and satellite-based remote sensing data. I then link the wildfire data to monthly visits to each national park over a 25-year timespan. Visitation data from the National Park Service (NPS) collects observed visitor counts at both the entire park and the entrance station levels. This enables me to estimate the causal effects of wildfires on visitation that are representative of all national park visitors by exploiting large variations in the time, size, and location of wildfires and smoke at a granular level.

This study addresses three primary research questions. First, I exploit month-over-month variation in fire events within or proximate to each park to provide a large-scale estimate of the impact of wildfires on outdoor recreation. In this way, my empirical strategy is most similar to that of Keiser et al. (2018), who use year-to-year variation in in-park ozone concentration to estimate the impact of air pollution on park visitation. Second, I explore potential channels for the negative relationship between fire and visitation, including emergency fire closures and an increase in wildfire risk. Third, I disentangle the effect of wildfire-induced smoke from the direct effect of local fires so that the smoke effects can be directly compared to the overall fire effects.

This study yields three substantive findings. First, using a representative sample of all park visitors, I consistently find a statistically negative relationship between wildfire activities and monthly park visitation. Following the previous literature (Duffield et al., 2013; Kim and Jakus,

2019), I measure wildfires near national parks as the total acres of fires burned within a certain radius of the park's geographic centroid. A primary identification concern that may arise when estimating fire-visitation effects is that because wildfires may not be randomly assigned, existing estimates may be biased by reverse causality, whereby increasing visitation could increase human-started fires or omitted variables introduced by spatial spillovers from fire activities at nearby recreation areas. Therefore, I overcome this challenge by exploiting the plausibly random variation in lightning-caused fires and accounting for potentially confounding spillover effects. I show that, on average, the monthly visitation loss is 0.064 percent per thousand acres burned, and the impact carries over to the following month, after flexibly controlling for differences in local weather, seasonality, and time-varying unobservable factors in each park. This translates to an annual visitation loss of about 700,000 visits for 32 western parks in my sample, suggesting a smaller impact than Gellman et al. (2022)'s study using campsite reservation data.

Second, I present evidence regarding the channels to explain fire-induced visitation impacts in response to past work that has suggested access (Scott et al., 2007) and risk perceptions (Martin et al., 2009) are two of the primary drivers of impacts incurred from wildfires. Specifically, lack of access due to emergency closures accounts for a large share of the decrease in visitation and shows impacts nearly 10 times larger than fires not causing any closures. Because parks are rarely closed entirely, I obtain daily measures of emergency fire closures for every park entrance station from the park's social media posts and State Departments of Transportation and find that monthly visitation falls by 2.27 percent for an additional park closure day in a month, suggesting that prolonged site closure may generate a relatively large loss in tourism. The gravity of decisions to close a park unit should be carefully considered by decision-makers (Gordon et al., 2013; Larsen et al., 2021). Furthermore, I confirm park visitors do respond to fires that are either too small in

scale or burning too remotely to cause any closure events, despite in a much smaller magnitude. The channel for these smaller fires is possibly due to health concerns or reduced visibility, which has been well documented in the literature (Schultze et al., 1983; Thapa et al., 2013). To distinguish the effects of such small fires from fire closure, I exclude dates with closure events and measured wildfire risk by the number of days with fires but no closure in place, arguing that significantly negative but much smaller effects for wildfire risk would emerge, as evidenced by a nearly 25 percent decrease in visitation for an additional “small fire day”.

Third, I observe that wildfire smoke alone has a weak and statistically insignificant impact compared to local fire effects. My estimation approach includes both measures of monthly fire days and smoke days constructed from the National Oceanic and Atmospheric Administration (NOAA) Hazard Mapping System Fire and Smoke Product (HMS). The coefficient for the fire variable is insensitive to the exclusion of the smoke variable from the model, and these two variables are highly correlated. These findings raise the concern that simply including both fire and smoke variables in the model would exacerbate the attenuation bias due to potential measurement error in the smoke variable and collinearity between smoke and fire measures (Wooldridge, 2010). To address this concern, I follow a procedure adapted from Brey and Fischer (2016) and restrict smoke days to dates with above-average ozone concentrations and no active local fires. After adjusting the smoke variable to account for smoke days in the absence of local fire, I find that the estimated impact of smoke days on visitation remains small and statistically non-significant, while the coefficient on the fire variable is always significant and fairly stable.

This study contributes to the literature by providing the first large-scale estimation of fire-visitation effects that are representative of the affected population of national park visitors. Existing quantification of the wildfire impact on recreation demand has generally relied on a

limited sample from a specific type of recreation users or a specific region (e.g., Loomis et al., 2001; Gellman et al., 2022). The external validity of their estimates is therefore worth questioning. Previous research has also indicated that we should expect campers' and backcountry users' responses to differ from those of day users who may just take day trips along scenic park roads (Manning, 1985; Rice et al., 2020; Riper et al., 2020), yet previous work has yet to nuance these differences with respect to wildlife risk perceptions. By contrast, my analysis leverages the only available comprehensive visitation data from NPS and takes advantage of the high granularity of my data to overcome the identification challenges in the two prior time-series studies on contemporaneous fire effects (Duffield et al., 2013; Kim and Jakus, 2019). My research finds a robust estimate of the wildfire-visitation relationship that better represents the park recreation in the American West. Compared to Gellman et al. (2022), a high-profile recent study using campground use data, I estimate a much smaller fire-induced visitor loss, which suggests that the impact evaluation based on a limited sample could be misleading.

Previous research suggests that there are two predominant channels by which wildfire activities may cause tourism loss.: First, fire-induced closures have emerged as an important source of impact on visitation and visitor experiences more broadly (Scott et al., 2007). Secondly, risk perceptions of wildfire – defined as an individual's concern about hazards they associate with an issue (Martin et al., 2009) – are also instrumental in explaining tourism impacts from wildfire. I find restricted access due to emergency fire closures throughout the season is a more direct block and impact on visitation, and visitors may also respond to their risk perceptions of relatively small or remote fires with a much smaller magnitude. Such evidence may be useful for park managers facing the challenge of allocating limited resources to mitigate wildfire risk when the park is overrun with visitors (Gordon et al., 2013; Kim and Jakus, 2019; Larsen et al., 2021). Given that

record-breaking wildfires are expected to become more frequent, the results suggest a greater need for timely adjustments for incident planning and park management. For example, the impacts of fire closures can be mitigated by providing timely information on fire restrictions and encouraging visitors to explore other unrestricted regions of the park that meet desired safety and health conditions.

Finally, my findings speak directly to the ongoing and contentious debate on the use of both prescribed fires and those naturally ignited fires that will be allowed to burn as a proactive fire prevention policy (Altangerel and Kull, 2013). Despite the efficacy of prescribed burning in reducing the size of wildfires (Cochrane et al., 2012), the primary concern surrounding its use centers on potential risks arising from externalities such as smoke from prescribed burning may have a substantial impact on downwind air quality. Indeed, a growing body of literature suggests that wildfire smoke is a growing public health issue (Black et al., 2017). My findings suggest that the effects of large catastrophic fires burning in heavily visited areas drive a large fraction of visitor loss, while smoke from distant sources creates somewhat small spatial externalities in local tourism flows. This information will shed light on the potential benefits of the use of off-season prescribed burning, especially for those tourism-dependent gateway communities that are vulnerable to shifting visitation patterns in the face of natural disasters (Smith et al., 2016; Kim and Jakus, 2019). Carefully planning prescribed burning in heavily visited areas at the right time can, to a large extent, mitigate the risk of a catastrophic fire, which causes most of the visitation loss and thus tourism loss. Meanwhile, reductions in fire-related externalities to nearby communities can be minimized to the extent possible.

The remainder of this paper is organized as follows. The next section describes the data sources and construction of the dataset. Section 2.3 explains my empirical strategy in detail.

Section 2.4 presents the main results on visitation impacts of wildfires, investigates the potential channels, estimates the effect of wildfire smoke, provides a back-of-the-envelope calculation, and compares it with existing estimates in the literature. Section 2.5 discusses the findings and concludes.

2.2 Data

This section summarizes data sources and defines key variables used in the analysis. A comprehensive description of data construction can be found in Appendix Table A.1. I link historical records on wildfire occurrences to monthly visits to 32 national parks (Figure 2.1) to produce a dataset that spans from 1993 to 2019 in the western US (Arizona, California, Colorado, Idaho, Montana, New Mexico, Nevada, Oregon, Texas, Washington, and Wyoming). I restrict my analysis to parks in these states for two reasons: First, lightning-started fires are relatively more common in the western US than in the eastern US, which is dominated by human-started wildfires (Nagy et al., 2018), so this focus alleviates the concern about reverse causality as discussed in Section 2.3. Second, visitation to western national parks collectively accounts for about 70% of the annual national park attendance.

2.2.1 Wildfire and smoke data

To create a metric on fire activity, I start with ground-based wildfire records. I use fire records from the US Forest Service (USFS) Fire Program Analysis's fire occurrence database (Short, 2017), the most comprehensive database available for wildfire records acquired from local, state, and federal agencies. This dataset reports geo-coordinates of wildfire ignitions and includes attributes such as final burned area, discovery, and end date of the fire. For each fire event, I calculate the distance from the fire origin to the centroids of national parks. I define a park's

wildfire size as the total acres burned by all fires within a certain radius of that park centroid in a given month. I chose a radius of 50 miles, following Duffield et al. (2013) and Kim and Jakus (2019). An example case of Yellowstone National Park in September 2003 is provided in Figure 2.2. Wildfire size serves as the primary measure of monthly fire activity in the analysis.

In order to measure the difference in park accessibility under wildfire emergency conditions, I utilize two archival sources to identify the consequential park closure days. The first source for closure records is the official news releases from the park service and historical social media posts from Twitter and Facebook. When the park closes to traffic for any reason such as fire, weather, or road construction, each park's social media team will announce the closure to visitors by putting out a press release on the park website and the Twitter and Facebook home pages. Since many park closures are not related to wildfires, I filter the text from news and social media that contains keywords such as "wildfire" or "fire". While this method provides me with accurate information on affected areas and closed segments of the park roads, the search generates only a limited subset of records on traffic conditions on state highways used to enter the park. I therefore also obtain historical state route closures related to wildfires from the state departments of transportation. Together, these two sources allow me to construct a proxy measure for fire closure days that is available since 2011. I define "fire closure days" as the number of days within a month when a park entrance is closed due to wildfires.

Data from satellite-based estimates of wildfire smoke plumes and fire activity are downloaded from the National Oceanic and Atmospheric Administration (NOAA) Hazard Mapping System (HMS). Both datasets are developed by HMS specialists who monitor and analyze sub-daily visible imagery from seven NOAA and National Aeronautics and Space Administration (NASA) satellites. The HMS spatial data files include the daily polygons of the

smoke plumes and spatial points of active fires detected during daylight hours since August 2005. I define a smoke-impacted day at a park as any day when a smoke plume intersects the park boundary. To construct a smoke exposure measure, I aggregate the number of smoke-impacted days to the month level to obtain the measure “smoke days” in a given month. Section 2.4.4 provides a detailed discussion of the stability of the estimated effect of smoke days. To compare the impacts of smoke to the impacts of fires, I also create the metric “fire days” by following the approach employed by (Burke et al., 2021). I begin by grouping all nearby fire points on a given day by placing a 375-meter buffer around each fire point and merging overlapping buffers into hotspots. I then estimate the daily fire size as the total acres of merged buffers. I filter out days with fire sizes smaller than 300 acres and count the total number of the remaining days within a month as the “fire days”.

2.2.2 Park visitation data

Data on park visitation comes from NPS Visitor Use Statistics. For each NPS-managed unit (e.g., national park and national monument), NPS maintains historical data on monthly visitation rates for different categories of user groups, including recreation visits, non-recreation visits, and overnight stays. A “visit” is generally defined as the entry of a visitor into an NPS-managed site (Manning, 2011). The total number of visits is primarily collected via automated vehicle counters set up at entrance roads and estimated by multiplying the number of vehicles by a person-per-vehicle multiplier (Ziesler and Pettebone, 2018). I focus on recreation visits because these data are the most widely used and broadly applicable statistics to reflect visitors’ demand for national park visits (Bergstrom et al., 2020). Large, iconic parks report vehicle counts for each individual entrance station. Such vehicle entrance counts are also obtained and used as weights to build weighted-average fire closure days at the park level.

2.2.3 Weather and travel costs

The analysis includes a flexible set of control variables including temperature, precipitation, vapor pressure deficit (VPD) (a measure of atmosphere dryness), and real gasoline price (a proxy for travel cost). Specifically, I extract gridded data on monthly maximum temperature, minimum temperature, total precipitation, minimum VPD, and maximum VPD from PRISM Climate Group (PRISM Climate Group, 2004), which provides gridded data at a 4 km spatial resolution. To construct weather conditions at the park level, I average the monthly measures over all grid cells within the park boundaries. Furthermore, I obtain monthly gasoline prices for all western states from the US Energy Information Administration (EIA, 2020). Following Duffield et al. (2013) and Kim and Jakus (2019), I divide the gasoline price by real per capita personal income to proxy state-wide travel costs at all distances.

2.3 Model

The first step in my empirical analysis is to establish the causal relationship between wildfires and national park visitation. To identify the effect of fire activities on visitation, I estimate a Poisson fixed effects panel model using the equation:

$$\mathbb{E}(\text{visitation}_{iym}) = \exp(\beta \cdot \text{fireSize}_{iym} + \mathbf{X}'_{iym}\boldsymbol{\gamma} + \alpha_{iy} + \delta_{im} + \theta_{ym} + \epsilon_{iym}) \quad (2.1)$$

where the dependent variable visitation_{iym} denotes the monthly recreation visits to park i in a year y and month m . The variable of interest fireSize_{iym} is constructed as the total acres burned by all fires within a 50-mile radius of the park in the preferred specification. To address unobservable factors that are correlated with both wildfire and visitation, I include a key set of controls $\mathbf{X}_{iym}$. This set includes a vector of weather variables – ten bins of monthly average temperature, total precipitation, and average VPD – to account for the nonlinear effect of weather

on visitation found in prior studies (Fisichelli et al., 2015; Keiser et al., 2018). My estimates are robust to less flexible functional forms of weather. I also control for income-adjusted gasoline price to capture the variation in travel costs and income determinants of demand for national park visits, as previous studies provide evidence that visitation is positively correlated with personal income and negatively associated with fuel travel costs (Bergstrom et al., 2020).

Eq. (2.1) also includes a rich set of fixed effects, including park-by-year (α_{iy}), park-by-month (δ_{im}), and month-of-sample (θ_{ym}). Specifically, park-by-year fixed effects capture within-year variations in park-specific factors that determine recreation visits but are not captured by covariates, such as park entrance fees, free time for leisure and demographic characteristics. Park-by-month fixed effects control for seasonal unobservables across parks, such as changes in regional fire suppression efforts and differences in peak/off-season visitation. Lastly, month-of-sample fixed effects pick up the time-varying shocks that are common in each month, such as economic recessions and the rise of social media. In sum, these fixed effects allow me to compare a park to itself in the same month of the year as well as across years with different levels of monthly wildfire activity.

Because $visitation_{iym}$ is a count variable with non-negative values and positively skewed distributions, all panel models in my analysis are estimated using Poisson Pseudo Maximum Likelihood (PPML) rather than a log-linearized approach. As a robust alternative to log-linear regressions, PPML can handle incorrect inference due to heteroskedasticity¹ and gives consistent estimates of slope parameters without any distributional assumptions about the data (Silva and Tenreiro, 2006, 2011; Wooldridge, 2010). The coefficient of interest in the Poisson model β can

¹ I conduct modified Wald tests for groupwise heteroskedasticity on both $visitation_{iym}$ and log-transformed $visitation_{iym}$. The null hypothesis of homoskedasticity was rejected, suggesting that visitation series exhibits groupwise heteroskedasticity.

be interpreted as the percentage change in visits to park i resulting from an additional acre burned in proximity to the park. Because of the long time-series nature of my panel data, the error term, ϵ_{iym} , may exhibit month-over-month serial correlation within parks. To correct for autocorrelation in monthly visitation and wildfire size, all standard errors are clustered at the park level.

My identification strategy crucially relies on the assumption that unobserved determinants of visitation, ϵ_{iym} , are independent of variation in wildfire size, conditional on covariates and fixed effects included in eq. (2.1). However, such an assumption can be violated due to reverse causality induced by human-started fires, thereby biasing the estimates of β . There is strong evidence that the vast majority of fires in the US are caused by humans (Balch et al., 2017). If large wildfire events discourage visits while crowds of visitors increase the frequency of human-caused fires at the same time, a potential identification concern arises. I address this concern by restricting the fire sample to lightning fires, as the variations of natural fires are plausibly as good as random. I therefore ensure that preferred estimates are free from reverse causality.

Another primary challenge in consistent estimation is omitted-variable bias. My fixed-effect set will flexibly control for the omitted determinant of visitation that is either varying within a park annually (e.g., park entrance fees), recurring seasonal trends within park, or time-varying trends that are common across all parks. However, it's well established in the recreation demand literature that the availability of substitute and complementary recreational opportunities is an important determinant of demand for visits (Englin et al., 2008). Since the fire activities in park i and at nearby alternative sites can be highly correlated and both could impact visitation, failing to control for such spatial spillovers from neighboring recreation areas might introduce omitted-variable bias. My fixed-effects estimates are thus expected to pick up the potential spillover effects and under- or over-estimate the true effect depending on whether visitors perceive nearby parks as substitutes

or complements on average. To probe this possibility, in my preferred specification, I have included an additional control variable that is defined as the total fire size burned in all other nearby NPS units (not limited to national parks) within 80 miles of park i . I provide several robustness checks and show that my results are not confounded by the spatial spillover from nearby destinations.

2.4 Results

2.4.1 Visitation and burned acres

I find a strong and negative relationship between wildfire size and observed aggregate visitation. Table 2.1 presents PPML estimates of different versions of eq. (2.1) with alternative sets of three-way interacted fixed effects² and control variables. All models include park-by-year fixed effects. The model in column (1) includes park-by-fire-season³ and fire season-by-year fixed effects. Column (2) uses park-by-month and month-of-sample fixed effects to more flexibly account for common or individual shocks to the demand for western national parks. If the wildfire activities are confounded by time-varying unobservable factors, failure to control for them could lead to biased estimates. The inclusion of more granular fixed effects than those in column (1) should adequately mitigate such bias by fully absorbing variations in visitation caused by these confounders. Indeed, I find a slight drop in the magnitude of coefficient estimates on burned acres

² A growing literature (e.g., Callaway and Sant’Anna, 2021; De Chaisemartin and d’Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021) has identified issues with commonly used two-way fixed effects (“TWFE”) models. The primary concern is that the TWFE estimator is a weighted average of treatment effects for all observations and can apply negative weights in the presence of heterogeneous treatment effects, thus resulting in biased estimates of the average treatment effect. However, the existing studies haven’t yet reached a consensus on the methods to correct the potential bias in a complicated design with non-binary treatments switching on and off, which is the setting for my fire measures, $fireSize_{iym}$. Callaway et al. (2021) propose alternative estimation methods for difference-in-differences setups with a continuous treatment. Future versions of this paper will consider these newly developed methods to test the robustness of my estimates in the presence of heterogeneous treatment effects.

³ Fire season is defined from June to September, when wildfires are most likely to occur, spread and threaten recreational resources in the western US.

from column (1) to column (2), but the estimates stay negative and significant. However, the results are less precisely estimated with the most exhaustive fixed effect, suggesting that such attenuated estimates may suffer from reverse causality and omitted variable bias that persists even after conditioning on controls and fully interacted fixed effects.

I further test whether the effect estimated in eq. (2.1) is threatened by reverse causation. If considerable variation in wildfire events is directly correlated with excessive human use in the most visited areas, ignoring the potential endogeneity could lead to an underestimate of the damages caused by wildfires. Column (3) investigates this possibility by restricting wildfire size to the acres burned by lightning-caused fires. The coefficient estimate is substantially larger (1.5 times) than that in column (2), estimating that a thousand-acre increase in total lightning-caused fire size is associated with 0.063 percent monthly visitor loss. The larger point estimate could be a result of leveraging the presumably random variation in lightning fires, which burned three times larger and longer, on average, than human-caused fires. Most notably, this finding demonstrates the downward bias by reverse causality and implies that wildfire cause, which has been largely overlooked by previous studies, should be considered when looking at its effect on outdoor recreation⁴.

The final column of Table 2.1 presents the results from my preferred specification, which also accounts for the spatial spillovers from neighboring parks managed by NPS. The coefficient of lightning-caused fire size remains relatively stable after the inclusion of burned acres on all

⁴ I address endogeneity by restricting the sample to fires started by lightning. A common approach to deal with endogeneity is an Instrumental Variable (IV) approach. However, when I instrument for $fireSize_{iym}$ using the monthly counts of cloud-to-ground lightning strikes, the coefficient for wildfire size is imprecisely estimated and the IV regression cannot pass the weak instrument test (Kleibergen-Paap F statistics < 10). In addition, I experiment with multiple fuel aridity indexes suggested by Abatzoglou and Williams (2016) and find the IV estimates still suffer from weak instrument bias. One possible explanation is that wildfire is a complex ecological process influenced by many factors that affect its ignition, spread, and containment. It's difficult to identify a single variable that is strongly correlated with total fire size.

other NPS-managed units within 100 miles (“nearby parks” variable). The estimate reported in column (4) is not sensitive to alternative definitions of “nearby parks”, such as one using an alternative radius or including protected areas managed by USFS and Bureau of Land Management (BLM) (Appendix Figure A.1). Besides, my results suggest there is a lack of evidence of spillover effects in the nearby protected areas; the point estimates of neighboring parks’ burned acres are very close to zero and statistically insignificant across all specifications. Thus, the decrease in visitation is not driven simply by substitution or complementary effects – travelers are diverted away from the local park because of the changes in fire activities in neighboring parks.

I perform several additional robustness checks on my preferred specification—i.e., column (4) of Table 2.1. Appendix Table A.2 shows the robustness of my main results when using alternative model specifications, including the log-linearized and the inverse hyperbolic sine transformation (arcsinh) (Bellemare and Wichman, 2020) of my outcome variable⁵. My results are also robust to whether standard errors are two-way clustered at the park and year level, clustered at the park using a 500-times bootstrap, or adjusted for spatial correlation in the data (Conley, 1999). Appendix Table A.3 presents additional robustness checks. First, columns (1)-(2) explore the impact of different wildfire zone definitions used to count the total acres burned and find that using an alternative definition of total wildfire size yields consistent coefficient estimates. Note that the coefficient on wildfire size using both smaller wildfire zones is larger in magnitude than the 50-mile buffer zone in Table 2.1 for a smaller average wildfire size, indicating that nonlinearities may exist in the wildfire-visitation relationship. Second, columns (3)-(4) indicate these findings are also robust to alternative functional forms of weather controls. Overall, those

⁵ Although compared to PPML, log-linear and inverse hyperbolic sine models generate relatively smaller point estimates. The difference in the magnitudes of the estimates between log-linear and PPML is well documented in international trade literature. Based on the Monte Carlo simulation conducted by Silva and Tenreyro (2006), larger estimates are more likely to be found when the model is estimated by PPML.

estimates suggest that my results are not driven by confounding variation between wildfire activities and visitation. Taken together, I conclude that the average monthly visitation loss is 0.064% per thousand acres burned.

2.4.2 Temporal spillovers and nonlinear impacts

This subsection expands on the main results in two ways: estimating a dynamic version of eq. (2.1) with leads and lags of wildfire size and examining the nonlinear relationship between visitation and burned area from wildfires. Beyond the contemporaneous effect of wildfire activities on visitation reflected in Table 2.1, one might expect impacts to persist over time. I estimate a dynamic version of eq. (2.1) that includes three leads and lags of the acres burned by lightning-caused fires. This also serves as a placebo test of my identification strategy, as current month visitation shouldn't be affected by fire events that will happen in the future. The coefficient plot of leads and lags in Figure 2.3 depicts insignificant coefficients of all leads, which supports the identification assumption. On the other hand, I find a negative and significant effect of fires in the prior month that is similar in magnitude to the estimated contemporaneous effect. Visitors appear to delay their visits to the third month, but such a result could also be caused by a serial correlation in wildfire size, as I find a null effect of the second lag. These findings are consistent with those of Duffield et al. (2013) and Kim and Jakus (2019), who find the impact of wildfires carries over to the following month.

Next, I examine the nonlinearities in the visitation impacts of wildfire activity. Panel (a) of Figure 2.4 plots the results by regressing a binned version of eq. (2.1) rather than restricting the response to be linear. The blue dots denote coefficients of indicators for 40 bins of total fire size, and the shaded blue area corresponds to the 95% confidence interval. It should be noted that the reference bin is zero burned acres, represented by the first 24 bins that are combined into a single

bin. I also report a histogram of all bins at the bottom. The distribution of lightning-caused fire size within the 50-mile buffer around each park is highly skewed because of the frequent low levels of monthly fire activity during my study period. Only 11.5% of the park-month observations experience greater than 300 aggregate burned acres. The effect of wildfire fluctuates around zero and is imprecisely estimated through the lower 90th percentile (36th bin). Above the 90th percentile, however, the effect drops sharply. The visitation loss during the month with a wildfire size in the 97.5th percentile (40th bin) is 2.39%, which is about 10 times as much as the month in the 90th percentile. These findings do provide evidence of a nonlinear effect for wildfire size in the highest percentiles where wildfires burned at least 1000 acres.

Another way to demonstrate the damaging impact of large fires is to compare the visitation loss arising from a typical fire across different fire size categories. I first group wildfires by size class as defined by National Wildfire Coordinating Group (NWCG), and then simultaneously estimate the effects of the total burned area by each class. Multiplied by the average wildfire areas in each category, the results in panel (b) of Figure 2.4 are depicted side by side in a similar fashion to panel (a). The visitation loss associated with fires over 5000 acres (NWCG class-G) is more pronounced than fires in all other categories. A mean sized wildfire of class G will reduce monthly visitation by 2.43%. Although class G fires are rare, they are very destructive, having burned up to 83% of all wildfire acreage in the sample. Therefore, the reported estimates of the average response of visitation to wildfire size in Table 2.1 are most likely driven by large fires. This finding is consistent with the previous literature that finds that fire-induced visitation impacts depend on the intensity of fire events (Bawa, 2017). The profound effect of severe wildfires on visitation is likely explained by the fact that larger fires are more likely to affect the heavily visited areas in a

park regardless of their origin. And the most disastrous fires are also most likely to trigger park closures.

The results so far show that the effect of wildfires depends on their sizes. I now consider how this effect might vary with its distance from the park centroid. Figure 2.5 shows the results from simultaneously estimating the effect of the total fire size within each 5-mile distance bin out to a distance of 50 miles. For example, the first blue bar represents the impact of total burned area in the five-mile buffer zone. Despite the clear evidence of distance decay beyond 15 miles, the impact of acres burned within 5 – 15 miles of the park center is about two times higher than that of the total area of fires burning less than 5 miles from the park center. One possible explanation for this result is that the 5-mile buffers around park centroids encompass relatively remote areas within parks that tend to be far away from the most heavily visited and developed areas, since visitor centers, main roads, and campgrounds are often located near the park boundaries.

2.4.3 Channels behind the findings

As noted above, fires burning 5000 acres or larger and those burning within 5 – 15 miles of a park's geographical center have the largest effects on monthly visitation. One possible explanation for these findings is that severe wildfires, even those originating from remote locations, are often accompanied by high winds and are more likely to quickly spread to heavily traveled areas, thus prompting evacuation and park closures. Additionally, fires starting within 5 – 15 miles of the park centroid have a high chance of burning near park facilities like visitor centers and headquarters and forcing the closure of the main road used to enter the park. Figure A.2 plots the distribution of the distances from visitor centers to park centroids. On average, a visitor center in my sample is 12.27 miles away from the park centroid, which happens to fall inside the 5 – 15 mile bin. The presence of a fire close enough to visitor centers or to the main access road to the facilities would

likely close access to the surrounding area and travel networks. That is, the previous results could be explained if the fire-caused closure is one of the main channels of visitation loss. As such, I now examine the relative importance of the channels behind the fire-induced visitation loss, especially the degree to which visitors respond to emergency fire closures, even when the park is only partially closed.

It is hypothesized that the closure channel should account for the majority of lost visitors in comparison to a fire scenario in which full access to the park is granted but underlying fire risks are present. I test this hypothesis by estimating a version of eq. (2.1) where the variable of interest is replaced by a proxy for “fire closure days”⁶, as described in Section 2.2.1. I also control for “fire days,” which account for the number of days when wildfires are burning either too small in scale or too far away from the populated areas to cause any closure events. The estimation equation is:

$$\begin{aligned} \mathbb{E}(\text{visitation}_{iym}) = & \exp (\beta_1 \cdot \text{closureDays}_{iym} + \beta_2 \cdot \text{fireDays}_{iym} \\ & + \mathbf{X}'_{iym}\boldsymbol{\gamma} + \alpha_{iy} + \delta_{im} + \theta_{ym} + \epsilon_{iym}) \end{aligned} \quad (2.2)$$

where the coefficient of interest β_1 now describes the effect of an additional day of fire closure on monthly visits. Note that closureDays_{iym} and fireDays_{iym} are both measured in days so that β_1 and β_2 can be directly comparable. All other covariates and fixed effects in the preferred specification from eq. (2.1) are included.

The results are shown in Table 2.2. A comparison of the first row and second row in column (1) suggests that fire closures clearly result in relatively sizable reductions in visitation. The coefficients on closureDays_{iym} and fireDays_{iym} both have the expected sign (negative) and are statistically significant. In terms of the magnitude of the estimated effect, the coefficient of fire

⁶ Currently, the closure data is only available for western national parks receiving more than one million annual visits (excluding parks in California). These parks include Arches, Bryce, Capital Reef, Canyonlands, Crater Lake, Glacier, Grand Canyon, Grand Teton, Mount Rainier, Olympic, Rocky Mountain, Yellowstone, and Zion national parks.

closure days is at least a magnitude larger than that of fire days. The difference is even greater if I adjust the calculation of fire days to exclude dates with closure events (column (2)) and if I use an alternative definition of closure days that includes only the closures caused by lightning fires (column (3)). Focusing on column (3), an additional day of fire closures in a month would reduce visitation by 2.27 percent, while the effect is 10 times smaller – 0.24 percent – for each fire day without any consequential closure in place⁷. Taken literally, these estimates highlight that lack of access due to emergency closures is indeed the main channel to explain fire-induced visitation impacts. As more than half of the fires resulting in closures burned more than 5000 acres, the substantial impacts of fire closure further confirm my previous finding of the nonlinear effects of wildfire size. Interestingly, the significant effects of fire days indicate that visitors also respond to risk exposure of relatively small or remote wildfires, even if they did not cause any fire closure.

2.4.4 Smoke effect

Increasing wildfire activities are often accompanied by a number of days of smoke and haze in the air. Travelers could still change their travel plans due to smoke-related health concerns, even when there is no fire closure imposed in the vacation region. Many of the previous studies on air pollution in scenic areas have shown that the public places great value on preserving visibility (Schultze et al., 1983; Chestnut and Rowe, 1990; Smith et al., 2005) and is more concerned about the health impacts of increased haze and smoke from wildfires, particularly for less healthy individuals (Haider et al., 2019; Thapa et al., 2013). As a result, this concern may be intermingled with the influence of other wildfire outcomes, including closures and high fire risk in the destination park, making it difficult to distinguish the independent smoke effects. As recent research utilizing satellite-based smoke data indicates that wildfire smoke can travel long distances from the source

⁷ I also conduct a placebo test by adding leads and lags of fire closure days and find that the effect of fire closures does not persist over time (Appendix Figure A.4).

fires (Rolph et al., 2009; Miller et al., 2017; Borgschulte et al., 2022), considerable global externalities could be associated with wildfires—namely, the externalities that arise if visitors respond to smoke and haze coming from fires thousands of miles away. If distant wildfire smoke creates substantial global externalities compared to the effects of local fires, little could be done for the local park managers to mitigate the damaging impact of smoke without partnering with other agencies.

This subsection considers this specific externality and estimates how it affects visitation, given the presence of local fires. To the best of my knowledge, this is the first study attempting to isolate the smoke effects in evaluating the impact of wildfires on outdoor recreation. To explicitly separate smoke effects from the influence of local fire activity, the estimation equation is specified in a manner similar to eq. (2.2) by replacing $closureDays_{iym}$ with the smoke measure:

$$\begin{aligned} \mathbb{E}(visitation_{iym}) = & \exp(\beta_1 \cdot fireDays_{iym} + \beta_2 \cdot smokeDays_{iym} \\ & + \mathbf{X}'_{iym}\boldsymbol{\gamma} + \alpha_{iy} + \delta_{im} + \theta_{ym} + \epsilon_{iym}) \end{aligned} \quad (2.3)$$

where $smokeDays_{iym}$, my measure of smoke exposure, is the number of days in which the park is covered by smoke, as described in Section 2.2.1. The results are displayed in Table 2.3. Columns (1) and (2) estimate the effect of fire days and smoke days separately, while column (3) estimates them simultaneously. The estimated effect of fire days is always significant and is not sensitive to the inclusion of smoke days in the model. However, the coefficient of smoke days is imprecisely estimated in column (3), suggesting that visitors, on average, are not responsive to wildfire smoke.

However, two major concerns might arise regarding the smoke measurement. The first concern is that the smoke effect estimated by simply adding the smoke variable in the regression might be confounded with the direct influence of local wildfire burning. This might be less of a concern if the measures of fire and smoke are not correlated. Yet my data reveals that the number

of fire days and smoke days are highly correlated ($\text{corr}(\text{fireDays}_{iym}, \text{smokeDays}_{iym}) = 0.549$). To provide a sense of the correlation between fire and smoke measures, Figure 2.6 further explores a few of their key features. Panel (a) compares the time series of the monthly number of fire days and of smoke days for the entire period, 2006 – 2019. As can be seen, two series move up and down together, and the number of smoke days greatly exceeds the number of fire days for most of the months. Because the fire variable is designed to only capture local wildfire activities, such a difference echoes findings from previous studies that suggest a significant share of smoke plumes come from distant sources. Smoky days usually come earlier in the spring and peak in mid-summer, and they may stay in the region until the end of the year. In panel (b), I break the smoke days into two components: non-fire smoke days and smoke days with active local fires. On average, non-fire smoke days (blue area) contribute to more than 80% of the total. In contrast, when I break down the fire days into two components in a similar manner in Panel (c), I find that the majority of fire days also experience smoke (red area). Overall, these findings imply that fire days are likely a subset of smoke days, and it is important to focus on the effect of non-fire smoke days to avoid attributing the direct effect of wildfire burn to the effect of distant smoke.

To address the issue of the high correlation between fire and smoke measures, I re-estimate eq. (2.3) by restricting the calculation of smoke days to dates with no active local fires. Thus, the coefficient estimates of β_2 can be interpreted as the effect of global externalities generated by wildfire smoke. The results are shown in column (4) of Table 2.3. Compared to column (3), the coefficient of smoke days is still small and statistically insignificant, while the estimated effect of fire days is fairly stable. These results indicate that the null effect of smoke days is not driven by the collinearity between fire and smoke measures.

Another potential concern is the measurement error in the satellite-based smoke data, as the extent of smoke plumes only approximates the areas with heavy smoke. The detection of any plume is based on the elevated smoke concentration in the atmospheric column well above the surface. The presence of a smoke plume covering the park does not necessarily indicate a strong increase in the surface air quality (Rolph et al., 2009; Ford et al., 2017; Brey et al., 2018; Burkhardt et al., 2019). Consequently, simply overlaying the smoke polygons with the park boundary to determine if the park is exposed to smoke may overcount the number of smoke-impacted days. To address this measurement error, I employ a procedure adapted from Brey and Fischer (2016), who refine the smoke measure using ground-based pollution readings. Specifically, I first obtain data on ozone concentration (O3)⁸ from the Environmental Protection Agency (EPA)'s Air Quality System (AQS) database and calculate the park-specific seasonal mean and standard deviation of ozone for days with no overlapping HMS smoke plume. Next, I define a park as smoke-impacted on a given day if (i) any part of it intersects smoke plumes on that day, and (ii) the ozone concentration for that day is more than one standard deviation above the park-specific seasonal mean. By restricting the definition of smoke days, I am able to reduce the possibility of misclassifying a day with little surface smoke concentration as smoke-impacted, thereby reducing the measurement error.

Using this refined measure of smoke days, I replicate Table 2.3 and report the result in Table 2.4. Column (4) of Table 2.4 corresponds to my preferred estimates, as it addresses two aforementioned concerns and reports the coefficient estimates of non-fire smoke days, which are

⁸ I use ozone as the primary measure of surface-level air pollution because it is the most widely monitored pollutant in national parks and has daily measurements available for most of the parks (> 22 parks) in my sample. These data preserve my sample to the largest extent possible, whereas PM2.5 measures are available for only 9 parks and focus on urban locations, and visibility measures from the Interagency Monitoring of Protected Visual Environments (IMPROVE) program only report data every third day or every sixth day.

adjusted by O3. Comparing the preferred estimates with the initial estimates (column (3) of Table 2.3), while the estimated effect of fire days slightly decreases, I find that it remains highly significant. On the other hand, the magnitude of the smoke effect drastically drops. The results after addressing two concerns lend further credence to my initial findings that park visitors, on average, do not respond to smoke from distant sources to a statistically significant degree. I found similar results using different methods to adjust the smoke day (see Appendix Table A.4). Therefore, the fire-induced visitation loss is primarily attributed to local fires.

2.4.5 Local economic impacts of wildfires

I report the relationship between local wildfire activity and visitation in Sections 2.4.1– 2.4.2 and provide evidence that visitation is not responsive to smoke from a remote source in Section 2.4.4. Based on those findings, this section provides back-of-the-envelope estimates for the implied change in visitation and economic loss due to wildfires for the local economy⁹. To account for the within-year substitution pattern, the estimated concurrent effect (-0.045% per thousand acres) and lagged effect (-0.046% per thousand acres) are both used to calculate the annual visitation loss. The mean annual visitation loss is computed as the product of the visitation reduction effect of wildfires, park-level annual mean wildfire size, and the baseline annual visitation of each park.

I find that the visitation loss for an average fire year is 717,013 visits or 1.35% of the annual average attendance to 32 parks in my sample. Figure 2.7 shows the heterogeneity in the loss of visitation across parks. It's noteworthy that these estimates assume the same elasticity of visitation with respect to wildfire size across different parks, so the variation in the park-specific visitation losses stems from the difference in the mean wildfire size and the baseline visitation of each park. Correspondingly, the most affected parks happen to be the most heavily visited parks with a high

⁹ Local gateway regions are defined as all counties within or intersecting a 60-mile radius around each park boundary.

level of wildfire activities. In particular, Yosemite NP has the largest visitation loss, which amounts to 3.51% of the annual average Yosemite visitation.

I compare my estimates of visitation loss to the numbers reported by three previous studies. For Yellowstone NP, Duffield et al. (2013) estimate that the visitation loss due to wildfires in a typical year for the period 1986 – 2011 is 1.3%, which aligns closely with my estimate for Yellowstone NP (1.59%). A comparison with Kim and Jakus (2019) reveals that my estimates for Arches, Canyonlands, and Capitol Reef NPs are about 2 times bigger than their estimates, but I obtain a slightly smaller visitation loss for Bryce Canyon NP. Such a discrepancy might be due to Kim and Jakus’ decisions to trim the data to only include summer months and rely on park-level time series analysis instead of the fixed-effect strategies employed in this study.

A much more meaningful comparison should be with a recent study, Gellman et al. (2022), using an empirical strategy similar to mine. Instead of using NPS’s official visitation count data, Gellman et al. (2022) utilize rather new daily data on campground use from Recreation.gov from 2008 to 2017. They estimate that around 1.39 million annual visits are affected by wildfires and smoke in national parks¹⁰, which is approximately one time larger than my estimate (717,013 visits per year¹¹). There are at least two reasons why Gellman et al. (2022) may overestimate the wildfire impacts on the affected visitor population: First, campers, as a specific sub-population of all national park visitors, might have a substantially larger elasticity of demand with respect to burned acres, which has been documented by Duffield et al. (2013). I re-estimate my preferred model specification in Section 2.4.1 by restricting the NPS total visitation data to the “backcountry users

¹⁰ Gellman et al. (2022) estimate the visitation loss by multiplying the total number of campers affected by fires and smoke at each national park by a ratio of all national park visitors to overnight users at each park for a given month. Therefore, they scale the relationship learned from a narrow group of visitors to the general population of all park visitors.

¹¹ The standard error associated with my estimate of 717,013 visits is 280,565. Gellman et al. (2022) didn’t included their standard error in the their study.

and campers” category as the dependent variable, which is a more comparable subsample to campers in Gellman et al. (2022). Using the restricted sample, I find greater responsiveness of backcountry users and campers to the change in wildfire size: one thousand acres of increase in wildfire size decreases overnight visits by 0.104%, compared to my original estimate of 0.064% for total visitation. If I use the marginal effect estimated based on the restricted sample in the back-of-the-envelope calculation, an impact of 1.17 million visitation losses due to wildfire would be estimated, which is much closer to their estimate of 1.39 million. Furthermore, compared to day users, overnight campers may also be more sensitive to bad air quality due to high smoke exposure, given their many hours spent in the backcountry. And I demonstrate that national park visitors, comprised of both day users and campers, are surprisingly resilient to smoky days. Combining these two factors into consideration leads me to believe that running an analysis on campers alone might overstate the visitation loss attributed to wildfire activities. On the other hand, my approach based on the entire population of all national park visitors, on the other hand, provides more representative estimates of visitation loss across all types of recreation users and regions.

Is a loss of 700,000 visitors per year economically meaningful to warrant a change in mitigation planning? The answer to this question is location-specific. To demonstrate this point, I convert my estimates of park-specific visitation loss to the park’s monetary contribution to the local economy (including visitor spending, labor income, and total economic output) based on the multipliers estimated by NPS. My estimates imply a \$66.71 million reduction in visitor spending, a \$29.61 million loss in labor income, and a total loss of \$86.22 million in local economic output per year. Among all the local communities around the national parks, the most vulnerable local economies appear to be the ones that heavily rely on tourism, for example, the gateway communities around Redwood, Crater Lake, and Glacier NPs. In those communities, the direct

loss in labor income and economic output due to the wildfire-caused visitor reduction could account for up to 5%, which translates to an annual loss of about \$2,200 to local workers. Furthermore, it is worth noting that such an estimate could likely represent a lower bound of the total economic losses to local communities because it does not account for region-wise and industry-wise spillover effects. Last but not least, a complete cost-benefit assessment should incorporate the wildfire damage to property and ecosystems, as well as the estimated costs of firefighting. However, such an estimation is beyond the scope of the current study.

2.5 Conclusion

This paper presents a large-scale analysis of the impacts of wildfires on visitor use in national parks across the western US. I compile a comprehensive dataset on wildfire occurrences, smoke, and visitation from several sources at the monthly-park level spanning from 1993 to 2019. Building on prior research evidence, I perform a series of tests to address the concern of reverse causality and ensure my results are not confounded by spatial spillovers from neighboring recreation areas. I consistently find a statistically significant and negative effect of burned area on aggregate monthly visitation. My preferred estimates show that an additional thousand acres burned are associated with a 0.064% decrease in the current month visitation. I also find evidence of persistent impacts on the subsequent month's visitation and prominent impacts for large fires and/or fires burning in the heavily visited areas.

I then highlight the importance of the main channel behind the above findings: emergency fire closures, which account for the majority of visitation loss and have impacts nearly 10 times larger than fires not causing any closures. I show that an additional ten days of park closures due to wildfires will lead to a total visitation loss of 22.73% per month. My results support and extend

previous literature on assessing the non-market value of access to recreation sites in a regional context.

I also present estimates of global externalities associated with wildfire smoke that, to my knowledge, is the first paper to disentangle the non-local smoke effects from the local fire effects. I find that without the presence of local fires, the distinct impact of smoke from remote sources on visitation is surprisingly small. This is likely due to the fact that, after the past five years of severe fires with popular parks getting overcrowded, visitors would rather continue their plans even under the smoky conditions than forego a visit and cancel their bookings made a year ago. Unless threatened by active local fires and disrupted from accessing the park, visitors are willing to travel despite the heavy smoke from distant sources.

Overall, my results underscore the importance of local efforts in combating the damaging impacts of wildfires on national park gateway communities, as the climate warms. To mitigate the visitation loss due to mandatory evacuations and park closures, it's crucial to factor in the potential economic impacts of reduced visitors on local tourism into the wildfire management plan. For example, park managers can promote the less-visited areas and prioritize wildfire prevention and suppression efforts on “visitor hubs” early in the season. In addition, the small effect of smoky days indicates there may be considerable benefits from the use of off-season prescribed burning.

2.6 Tables

Table 2.1 Effect of burned acres on monthly recreation visits

	(1)	(2)	(3)	(4)
Fire size within 50 miles (thousand ac)	-0.0527* (0.0275)	-0.0413 (0.0270)	-0.0634** (0.0291)	-0.0636** (0.0292)
Observations	8,736	8,736	8,736	8,736
Number of Parks	32	32	32	32
Mean fire size (thousand ac)	2.312	2.312	1.687	1.687
Controls	Yes	Yes	Yes	Yes
Park × Year FE	Yes	Yes	Yes	Yes
Park × Fire Season FE	Yes			
Fire Season × Year FE	Yes			
Park × Month FE		Yes	Yes	Yes
Month-of-Sample FE		Yes	Yes	Yes
Lightning fires only			Yes	Yes
Nearby Parks				Yes

Note: This table reports the effect of the total burned acres (in thousand acres) on monthly recreation visits from Poisson pseudo-maximum-likelihood estimates. All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. All regressions include 10 bins of monthly mean temperature, monthly precipitation, and monthly mean vapor pressure deficit (VPD) as well as income-adjusted gasoline price. Fixed-effects strategies and additional controls are listed at the bottom of this table. Column (4) is the preferred specification. Standard errors are clustered at the park level. ***p<0.01, **p<0.05, *p<0.1.

Table 2.2 Effect of fire closures on monthly recreation visits

	(1)	(2)	(3)
Fire closure days	-1.856*** (0.358)	-2.007*** (0.330)	-2.273*** (0.300)
Fire days	-0.251*** (0.0942)	-0.247*** (0.0949)	-0.241*** (0.0923)
Observations	1,560	1,560	1,560
Number of Parks	13	13	13
Mean fire closure days	0.0677	0.0677	0.0477
Mean fire days	0.828	0.787	0.793
Controls	Yes	Yes	Yes
Park $\times$ Year FE	Yes	Yes	Yes
Park $\times$ Month FE	Yes	Yes	Yes
Month-of-Sample FE	Yes	Yes	Yes
Lightning fires only			Yes

Note: This table reports the effect of monthly fire closure days and fire days on monthly recreation visits from Poisson pseudo-maximum-likelihood estimates of eq. (2.2). All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. All regression includes 10 bins of monthly mean temperature, monthly precipitation and monthly mean vapor pressure deficit (VPD) as well as income-adjusted gasoline price. Fixed-effects strategies and additional controls are listed at the bottom of this table. Column (2) instead uses the adjusted fire days, which exclude dates with closure events in the calculation. ***p<0.01, **p<0.05, *p<0.1.

Table 2.3 Effect of wildfire smoke on monthly recreation visits

	(1)	(2)	(3)	(4)
Fire days	-0.370*** (0.0876)		-0.433*** (0.104)	-0.338*** (0.108)
Smoke days		-0.0920 (0.0615)	0.109 (0.0734)	0.111 (0.0796)
Observations	5,376	5,376	5,376	5,376
Number of Parks	32	32	32	32
Mean fire days	0.656		0.656	0.656
Mean smoke days		2.400	2.400	1.956
Controls	Yes	Yes	Yes	Yes
Park $\times$ Year FE	Yes	Yes	Yes	Yes
Park $\times$ Month FE	Yes	Yes	Yes	Yes
Month-of-Sample	Yes	Yes	Yes	Yes
Non-fire smoke days only				Yes

Notes: This table reports the effect of monthly fire days and smoke days on monthly recreation visits from Poisson pseudo-maximum-likelihood estimates of eq. (2.3). All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. The “smoke days” is calculated as the number of days in which the park intersects any smoke plume. Column (4) instead uses the non-fire smoke days which exclude dates with active local fire in calculation. All regression includes 10 bins of monthly mean temperature, monthly precipitation, and monthly mean vapor pressure deficit (VPD) as well as income-adjusted gasoline price. Fixed-effects strategies and additional controls are listed at the bottom of this table. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2.4 Effect of wildfire smoke on monthly recreation visits (Adjusted by ozone)

	(1)	(2)	(3)	(4)
Fire days	-0.370*** (0.0876)		-0.367** (0.159)	-0.368*** (0.134)
Smoke days		-0.204 (0.149)	-0.0201 (0.170)	0.0399 (0.238)
Observations	5,376	3,381	3,381	3,381
Number of Parks	23	23	23	23
Mean fire days	0.656		0.656	0.656
Mean smoke days		0.833	0.833	0.651
Controls	Yes	Yes	Yes	Yes
Park $\times$ Year FE	Yes	Yes	Yes	Yes
Park $\times$ Month FE	Yes	Yes	Yes	Yes
Month-of-Sample	Yes	Yes	Yes	Yes
Non-fire smoke days only				Yes

Notes: This table reports the effect of monthly fire days and smoke days on monthly recreation visits from Poisson pseudo-maximum-likelihood estimates of eq. (2.3). All coefficient estimates are multiplied by 100 to demonstrate the effect in percentage points. The “smoke days” is calculated as the number of days in which (i) the park intersects any smoke plume, and (ii) the ozone concentration for that day is more than one standard deviation above the park-specific seasonal mean. Column (4) instead uses the non-fire smoke days which exclude dates with active local fire in calculation. All regression includes 10 bins of monthly mean temperature, monthly precipitation, and monthly mean vapor pressure deficit (VPD) as well as income-adjusted gasoline price. Fixed-effects strategies and additional controls are listed at the bottom of this table. ***p<0.01, **p<0.05, *p<0.1.

2.7 Figures

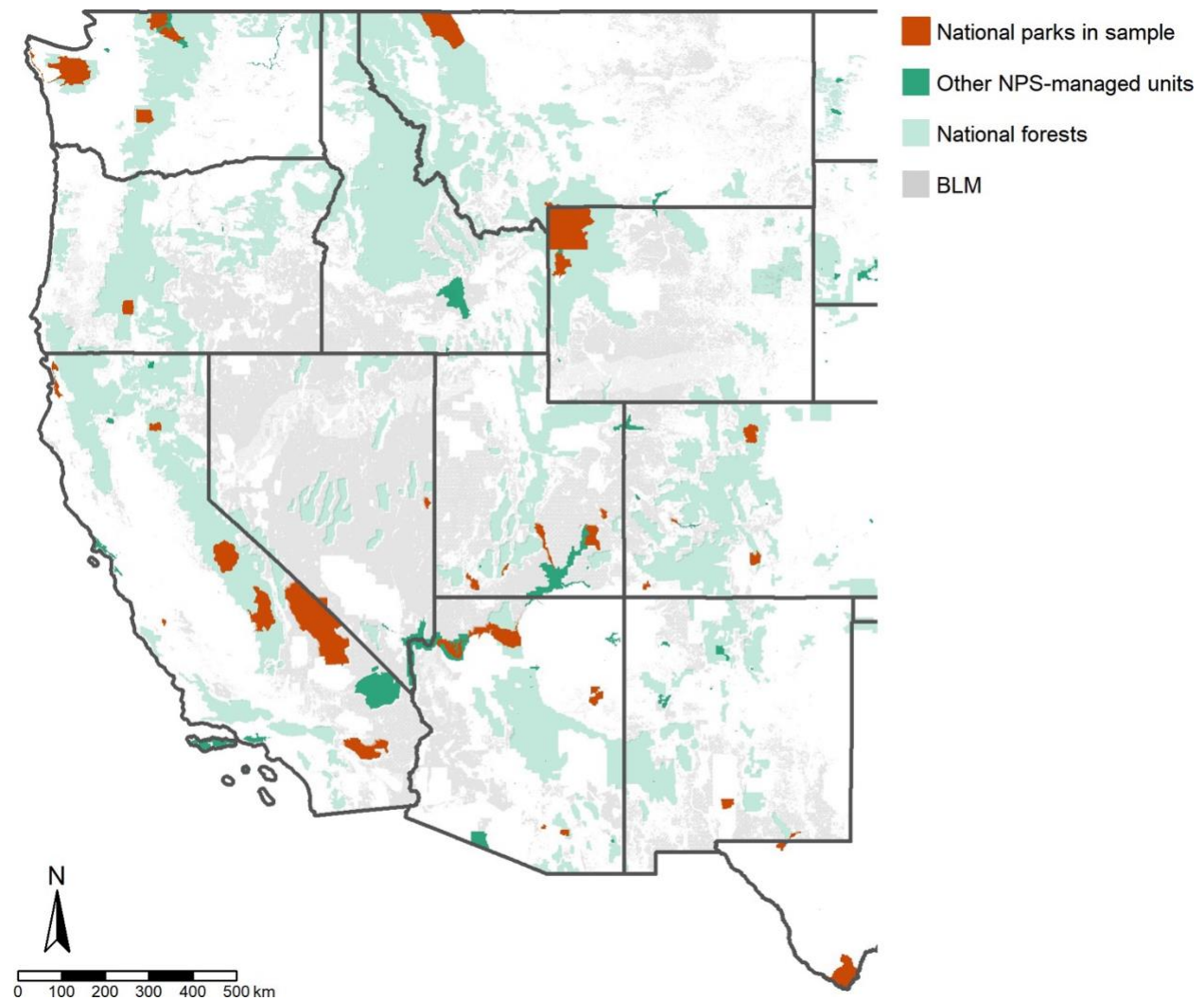


Figure 2.1 Locations of 32 western national parks in my sample

Note: This figure plots the locations of the 32 western national parks considered in this study.

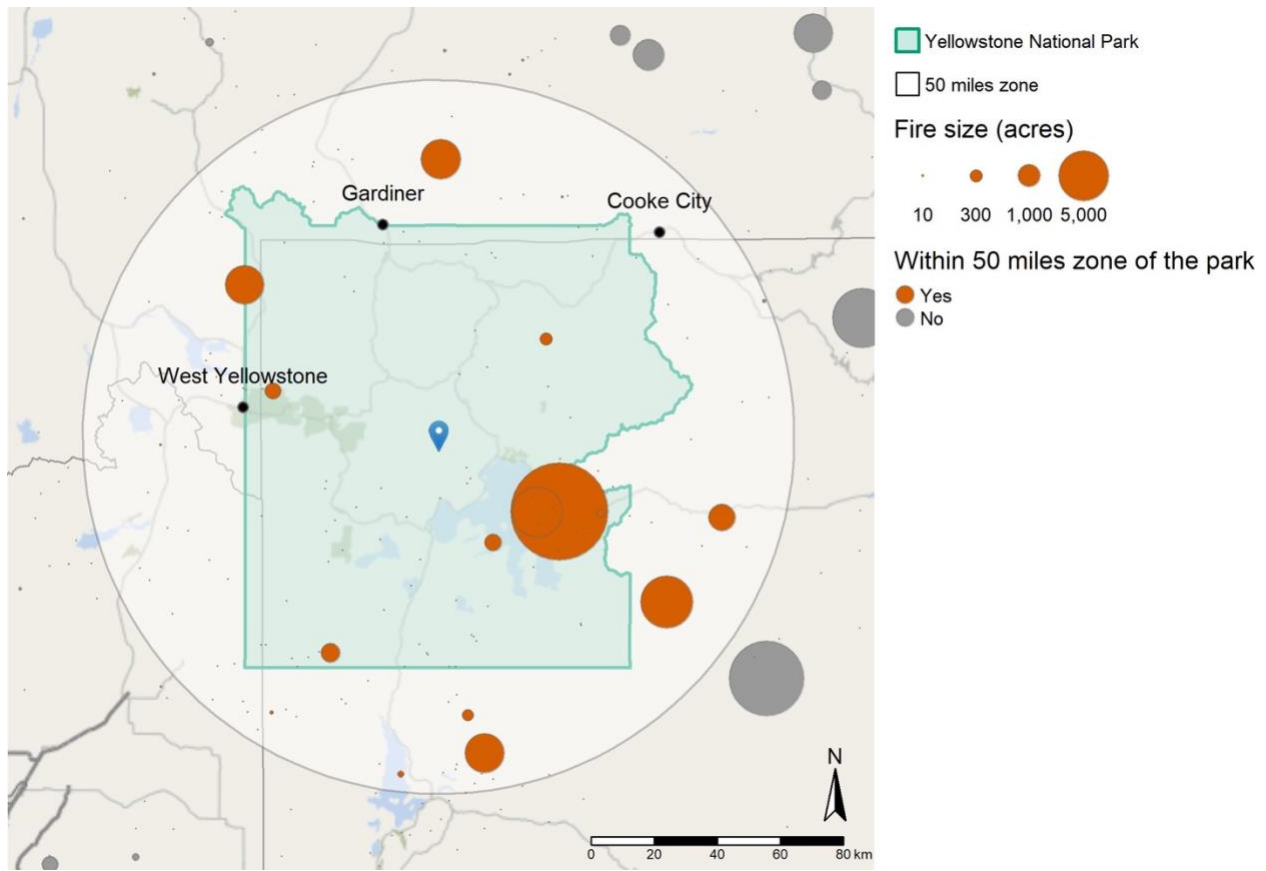


Figure 2.2 Illustration describing the construction of the total fire size

Note: This figure illustrates all wildfire events within a 50-mile radius around the centroid of Yellowstone National Park in September 2003. Each red bubble represents a wildfire event. The location of the bubble shows the ignition point, and the size of the bubble denotes the size of the wildfire event in terms of the total burned area within the final perimeter.

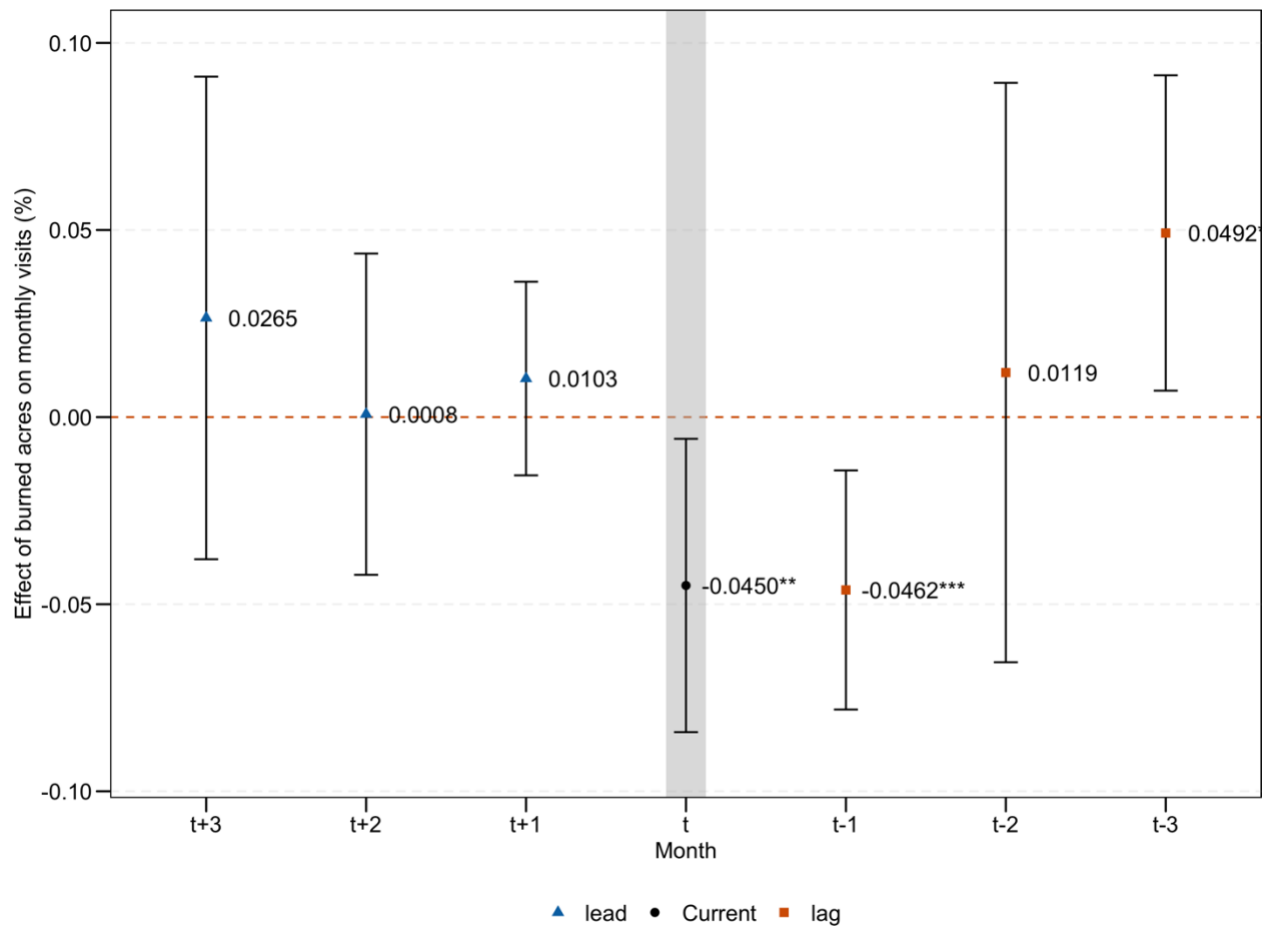


Figure 2.3 Estimated effects of leads and lags

Note: This figure plots estimated coefficients from a lead-and-lag test where the independent variables are the wildfire sizes in month $t + k$ ($k = 3, 2, 1, 0, -1, -2, -3$). The whiskers represent the 95% confidence intervals based on standard errors clustered at the park level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The insignificant coefficients of all leads suggest the estimation strategy passes the placebo test. The significant coefficient for the first lag term means the impact of wildfire on visitation may persist in the following month.

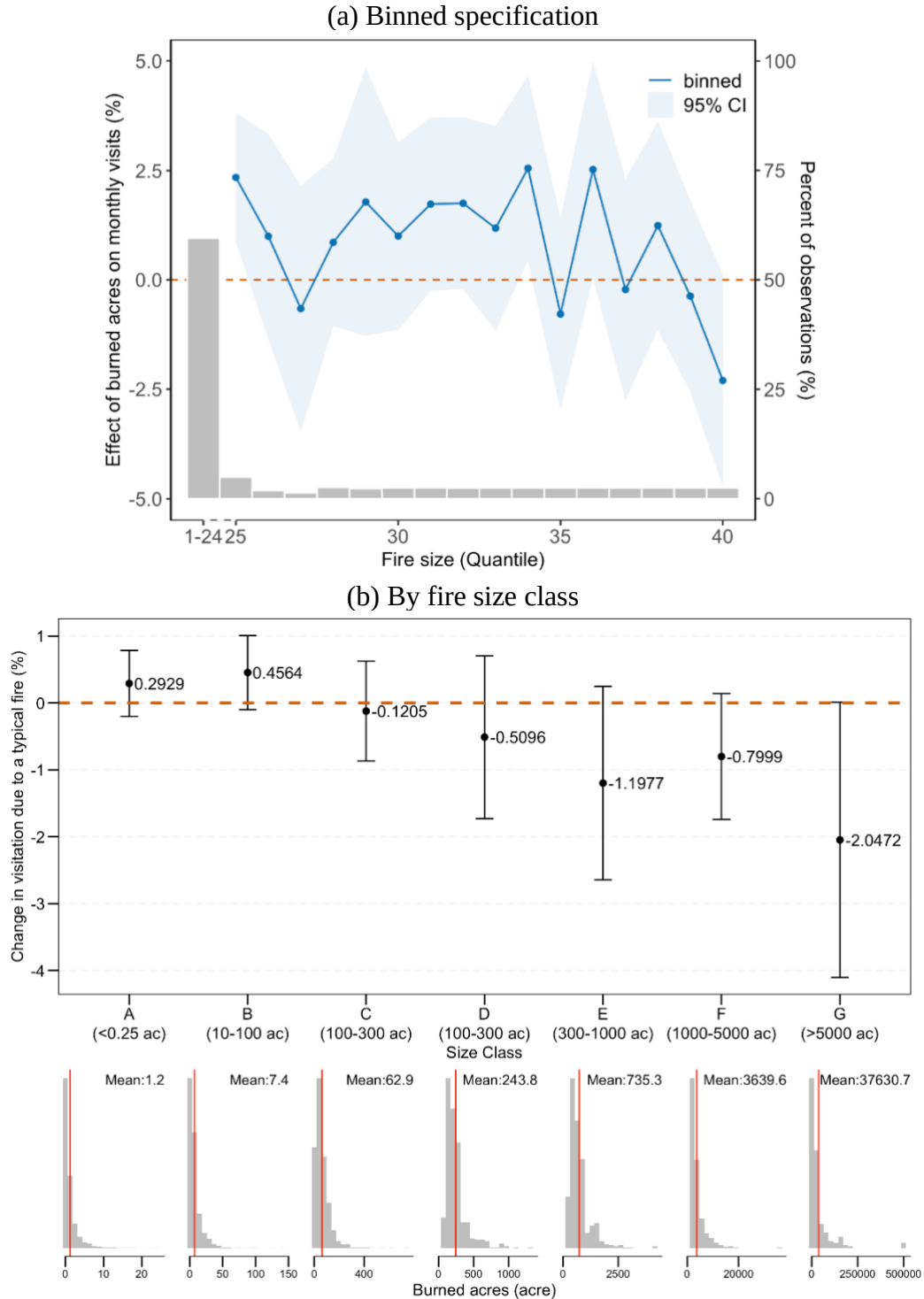


Figure 2.4 Non-linear effects of burned acres on visitation

Note: Panel (a) shows the estimated coefficients of indicators for 16 bins of total fire size. The shaded blue area is the 95% confidence interval, where the standard errors are clustered at the park level. The grey histogram shows the proportion of observations in each bin. Panel (b) shows the point estimates and 95% confidence interval of the response of visitation to a mean wildfire in size class A – G. The distributions of burned acres and average wildfire size in each size class are shown at the bottom.

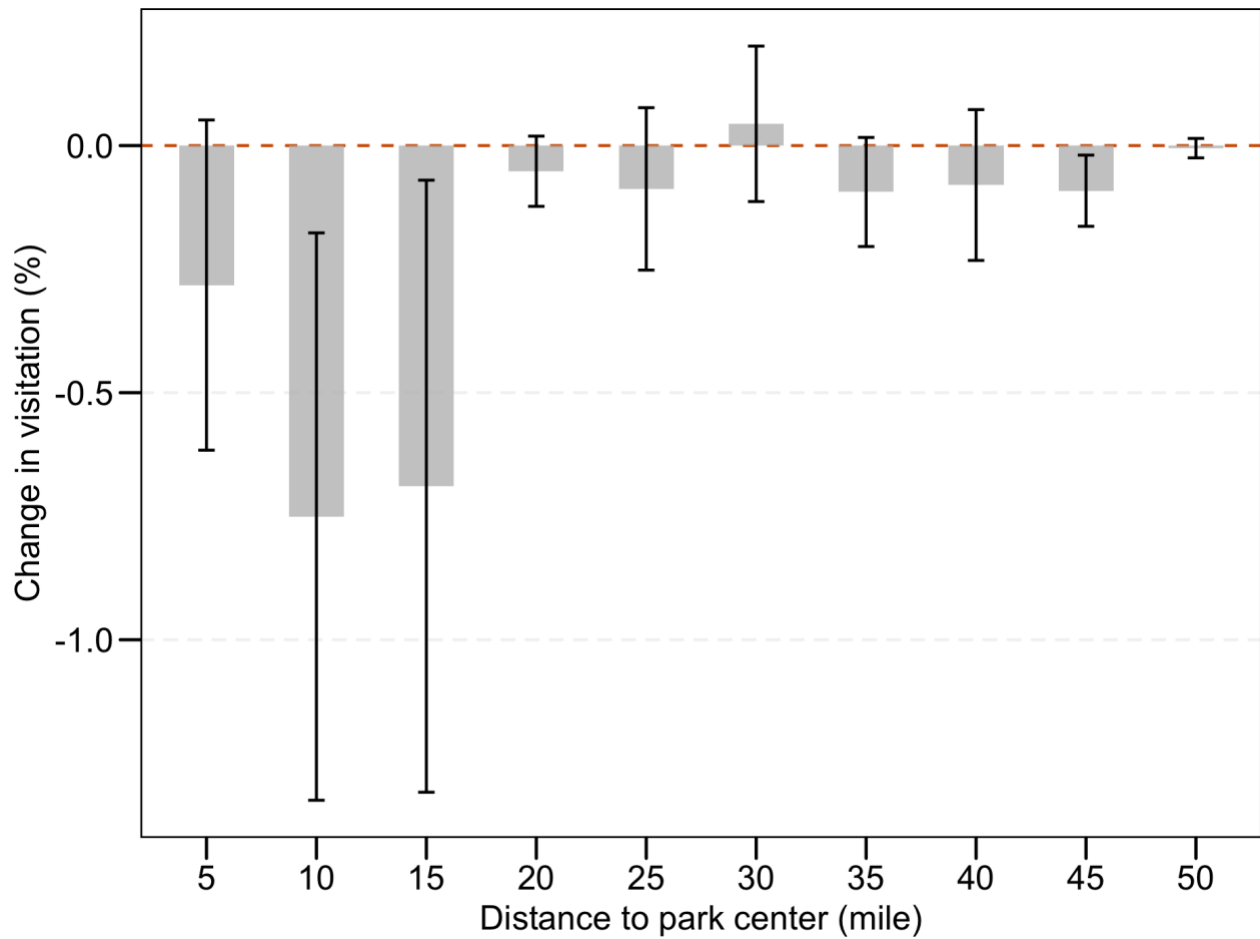


Figure 2.5 The effect of burned acres by distance to the park center

Note: This figure plots the point estimates and 95% confidence intervals of the marginal effects of total fire size within certain distance bins to the park's geographic centroid. These marginal effects are estimated by regressing visitation on the total fire size within each 5-mile distance bin. Distance is shown on the x -axis, such that the point estimates located at $x = 10$ can be interpreted as the impact of the total burned area within 5 – 10 miles from the park center.

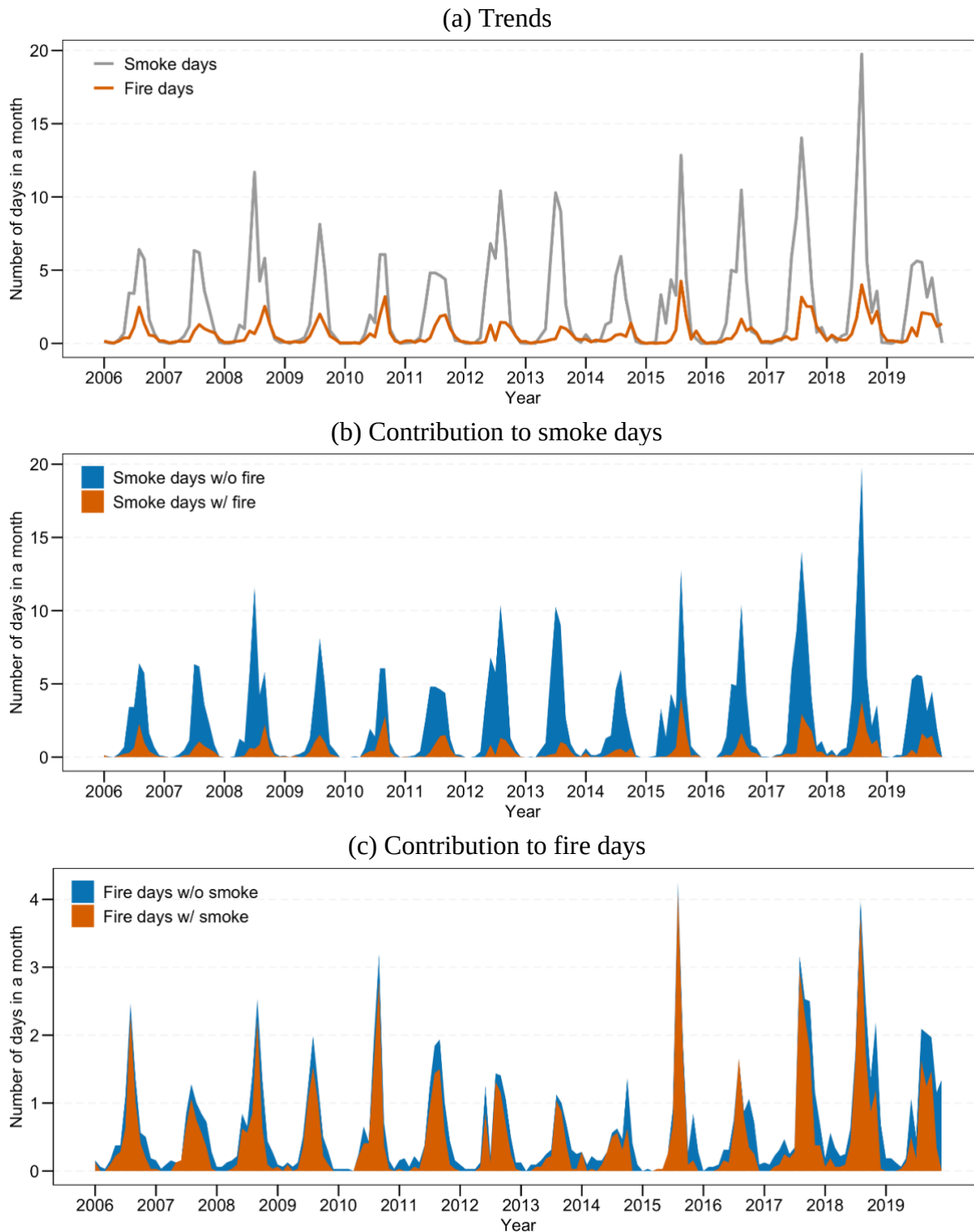


Figure 2.6 Monthly fire days and smoke days with their components

Note: Panel (a) plots the trends in the number of fire days in each month and the number of smoke days in each month from 2006 to 2019. Panel (b) decomposes the number of smoke days into two components: the number of non-fire smoke days (blue area) and the number of smoke days with active local fires (red area). Panel (c) decomposes the number of fire days into the number of fire days without smoke (blue area) and the number of smoke days with smoke (red area).

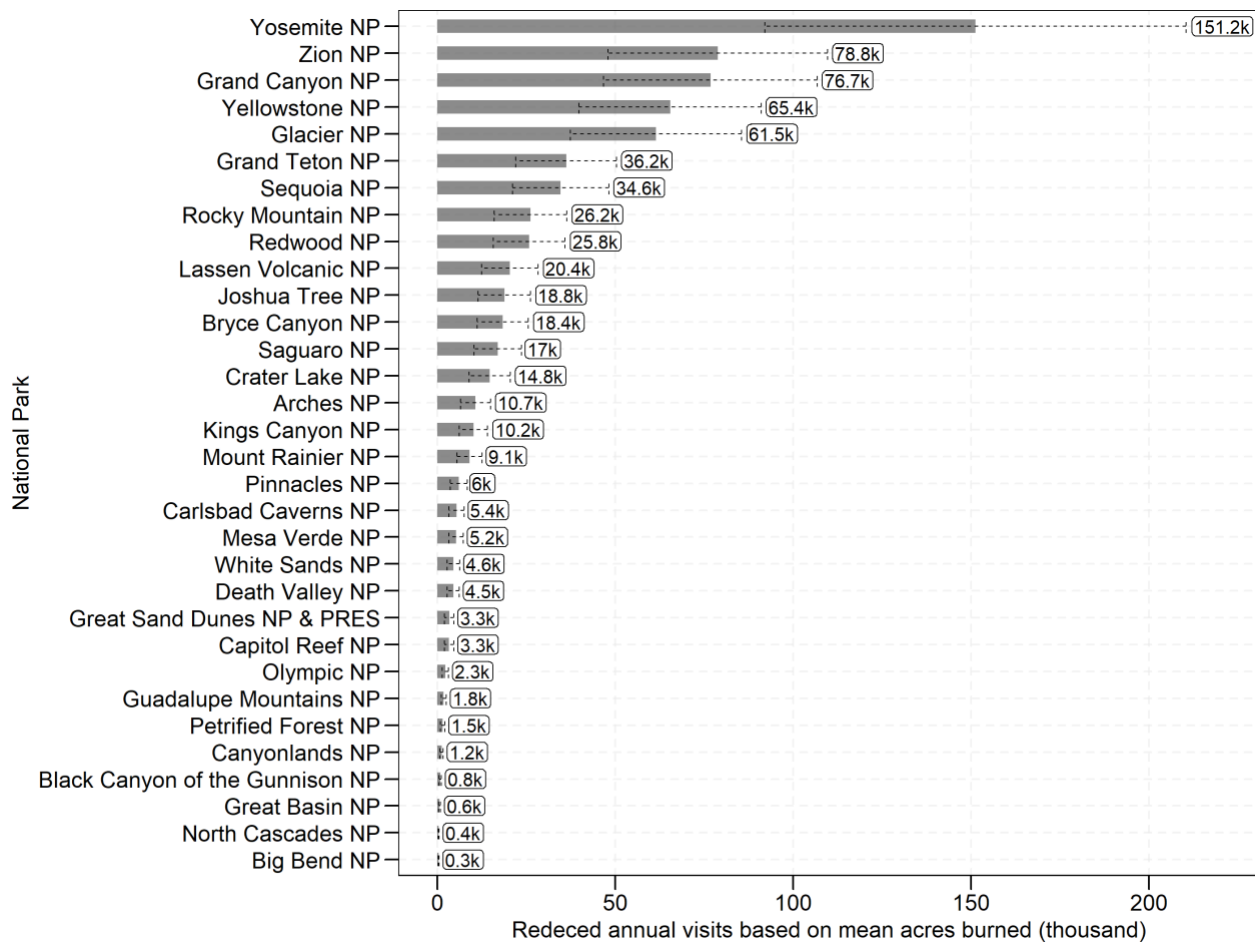


Figure 2.7 Visitation Loss due to wildfires

Note: This figure plots the estimates and 95% confidence interval from a back-of-the-envelope calculation of implied reductions in annual visitation due to wildfires

CHAPTER 3:

TEMPORAL DISAGGREGATION OF VISITATION TIME SERIES IN U.S. NATIONAL PARKS USING MOBILE PHONE GPS DATA

3.1 Introduction

Parks and protected areas not only preserve the natural environment and human history but also provide irreplaceable recreation opportunities (Manning, 2011). Nature-based tourism and recreation are particularly important components in many local economies, contributing up to 4.7% of the states' gross domestic product in 2019 (BEA, 2020). Recent evidence suggests that recreation use in protected areas will continue to increase rapidly in response to climate change (Fisichelli et al., 2015; Smith et al., 2016; Kubo et al., 2020), social media growth (Alonso-Almeida et al., 2019), and the COVID-19 pandemic (Rice and Pan, 2021; Yang et al., 2021). Increased visitor use without targeted monitoring and management can introduce a series of unintended consequences, including crowded conditions, ecological degradation, visitor safety, and a decrease in the quality of visitor experience (Manning et al., 2011; Manning, 2011). To address increasing visitor use, the National Park Service (NPS) has employed a range of management strategies to keep open spaces accessible, enjoyable, and safe while also protecting natural resources (Cahill et al., 2018). These strategies have evolved from shuttle bus services and decade-long permit systems for backcountry areas and wilderness to a more recent timed-entry system (US Department of the Interior, 2021). The time-entry system is designed to spread visitation throughout the week or the day, and some of the daily reservation systems impose varying limits based on the day of the week. A notable example is Yosemite National Park's day-

use reservation, which was introduced in 2021 and established a daily vehicle limit of 2600 for weekdays and 1900 for weekends¹².

However, little is known about visitors' responses to such management strategies prior to their implementation. These uncertainties include whether day use permits successfully displace visitors from weekends to weekdays or instead disrupt visitors' travel plans and result in forgone visits. To understand the costs and benefits of management actions, researchers and park managers usually rely on historical visitation data to conduct analyses (Loomis, 2000). Although the National Park Service (NPS) monitors visitor use and publishes basic counts of park visitors consistently, such releases are typically available on a monthly basis, thus precluding fine-scale analysis of a time-varying policy like day use permits (Freeman III et al., 2014). Because it's not clear when individual trips are taken in a month beyond the simple headcount of visitors entering a park, more granular daily data is needed to accurately estimate the effect of day use permits (Sessions et al., 2016).

Despite the fact that scholars have long recognized the importance of having daily data on visitor use, collecting and maintaining such a database across a dispersed area like national parks is time-consuming and expensive (Moeltner and Englin, 2004; Freeman III et al., 2014; Sessions et al., 2016). A growing body of literature has utilized quality big data to supplement the traditional visitor counts (see reviews by Li et al., 2018 and Li et al., 2021). However, all previous studies only used NPS visitation counts at the annual or monthly scale to validate their alternative measures and have been generally unable to generate useful results at a finer temporal scale (e.g., Wood et al., 2013; Sessions et al., 2016; Sonter et al., 2016; Tenkanen et al., 2017; Fisher et al., 2019), though their proxy and observed visitation may have vastly different scales and diverging

¹² Source: Frequently Asked Questions about Visiting Yosemite in 2022 - Yosemite National Park <<https://www.nps.gov/yose/planyourvisit/faq.htm>>

trends over time. Research efforts to improve the temporal granularity of visitation are, hence, lacking. To the best of my knowledge, little empirical research has taken a step further to make the best use of NPS visitation to reconstruct daily visitor numbers.

This study seeks to address this challenge of downscaling the monthly official visitation time series using the Temporal Disaggregation (TD) approach from the time series literature to downscale the monthly official visitation time series. In macroeconomic studies, TD is a well-established time series approach for disaggregating lower frequency macroeconomic statistics into higher frequency statistics, such as estimating the monthly GDP from a quarterly GDP time series (Chen, 2007; Chamberlin, 2009; Pavía-Miralles, 2010; Quilis, 2018). TD perfectly fulfills the purpose of disaggregating monthly visitation, and it's also easy to implement in practice. The implementation of TD often requires one or more indicator series as inputs. The indicator series needs to be a high-frequency indicator and has the capability of influencing the short-term temporal changes of the low-frequency time series that researchers wish to disaggregate, often called “target series” in the TD literature. The target series in my case is the monthly visitation. For the indicator series, I take advantage of daily foot traffic records sampled from mobile devices and provided by a commercial company named SafeGraph. While other open-sourced digital data (e.g., social media, Google Trends, etc.) may be qualified for the indicator series, these data often lack sufficient daily observations (Tenkanen et al., 2017). In contrast, the SafeGraph dataset, which is gathered and updated at a daily timestep on a consistent basis, has better temporal resolution than other data sources and is therefore reasonably qualified to serve as the indicator series. Additionally, the SafeGraph dataset is easy to use and freely accessible for academic research. More importantly, several previous studies suggest that SafeGraph data can be considered a

surrogate for measuring visitor use in national parks (Juhász and Hochmair, 2020; Kupfer et al., 2021; Liang et al., 2021).

In light of the potential value of daily visitor use data, this study investigates the possibility of using SafeGraph POI-level visitation data in the TD framework to reconstruct historical daily visitation series. Specifically, the research objective is not limited to a validation of the reliability of SafeGraph data as a good surrogate for NPS visitor counts, but further downscales NPS monthly data to the daily level and identifies the best practices for achieving a satisfying TD model performance. Such practices can be widely applied to any protected area with long-term visitation data. To investigate these questions, I leverage the daily variations in the observed visitor volume from SafeGraph data and conduct temporal disaggregation for 31 national parks in the western US from Jan. 2021 to Dec. 2021. To improve the model performance, I develop an automated search algorithm that simultaneously refines SafeGraph raw data and selects the best park-specific model among four different TD implementation methods. This algorithm allows me to expand the analysis from park-based case studies to a large-scale investigation.

Overall, the results from my large-scale analysis indicate that SafeGraph data is a helpful proxy for daily visitation and can be used to estimate daily visitation by disaggregating widely available monthly NPS visitation data. My best-fit TD model can accurately estimate at least 75% of the variation in monthly NPS visitation in 25 out of 31 national parks (adjusted R-square > 0.75). The performance of the best-fit model varies across parks, and it largely depends on how closely first-differenced SafeGraph-derived monthly aggregated series are associated with first-differenced NPS series. My TD models perform well for high-elevation and high-latitude popular parks with large, developed areas, but tend to overestimate the summer visitation for low-latitude desert parks.

One key component in my automated search algorithm is the POI pre-selection, which aims to improve the correlation between the SafeGraph visitor counts estimated from the selected POIs and the NPS data. I show that pre-selection of POIs is crucial in achieving the best model performance. Overall, the best practices for disaggregating monthly visitation based on SafeGraph data require that: (1) the indicator series derived from foot traffic to within-park SafeGraph POIs represent the total volume of visitors to the national park; and (2) the POI pre-selection provides a viable solution for the low-quality SafeGraph data, but compelling justification should be provided to ensure that the TD model will estimate daily visitation consistently for the newly released SafeGraph data.

This study contributes to the rapidly expanding body of literature that assesses the usefulness of important big data to measure visitation by utilizing the temporal disaggregation approach. Unlike previous studies (Wood et al., 2013; Sessions et al., 2016; Clark et al., 2019) that focused on searching, validating, and comparing alternative data sources that could be used to infer park visitation, the TD approach methodologically provides researchers with a framework to integrate all available information from both newly emerged park-related data sources and traditional NPS visitor counts. TD allows researchers to estimate the daily visitation directly based on those new data sources under the temporal aggregation constraints defined by the low-frequency official records. Under the TD framework, new data sources are no longer just a substitute for official visitation counts, and the official records on the ground are no longer just a benchmark to validate the new data sources. In contrast, I show that the final outcome of the analysis is an estimated daily visitation series that is comparable in magnitude to the official park use records rather than an index of recreation use.

The analysis in this paper expands and builds upon the consensus that POI filtering and selection would be required before modeling to guarantee a strong correlation between SafeGraph and NPS datasets (Liang et al., 2021; Yang et al., 2021; Alba et al., 2022). Previous studies typically relied on manual filtering POIs or subjective judgement based on a cut off of the POI's popularity or the North American Industry Classification System (NAICS) category assigned by SafeGraph. The automated POI pre-selection algorithm I adopted can reproducibly find the best subset of POIs for maximizing the correlation between the SafeGraph indicator series and the NPS visitation series. This algorithm was then carried out and assessed for 31 national parks on a large scale.

This paper proceeds as follows. The next section describes the data sources and explains the variables. In Section 3.3, I discuss the fundamental framework of temporal disaggregation and how I address the challenges of applying temporal disaggregation. In Section 3.4, I present the results. The last two sections suggest best practices for disaggregating monthly visitation using SafeGraph data and conclude the paper.

3.2 Data

This study considers 31 national parks spanning the western US as its sample (Arizona, California, Colorado, Idaho, Montana, New Mexico, Nevada, Oregon, Washington, and Wyoming). These national parks combined account for 65% of the nationwide annual visitation to national parks in 2019 and include iconic parks such as Glacier, Grand Teton, Yellowstone, Yosemite, and Zion, which have experienced record demand and faced the challenge of overcrowding in recent years.

3.2.1 SafeGraph data

I construct daily park-level visitation data from the SafeGraph patterns dataset (www.safegraph.com), a widely used Point-of-Interest (POI) dataset (Chiou and Tucker, 2020; Grossman et al., 2020; Goolsbee and Syverson, 2021). A POI indicates a specific physical location that someone may find interesting to visit. SafeGraph derives visitation pattern information from a panel of over 47 million smartphone devices in North America. The company collects GPS location records of these devices anonymously after their users give permission for their precise location to be tracked by third-party applications. The dataset has been aggregated to the POI level and allows users to observe how many visitors with smartphones visited these POIs each day, when they visited, and where they came from. I identify all POIs that are located within the boundaries of each national park based on the rationale that visitors would have to pass through a park entrance to access POIs in the park. I then aggregate the daily visitation patterns of those POIs by the park and construct my indicator series of interest for each national park.

Since SafeGraph began reporting POI data in 2018, temporal disaggregation is carried out using the entire timespan from January 1, 2018, to December 31, 2021. I combine the SafeGraph data with official visitor counts from NPS Visitor Use Statistics for all parks. I use monthly recreation visits as the aggregated target series that will be disaggregated into daily series. Although some reported visitation counts are susceptible to measurement error (for example, fixed estimates of visitor use in remote areas and inaccurate counts due to automated counter errors), I believe the visitation data during peak season for popular national parks is reliable because these parks have multiple staffed traffic monitoring stations to capture use (Ziesler and Pettebone, 2018). Therefore, I take the NPS data as the ground truth for monthly visitations in each park in the sample.

SafeGraph visitation patterns have been previously utilized to understand human mobility patterns (Kang et al., 2020; Chang et al., 2021) and social distancing behaviors in response to the coronavirus (COVID-19) pandemic (Goolsbee and Syverson, 2021). A growing body of literature suggests that SafeGraph data can be considered a surrogate for identifying temporal and spatial variations in national park visitation (Juhász and Hochmair, 2020; Kupfer et al., 2021; Liang et al., 2021). However, how well these data can be used to downscale and predict daily visitation numbers in protected areas remains largely unstudied. I believe SafeGraph data has the potential to be an ideal indicator series in temporal disaggregation because it can provide not only very granular (i.e., daily) data for measuring visitor use patterns but also has a large sample size compared to alternative data sources such as Google Index and social media data. In addition to studies specifically utilizing SafeGraph data, recently there has been an increase in research that has exploited various mobile device location datasets from an array of providers (e.g., Streetlight, Near, etc.) and validated the quality of measures on visitor volume and demographics extrapolated from mobile device data with on-the-ground records such as survey data and actual traffic counts (Monz et al., 2019, 2021; Kubo et al., 2020; Merrill et al., 2020; Creany et al., 2021; Rice et al., 2021). The SafeGraph dataset, like other mobile device data providers, pulls in records from a variety of different sources of location information available at any given time, including GPS, Wi-Fi proximity, Bluetooth proximity, and cell tower signaling (SafeGraph, 2022). Thus, it is less dependent on one carrier and cellular tower coverage and has fewer systemic coverage gaps. Given mobile device data's temporal granularity and spatial coverage, there is strong potential to use SafeGraph data to improve the current temporal resolution of aggregate recreation counts over a dispersed area, a process that has been proved challenging for other data sources.

Nevertheless, I acknowledge that the early use of the newly available SafeGraph data may introduce a range of sampling biases, including but not limited to geographic and demographic biases. In terms of geographic bias, SafeGraph (2019) shows that population density derived from SafeGraph data aligns well with county-level US Census data, as the correlation across all counties is 0.97. Of particular interest is the potential sampling bias on demographic characteristics of recreation users. Although SafeGraph gathers data from 10% of devices in the United States (SafeGraph, 2019), the sample conditioned by mobile phone ownership may not be representative of young, elderly, and low-income groups (Lawson, 2021). There is minimal research assessing demographic disparities in SafeGraph data, but one notable exception, that compare the visitor demographics with Yellowstone National Park survey data and surprisingly finds that SafeGraph disproportionately oversamples the oldest age group (75 or older) and household income groups less than \$50,000 (Liang et al., 2021). However, when comparing demographic variables against the census block group (CBG) population at the national level, SafeGraph (2019) suggests that their data is “well-sampled on most demographic dimensions,” including household income. I acknowledge that demographic distribution may not fully represent the greater population of visitors; therefore, in Section 3.5 I propose the conditions when the SafeGraph sample is representative of the total visitor volume.

3.2.2 Other indicator series

In addition to SafeGraph data, I also consider a couple of daily indicator series that are relevant determinants of visitation. Temperature has been shown to have an inverse u-shape relationship with visitation (Fisichelli et al., 2015; Smith et al., 2016), and heavy precipitation decreases visitation (Keiser et al., 2018). Daily maximum temperature and total precipitation grid data are obtained from PRISM Climate Group (PRISM Climate Group, 2004). I also construct an indicator

series of large-scale national park closures in response to the COVID-19 pandemic, which is a dummy variable representing the period from April 1, 2020, to May 30, 2020. To capture the seasonal variation in visitor volumes, I consider using the access levels and seasonal road closures to classify the access and the winter season (typically December to April), which is a classification preferable to conventional meteorological seasons for accommodating the climatically and geographically diverse national parks. Seasonal road closure records are compiled from each park's website and past Twitter posts.

3.3 Method

The main methodology adopted in this study is called Temporal Disaggregation (hereafter TD). TD is a well-established time series approach for tracking when the time series have different frequencies (Ciammola et al., 2005; Chamberlin, 2009; Pavía-Miralles, 2010). It is widely used by macroeconomists and national statistical agencies to downscale and forecast important macroeconomic series, such as the GDP series (Abeyasinghe and Lee, 1998; Chen, 2007; Islam, 2009; Quilis, 2018; Rossi and Zanetti Chini, 2021). To my best knowledge, the TD method has not been applied to disaggregate recreational visitation data in previous literature.

3.3.1 The Chow-Lin-Fernández framework

The classic TD method is based on the work of Chow and Lin (1971), which was further developed by Fernandez (1981). I will start with a brief outline of the Chow-Lin-Fernández framework (hereafter CLF framework). Unlike a standard OLS regression setting in which the goal is to estimate a vector of coefficients of the linear regression model, the target of TD is to estimate a high-frequency time series from a lower frequency series. To make this point precise, I will introduce Y_h to stand for the high frequency (hereafter HF) outcome or target series. In my case,

Y_h is the daily visitation count in a specific national park, that is unobservable to researchers and needs to be estimated. To implement the CLF framework, two inputs are needed. One is the low frequency (hereafter LF) time series of the outcome variable, denoted as Y_l in this case, monthly visitation. The other is the HF time series of indicator variables, denoted as X_h . According to Chow and Lin (1976), any series that is capable of influencing the temporal changes of the targeted low-frequency series is qualified as an indicator. In my case, SafeGraph daily visitation data serve as the main HF indicator series.

The CLF framework assumes a linear relationship between outcome and indicator series at the HF level, i.e.,

$$Y_h = X_h \beta + u_h \quad (3.1)$$

Where u_h is the HF disturbance series, as I will discuss in the TD method selection in section 3.3.2. The main difference between variants of the CLF type of TD method is the assumptions regarding the data generating process of u_h . To make the discussion complete, I assume a generic structure of u_h as (1) a conditional mean zero, i.e., $E(u_h|X_h) = 0$, and (2) an unspecified form for the conditional covariance matrix, i.e., $E(u_h u_h' | X_h) = V_h$.

Note that though eq. (3.1) is conceptually valid, the unknown parameter β can not be estimated directly, as Y_h is also unknown to the researchers. Therefore, it is wise to also study the LF counterpart of eq. (3.1):

$$Y_l = X_l \beta + u_l \quad (3.2)$$

Where Y_l is the observable LF outcome series, and X_l can be attained by aggregating the HF indicator. Mathematically, it can be achieved by pre-multiplying X_h by an aggregation or conversion matrix, usually denoted as C in the TD literature, i.e., $X_l = C X_h$. Given the length of the HF series as T and the length of the LF series as n , the aggregation matrix C is a n by T matrix.

Finally, u_l is the LF disturbance series. Its distributional property can be easily derived by combining the assumptions on its HF counterpart and the aggregation matrix, namely, $E(u_l|X_h) = 0$ and $E(u_l u_l' | X_h) = C V_h C' = V_l$.

Unlike eq. (3.1), one can easily estimate the coefficient β in eq. (3.2) using the classic GLS technique. And, by the linearity assumption, equations (3.1) and (3.2) share the same coefficient β because eq. (3.2) is attained by pre-multiplying C by every term in eq. (3.1).

The last step is to get the optimal (in the sense of BLUE) estimates of Y_h , denoted as $\hat{Y}_h$. Chow and Lin derived the following expression by minimizing the variance of the predicted error¹³ under unbiasedness and temporal aggregation constraints:

$$\hat{Y}_h = X_h \hat{\beta} + V_h C' V_l^{-1} (y_l - X_l \hat{\beta}) \quad (3.3)$$

$$\hat{\beta} = [X_l' V_l^{-1} X_l]^{-1} [X_l' V_l^{-1} Y_l] \quad (3.4)$$

Two remarks on the above expression are worth noting. First, Sax and Steiner (2013) provided a useful interpretation of eq. (3.3). They called the first item in the RHS a “naive” estimate of the HF outcome series, which is obtained by directly multiplying the HF indicator series X_h by $\hat{\beta}$. The second item in RHS can be further decomposed into two parts; $(y_l - X_l \hat{\beta})$ is the residual of eq. (3.2), i.e., the LF regression, and the matrix $[V_h C' V_l^{-1}]$ is called the distribution matrix. It distributes the residuals of the LF regression into each HF period. In my case, this distribution matrix distributes the monthly visitation residuals for a specific month into each day of that month. Therefore, the RHS of eq. (3.3) can be interpreted as an adjusted version of the “naive” estimates. Second, eq. (3.4) is nothing but a standard GLS formula given the error term

¹³ Predicted error is defined as the difference between an estimated HF outcome series and its unobservable ground-truth. Therefore, unbiasedness enters Chow-Lin’s framework as a constraint, not a property. Given unbiasedness, the variance of the predicted error is *not* a function of unobservable ground truth. So, the minimization problem is well defined. The original paper by Chow and Lin (1971) remains the most valuable resource for understanding this idea.

has a zero mean and a covariance matrix V_I , which one can easily find in standard econometrics textbooks such as Wooldridge (2010) or Greene (2018). In practice, when a specific parametrization of the covariance matrix is assumed, the hyper-parameters in the covariance matrix can be estimated, and a Feasible GLS procedure will be applied.

3.3.2 Two challenges for implementing TD methods

Two modeling challenges should be addressed when implementing TD methods with the SafeGraph daily visit data to disaggregate the monthly park visitation into its daily counterpart: pre-selection of POI and TD method selection.

Pre-selection of POI

The first practical challenge is that how to improve the quality of the indicator series. Despite the theory of Chow and Lin (1976) suggesting that any series with an impact on daily visits might serve as an indicator series in the CLF framework, the quality of the final HF estimation of the target series largely depends on the quality of the HF indicator series (Santos Silva and Cardoso, 2001; Mosley et al., 2021).

In this manuscript, the indicator series of interest is the daily visitation series derived from SafeGraph, which is constructed by aggregating all visits at SafeGraph POI sites located within the park boundaries.¹⁴ One important data pre-processing step, therefore, is to eliminate low-quality and irrelevant POIs before aggregation. Typical examples of poor-quality POIs are elementary schools or research stations, as there is no reason to expect the visitors to these POIs belong to the population of park visitors. Despite the fact that previous researchers have manually filtered and removed the poor-quality POIs before their subsequent analyses (e.g. Liang et al.,

¹⁴ I also extended the search radius to include 1, 2, 5, and 10-mile buffers around the park boundary to extract POIs outside of park polygons.

2021), I decided to adopt a more automated and reproducible approach to achieve: (1) scalability as opposed to manually checking hundreds of POIs around over 30 national parks in the western US; (2) objectivity because manually checking inevitably involves subjective judgments on the decision rule; and (3) accuracy to avoid the potentially serious issue of inaccurate location names and NAICS codes provided by raw SafeGraph data. My quality validation of the raw SafeGraph data reveals that such labeling errors do exist in several parks. As a result, decision-making based on these variables may propagate such labeling errors to subsequent analyses.

I develop the following POI pre-selection algorithm: I first aggregate the time series of all POIs within a distance threshold to generate the raw indicator series of interest, and then I calculate the Pearson correlation coefficients between the first-differenced raw indicator and the outcome series¹⁵. The Pearson correlation calculated here serves as a benchmark for potential improvement. Next, I calculate the Pearson correlation coefficients between the outcome series and each POI time series separately. I use a correlation threshold to filter out the POIs with correlation coefficients lower than the threshold. I then aggregate the remaining POIs to form a refined indicator series of interest and calculate the correlation between the refined indicator series and the outcome series again. Lastly, because the correlation threshold determines the set of remaining POIs and, hence, the performance of the refined indicator series in terms of correlation coefficients, I conduct a grid search to determine the optimal threshold, which maximizes the correlation coefficients between the outcome series and the refined indicator series generated by that threshold.

Two points are worth mentioning. First, my automatic pre-selection rule is inspired by and consistent with the underlying assumptions of the CLF framework. As shown in the preceding

¹⁵ To prevent the potential false correlation due to time trends, the first difference operation is always applied to the time series of interest before any calculation of correlation coefficients. I conduct Augmented Dicky-Fuller tests on NPS visitation series and raw SafeGraph series for 27 parks in the sample (excluding 4 parks without enough data) and find that 6 out of 27 NPS visitation series and 13 out of 27 raw SafeGraph series are non-stationary.

section, the CLF framework heavily relies on the linearity of the model structure to infer the HF relationship by examining its LF counterpart. This allows researchers to bypass a fundamental challenge in TD: the unobservable HF target series. Hence, it's reasonable to evaluate the quality of the indicator series by examining the similarity between its LF series and LF target series.

Second, I choose the Pearson correlation coefficient as the similarity metric between the indicator series and the outcome series after having experimented with multiple options proposed in the literature. Because my use case is to compare the exact same month in SafeGraph and NPS visitation records, measures frequently employed in signal processing literature that tolerate the phase shifts in two time series, such as Dynamic Time Warping, are not considered. An alternative to Pearson correlation, therefore, is the Euclidean distance between the z-transformed indicator and the outcome series. A comparison of the Pearson correlation and Euclidean distance can be found in Figure B.1 in the appendix. It shows how effectively the Pearson correlation predicts a high adjusted R-squared for the LF regressions in the CLF framework.

TD method selection

Another challenge is to select the appropriate TD method among the rich family of variants of the CLF framework. Specifically, I choose from (1) standard Chow and Lin (1971), (2) Fernandez (1981), (3) Litterman (1983), and (4) Santos Silva and Cardoso (2001)¹⁶. As briefly mentioned in Section 3.3.1, the key difference among models (1) - (3) is the assumption about the data generation process of the HF disturbance series. Standard Chow-Lin assumes a first-order autoregressive process with a coefficient ρ , i.e., $u_{h,t} = \rho u_{h,t-1} + \varepsilon_{h,t}$, where $\varepsilon_{h,t}$ always stands

¹⁶ I decided not to implement Denton (1971), another famous and widely used TD method, in this manuscript for three reasons: First, Denton's approach cannot allow more than one indicator series. Second, there is no regression-based performance metric in Denton's approach. The quality of estimates is mainly determined by the similarity between the HF indicator series and HF target series, which cannot be tested. Finally, one can view Denton as a special case of Fernandez (1981). In his influential paper, Fernández proposed a reinterpretation of Chow-Lin's framework that incorporates Denton's approach as a special case in the unified framework.

for white noise. Fernández assumes it as a unit-root process, i.e., $u_{h,t} = u_{h,t-1} + \varepsilon_{h,t}$. Litterman assumes a more complicated “Random Walk + Markov” model, i.e., $u_{h,t} = u_{h,t-1} + v_{h,t}$; $v_{h,t} = \eta v_{h,t-1} + \varepsilon_{h,t}$. Santos Silva and Cardoso (2001) proposed a so-called “Dynamic Chow-Lin” approach that introduces an autoregressive structure to eq. (3.1), the HF regression equation. The dynamic version of eq. (3.1), $Y_{h,t} = \kappa Y_{h,t-1} + X_{h,t} \beta + u_{h,t}$, is estimated by an approach invented by Klein (1958). Figure 3.1 summarizes the relationship between these four models.

The unobservable HF target series poses a fundamental challenge in selecting from different TD methods. The standard model selection rules, such as minimizing out-of-sample RMSE, are infeasible in the TD setting. Alternatively, previous literature suggests numerous indirect metrics to guide the selection of TD methods. The most common one is based on the adjusted R-squared of the LF regression, i.e., eq. (3.2). The rationale for this rule is again tied to the linearity assumption of the CLF framework (Chamberlin, 2009; Pavía-Miralles, 2010; Quilis, 2018). A similar measure in the same direction is the out-of-sample prediction error of the LF target series (Islam, 2009). Such a measure, however, confounds disaggregation performance with forecasting performance. Finally, Chen (2007) proposed several quality standards for the estimated HF outcome series, including (1) satisfying temporal aggregation constraints; (2) preserving short-term movement in the indicator series; (3) avoiding drastic distortion at the breaks of LF series, and (4) avoiding drastic distortions at the beginning or ending periods of the sample. However, Chen’s quality standards heavily rely on subjective judgments about the relative importance of each quality standard. There is no systematic way to combine them as a holistic score. Therefore, in this study, I decided to adopt the standard approach of selecting the best TD methods based on

the adjusted R-squared of eq. (3.2). Figure 3.2 summarizes the procedure followed to pre-select POI and select the best TD method.

3.3.3 Two data quality concerns and their implications for TD method selection

Two limitations are worth noting when it comes to the quality of SafeGraph data. Depending on how important these two problems are perceived to be, the best TD method choice may diverge from the one selected by maximizing adjusted R-squared. The first quality issue is unstable data volume. SafeGraph has continually sourced new mobility data from users and updated their visitation datasets and POI geometries, which brings up a potential issue of unstable data volume over time. According to SafeGraph's documentation¹⁷, the number of devices in Washington state has increased by 20% in November 2021. Since the visitors to POIs in parks are just a subset of the SafeGraph panel of all devices, there can be an artificial increase in the aggregated park-level visitor counts derived from SafeGraph data, which may not reflect the ground truth.

The issue of unstable data volume raises serious concerns about the linearity assumption between the indicator and target series. In general, there are two types of data volume instability: differing time trends between NPS visitation and SafeGraph data, and outlier POIs that can cause a substantial jump or drop in SafeGraph time series. The latter should be largely mitigated by the above-mentioned POI pre-selection procedure. Therefore, I will focus on the former type below.

The issue of different time trends will violate the linearity assumption. For example, if SafeGraph data increases at a faster rate than NPS visitation data over the study period, then when we plot the SafeGraph data and NPS visitation on the scatterplot, the underlying relationship between those two will be captured by a concave-shape curve rather than the linear curve that the CLF framework assumes. The degree to which the linearity assumption is violated is determined

¹⁷ Source: Known Issues and Data Artifacts – SafeGraph <<https://docs.safegraph.com/docs/known-issues-or-data-artifacts>>

by the difference between the two time trends. The larger the difference, the higher the chance the linearity assumption will be violated. Under the CLF framework, the consequence of neglecting this issue is as follows: because of this model misspecification, certain periods will have higher residuals than others, resulting in greater artificial fluctuations in estimated daily series for those periods with higher residuals. A possible solution to this issue is to adopt the following procedure: first, estimate and compare the difference in time trends between daily SafeGraph and monthly NPS series, and if the difference is statistically significant, then follow either of the Fernández or Litterman approaches, which both use the first difference series that is much less likely to be affected by different time trends.

The second potential quality issue lies in whether the SafeGraph daily visit time series is representative of the underlying daily distribution of park visitation. Clearly, given that the ground-truth of daily visitation is unobservable, it is impossible to definitively evaluate the severity of this issue or provide an easy fix when the bias likely exists. Hence, I assume that the observed distribution of daily visitor volume based on SafeGraph data is representative of population-wide travel. However, the researchers' subjective views on this issue might influence the selection of the optimal TD method. Recall that one of the quality standards proposed by Chen (2007) is that the estimated HF outcome series should preserve the short-term variations in the HF indicator series as much as possible. Consideration of this quality standard clearly depends on one's perception of the representativeness of the SafeGraph daily series. The greater the confidence one has in the SafeGraph series, the greater weight one should give to the property of preserving short-term variations.

3.4 Results

This section begins with a discussion on the optimal model selected for each park for adjusted R-sq maximization. Figure 3.3 summarizes the outcome of the POI pre-selection and the optimal TD method choice for each park. The POI pre-selection process itself consists of two components. First, the irrelevant POIs are filtered out using an optimum correlation cutoff, which is determined by a grid search designed to maximize the correlation coefficient between the aggregated monthly visits derived from selected POIs and the NPS official visitation series (see section 3.3.1 for more details on the algorithm). Second, the search radius is extended beyond park boundaries to increase the number of nearby POIs considered.

As we can see from Figure 3.3, POI pre-selection indeed improves model performance for the vast majority of parks in the sample (27 out of the 31 parks). The baseline model is only optimal for four parks, namely Yellowstone, Mesa Verde, Channel Island, and Zion. Expanding the search radius to include more POIs is crucial for improving model performance, since 20 out of the 31 parks fall into more refined, higher rows 3 – 6 of the figure. Due to the scarcity of SafeGraph POIs inside the boundaries of four parks (BLCA, NOCA, GRBA, and WHSA), increasing the search radius beyond their boundaries is especially important for building the indicator of interest. Therefore, I conclude that the pre-selection of POIs is a key component in achieving the best model performance.

Additionally, Figure 3.3 reports the results of the best TD method chosen for each park. The Dynamic Chow-Lin method is the most prevalent of the four alternatives (accounting for 23 out of the 31 parks). The standard Chow-Lin is selected six times, Fernández is selected twice, and Litterman was never chosen as the best TD method, despite being theoretically appealing. The dominance of the Dynamic Chow-Lin is expected because I use the adjusted R-sq to evaluate

model performance. Adapting the Chow-Lin method to a dynamic setting improves its model fit and generates results more robust to the untestable assumption on the random components of HF equations in eq. (3.1). As the inventors of Dynamic Chow-Lin, Santos Silva and Cardoso (2001) explicitly state in their original work, “... the better the model fits the data, the less important the correct specification of Ω (the error structure of HF equations) is” (Santos Silva and Cardoso, 2001, p. 273). Since the visitation series often follows an autoregressive process¹⁸, including a dynamic structure in the TD framework should, in general, help improve the goodness-of-fit metric.

3.4.1 Heterogeneity in model performance

Figure 3.4 shows how much the model has improved following the POI pre-selection for each park. The baseline model’s adjusted R-squared is illustrated in blue, whereas the best model is shown in red. The changes in adjusted R-squared are reported in the right margin. I find that pre-selection of POIs together with model selection, on average, raises the adjusted R-squared by 0.067 (median = 0.035), which does not incorporate the four parks that do not have a baseline model. The greatest improvement is 0.4 for Saguaro National Park (SAGU), a park near Tucson, AZ. Seven parks exhibit no or minimal differences between the baseline model and the optimal model. Furthermore, I find that the magnitude of the improvement depends on the initial model performance of the baseline model. On one hand, the parks with poor initial performances are more likely to see considerable gains post POI pre-selection, for example, Saguaro, Pinnacles, and Death Valley NPs. The parks with already-high initial performances, on the other hand, are more likely to witness a small rise in adjusted R-squared. For example, all seven parks with no improvement have an initial adjusted R-squared of above 0.8.

¹⁸ I run autoregressive models on monthly park visitors for each park, and find all parks have significant AR(1) terms.

Additionally, I categorize 31 parks into 3 groups based on their best model performance: poor ($\text{Adj. R-sq} < 0.70$), moderate ($0.70 < \text{Adj. R-sq} < 0.9$), and good ($\text{Adj. R-sq} > 0.9$). The "good" group includes fourteen parks that account for almost 60% of overall visitors in the study. Among them are Yellowstone, Grand Teton, Rocky Mountain, Yosemite, Glacier, and Olympic NPs, which are six of the top ten most visited western US parks in 2021 (see Table 3.1). The majority of parks in this category are high-latitude, high-elevation mountain parks with a single peak season in the summer. These parks are also likely to have high visitation volumes and more developed areas. When these features are combined, SafeGraph pattern datasets collected within the park boundaries generally find more POIs, capture a substantially larger amount of data, and hence are more representative of the visitor population. Consequently, the high-quality SafeGraph data ensures that TD models perform well.

Twelve parks are classified in the "moderate" category, which includes some very popular parks in low-latitude and low-elevation deserts and semi-arid regions, such as the Grand Canyon and Zion National Park. The summer peak seasons in low-latitude desert parks are less defined than those in high-latitude parks; such parks may experience an early peak season in spring and a second visitation peak in late fall or even early winter. My studies show that SafeGraph data, in general, is more likely to capture visits during the summer than in late spring and early autumn, which is probably because visitors are more likely to stay longer in SafeGraph POIs such as restaurants or hotels during the hot weather.

In addition, I find that the three parks located near major urban centers, Joshua Tree, Saguaro, and Channel Islands, are also in the "moderate" category. One extra challenge of using SafeGraph data in these parks lies in how to distinguish local visitors from park visitors. If there is a relatively

large proportion of local resident visits recorded by SafeGraph, then a moderate model performance should be expected.

The remaining five parks are in the “poor” category, namely Black Canyon, Pinnacles, Death Valley, White Sands, and Redwood NPs. The five parks together account for barely 5% of the total visitation in the sample. Parks in this category are typically small in terms of the size of both the annual visitation and developed areas, resulting in poor quality of SafeGraph data within the park boundaries. There is no POI with complete data within the boundaries of White Sands and Black Canyon NPs. For the other three parks, the correlations between SafeGraph data within the park and NPS’s visitation numbers are all very weak (the correlation coefficient is less than 0.3).

3.4.2 Low-frequency GLS coefficient estimates

Figure 3.5 presents the coefficient point estimates of LF GLS regression (i.e., eq. (3.2)) for the best model chosen. Panel (a) shows the coefficient estimates for the indicator series of interest, monthly aggregate daily visits to all selected SafeGraph POIs within the best model’s selected distance cutoff (see Figure 3.3) near the park boundaries. The magnitude of coefficients is relatively less important than their significance, as it depends on both the relative size of NPS visitation and SafeGraph data as well as different TD methods¹⁹ for each park reported in Figure 3.3. The results show that the coefficients associated with the SafeGraph series are positive and highly significant in all but one park, White Sands National Park (WHSA) in New Mexico. It is a fairly new park and was designated as a national park only three years ago in 2019. There is no SafeGraph POI within its boundary. There are only a few POIs within the 10-mile buffer zone around the park boundaries, and those POIs are only weakly associated with official NPS visitation numbers. In terms of adjusted R-squared, WHSA is likewise the park with the poorest performance.

¹⁹ The Dynamic Chow-Lin model usually generates smaller coefficients than the other three alternative TD methods because it includes the lagged dependent variable as an additional regressor.

There is no doubt that the monthly aggregated HF SafeGraph visitor count is the most important factor in determining TD performance.

The next two panels present the results for two weather variables. Panel (b) shows the coefficient estimates for the count of days with a daily temperature greater than the 75th percentile of the park-specific temperature distribution. Only one-third of the parks (11 out of 31) have positive and significant coefficients, and the majority of the parks with significant coefficients are located in high latitudes or high elevations because visitors usually flock to these parks on hot summer days. Consistent with this behavior, I found a negative (albeit not statistically significant) coefficient for several parks in southern California, Nevada, and New Mexico, which is consistent with the fact that visitors to those areas prefer cooler weather. Panel (c) shows the coefficient estimates for the count of days with daily rainfall greater than the 75th percentile of the park-specific precipitation distribution. The majority of the parks have insignificant negative precipitation coefficients. As expected, experiencing more heavy precipitation days is associated with fewer visitors. It is worth noting that the prevalence of insignificant coefficients associated with the weather variables among parks is likely explained by the fact that SafeGraph data absorbs the majority of the variations in park visitation.

The last two panels present the coefficient results for two exogenous shocks that directly impact park accessibility. Panel (d) displays the coefficients of the indicator series for nationwide COVID-related park closures in April and May 2020. It is no surprise that this nationwide park closure has a negative impact on visitation for about one-third of the parks in the sample. Panel (e) shows the coefficients for monthly indicators of full road access to the park (i.e. no seasonal road closure). Many parks have positive coefficients, as one would anticipate. Again, the insignificant

results are mainly attributed to the fact that SafeGraph data is a close proxy of the actual visitation records.

Figure B.2 shows the final output of the TD analysis: estimated daily visitation series for all 31 parks for the selected optimal model specifications. Compared to an analysis of individual indicators of interest (Figure 3.5), one significant benefit of a temporally disaggregated series is that, by design, they satisfy the temporal aggregation constraints that daily estimates sum to the NPS monthly visitation series (Chen, 2007; Denton, 1971). Therefore, it is more suitable for analyses of policies for recreation use than the raw SafeGraph series that many previous researchers have used (Jay et al., 2021; Kupfer et al., 2021; Liang et al., 2021; Yang et al., 2021; Alba et al., 2022).

3.4.3 Weekend and holiday effects on daily visitation series

To conclude the result section, I conduct an exercise examining the effects of weekends and holidays on the daily visitation series. Specifically, I examine whether the distribution of the estimated daily visitation series is sensitive to the day of the week being on a weekend or holiday when visitor increases are often observed in recreation areas (Hewer et al., 2016; Gellman et al., 2022). I hypothesize that weekend and holiday impacts are likely to be larger at parks located near a major metropolitan center. The goal of this exercise is twofold: First, the results of this exercise could help to check the assumption made in Section 3.3 that the observed distribution of SafeGraph daily visitor volume is representative of population-wide travel; second, the results of this exercise could serve as a simple quality check for the estimated daily visitation series, as the conventional metrics for estimation accuracy such as RMSE are not applicable in this case with no available ground truth data for daily visitation.

I regress the estimated daily visitation series on weekend and holiday dummies²⁰, complementing year-by-month fixed effects used to control time trends and seasonal fluctuations. Figure 3.6 shows the coefficient estimates for the weekend and holiday dummies for all 31 parks. Panel (a) shows that 19 out of 31 parks have positive and significant coefficients for the weekend dummy. Panel (b) shows that 22 out of 31 parks have positive and significant coefficients for the holiday dummy. The fact that the majority of the parks have positive and significant coefficients is consistent with the conventional wisdom that more tourists visit parks during weekends and holidays than on non-holiday weekdays. Equally intuitive is the finding that parks located near large metropolitan areas, such as Joshua Tree, Rocky Mountain, and Saguaro NPs, have the largest positive coefficients, consistent with parks that have the largest numbers of nearby city residents experiencing higher visitation during weekends and holidays. These findings, together with the strong performance of TD model fitting consistent with monthly park visitation data, increase confidence that the estimated daily visitation series based on SafeGraph data captures intra-week patterns of park visitation.

3.5 Discussion

In light of the findings discussed above, I will begin this section by recalling the research questions raised in the introduction: Is SafeGraph a valid source for estimating the historical daily visitation series? If it is, then what are the best practices for disaggregating monthly visitation using SafeGraph data (or similar) georeferenced POI-level mobile device location data?

The short answer to the first question is yes. My results confirm that in many parks SafeGraph data can be used to estimate daily visitation by disaggregating widely available monthly NPS

²⁰ I use NYSE's list of US holidays to construct the holiday dummy.

visitation data. Previous studies' practices of refining the SafeGraph POIs within the park boundary were found to be a valid approach to reducing the noise from irrelevant POIs. Furthermore, the automated POI pre-selection algorithm I developed allows me to expand the study regions from a single park to all 31 national parks in the western US. This large-scale analysis gives me a more holistic understanding of the conditions under which SafeGraph data may achieve better model performance and the strategies that can help improve it. In short, I find that the model performance is mostly driven by the quality of POIs *within* the park boundary. While it is possible to incorporate POIs beyond park boundaries that have a strong link to national park tourism, such practices should be carried out with caution and on an individual POI basis. I elaborate on these two primary lessons learned by making two suggestions for “best practices” for future research using such POI-level visitation data.

Best Practice No. 1: The quality of SafeGraph data in the temporal disaggregation should be largely dictated by the correlation between the indicator series derived from within-park POIs and the NPS official visitation series.

It stands to reason that within-park POI foot traffic deserves special attention because it contributes directly to the official visitation records collected at park entrance stations. All visitors that enter parks via entrance stations equipped with traffic monitoring devices are counted before they arrive at POI destinations within park boundaries, where these visitors are recorded by the SafeGraph algorithm. As a result, SafeGraph visitor counts collected at within-park POIs are inherently a subset of NPS visitation statistics. My large-scale analysis suggests two necessary conditions for achieving a high correlation between within-park aggregate daily POI foot traffic and NPS monthly visitation: First and foremost, the park must have a large developed area(s) where the majority of visitors spend some time during the trip (e.g., park visitor centers or transit

stations in a park with a shuttle system). Because the majority of SafeGraph POIs are located in developed areas, a visitor with a longer dwell time at a POI has a higher chance of being recorded in the SafeGraph database²¹. A well-developed park is expected to have a sufficient number of within-park POIs with a high daily average volume of visitors. The second factor is the popularity of the park. SafeGraph (2019) demonstrates that their data has better demographic representation at county and state levels compared to census block groups. I expect that SafeGraph is a more accurate representation of the visitor population in parks with the highest visitation. Yellowstone and Yosemite NPs are two examples that meet both of the aforementioned conditions; the high quality of within-park POIs results in strong TD model performance.

However, there may be exceptions to the first best practice. Exceptions will arise when there are singularities in park boundaries for historical reasons (Runte, 1997). In my sample, I observe two types of singularities (Figure 3.7). On one hand, filtering POIs based on park boundaries may exclude important POIs that tourists will visit. For example, North Cascade National Park's boundary does not include the popular scenic drive passing through the park (WA SR-20), most likely due to the long-standing disputes between outdoor recreation activities and the local community²². As a result, some of the famous park pullouts and viewpoints are excluded from the park boundary. On the other hand, filtering by park boundaries may also incorporate irrelevant POIs where the majority of the visits recorded do not belong to park visitors. For instance, Jackson Hole Airport has a disproportionately large number of SafeGraph records and is located inside the Grand Teton National Park boundaries²³. However, POI visitor counts recorded at the airport do

²¹ Source: SafeGraph Community FAQs – SafeGraph <https://docs.google.com/document/d/1h-pkpIZWeynF3_BcylRmgeWS7282kIUGoM0TSSIUhgM/edit>

²²Source: President Johnson signs bill creating North Cascades National Park on October 2, 1968. – HistoryLink.org <<https://historylink.org/File/9789>>

²³ The SafeGraph POI at Jackson Hole Airport accounts for 19% of total SafeGraph visits to Grand Teton in 2018, 14.6% in 2019, 64.3% in 2020, and 60% in 2021.

not necessarily reflect visitors to Grand Teton NP, especially during the ski season. Because Jackson Hole is a popular skiing and snowboarding destination in the US, and the Jackson Hole Airport is the main public airport serving the surrounding area²⁴. This POI is a common destination for many winter recreationists visiting nearby ski resorts rather than the park. In the case of both North Cascade and Grand Teton, my automatic algorithm can help to discover both relevant POIs outside the park boundary and irrelevant ones inside the park boundary.

Best Practice No. 2: Incorporating POIs beyond park boundaries may increase the quality of the SafeGraph data, but careful justification should be provided to ensure the recorded visits at those outside-park POIs mainly come from park visitors.

I find that when in-park POIs do not have a sufficiently high correlation with NPS visitation data, expanding the search radius and including POIs beyond park boundaries may be a viable solution. The automated search algorithm developed can identify the best subset of outside-park POIs for maximizing the correlation between the SafeGraph indicator series and NPS visitation series. However, there should be a clear justification for why an outside-park POI should be included. Without compelling justification, it's uncertain whether the correlation observed in the historical time series will persist into the future. Next, I present two examples of compelling justifications for including an outside-park POI. First, my algorithm selects Dead Horse Point State Park as an outside-park POI for Canyonlands NP. Dead Horse Point State Park is usually considered a companion tourist attraction to Canyonlands NP because both parks share a common entry route (UT-313). Visitors will pass by Dead Horse Point State Park on their way to Canyonlands' main visitor center at Island in the Sky. Second, for Bryce Canyon NP, my algorithm selects multiple hotels, restaurants, bike rentals, and gas stations in Bryce Canyon City (population

²⁴Source: Jackson Hole Airport – Wikipedia <https://en.wikipedia.org/wiki/Jackson_Hole_Airport>

= 228 in 2019) and Tropic (population = 485 in 2019) as outside-park POIs for Bryce Canyon NP (Figure 3.7C). Compared to Bryce Canyon NP's over 1.5 million annual visits since 2015, one could reasonably argue that customers at these establishments in these two isolated and low-population gateway communities are primarily park visitors, and these POIs, therefore, are potentially complementary sites to Bryce Canyon NP.

However, the automated algorithm does not always choose meaningful outside-park POIs. For instance, for Black Canyon of the Gunnison NP, the algorithm selects a Home Depot located in the commercial area of Montrose, CO (population = 20,291 in 2020) (Figure 3.7D). Though travelers could visit Home Depot to buy some hardware or other supplies, it is difficult to justify that the majority of the visits recorded at that Home Depot are from visitors to Black Canyon NP. The local tourist-to-resident ratio for this relatively new national park (1/15, compared to Bryce Canyon's 1/3000) fails to support the argument that the Home Depot is a meaningful outside-park POI. Related to this example and worth noting is the fact that previous studies (e.g. Jay et al., 2021) suggest that if a site is close to a city or a population center, the quality of the SafeGraph data will improve. However, I find that this trend is less relevant for identifying outdoor recreation patterns because it is difficult for researchers to separate local residents from park visitors when the POIs are close to a population center. In the case of Bryce Canyon, there are fewer than 1,000 local residents, and NPS visitation is much higher than in the case of Black Canyon. Recall that (see Figure 3.4) the parks near the major cities in my sample, especially Saguaro, Joshua Tree, and Channel Islands NP, do not yield very high model performance in general.

3.6 Conclusion

This study is the first attempt to implement the Temporal Disaggregation approach to improve the temporal resolution of recreation use data. By combining cellphone GPS data with the TD approach, I downscale the monthly visitation series for 31 national parks in the western US to the daily scale by extracting the information of short-term variation in SafeGraph's daily foot traffic records. My results suggest that SafeGraph data can be used as the indicator series for TD for the majority of the parks in my sample. High model performance could be expected if a strong correlation exists between the first-differenced aggregated SafeGraph series and NPS's official monthly visitation records. In order to maximize these correlations, I expanded the search radius beyond the park boundaries and developed an automatic search algorithm to pre-select SafeGraph POIs and identify the best subset of POIs. My findings suggest that high-quality within-park POIs are essential to guarantee a high-quality indicator series, while including outside-park POIs could moderately increase the model performance if compelling evidence could be provided to justify that the target audiences of those outside-park POIs are park visitors rather than local residents.

The ability to analyze historical daily visitation would also allow scholars and park managers to investigate whether visitation shifts are a direct result of management policy changes (e.g., the implementation of day use permits) or acute events (e.g., wildfires and media exposure). The estimated daily visitation can be used to capture the degree of potential inter-month temporal spillovers and spatial substitutions among nearby sites. The methodology presented in this study contributes to parks' existing analytic toolbox for reconstructing granular visitor use data. My approach focuses on an easy and reproducible implementation of downscaling visitation, which has the potential to assist park management decision-makers in evaluating the efficacy of past management strategies and further adjusting their current programs. Knowledge of finer-scale

information on visitation changes is useful for park managers to mitigate congestion and incrementally adapt their management actions to address the ever-changing visitation trends in the time of the COVID-19 pandemic.

While existing SafeGraph data is a helpful proxy for daily visitation, I should note that the estimated daily time series needs to be validated with ground-truth visitor counts. Although park-level daily visitation to NPS-managed units is not available for public access, other well-studied data sources on a daily scale create exciting alternative opportunities for validation purposes, for example, the daily data on camping reservations from Recreation.gov (Rice et al., 2021; Gellman et al., 2022). For parks with a sufficient sample of campsites, further validation of disaggregated visitation estimated from SafeGraph POIs nearby these campsites should be a high research priority. Another potential application of the proposed downscaling approach using SafeGraph data is the ability to apply my automated search algorithm not only to other NPS-managed units (e.g., national monuments and national recreation areas), but to any recreation area with long-term monthly or annual visitation data. Of considerable interest would be estimating temporal granular visitation data for national forests covered by National Visitor Use Monitoring (NVUM) data and state and city parks with consistent visitation statistics. This opens up the opportunity of investigating which types of recreation areas are suitable for using SafeGraph data to reconstruct finer-scale historical visitation data in broader terms.

3.7 Tables

Table 3.1 Summary Statistics

Code	Park Name	Annual Visitation	Annual SG visits	# of in-park POIs	Park Area (thous acres)	Altitude	Elevation	Eco-region
ARCH	Arches NP	1,433,041	14,890	6	76.4	38.7	1,454	North American Deserts
BLCA	Black Canyon of the Gunnison NP	269,810	-	1	31.2	38.6	2,313	North American Deserts
BRCA	Bryce Canyon NP	1,965,920	37,552	18	36	37.6	2,366	North American Deserts
CANY	Canyonlands NP	648,967	4,639	4	334	38.2	1,527	North American Deserts
CARE	Capitol Reef NP	1,012,016	3,683	15	244.6	38.4	1,860	North American Deserts
CAVE	Carlsbad Caverns NP	403,913	9,204	5	46.8	32.1	1,502	Temperate Sierras
CHIS	Channel Islands NP	313,952	1,409	4	245.8	34	77	Mediterranean California
CRLA	Crater Lake NP	633,201	2,945	8	182.2	42.9	1,841	Northwestern Forested Mountains
DEVA	Death Valley NP	1,216,998	31,565	40	3434	36.5	971	North American Deserts
GLAC	Glacier NP	2,610,485	9,938	14	1005	48.7	1,717	Northwestern Forested Mountains
GRBA	Great Basin NP	127,429	-	3	76.9	38.9	2,774	North American Deserts
GRCA	Grand Canyon NP	5,127,243	154,469	72	1202.9	36.2	1,493	North American Deserts
GRSA	Great Sand Dunes NP & PRES	397,864	4,084	1	149.2	37.8	2,615	North American Deserts
GRTE	Grand Teton NP	3,199,407	50,799	36	309.4	43.8	2,345	Northwestern Forested Mountains
JOTR	Joshua Tree NP	2,314,889	4,166	14	797.7	33.8	875	North American Deserts
KICA	Kings Canyon NP	573,849	5,500	6	458.8	36.9	3,006	Northwestern Forested Mountains
LAVO	Lassen Volcanic NP	469,784	879	9	107.4	40.5	2,091	Northwestern Forested Mountains
MEVE	Mesa Verde NP	515,088	3,314	7	53.7	37.2	2,248	North American Deserts
MORA	Mount Rainier NP	1,332,293	6,541	16	235.8	46.9	1,567	Northwestern Forested Mountains
NOCA	North Cascades NP	26,911	-	0	499.8	48.8	1,453	Northwestern Forested Mountains
OLYM	Olympic NP	3,077,846	17,632	27	912.5	47.8	930	Marine West Coast Forest
PEFO	Petrified Forest NP	647,389	12,013	6	223	35	1,684	North American Deserts
PINN	Pinnacles NP	222,818	835	2	26.7	36.5	531	Mediterranean California
REDW	Redwood NP	437,180	73,833	15	140.2	41.4	290	Marine West Coast Forest
ROMO	Rocky Mountain NP	3,976,682	11,091	27	266.7	40.4	3,198	Northwestern Forested Mountains
SAGU	Saguaro NP	834,439	65,520	7	92.5	32.2	1,349	North American Deserts
SEQU	Sequoia NP	1,102,968	4,078	14	406.7	36.5	2,673	Northwestern Forested Mountains
WHSA	White Sands NP	551,700	-	1	148.3	32.8	1,206	North American Deserts
YELL	Yellowstone NP	3,942,249	79,050	67	2194.8	44.6	2,447	Northwestern Forested Mountains

YOSE	Yosemite NP	3,893,142	100,740	94	745.5	37.8	2,450	Northwestern Forested Mountains
ZION	Zion NP	3,885,887	52,976	11	148.6	37.3	1813	North American Deserts

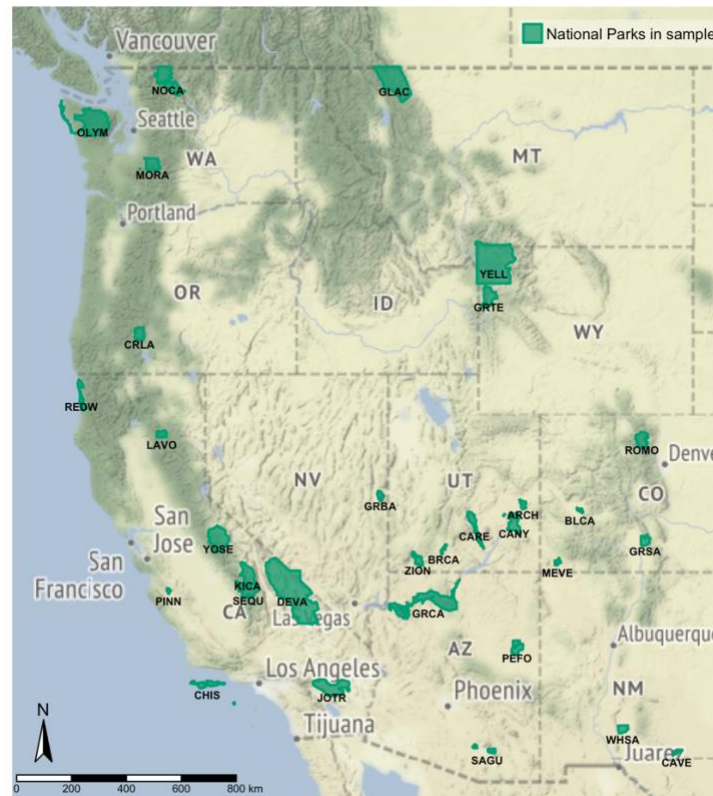


Table 3.1 (cont.)

3.8 Figures

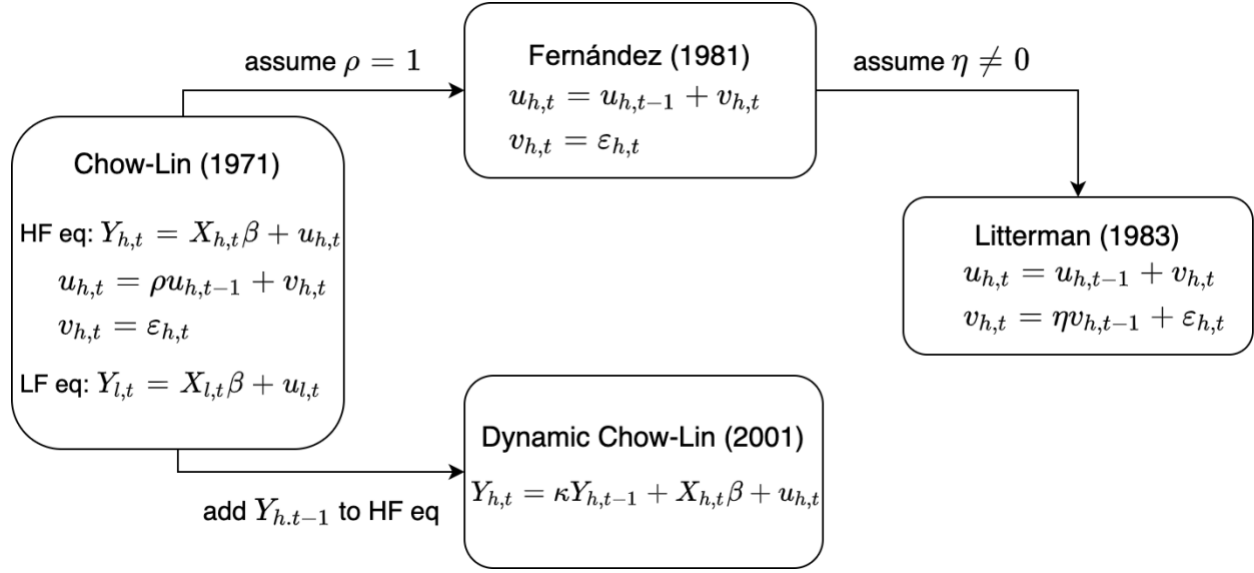


Figure 3.1 Relationship among four Temporal Disaggregation (TD) methods

Note: This figure summarizes the relationship among the four Temporal disaggregation methods. Please refer to Section 3.3.2 in the main text for a detailed comparison between those four methods.

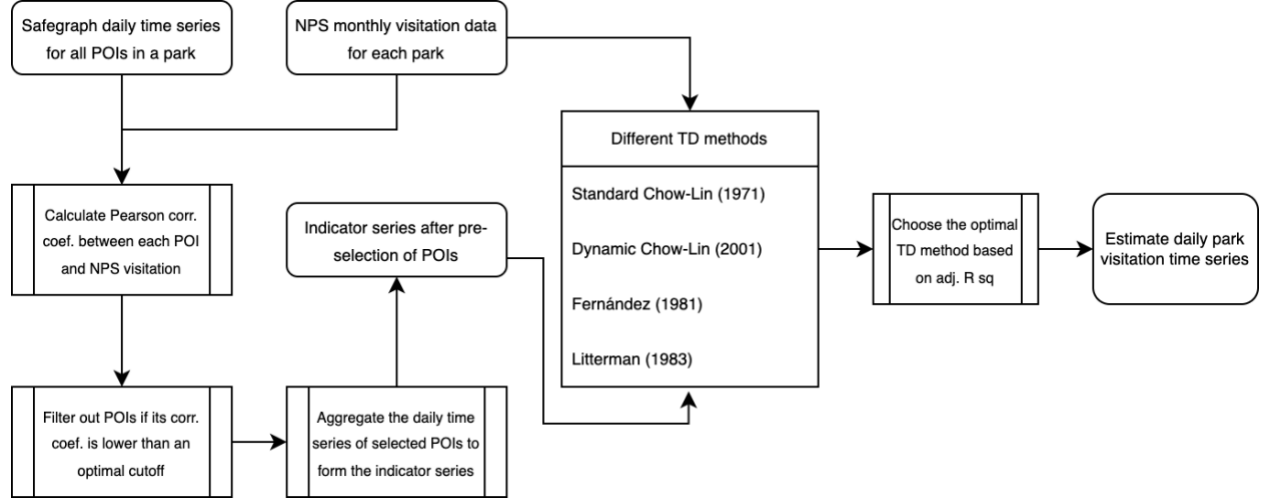


Figure 3.2 A flowchart of automated Point-of-Interest (POI) pre-selection and Temporal Disaggregation (TD) method selection

Notes: This figure summarizes the automatic algorithm conducting Point-of-interest (POI) pre-selection and the Temporal Disaggregation (TD) method selection. Please refer to Section 3.3.2 in the main text for a detailed description of the algorithm.

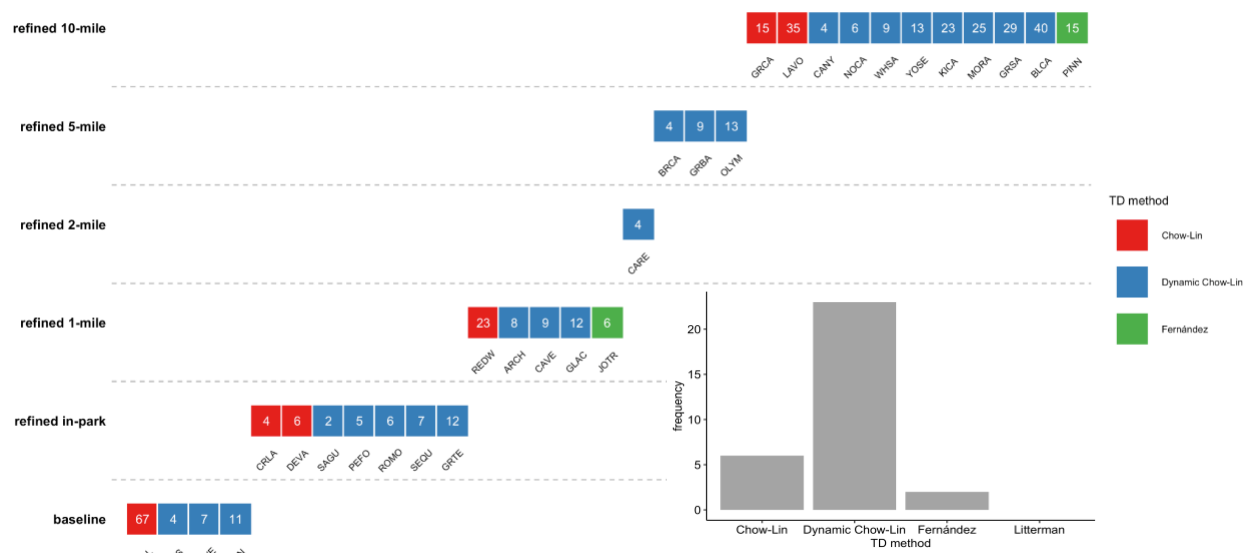


Figure 3.3 Point-of-Interest (POI) pre-selection and optimal Temporal Disaggregation (TD) method selected for each park

Note: The 31 parks (park codes defined in Table 3.1) in the sample fall into 6 different rows. The bottom row represents the baseline model without any pre-selection of POIs. Rows 2-6 (from bottom to top) represent the pre-selection of POIs at different search radii. For example, Row 2, “refined in-park,” represents the pre-selection of POIs applied to all POIs within the park boundaries. Row 3, “refined 1mi,” refers to the pre-selection of POIs applied to all POIs within the 1-mile buffer zone (including the interior of the park). The rows “refined 2mi,” “refined 5mi,” and “refined 10mi” can all be interpreted similarly. The number in the box represents the number of POIs selected to form the final indicator series of interest. The color of the boxes indicates the chosen TD method, and the histogram at the right-bottom corner summarizes the distribution of the chosen TD method.

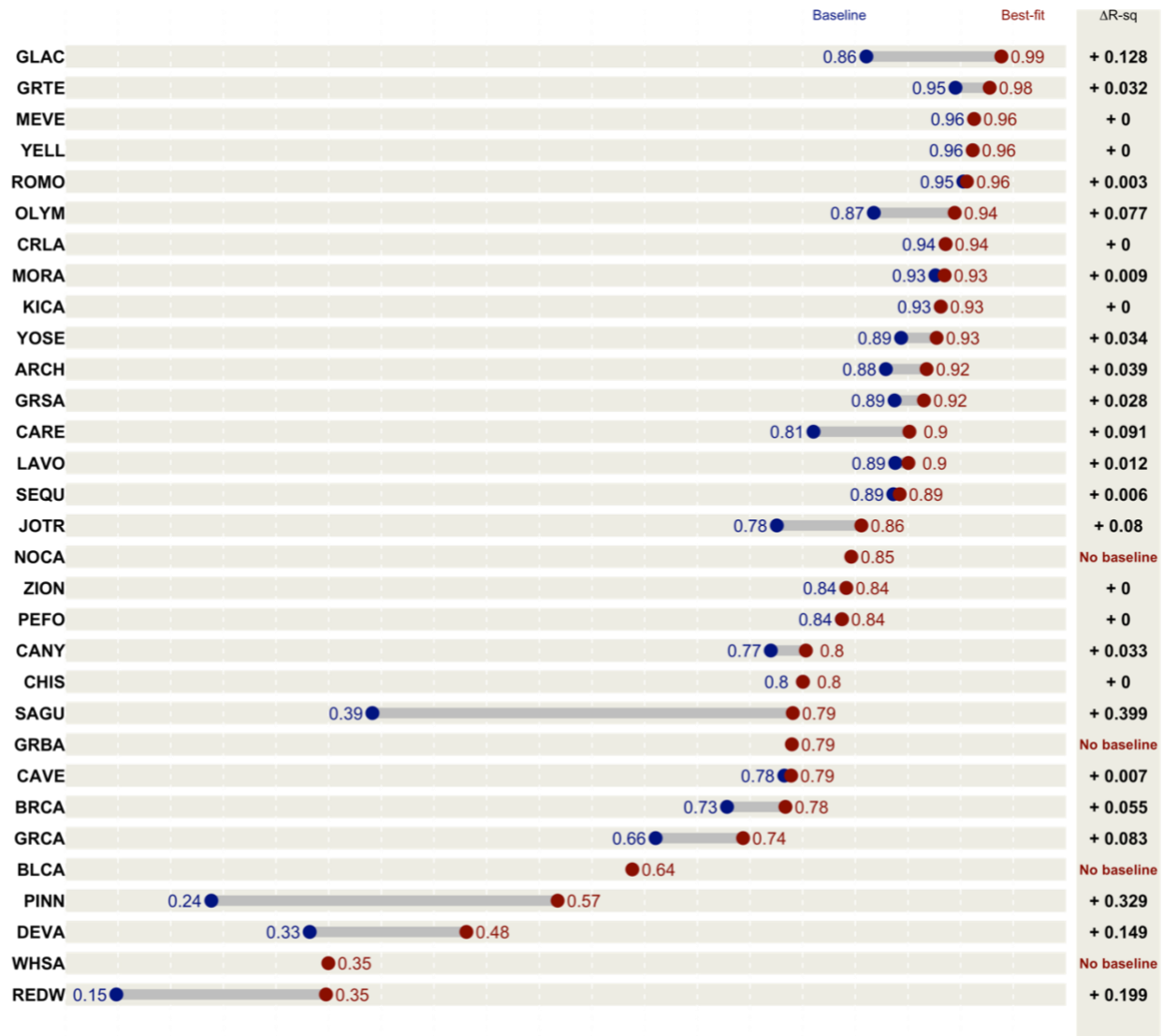


Figure 3.4 Improvement of model performance by park

Note: For each park, the figure shows adjusted R-sq for both the baseline model (in blue) and the best model (in red), with the right margin displaying the performance improvement in terms of changes in adjusted R-sq. Parks are ranked by the adjusted R-sq of their optimal models.

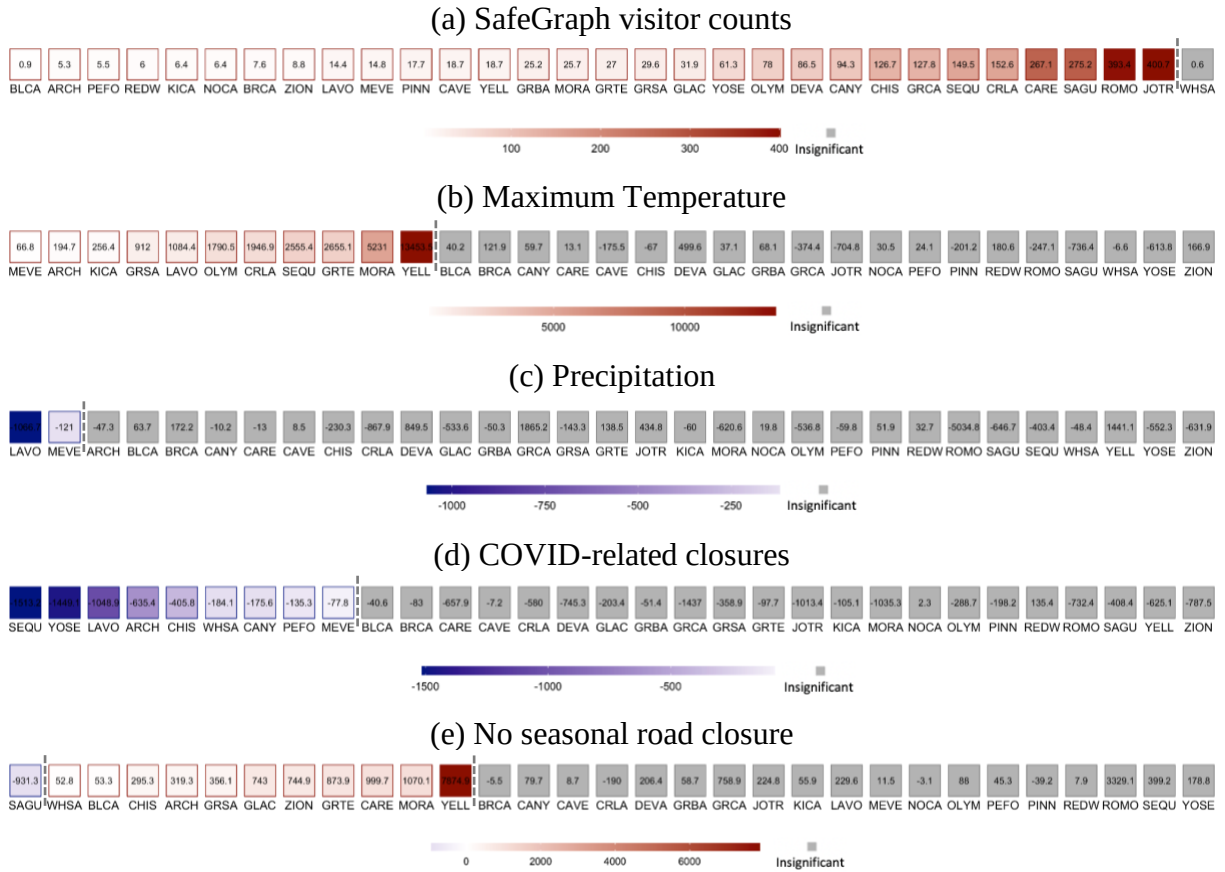


Figure 3.5 Coefficient estimates from the optimal Temporal Disaggregation models by parks

Note: Each panel of the figure displays the coefficient estimates for one indicator series for all 31 national parks in the sample. The park-specific coefficients are ranked by their magnitude from low to high, reflected by the color gradient. A red box denotes a positive coefficient, while a blue box denotes a negative one. Parks with statistically significant coefficients are listed first, followed by parks with insignificant coefficients, which are represented by gray boxes.

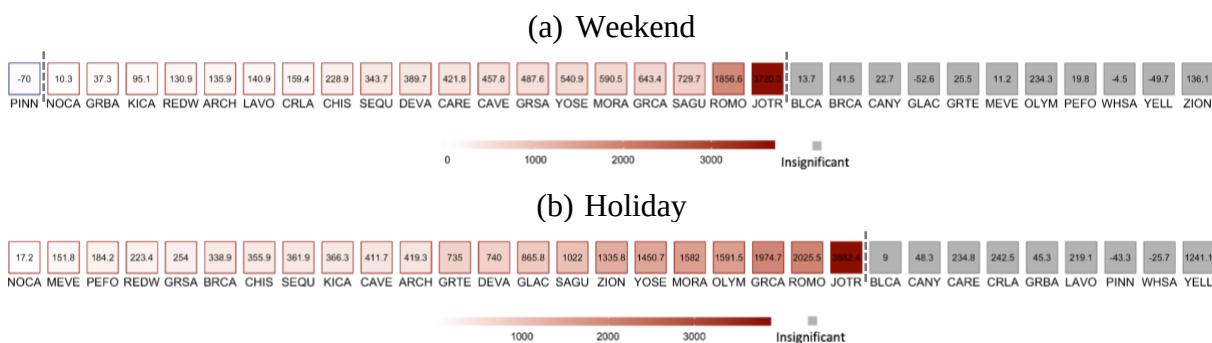


Figure 3.6 Coefficient estimates of the weekend and holiday dummies by parks

Note: Each panel of the figure displays the coefficient estimates from regressing estimated daily visitation on weekend and holiday dummy variables for all 31 national parks in the sample. The park-specific coefficients for each binary indicator variable are ranked by their magnitude from low to high, reflected by the color gradient. A red box denotes a positive coefficient, while a blue box denotes a negative one. Parks with statistically significant coefficients are listed first, followed by parks with insignificant coefficients, which are represented by gray boxes.

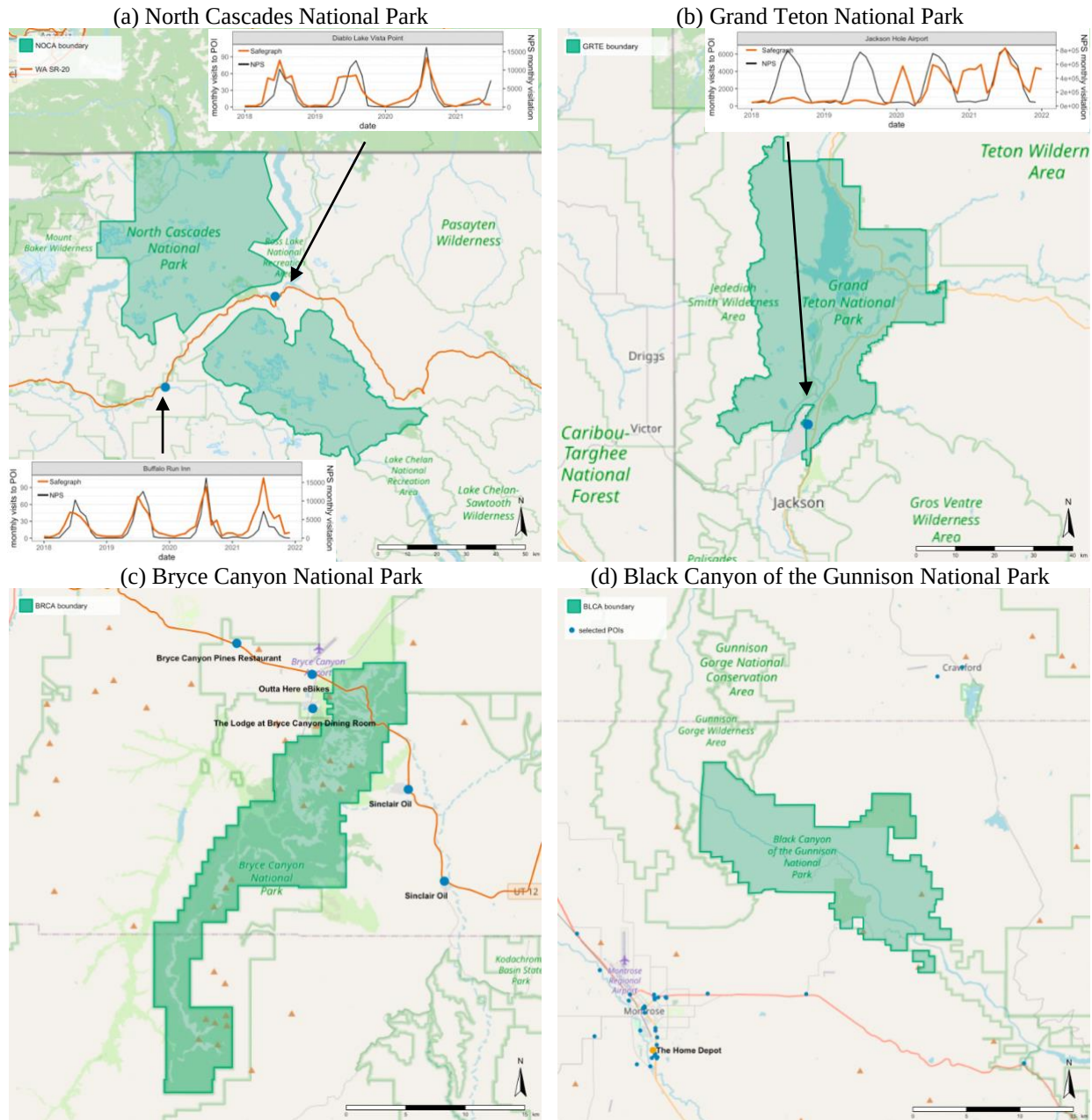


Figure 3.7 Example National Parks with Point-of-Interest (POIs)

Note: Each panel shows a single national park in green. Blue points represent SafeGraph POIs. Panel (a) is the map of North Cascades National Park. The red line is the scenic drive passing through the park (WA SR-20), and two outside-park POIs have a very high correlation with NPS visitation. Panel (b) is the map of Grand Teton National Park. The blue point shows Jackson Hole Airport, which is a within-park POI that has a dominating number of SafeGraph records but is poorly correlated with NPS visitation. Panel (c) is the map of Bryce Canyon National Park. Blue points show several amenities located in two low-population gateway communities that could be considered mainly serving the park visitors. Panel (d) is the map of Black Canyon of the Gunnison National Park. Blue points show the selected POIs in the business district of Montrose, a city with over 20,000 residents. The orange point highlights a Home Depot that is selected by the automatic algorithm. However, it is not easy to detect if the target audience of this POI is park visitors or local residents.

CHAPTER 4:

ON THE EVALUATION OF HETEROGENEOUS CLIMATE CHANGE IMPACTS ON US AGRICULTURE: DOES GROUP MEMBERSHIP MATTER?

4.1 Introduction

With mounting evidence that increased concentration of greenhouse gas will lead to higher temperature and shifting climate pattern, numerous research activities have emerged to estimate the projected effect of climate change on economic well-being. Agriculture, an important sector in the US economy, is a key focus since agricultural production intrinsically depends on climate conditions such as temperature and precipitation. Despite substantial coverage, the literature has not reached a consensus on the impacts of climate change on US agricultural land value. For instance, while Mendelsohn et al. (1994) conclude with an expected impact of +\$34.5 billion, Schlenker et al. (2005) show change in farmland value of −258.9 billion, with a wide range of impact estimates on agricultural gain or loss that often diverge (Adams, 1989; Mendelsohn et al., 1994; Schlenker et al., 2006, 2005; Deschênes and Greenstone, 2007; Schlenker and Roberts, 2009).

The Ricardian approach (Mendelsohn et al., 1994) has become one of the most popular frameworks for the evaluation of climate change impacts on agriculture (for a review, see Mendelsohn and Massetti, 2017). It is based on a simple hedonic equation that regresses farmland value or agricultural profits on climate variables, such as long-term averages of temperature and precipitation, along with other controls that include the soil characteristics and various socioeconomic conditions. Based on the estimated marginal effects of the past climate variables, one can make quantitative projections about the expected changes in agricultural profits given the

projected value of the future climate conditions. In early Ricardian studies, most researchers pooled the various spatial units together to estimate a space-invariant relationship between climate and the agricultural outcome of interest for their entire sample, which was usually the country as a whole. In this framework, a 1% increase in winter temperature is assumed to have the same impact on the northern states—say, Montana—and on the southern states, like Florida. Yet, the appropriateness of this simplifying assumption has been challenged. As Schlenker et al. (2005) point out, when one splits the US sample into rain-fed and irrigated counties, the marginal effect of the climate variables on agriculture differs significantly across groups. As a result, the sign of the projected impact of climate change on the rain-fed counties changes from slightly positive in Mendelsohn et al. (1994), where this type of spatial heterogeneity is disregarded, to unambiguously negative.

Recent years have seen a handful of new contributions challenging the spatial homogeneity assumption in the USA. The “standard” approach to uncover spatial heterogeneity consists of grouping the observations and testing the relevance of these groups. For instance, the entire county-level sample is first partitioned into several mutually exclusive groups known as regions. An example of this approach is that of Schlenker et al. (2005) who define the irrigated region as contiguous counties having more than 20% of harvested cropland irrigated. Next, the different groups are allowed to exhibit distinct marginal effects of climate in hedonic estimation and statistical testing based on the Chow test (Chow, 1960), which is then conducted to determine whether the difference between the region-specific coefficient estimates is indeed significant.

There have been notable variations in the way in which the is are partitioned in the initial step of this approach. The existing literature tends to rely largely on prior knowledge and the ad hoc judgment of the researcher to cluster the counties into regions. The current Ricardian literature

applied to the USA indicates that the grouping can be defined based on agriculture practices such as irrigation versus dryland (Mendelsohn and Dinar, 2003; Schlenker et al., 2006, 2005), on administrative boundaries such as the states used in Deschênes and Greenstone (2007), or on topological features—more especially high versus low elevation—such as those used in Dall’erba and Domínguez (2016).

If the impact estimates were insensitive to the choice of grouping, grouping uncertainty would be far less of a concern. However, if increasing grouping uncertainty leads to considerable variation in impact estimates, and if existing evaluation methods offer little evidence on the superiority of one grouping over the others, then grouping uncertainty is of analytical significance and deserves careful examination. Indeed, the existence of grouping uncertainty reduces the reliability of group-specific impact estimates and may result in potentially poor policy guidance for regional planners.

In light of the importance of these questions, this paper is composed of two stages. In the first stage, we examine the grouping definitions currently used in the Ricardian literature. Considering that part of the variation in group-specific impact estimates stems from the choices the authors make on the sample and the econometric specification, our replication exercises rely on the same sample, Ricardian framework, and model specification, but differ in their definition of spatial heterogeneity based on the groups chosen in the literature. This approach guarantees that the variability in group-specific predicted impacts is exclusively driven by grouping uncertainty. The second stage will evaluate whether one grouping is a better fit than the others. Two procedures, the mean squared prediction error (MSPE) via cross-validation (CV) and the pairwise comparison via the J test, will be used for that purpose.

Once grouping uncertainty is accounted for, we find that the results from stage 1 indicate a high level of variation in the estimated impact of climate change in the US primary agricultural areas (the Great Plains and Northeast). Even a modest change in group structure leads group-specific impact estimates to shift up and down by as much as 25%. As a result, neglecting grouping uncertainty could underestimate the worst-case climate projections (Burke et al., 2015). Additional results from stage 2 complement our findings on the importance of grouping uncertainty. Indeed, the results from both the CV and J test provide limited evidence that any single grouping is superior to its current competitors. In summary, these results urge researchers to consider grouping uncertainty when assessing the heterogeneous impacts of climate change on US agriculture. Future research should consider moving away from pre-determined classifications and towards a statistical identification of the spatial heterogeneity as suggested by recent developments in C-Lasso (Su et al., 2016) and causal forests algorithm (Wager and Athey, 2018).

The paper proceeds as follows: The next section reviews the conceptual basis for the presence of heterogeneity in the Ricardian analysis. Section 4.3 describes the data sources. Section 4.4 presents the econometric specification, while Section 4.5 describes the results based on the groupings used in the literature and compares their predictive capacity with each other's and with the one of a grouping based on climate regions. Section 4.6 offers two alternative methods of reducing the grouping uncertainty. Finally, the last section summarizes the most important findings and provides some concluding remarks.

4.2 A review of the conceptual framework

4.2.1 Ricardian analysis with average slope coefficients

The traditional Ricardian approach relies on cross-sectional variation in the climate conditions to explain the differences in net farmland value. In competitive markets, farmland value (V) should reflect the discounted sum of future net returns by allocating farmland to its optimal use (Mendelsohn et al., 1994; Plantinga et al., 2002):

$$V_i = \int \pi_i^*(t) e^{-rt} dt \quad (4.1)$$

where π_i^* is the optimal net returns to land at location i . According to Schlenker et al. (2006), the latter corresponds to the envelope of the return function associated with each potential use of the land, i.e.:

$$\pi_i^* = \max_{k \in \{1, \dots, K\}} \pi_{ik}(P_k, \mathbf{w}, \mathbf{z}_i) \quad (4.2)$$

where π_{ik} is the land return when the land at location i is devoted to its k th potential use; P_k is the market price of output associated with that land use; $\mathbf{w}$ is a vector for the price of the purchased inputs, while $\mathbf{z}_i = (\mathbf{C}_i, \mathbf{G}_i, \mathbf{S}_i)$ denotes a matrix of exogenous inputs at location i . It consists of a vector of climate attributes, $\mathbf{C}_i$; a set of socio-economic variables, $\mathbf{G}_i$; and a set of soil quality indicators, $\mathbf{S}_i$. The Ricardian model is thus a reduced form model where farmland value $V_i(\mathbf{P}, \mathbf{w}, \mathbf{z}_i)$ depends only on a set of exogenous variables, $\mathbf{z}_i$. The nonlinear relationship between climate data and farmland value is ubiquitously approximated by a quadratic formulation²⁵ of inputs as demonstrated by Hsiang (2016):

$$V_i = \beta_0 + \beta_1 \mathbf{C}_i + \beta_2 \mathbf{C}_i^2 + \delta \mathbf{G}_i + \theta \mathbf{S}_i + \varepsilon_i \quad (4.3)$$

²⁵ The nonlinear effect of climate can also be modeled through M th order polynomials (Schlenker and Roberts, 2009) or a M -piecewise “binned” function (Barreca et al., 2016). Since our focal point is grouping uncertainty, it is beyond the scope of this paper to investigate whether adding nonlinear effects would lead to an efficiency gain.

Since the response function in eq. (4.3) is quadratic, the average marginal effect of climate, C_i , on farmland value depends not only on linear and quadratic coefficients estimated on the overall sample, but also on the climate conditions at location i (Mendelsohn and Dinar, 2003; Vanschoenwinkel et al., 2016). Indeed, Eq. (4.4) reports the following marginal effect:

$$\frac{\partial V_i}{\partial C_i} = \beta_1 + 2\beta_2 C_i \quad (4.4)$$

This design has been extensively used in early Ricardian research to inform the direct impact of the intertemporal change in climate by comparing historical and projected farmland values.

4.2.2 Ricardian analysis with local slope coefficients

Applications of the Ricardian approach that quantify the predictions of agricultural outcomes have been mostly criticized for the framework's implicit assumption of homogeneity of the coefficients β_1 and β_2 across spatial units. Spatial variation in the marginal returns only comes from location-specific past and future climate values, C_i . Polsky and Easterling (2001) offer the first attempt at introducing slope heterogeneity in the Ricardian literature. Their multilevel models demonstrate significant variation in Ricardian climate sensitivity across different spatial scales (county-level vs. district-level). Results show that county-level Ricardian climate sensitivities are partially determined by the climate variability of the agro-climatic zones in which the counties are located. A few years later, Polsky (2004) considers a model that includes the spatial lag of the dependent variable. He concludes that climate sensitivity varies significantly across space even though he does not readily report the marginal effect of the cross-partial derivatives as has become standard since LeSage and Pace (2009). Furthermore, Timmins (2006) clearly articulates that traditional Ricardian regressions might yield biased predictive results when the assumption of slope

homogeneity is violated. His work controls for the land use share of each location²⁶. When the 1,000 locations that compose his sample are shocked with a 5°C increase in temperature, he finds that the traditional Ricardian evaluation would overstate the value of global warming compared to the true marginal effect obtained with simulated data. These results confirm the endogenous and heterogeneous behavior of landowners' decision-making in Timmins's discrete choice approach. Therefore, the results emphasize the need to control for the spatial heterogeneity in the slope coefficients.

Since then, several authors have challenged the slope homogeneity assumption by testing the climate sensitivity of irrigated versus rain-fed counties (Darwin, 1999; Mendelsohn and Dinar, 2003; Schlenker et al., 2005, 2006). Their underlying assumption is that the estimated coefficient of temperature would be biased if both rainfed and irrigated counties are pulled together. Figure 4.1 illustrates the rationale.

The net return, π_i^* /profit π_{ik} , is on the y -axis, and temperature is on the x -axis²⁷ (Figure 4.1A). The three production-specific profit functions provide a simplified example of the relationship between profits ($\pi_{Ak}, k = 1,2,3$) and temperature (T) in rain-fed location A when crops 1, 2, and 3 are chosen. These profit functions are denoted by the gray lines. Now consider an incremental increase in temperature from T_1 ²⁸: the use of irrigation water improves the heat tolerance of crops, so the detrimental effect of warmer temperatures in irrigated locations is expected to be smaller than in rain-fed locations. Furthermore, since farmland value is the present value of future net returns, a wider choice of adaptation adjustment within each location leads to a

²⁶ In Timmins's (2006) approach, the categories of land use are forest, pasture, temporary crops and permanent crops.

²⁷ For simplicity, we assume here that the effect of precipitation and all other exogenous determinants of net return or profits are appropriately controlled for.

²⁸ Figure 4.1 is adapted from Mendelsohn et al. (1994). We choose temperature on the x -axis to illustrate the rationale of varying slope in the Ricardian framework. A similar conceptual framework could have been done for precipitation.

flatter and smoother response of farmland value to temperature (Figure 4.1B). Irrigation changes the response between farmland value and temperature. The marginal effect of temperature in locations A or B is graphically identified by distinct slopes of a tangent line at $T = T_1$.

A series of studies have introduced slope heterogeneity by partitioning the data into distinct groups to generate group-specific projections on the US agriculture. A summary of their modeling details, including the choice of the outcome variables, the climate/weather variables, the sample, and the grouping choices can be found in the appendix Table C.1. Cross-sectional models examine the long-run relationship between agricultural outcomes and climate conditions by incorporating adaptation. However, this approach is most susceptible to the omitted variables bias. Panel models can reveal a causal relationship by relying on exogenous weather shocks and controlling for unobserved spatial differences, but they do not explicitly model long-term climate conditions. It is beyond the scope of this study to determine which serves as a better linear approximation of the underlying relationship, and which model's prediction would uncover the "true" climate effect. This important question is still under debate.

Figure C.1 in the appendix maps the membership used in each grouping. The first is the seminal paper by Schlenker et al. (2005) (SHF) that highlights the importance of separating the rain-fed from the irrigated counties. The latter are defined as the counties in which more than 20% of the harvested cropland is irrigated. Once irrigation is accounted for, the authors find that the projected climate impact estimates for rain-fed areas converge to an annual loss of \$9.62 billion²⁹. Building on SHF, Deschênes and Greenstone (2007) (DG) use a panel approach to study the impact of year-to-year weather fluctuations on agricultural profits. Their work uses a cutoff of 10% of a county's cropland area under irrigation to define the irrigated counties. Their findings generate an

²⁹ All figures are in 2012 constant US dollars.

impact of future weather conditions estimated at $-\$3.23$ billion. Note that neither DG nor SHF provide a clear explanation about the choice of the cutoff. Yet, SHF make use of the Chow-Wald test to verify that their results are not sensitive to other cutoffs (at 5%, 10%, 15%). In addition, DG look at the different responses to climate across states and allow the climate coefficients to fully interact with state dummies³⁰. Their results indicate that North Carolina and California are expected to experience a severe loss, while South Dakota and Georgia will significantly benefit from climate change.

More recently, Dall’erba and Domínguez (2016) offer a Ricardian evaluation of the counties in the Southwest USA, one of the sunniest and driest regions in the country where farming nevertheless occupies 20% (Utah) to 55% (New Mexico) of the land. They split their sample into two groups based on the median value of elevation in order to separate the marginal effect of the regressors in “the mountains, high plains and plateaus of Colorado, northern New Mexico, northern Arizona and the Northeastern part of Utah from the marginal effect in the less elevated counties of southern Arizona, southern New Mexico and the eastern part of Colorado” (Dall’erba and Domínguez, 2016, p. 54). Their results indicate that the probability of having a greater negative projected impact is larger in the highland counties than in the lowland counties, as the former are expected to see more important changes in climate conditions and rely less on irrigated water to compensate for them. Note that for our analysis, we use the median elevation level of the continental counties.

Given the large discrepancy between the groupings available in the literature, the practitioner is faced with the challenge of choosing one option over the others. The fact that the Chow test is not sensitive to other irrigation cutoffs, as reported in SHF, makes it insufficient for

³⁰ Some states are grouped together though. One group is composed of AZ and NV. Another one includes CT, DE, MA, MD, ME, NH, NJ, RI, and VT. The authors do not explain the reason(s) for such groupings.

establishing the superiority of one grouping over the others. Therefore, the question of which grouping should be preferred remains largely unresolved, and, if the uncertainty associated with grouping choice is not addressed, the actual confidence interval associated with climate change impact estimates may be much larger than that reported in each individual study.

4.3 Data

The sample used in this paper encompasses the 2,767 continental US rural counties³¹. We disregard the urban counties since farmland value in the urban counties depends largely on urban growth at the urban fringe (Plantinga et al., 2002). The dataset covers the four most recent USDA agricultural census years: 1997, 2002, 2007, and 2012. An overview and detailed description of all the variables and their source can be found in the supplementary materials Table C.3.

Agricultural outcomes—The outcome variables differ in the cross-sectional model and panel model. In the cross-sectional model, the dependent variable is the county’s average farmland value per acre. In the panel model, the outcome of interest is the county-level agricultural profits from farming operations per acre. Following Schlenker et al. (2006) and Deschênes and Greenstone (2007), we construct the value of profit as the difference between revenue from agricultural products and farm production expenses, and we divide it by the number of acres of farmland. All values have been converted to 2012 constant dollars using the PPIs for farm products collected by the Bureau of Labor Statistics.

The independent variables can be classified into three categories:

³¹ The rural counties are defined as the counties having a population density below 400 inhabitants per square mile (Schlenker et al. 2006; Dall’erba and Domínguez 2016). We drop 34 rural counties that have missing socio-economic or soil data for 1 or more years. In addition, there are 143 instances missing agricultural profit observation, which are replaced with the average of the remaining years’ values. The sample restrictions are imposed to provide a balanced panel of counties from 1997 to 2012 for the subsequent analysis.

Climate and weather data—Both climate and weather data are drawn from the National Centers for Environmental Prediction (NCEP)’s North American Region Reanalysis (NARR) database. NARR provides high-resolution climate data on both spatial (30 km by 30 km) and temporal (3-h) aspects. Most cross-sectional studies find that seasonal measures are important, so we calculated the previous 20 years’ (1993–2012) average monthly temperature (°C) and precipitation (mm/day), then aggregated them into four seasons: winter (Dec.–Feb.), spring (Mar.–May), summer (Jun.–Aug.), and autumn (Sep.–Nov.). Instead, panel studies suggest using degree days to model the effect of annual weather fluctuation (Deschênes and Greenstone, 2007; Blanc and Schlenker, 2017). In order to construct the growing season degree-days and the growing season precipitation, we collected the daily average temperature and daily precipitation for each census year. The degree-days are calculated so that a day with a mean temperature at or below 8°C results in 0 degree days; a day with a mean temperature between 8°C and 32°C constitutes the number of degrees above 8°C; a day above 32°C results in 24 degree-days³². We calculate the growing season degree-days by summing up these daily degree-days over the entire growing season (April 1 to September 30). Growing season precipitation is calculated as the sum of precipitation over the growing period.

Socioeconomic controls—As suggested by earlier studies (Deschênes and Greenstone, 2007; Fisher et al., 2012), per-capita income and population density influence agricultural farmland values because a higher population density proxies for the larger likelihood of conversion to urban use and higher per capita income translates into higher local demand for agricultural goods.

³² Recent research shows that the threshold to calculate growing season degree-days is broadly representative of major crops such as corn, soybeans, and cotton (Ritchie and NeSmith 1991). Some crops and varieties have different growing seasons. For example, winter wheat is planted in September through November and harvested in late Spring and early Summer.

The density and income data are all collected from the Bureau of Economic Analysis (BEA) for each census year.

Soil quality data—The soil quality variables come from the US Geological Survey (USGS). They include salinity, flood frequency ratio, wetland ratio, slope steepness, soil erosion, sand content, clay content, moisture capacity, and permeability³³. These soil quality variables are extensively used in Ricardian studies.

4.4 Research method

This section describes the econometric method for conducting replication analysis and performance evaluation. It is composed of four subsections. Section 4.4.1 briefly introduces the two types of models used in the Ricardian approach. Section 4.4.2 describes how homogenous models may produce biased results when heterogeneity is present. In section 4.4.3, we specify the fixed coefficient models that allow for group heterogeneity and that are used for replication purposes. We also discuss the impact of grouping uncertainty. Section 4.4.4 proposes two evaluation methods that attempt to address the grouping uncertainty statistically.

4.4.1 Homogenous benchmark model

We begin with the cross-sectional setting and then carry on with the panel approach commonly used in the Ricardian literature. The standard cross-sectional model is given by:

$$V_{it} = \beta_1 T_i + \beta_2 P_i + \beta_3 T_i^2 + \beta_4 P_i^2 + \delta G_{it} + \theta S_i + \eta_t + \varepsilon_{it} \quad (4.5)$$

where V_{it} is the value of farmland per acre in county i measured in a specific year t for $i = 1, \dots, N$, and t is the year of interest. The variables of interest are the measures of temperature and precipitation, where T_i captures, for each season, the average temperature over the 20 or 30 years

³³ The soil data and the weather conditions are initially two raster datasets. We aggregate each of them to the county level using the ArcGIS toolbox.

before t . $\mathbf{P}_i$ is the equivalent vector for precipitation. β_1 to β_4 are the main parameters of interest in this study. $\mathbf{G}_{it}$ is a vector of socio-economic variables usually composed of density and income, while $\mathbf{S}_i$ captures the soil determinants. ε_i is the error term with the usual assumptions.

Deschênes and Greenstone (2007) and Fisher et al. (2012) rely on the panel version of the Ricardian estimation where the core relationship is between the weather variables and the agricultural profits measured across $i = 1, \dots, N$ and $t = 1, \dots, T$ as follows:

$$y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 P_{it} + \beta_3 T_{it}^2 + \beta_4 P_{it}^2 + \delta \mathbf{G}_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad (4.6)$$

Two major differences arise in eq. (4.6) relative to eq. (4.5). First, the dependent variable is now y_{it} , which represents the per-acre profit of county i in year t . Second, the climate measures are replaced by growing season degree-days, T_{it} , and growing season precipitation, P_{it} . μ_i are the location fixed effects and η_t are the year fixed effects. ε_{it} is the error term with the usual properties. The county fixed effects control for the difference of all time-invariant factors between counties, including the soil characteristics. The year fixed effects are included to account for all the unobservable annual factors, such as the national variation in price across years.

4.4.2 Slope heterogeneity

This subsection illustrates the implications of slope heterogeneity on the consistency of the panel model only, as the derivation for the cross-sectional model is similar. The validity of the homogenous benchmark model captured in eq. (4.6) depends crucially on the assumption that the fixed-effects estimators will produce unbiased estimates of β , i.e., $E(\hat{\beta}^{FE} | T_{it}, P_{it}, \mathbf{G}_{it}) = \beta$ (Wooldridge, 2010). However, if the unknown slope coefficient β differs between individual counties due to unobserved heterogeneity, then $\hat{\beta}^{FE}$ is biased. To illustrate the consequence of erroneously neglecting such slope heterogeneity on the consistency of homogenous fixed-effect

estimators, we consider the following data-generating process with complete slope heterogeneity applied to one covariate only³⁴:

$$y_{it} = \beta_0 + \beta_{1i}T_{it} + \beta_2P_{it} + \beta_3T_{it}^2 + \beta_4P_{it}^2 + \delta G_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad (4.7)$$

$$\beta_{1i} = \beta_1 + \gamma_i \quad (4.8)$$

Eq. (4.7) has more coefficients than the homogenous benchmark model since the slope coefficient for linear terms of growing season degree-days, β_{1i} , is fixed over time but is allowed to vary freely across counties. Note that the homogenous benchmark model of eq. (4.6) is a special case of eq. (4.7) with the following restriction: $\beta_{11} = \dots = \beta_{1N} = \beta_1$. In eq. (4.8), a county-specific slope coefficient, β_{1i} , is composed of a common slope, β_1 , and of a fixed constant, γ_i , that captures the difference between the county-specific slopes of the linear terms of the growing season degree-days.

Pesaran (2015) proves that, if slope heterogeneity is present but is ignored, the application of fixed-effects estimators to the homogenous benchmark model in eq. (4.6) could be biased to a degree that depends on the heterogeneity of β_1 . The direction of such a bias follows the sign of the correlation between γ_i and the regressors C_{it} and G_{it} . Blanc and Schlenker (2017) provide an example of how pooling all the counties together can produce an upward bias of the marginal effect of temperature in the irrigated counties when the slope coefficient of temperature differs across rain-fed and irrigated counties, and the use of irrigated water is, as expected, positively correlated with temperature. In the context of eq. (4.8), this involves setting $\gamma_i = \pi\alpha_i$, where α_i is a binary variable indicating whether county i has access to irrigation or not, and π is a parameter to estimate. Thus, equations (4.7)–(4.8) can be rewritten as:

³⁴ Under this assumption, slope coefficients are treated as completely different across different individual units, and each β_i is estimated by a separate regression. Traditional panel models frequently assume complete slope homogeneity. Extension to other covariates is straightforward.

$$y_{it} = \beta_0 + \beta_{1i}T_{it} + \pi(\alpha_i \times T_{it}) + \beta_2P_{it} + \beta_3T_{it}^2 + \beta_4P_{it}^2 + \delta\mathbf{G}_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad (4.9)$$

Compared with the data-generating process in eq. (4.9), the homogenous benchmark model in eq. (4.6) fails to include the term $\pi(\alpha_i \times T_{it})$, which implies that the coefficient estimates based on eq. (4.6) are biased. Because the presence of irrigation water mitigates the detrimental effect of extreme heat on agricultural profit in irrigated counties, we anticipate a positive coefficient ($\pi > 0$). The fixed-effect estimator of β_1 in eq. (4.6), noted $\hat{\beta}_1^{FE}$, will pick up the beneficial effect of irrigation and overestimate the true effect of the linear terms of temperature. In other words, since α_i is positively correlated with T_{it} ($cov(\alpha_i, T_{it}) > 0$), and the coefficient of the interaction between irrigation and temperature is positive ($\pi > 0$) in accordance with the aforementioned assumption, we expect a positive correlation between γ_i and temperature T_{it} : $cov(\gamma_i, T_{it}) > 0$. Following Pesaran (2015)'s derivation, we thus anticipate an upward bias of the fixed-effect estimators of β_1 from eq. (4.6), i.e. $E(\hat{\beta}_1^{FE} | T_{it}, P_{it}, \mathbf{G}_{it}) > \beta_1$. Hence, if the slope heterogeneity of the climate variables is driven by the presence of irrigation or by differences in irrigation levels, then neglecting the slope heterogeneity can lead to overestimating the temperature coefficients.

4.4.3 Group heterogeneity

A traditional solution to potential slope heterogeneity is to run a separate regression for each county and allow the slope coefficients to vary completely across counties. Ideally, a county-specific estimator of coefficients from each separate regression should be unbiased. However, this approach is sensitive to the incidental parameter problem if location fixed effects μ_i are included. In addition, estimators of each county-specific slope are inconsistent when T is finite (Lancaster, 2000). Recent empirical studies consider approximating complete heterogeneity by group heterogeneity where the sample is divided into heterogeneous groups and slope parameters are the same within each group (Cheng, 2014; Su et al., 2016). Group heterogeneity is modeled by a fixed

coefficient model that expands the benchmark model by interacting the climate/weather variables with group indicators. In the cross-sectional setting, the fixed coefficient model is written as:

$$V_{it,j} = \beta_{1j}T_{i,j} + \beta_{2j}P_{i,j} + \beta_{3j}T_{i,j}^2 + \beta_{4j}P_{i,j}^2 + \delta G_{it} + \theta S_i + \eta_t + \varepsilon_{it} \quad \text{for } i \in j \quad (4.10)$$

while in a panel setting, it is specified as:

$$y_{it,j} = \beta_0 + \beta_{1j}T_{it,j} + \beta_{2j}P_{it,j} + \beta_{3j}T_{it,j}^2 + \beta_{4j}P_{it,j}^2 + \delta G_{it} + \mu_i + \eta_t + \varepsilon_{it} \quad \text{for } i \in j \quad (4.11)$$

where j indicates the groups (i.e. regions, $j = 1, \dots, J$). β s and β s are fixed constant but differ across groups. Each weather/climate variable is created as, for example, $T_{i,j} = T_i \times D_j$, where the dummy indicator $D_j = 1$ if $i \in j$. Note that any impact projection based on eq. (4.10) or (4.11) for different choices of grouping may yield different point estimates.

Given the fixed coefficient model with the group-specific parameters, the challenge lies in finding a correct grouping. Because the driving force of slope heterogeneity is unobserved, the true numbers of groups and group membership are entirely unknown to the researchers. As Section 4.2.2 highlights, since the existing definitions of group structure tend to rely on a combination of prior knowledge and the authors' judgment, a large degree of grouping uncertainty is present in the literature. Even when the source of slope heterogeneity is clearly defined as irrigation, grouping uncertainty on the choice of the correct irrigation cutoff remains.

4.4.4 Evaluating the relative performance of the existing groupings

In order to evaluate if one grouping is more “appropriate” than another, we rely on two methods, namely the leave-one-out cross-validation (LOOCV) technique and the non-nested tests. The LOOCV is closely related to out-of-sample prediction, a procedure widely used in the panel data literature to compare the performance of alternative specifications (Baltagi et al., 2014; Brücker and Siliverstovs, 2006). It involves first partitioning a panel dataset into two subsets, a training set and a testing set, which can be split either in the cross-sectional dimension or in the time-series

dimension. In the case of the cross-sectional dimension, a single county's observations $\{(y_{1t}, \mathbf{x}_{1t}), \forall t = 1, \dots, T\}$ are reserved as the testing set and the remaining observations $\{(y_{it}, \mathbf{x}_{it}), \dots, (y_{Nt}, \mathbf{x}_{Nt}), \forall i = 2, \dots, N, \forall t = 1, \dots, T\}$ comprise the training set. For each of the possible groupings, g , the fixed coefficient model is fitted on the remaining $(N - 1) \times T$ training set and T fitted values $\hat{y}_{it}(g)$ are computed for the reserved county using its observations $\mathbf{x}_{1t}$ for $t = 1, \dots, T$. A prediction error can be calculated as the difference between the observed value of the dependent variable and the fitted value based on the grouping g : $(\hat{y}_{it}(g) - y_{it})$ for $t = 1, \dots, T$. This procedure is repeated N times to produce $N \times T$ prediction errors. Following Herwartz and Xu (2009), the LOOCV criterion of mean squared prediction error (MSPE) is calculated as follows:

$$MSPE(g) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_{it}(g) - y_{it})^2 \quad (4.12)$$

Minimizing the above criterion function, $MSPE(g)$ w.r.t. g allows us to identify the preferred grouping $\hat{g}$. However, MSPE typically overweights the outliers. As an alternative criterion, we also consider the mean absolute percentage error (MAPE) as it is robust to a large number of outliers:

$$MAPE(g) = 100\% \times \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \left| \frac{\hat{y}_{it}(g) - y_{it}}{y_{it}} \right| \quad (4.13)$$

Since the fixed coefficient models based on alternative groupings are in fact equivalent to non-nested models, the non-nested J test (Davidson and MacKinnon, 1981) serves as a second method for evaluating the relative merits of the various groupings. The test is performed pairwise on two non-nested fixed coefficient models based on different choices of groupings. As Goodman and Dubin (1990) explain, alternative stratification imposed on Eq. (4.11) can be a special case of

non-nested models. Consider two sets of groupings based on different irrigation cutoffs, then we have two testable hypotheses:

H_0 : model A based on the 20% irrigation cutoff used in SHF (2005)

H_1 : model B based on the 10% irrigation cutoff used in DG (2007)

In testing H_0 over H_1 and vice versa, four outcomes may occur: reject both models, reject none, reject H_0 over H_1 or reject H_1 over H_0 .

4.5 Results

This section is divided into two subsections. The first subsection quantifies the importance of grouping uncertainty based on the “group and test” procedure. The second subsection applies the LOOCV technique and the J test to evaluate the relative merits of the alternative groupings used in the literature.

4.5.1 Replication

This subsection implements the “group and test” procedure using the grouping choices that appear in each of the studies reviewed in Section 4.2.2. It involves first partitioning the sample into groups based on the corresponding authors’ choices and conducting a Chow test to determine whether the climate/weather coefficients are similar across the study’s groups. Second, the fixed-coefficient model for each of the possible groupings is estimated in order to generate group-specific projected impacts using future climate data. Specifically, the replication results from the cross-sectional model in Eq. (4.10) are presented in the next subsection while the results from the panel model in Eq. (4.11) are presented in the following one.

When it comes to future weather/climate conditions, we rely on dynamically downscaled predictions from NARCCAP (the North American Regional Climate Change Assessment Program)

simulations. They downscale the global projections from a global climate model (GCM) to the local scale using a regional climate model (RCM). Such future climate data has a spatial resolution of 35–50km. The GCM-RCM combination that we use is the Mesoscale Model 5I (MM5I) paired with the Community Climate Systems Model (CCSM). The driving CCSM utilizes the A2 emission scenario from the fourth Intergovernmental Panel on Climate Change (IPCC) report. It assumes an increase in carbon dioxide to roughly 1.5 times and 3 times the current concentration for the middle and the end of the 21st-century.

Results from the cross-sectional model

Chow tests – Panel A of Table 4.1 summarizes the results of the Chow tests from Eq. (4.10). Our Chow test results suggest that the probability that the seasonal climate variables display dissimilar marginal effects is 75% when the grouping is based on the 10% irrigation cutoff or the median elevation value. When the state grouping is used, the proportion of significant tests reduces substantially to only 31%, which implies that there is no significant variation in several state specific weather coefficients. In fact, this large decline in the proportion of significant tests is in line with the results that will be presented further below, hence undermining the validity of grouping by state.

Group-specific projected impacts – Figure 4.2 shows, for each of the four groups, the cross-sectional impact of the predicted change in climate on county-level farmland values (measured in millions of 2012 dollars). Future climate is calculated as an average over the 2038–2070 period under MM5I-CCSM projections for the A2 scenario. The group boundaries used in each of the panels A–D appear in black. For example, panel A summarizes the group-specific projected impacts with the 20% irrigation cutoff used in SHF. These estimated impacts derive from the projected changes in climate conditions and the coefficients estimated for the irrigated and rainfed

groups in Eq. (4.10). The estimates are then adjusted by an implicit discount rate³⁵ of 2.9% and the share of acres of farmland in each county to generate the changes in total agricultural profits (Deschênes and Greenstone, 2007). This discount rate is chosen to follow SHF's figure and is applied to all specifications, so it does not affect our conclusions.

The results of Figure 4.2 indicate that the estimated impacts of climate change on profits per acre are sensitive to the grouping choice. If the extent of grouping uncertainty is relatively small, there would be very few differences in group-specific impacts generated by alternative classifications. However, tremendous variation emerges in the estimated impacts of climate change in profits per acre if the choice of grouping switches from one to another. For instance, when the irrigation ratio cutoff drops from 20% (Figure 4.2A) to 10% (Figure 4.2B), the estimated gain in profits due to climate change in the rain-fed sample increases from \$48.4 million to \$57.38 million. We will revisit this comparison between the two irrigation cutoffs in Figure 4.3A. Second, when the grouping is based on states as in DG (Figure 4.2C), the impact estimates vary substantially across state borders so that neighboring counties from contiguous states display extremely different values. An example is the southern part of Illinois where no change in profits is expected in the future while, across the border, counties in Missouri should expect large gains. This result undermines the validity of the grouping by states. Finally, the results in Figure 4.2D based on the elevation cutoff display a pattern that is similar to the one of Figure 4.2A–B. However, Figure 4.2D suggests a slight decrease in the overall impact estimates, especially in the Great Plains, the Mississippi Delta, and the Northwestern forested mountains. We complement these figures with the estimated impact of climate change on the entire nation in the supplementary materials Table

³⁵ The discount rate is the ratio of net farm income to aggregate farm value. It allows us to convert annual profits into land values.

S4. These figures complement previous observations about the role of the grouping choice on the predicted impact.

Due to the great similarity in the results based on the cutoff level of irrigation set at 10% or 20%, we investigate in Figure 4.3A the difference (in %) between the estimated impact of climate change in these two specifications. The bimodal histogram indicates that increasing the cutoff level from 10 to 20% results in 40% of the counties experiencing a decrease of 25% or more in the impact estimates (mostly in the upper Midwest and the Appalachia/Northeastern parts of the country), while 12% of the counties (West of the Great Plains and in Florida) experience at least a 25% increase. Part of the explanation is that 359 counties (13% of the sample, delimited by black contours in Figure 4.3A) switch from the irrigated to the rainfed group when the cutoff increases, and a large majority of them (in Texas, in the Southeast and Southwest of the country) display a drop in the estimated impact. In summary, the estimated impacts of climate change are very sensitive to the numbers of counties that are assigned to the irrigated and rainfed categories.

Results from the panel model

Chow tests – When we repeat the above exercise to the panel model with fixed-effect estimators as in Eq. (4.11), we find that the qualitative results of the joint Chow test stay unchanged (Table 4.1). However, the ratio of significant Chow tests indicates that each of the four weather coefficients (growing season degree day and precipitation) is significantly different at the 5% level for both the 20% and the 10% irrigation cutoffs. The ratio of significant tests is 42.8% for the state grouping while it was 31% in the cross-sectional setting. On the other hand, the specification based on elevation displays a decrease from 75% in cross-section to 50% in panel. As a result, and with the exception of the elevation setting, the results in cross-section are more sensitive to the choice of grouping than those in panel.

Group-specific projected impacts – The four panels of Figure 4.4 report the group-specific projected impacts based on the panel model. We note in Figure 4.3B that when the irrigation cutoff drops from 20% (Figure 4.4A) to 10% (Figure 4.4B), around one-fourth of the counties experience a decrease of at least 25% in the magnitude of the impact estimates. This change is particularly acute in the SouthWest. Another notable feature of the panel estimates is that the histogram of the difference in projected change between the two cutoffs above (Figure 4.3B) is much more centered around zero than those based on the cross-sectional model (Figure 4.3A). Therefore, estimates from the cross-sectional model are more sensitive to the definition of the irrigation cutoff level than the panel-based estimates.

Furthermore, our results in Figure 4.4C further demonstrate that state-specific projections display a broad range of values across the country and are very peculiar. For instance, the impact estimates vary dramatically across neighboring states like Arkansas and Texas or Illinois and Indiana, and nearly half of the sample display non-significant impact estimates (at the 5% level). Finally, Figure 4.4D displays an overall change that is greater than in Figure 4.4A–B (see Appendix Table C.4 for the exact value) and many counties in the Southern part of the country display non-significant changes in profits. For the others, the pattern of the projected change is in line with that of Figure 4.4A–B.

Overall, our results indicate that uncertainty associated with the appropriate definition of grouping increases the uncertainty around group-specific point estimates of climate change impact. This phenomenon is particularly relevant in the US primary agricultural areas of the Great Plains and the Northeast. This type of uncertainty is nontrivial for four reasons: first, for the latter areas, the change in impact estimates due to grouping uncertainty is always in one direction. This fact is especially true when switching from the grouping of SHF or DD to that of DG, which would result

in underestimating the severity of climate projections (Burke et al., 2015). Second, even when there is only a modest change in the rule of grouping (i.e., dropping the irrigation cutoff from 20 to 10%), the impact estimates are found to differ by as much as 25% for more than half of the US counties. Third, our results indicate that grouping by states is the least reliable option because adjacent counties that belong to neighboring states are found to display large discrepancies in their estimated impacts. Fourth, grouping uncertainty is more acute in the cross-sectional than in the panel approach. This result is largely driven by the fact that the coefficient estimates are more reliable in the panel model due to county fixed effects. As a result, the current “group and test” approach used in the literature disregards grouping uncertainty and generates regional impact projections that are not as reliable as they seem.

4.5.2 How good are the groupings?

Due to the latent group structure and the lack of evidence that any single grouping outperforms the others (besides the least believable results that the state-level grouping generates), this section offers an alternative grouping option and tests its predictive effectiveness compared to existing groupings. We will rely on climate regions (climate sections) as defined by the National Oceanic and Atmospheric Administration (NOAA); see Appendix Figure C.2. They are climatologically defined groups of adjoining states that are climatically similar from a historical perspective (Karl and Koss, 1984). Climate regions have been widely used in the climate economics literature (see, for instance, Barreca et al., 2016; Heutel et al., 2021) and in atmospheric science (e.g., Hoerling et al., 2016).

Following Section 4.4.4, Table 4.2 summarizes the predictive accuracy of four groupings based on cross-sectional models. We focus on the results from cross-sectional models rather than

the panel approach because the former is more sensitive to grouping uncertainty than the latter³⁶. Note that we also include the benchmark model with no grouping to report the efficiency gained from the fixed coefficient models. Two criteria are used to measure the predictive effectiveness. They are the mean squared prediction error (MSPE) and the mean absolute percentage error (MAPE). Both are commonly used in the literature (e.g., Brücker and Siliverstovs, 2006). The results indicate that any form of grouping besides the one based on elevation leads to a better predictive power than the model with no grouping. While the groupings based on irrigation and elevation display very similar predictive accuracy levels (similar adjusted R^2), we find that the grouping based on the climate regions has superior predictive accuracy compared to any other options. It achieves an average difference of only 31.14% between the predicted and the observed farmland values. However, the F-test on the R^2 associated with each model specification does not find that the difference with the benchmark model's R^2 is statistically significant for any specification. This idea is explored further next.

Table 4.3 reports the p-value of the J test, which is a pairwise comparison between two competing groupings. Each entry lists the test results from the H_0 grouping (by column) and the H_1 grouping (by row). All but one of the J test results are statistically significant at the 5% significance level, which means that none of the four competing groupings dominates the others. The only exception is the test between the climate region grouping and the alternative defined as the 10% irrigation cutoff. This finding indicates Climate Region is more appropriate than the classification by 10% irrigation cutoff, but none of the four alternative groupings dominates all others³⁷.

³⁶ The results in panel are available from the authors upon request.

³⁷ Table C.5 – Table C.8 report the results of robustness checks based on a log-linear functional form and the presence of temperature-precipitation interactions. These results confirm our main findings.

In sum, the two evaluation methods proposed here can help researchers eliminate less trustworthy groupings. If predictive accuracy is their main objective, then grouping by climate regions seems preferable. However, the results do not provide a comprehensive ranking of all the various options. In particular, we do not find any major difference in MSPE or MAPE between the two groupings based on the irrigation cutoff. The J test results further confirm that no definite conclusion can be drawn in terms of the superiority of a particular grouping. Therefore, grouping uncertainty should not be overlooked in climate impact models, as, until now, there has been no sufficient evidence that any specific grouping captures the “true” data-generating process better.

4.6 The way forward

As noted above, existing approaches investigate regional differences based on pre-determined groups of counties and membership. This section offers two alternative approaches. Both of them are based on a statistical identification of the latent group structure and the estimation of place-specific slope coefficients. The first of these options is the new variant of Lasso, C-Lasso (or classifier-Lasso) proposed by Su et al. (2016) for panel data sets. The advantage of C-Lasso is that it provides a general approach to achieve classification—without a priori knowledge of the group structure—and coefficient estimation simultaneously. To implement C-Lasso, Su et al. (2016) recommend a panel with a long time series ($T \geq 20$) in order to guarantee proper finite-sample performance. Our data do not allow us to directly apply this method since county-level measures of our outcome variable (farmland value/agricultural profit) are only collected once every 5 years in the agricultural census years (only 4 observations in the study period). However, we construct

a state-level dataset and apply C-Lasso to it³⁸. The results, reported in Figure C.3, indicate that two groups should be considered.

When it comes to cross-sectional datasets, one may consider Wager and Athey(2018)'s casual forests algorithm. Their method can adequately capture how the marginal effect of a covariate varies across counties and test if the presence of spatial heterogeneity is significant. There have been considerable applications of causal forests in the study of energy conservation (Knittel and Stolper, 2019; Miller, 2020), labor market (Davis and Heller, 2017), and public health (Wendling et al., 2018). Figure C.4 reports the spatial distribution of the slope coefficients for one covariate, summer temperature, as a way of depicting the place-specific marginal effects it leads to³⁹. The results reveal a small number of clusters across space. We also conduct a test for the presence of spatial heterogeneity in the climate effect and conclude that it is significantly present at the county level ($p\text{-value} < 0.05$). Although this method returns valid confidence intervals for local climate effects, there is less information on how to construct confidence intervals for projected impacts of climate change due to the nonlinearity of the model. Future research should explore the potential use of causal forests to improve climate change impact projections. Also, it is worthwhile to investigate common factors, i.e., the variables selected by the causal forest algorithm that have significant impacts on spatial heterogeneity, and interpret the variations in county-specific marginal effects to gain a better understanding of the heterogeneous climate change impacts on agricultural productivities.

³⁸Because the numerical algorithm in implementation of C-Lasso involves separate regressions at initial step and half-panel jackknife method to correct bias, a panel with long time series is recommended. Because no county-level measures of the dependent variable, profit per acre, exist beyond agricultural census years, we obtain raw data at the state level to construct annual profit per acre from USDA Agricultural Resource Management Survey (ARMS). We then aggregate all the climate and socioeconomic measures of all counties in a state. Using the time span 1993 - 2002, we are able to compile a balanced panel of 48 states. Note that the paucity of state-level panel data results in a classification on a coarser geographic scale than a county-level panel.

³⁹ A similar analysis can be done for any other climate variable.

Beyond the two methods introduced above, we also notice two other efforts in the literature by the scholars who are interested in dealing with the heterogeneity present in Ricardian studies. The first one consists in keeping the grouping variable continuous and interacting it with the covariates. However, consistent estimates of the coefficients require the grouping variable to be exogenous, which is hardly the case in agricultural applications. As we mentioned above, the most commonly used grouping variable in such studies is irrigation ratio. While the endogenous nature of irrigation renders the interaction endogenous too, Aghion et al. (2005) and Bun and Harrison (2019) demonstrate that an instrumental variable approach can be used and generate consistent estimates. The second stream of research is to adapt the geographically weighted regression (GWR) method developed by Brunsdon et al. (1996) and Fotheringham et al. (2003) to quantify the spatial heterogeneity. GWR extends the local polynomial regression on one dimension to a two-dimensional space. The local marginal effect for each spatial unit is estimated using the data in a neighborhood centered by the spatial unit for which the marginal effect is calculated. GWR can be used to generate a marginal effect for each spatial unit as well as for each covariate, and to test the equality across estimates. The local marginal effect estimates then can be used to suggest place-tailored climate change adaptation strategies.

4.7 Conclusion

While a growing number of studies provide an evaluation of the expected impact of climate change, the presence of spatial heterogeneity in the marginal effects has been systematically assessed in the recent literature focusing on the USA. Because various forms of grouping are available in the literature, the question of whether the grouping that is ultimately chosen may affect the value and the precision of the estimated impact of climate change has not been investigated so far. Yet,

grouping is systematically defined on the belief that the authors of each study have perfect knowledge of the true data-generating process and, in some cases, this belief is strengthened by a Chow-test of equality between the marginal effects measured across groups, as well as small changes in the cutoff value.

In this paper, we collect all the existing literature on the US case and apply all the available definitions of groupings on the same sample and same time period. Then, we run our estimates on the cross-sectional case (where the dependent variable is profit) and on the panel case (where it is land value). The results indicate that accounting for grouping uncertainty in these studies increases uncertainty around group-specific point estimates of the climate change impact in US primary agricultural areas. Existing “group and test” approaches that disregard grouping uncertainty may be unable to produce reliable regional impact projections.

Next, we evaluate the predictive accuracy of all the alternative groupings based on MSPE and MAPE, and discover that there is little evidence that any single grouping is more appropriate than any alternative. We reach the same conclusion when assessing the relative performance of a new grouping, the nine US climate regions commonly studied in climate economics (e.g., Heutel et al., 2021) and atmospheric sciences (Hoerling et al., 2016). As a result, we believe that grouping uncertainty is of analytical significance and deserves a more careful examination as its presence reduces the reliability of group-specific impact estimates and the efficiency of any climate adaptation policy based on them.

Overall, we believe future researchers who want to quantify the heterogeneous impact of climate change on agriculture should be cautious about the usage of any “group and test” strategy. We find that there is no superior grouping over the others so pre-determined groups might add significant uncertainty to the impact evaluation. As an alternative, we recommend researchers

consider data-driven methods such as C-Lasso and causal forests, and we demonstrate their potentials by quantifying the slope heterogeneity they generate. In our views, the most promising is the causal forest approach of Wager and Athey (2018) as it clearly generates county-specific marginal effects which can be used to suggest place-tailored climate change adaptation strategies. Future research could also investigate the role of networks. While this paper has dealt with structural instability defined on stock data, heterogeneity is also present when focusing on flow data. An example is a recent publication by Dall’Erba et al. (2021) who find that no two US states have the same trade partners of grains, fruits, and vegetables and that interstate trade mitigates the impact of climate change by \$14.5 billion by mid-century. The authors conclude that place-specific mitigation efforts have not only local but also system-wide impacts on the economy.

4.8 Tables

Table 4.1 Chow test results

Grouping	Irrigation > 0.2 (1)	Irrigation > 0.1 (2)	State (3)	Elevation > median (4)
<i>Panel. A Cross-sectional model</i>				
(1) F test that the 16 climate coefficients are identical (p-value)	0.000	0.000	0.000	0.000
(2) Equality of the climate coefficients across groups (5% level)	8/16 (=50%)	12/16 (=75 %)	3,683/11,865 (=31%)	12/16 (=75%)
<i>Panel. B Panel model</i>				
(1) F test that the four weather coefficients are identical (p-value)	0.000	0.000	0.000	0.000
(2) Equality of the weather coefficients across groups (5% level)	4/4 (=100%)	4/4 (=100%)	1,271/2,964 (=42.8%)	2/4 (=50%)

Note: This table presents the results of the Chow test from: (A) cross-sectional model, (B) panel model. The first row in each panel reports the p-value from the joint Chow test for the grouping noted in each column. The second row is in the format of “ n/m ”, where m is the total number of individual Chow tests conducted and n is the number of significant tests.

Table 4.2 Predictive accuracy of our alternative groupings

Grouping	No grouping = benchmark model	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate Region
# of groups	1	2	2	2	9
MSPE	2,384,716	2,371,418	2,384,917	2,514,031	1,860,730
MAPE	38.13%	37.72%	37.83%	39.11%	31.14%
Adjusted R^2	0.685	0.6954	0.6917	0.6852	0.7820
F test (H_0 : benchmark model)	-	0.000	0.000	0.000	0.000

Note: This table presents the results of cross-validation from estimating a fixed coefficient model in the cross-sectional setting, noted as eq. (4.10) in the text. The classifications to be evaluated are in the row headings. The first row reports the number of groups for each grouping. The second row reports mean squared prediction error (MSPE). The third row reports the mean absolute percentage error (MAPE).

Table 4.3 Results of the non-nested J test

Grouping pairs (H_0 by column vs. H_1 by row)	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate region
Irrigated ratio > 0.2	-	16.1508 (0.000)	18.8189 (0.000)	5.3317 (0.000)
Irrigated ratio > 0.1	6.6183 (0.000)	-	14.0135 (0.000)	0.3289 (0.742)
Elevation > median	4.7772 (0.000)	5.1889 (0.000)	-	79.0494 (0.000)
Climate region	76.4048 (0.000)	77.7504 (0.000)	4.6489 (0.000)	-

Note: This table reports the results from non-nested J test. Both the row and column headings list the competing classifications. The null hypothesis (H_0) is that grouping in the row outperforms the grouping in the column. Each entry lists both t-test statistic and p-value for the null hypothesis being rejected.

4.9 Figures

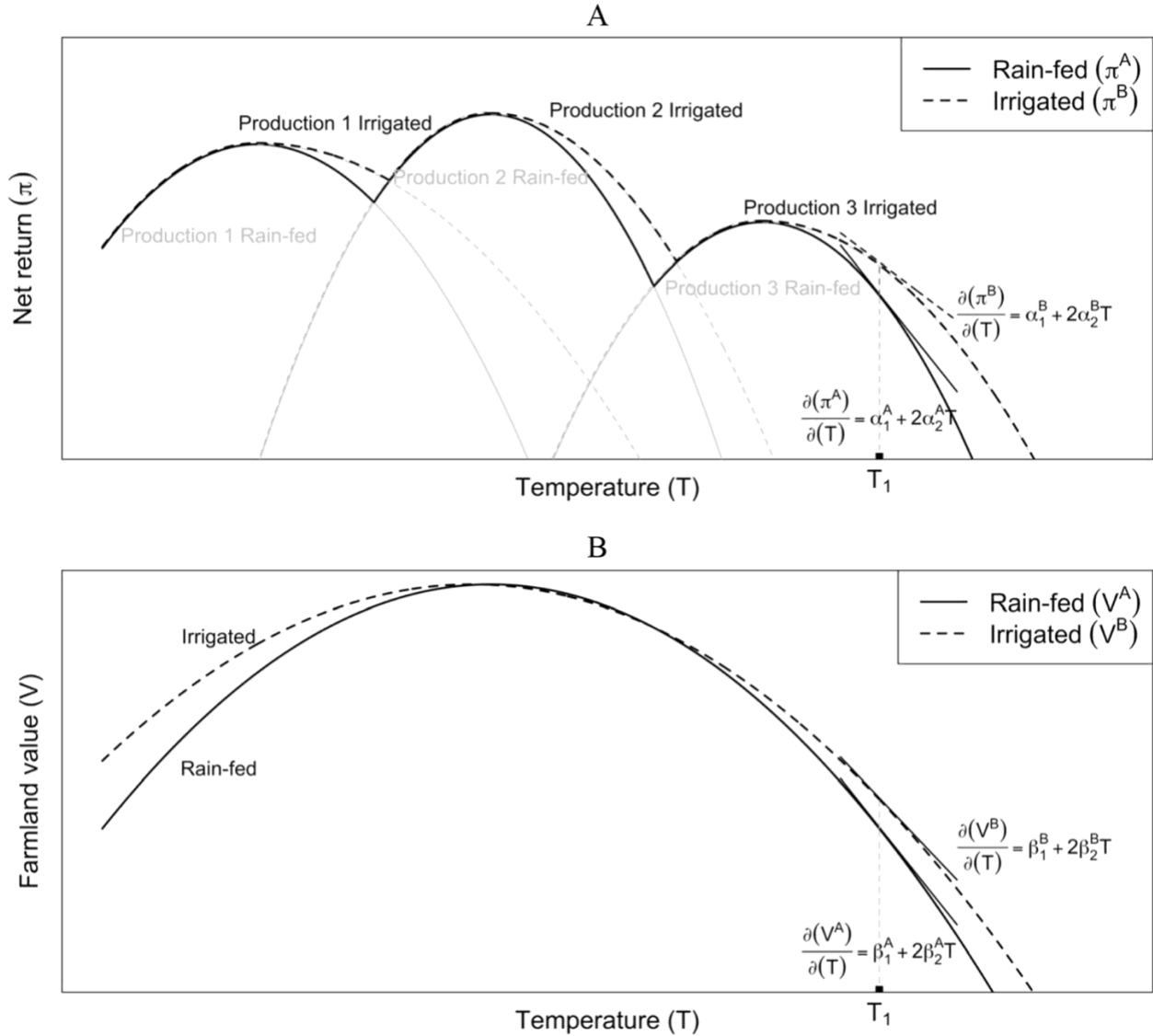
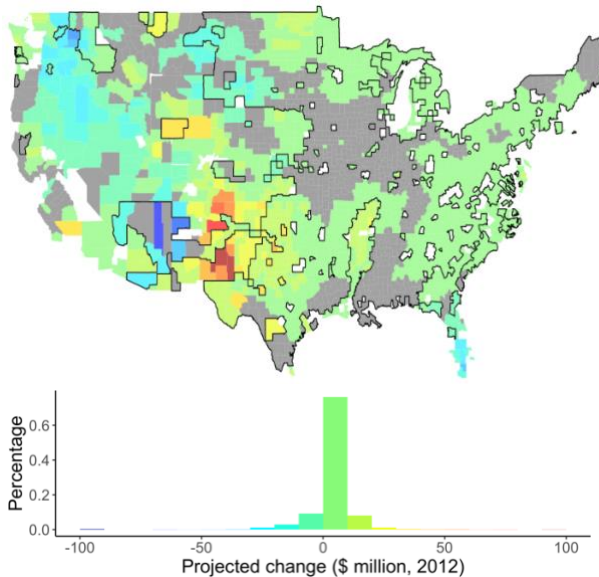


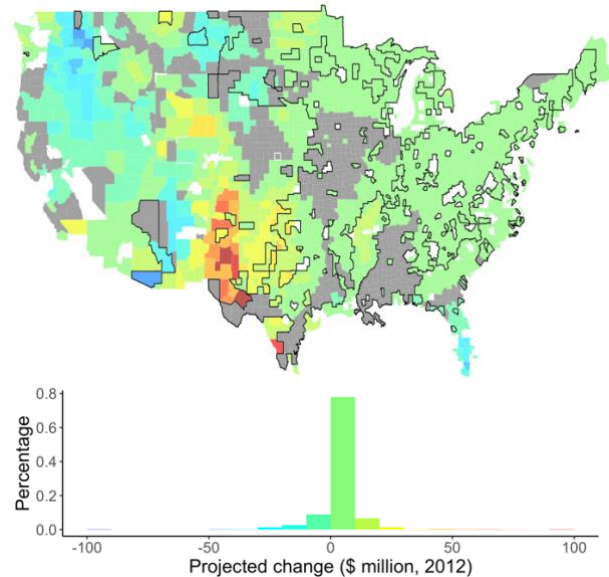
Figure 4.1 Shift in adaptation by irrigation technology adoption

Note: This figure illustrates the rationale for separating irrigated and rain-fed counties. In Panel A, the solid (dotted) line represents the production-specific profit function in a rain-fed location A (irrigated location B). The outer envelope of these profit functions identifies the hedonic equation, which is represented by the bold line. The tangent lines represent the slope of two hedonic equations at temperature T_1 for both rain-fed location A and irrigated location B. Panel B assumes the functional form of the two hedonic equations is quadratic. The solid curve represents the quadratic equation for rain-fed location A, while the dotted curve is for irrigated location B. When comparing the two slopes, the flatter slope for location B indicates that the marginal effect of temperature for the irrigated location is smaller.

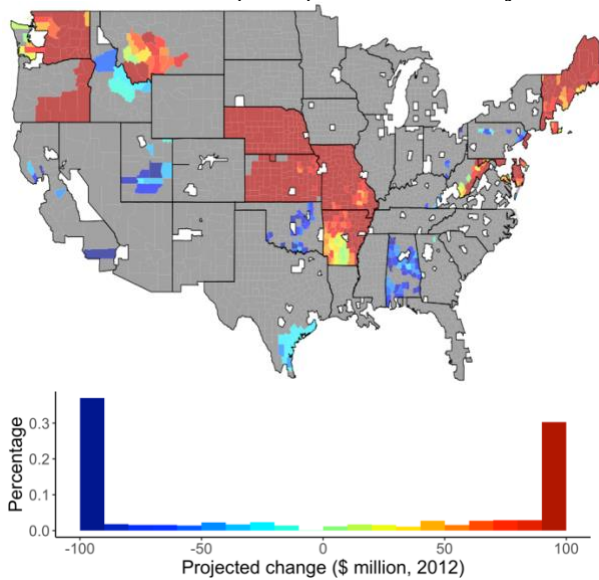
Panel A: Grouping used in Schlenker et al. (2005): irrigated ratio > 0.2



Panel B: Grouping used in Deschênes and Greenstone (2007): irrigated ratio > 0.1



Panel C: Grouping used in Deschênes and Greenstone (2007): state boundary



Panel D: Grouping used in Dall'erba and Domínguez (2016): elevation > median

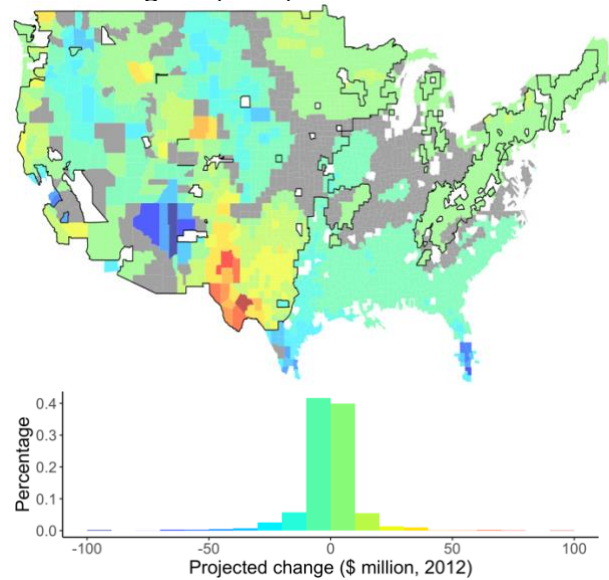


Figure 4.2 Group-specific projections of change in agriculture profits due to climate change based on the cross-sectional model

Note: Estimated impact of climate change in agriculture profits per acre is based on the cross-sectional model of eq. (4.10). The title of each panel indicates the grouping used in the respective map (top) and histogram (bottom). The black lines indicate the group boundaries. All dollar figures are in constant 2012 million US dollars. See the text for further details.

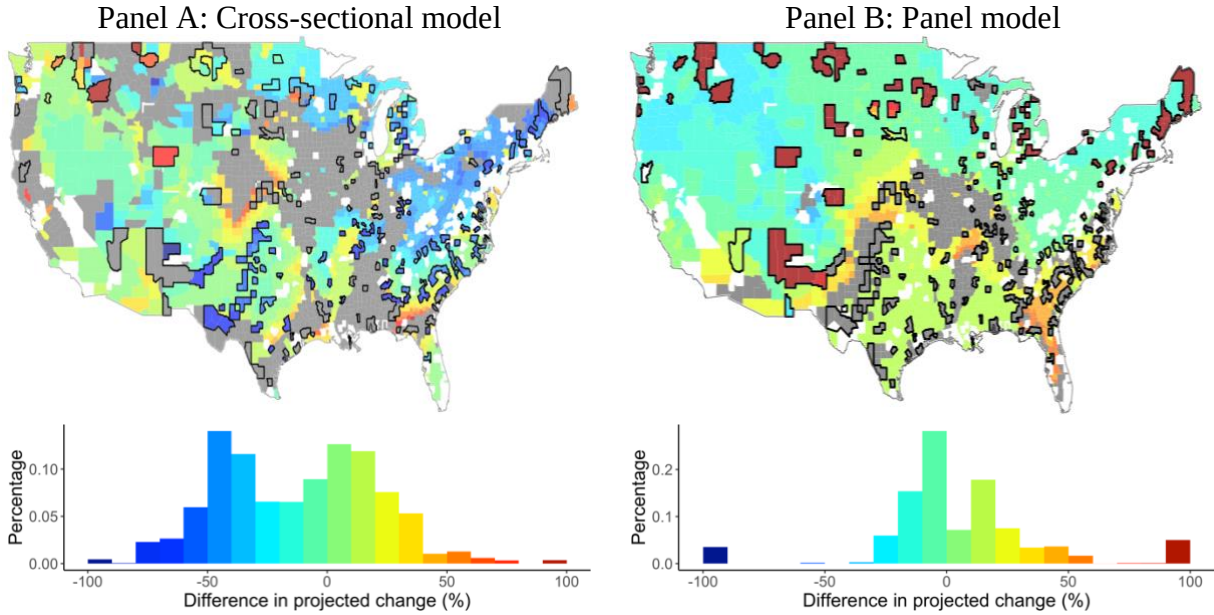


Figure 4.3 Percentage difference in the predicted change in annual profit between a grouping based on 10% and 20% cutoff

Note: Percentage change in the projected impact using a 10% vs. 20% irrigation cutoff in the cross-sectional model (panel A) and in the panel model (panel B). The black contours mark the counties that are usually assigned to the irrigated group based on the 10% cutoff but are now assigned to the rain-fed group with the 20% cutoff. See the text for further details.

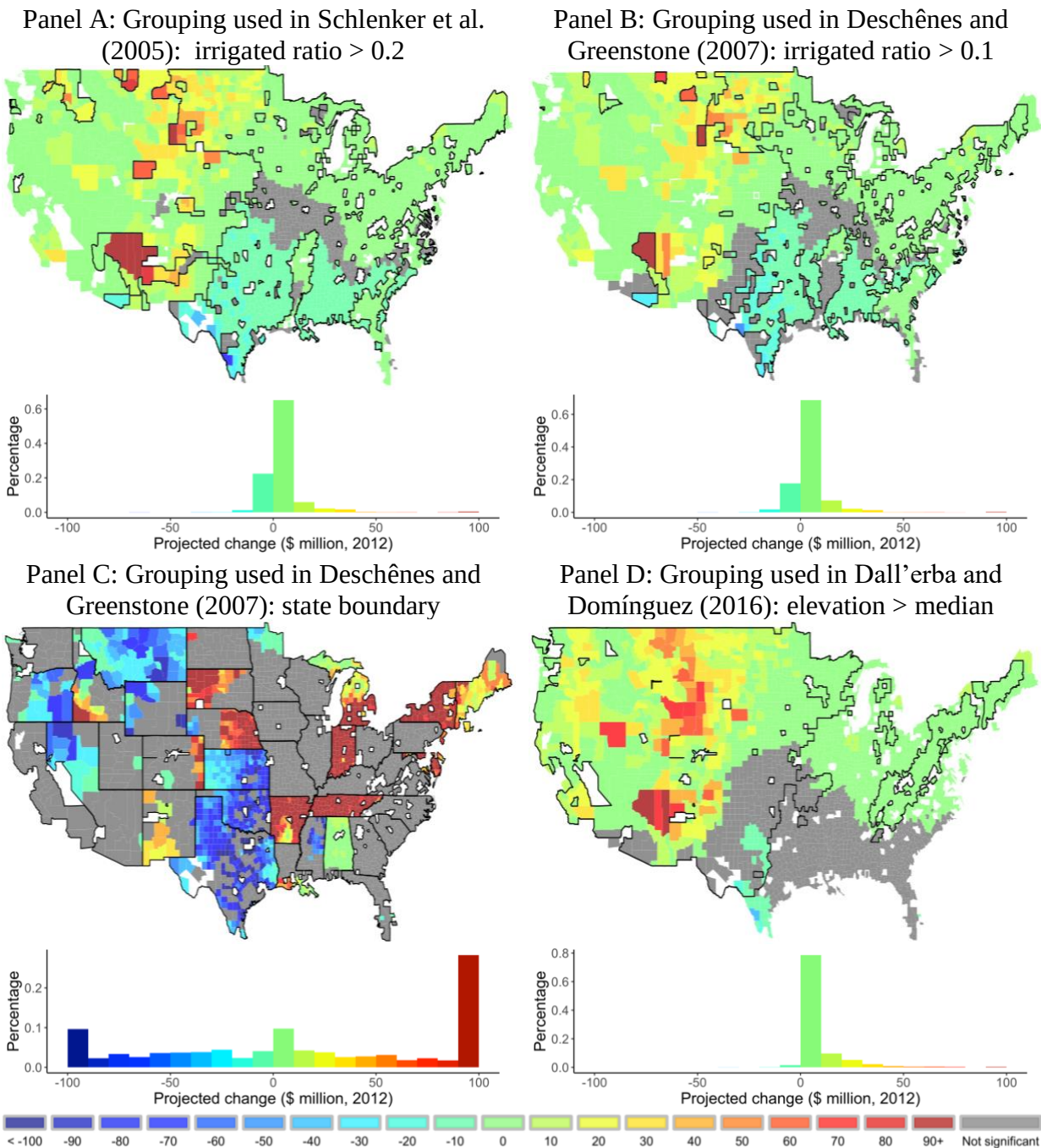


Figure 4.4 Group-specific projections of change in agriculture profits due to climate change based on the panel model

Note: Estimated impact of climate change in agriculture profits per acre is based on the panel model of eq. (4.11). The title of each panel indicates the grouping used in the respective map (top) and histogram (bottom). The black lines indicate the group boundaries. All dollar figures are in constant 2012 million US dollars. See the text for further details.

CHAPTER 5:

CONCLUSION

This dissertation integrates established empirical methods with newly emerging methods and datasets to study problems related to natural resource management, namely, recreation demand, park management, and climate change impacts on the agriculture sector. Throughout this dissertation, I emphasize the importance of spatial and temporal heterogeneity in understanding and evaluating the response of socioeconomic indicators to climate change, natural shocks, or policy-induced access restrictions, which provide some insight for designing effective management policies and preparing local communities for an ever-changing environment.

In Chapter 2, *Wildfire and Visitation in U.S. National Parks*, I quantify the response of tourists to increasing wildfire activities in national parks over the last three decades. I focus on the 32 national parks in the West as this can give me common trends that are representative of all western parks. Using fixed effect panel regressions at the park level, I exploit the annual variation of fire events in proximity to each park to estimate how park visitation responds to wildfires. I also allow this relationship to vary by the size and locations of wildfires. I find that wildfires reduce national park visitation by about 700,000 visits per year on average, which translate into approximately 1.35% each year. This reduction is disproportionately larger for catastrophic fires in the most heavily visited areas. Furthermore, I find that restricted access to parks due to fire closures is the most important channel driving these visitor losses, even when a park is partially closed. In contrast, visitors are surprisingly unresponsive to smoky conditions in the absence of local fires. More broadly, my research highlights the importance of local adaptation efforts in mitigating economic loss in tourism due to wildfires, and it suggests when and where to prioritize

wildfire prevention and suppression efforts in protected areas when updating park management plans.

Chapter 3, *Temporal Disaggregation of Visitation Time Series in U.S. National Parks Using Mobile Phone GPS data*, builds on chapter 2 as well as the consensus on the potential value of finer-scale visitation data. This study investigates the possibility of applying the Temporal Disaggregation approach to SafeGraph's daily foot traffic data to improve the temporal resolution of NPS monthly visitation statistics. My results confirm that in many parks SafeGraph data can be used to estimate daily visitation by disaggregating widely available monthly NPS visitation data. The performance of the best-fit model largely depends on how closely first-differenced SafeGraph-derived monthly aggregated series are associated with first-differenced NPS series. Moreover, this study suggests high-quality within-park POIs are essential inputs to modeling. The grid search design I adopted for POI pre-selection provides a viable solution for the low-quality SafeGraph data, but strong justification should be provided to ensure that the TD model will estimate daily visitation consistently for the newly released SafeGraph data. In short, the approach presented in this study provides an opportunity for park managers to reconstruct granular visitor use data.

Chapter 4, *On the Evaluation of Heterogeneous Climate Change Impacts on U.S. Agriculture: Does Group Membership Matter?*, examines how the choice of grouping can affect expected future agricultural profits when one wants to quantify the heterogeneous impact of climate change on agriculture. I find that the estimated impact of climate change on local agriculture varies widely if we choose one grouping rule versus another. Namely, a county could easily switch from a median winner to a huge loser, even just for a modest change in the assigned group membership. Furthermore, I assess the relative effectiveness of all available groupings and find that none of the grouping rules used in previous studies dominates the others in terms of out-

of-sample prediction performance. Allowing the climate change impacts to vary across space is crucial, because it can yield more accurate county-level impact projections. Such impact projections are especially important for local policymakers because they can make place-tailored decisions in areas where climate change is anticipated to pose the most serious challenges. Overall, this study suggests that a careful choice of grouping by combining prior knowledge and group structures revealed by data-driven methods such as C-Lasso can render more trustworthy guidance for local planners.

APPENDIX A:

SUPPLEMENTAL MATERIALS FOR Chapter 2:

Table A.1 Summary statistics, 1992 - 2019

	Mean	Std. Dev.	Min.	Max.	N
Recreation Use					
Recreation Visitation (times)	99,977	147,722	0	980,702	9,216
Backcountry visitation (nights)	21,074	49,531	0	416,510	9,216
Wildfire and smoke data					
Wildfire size within 50 miles (thousand acres)	22.14	169.20	0.00	5,038.33	9,216
Wildfire size of lightning fires within 50 miles (thousand acres)	16.02	152.05	0.00	5,028.33	9,216
Wildfire size of class A fires within 50 miles (thousand acres)	0.00	0.01	0.00	0.24	9,216
Wildfire size of class B fires within 50 miles (thousand acres)	0.02	0.06	0.00	1.40	9,216
Wildfire size of class C fires within 50 miles (thousand acres)	0.08	0.31	0.00	7.38	9,216
Wildfire size of class D fires within 50 miles (thousand acres)	0.15	0.69	0.00	13.13	9,216
Wildfire size of class E fires within 50 miles (thousand acres)	0.41	2.05	0.00	41.61	9,216
Wildfire size of class F fires within 50 miles (thousand acres)	2.05	11.29	0.00	358.52	9,216
Wildfire size of class G fires within 50 miles (thousand acres)	13.31	149.89	0.00	4,999.45	9,216
Wildfire size within 0-5 miles (thousand acres)	0.32	6.91	0.00	280.80	9,216
Wildfire size within 5-10 miles (thousand acres)	0.62	7.82	0.00	232.21	9,216
Wildfire size within 10-15 miles (thousand acres)	0.59	6.57	0.00	188.47	9,216
Wildfire size within 15-20 miles (thousand acres)	1.71	36.85	0.00	1,516.24	9,216
Wildfire size within 20-25 miles (thousand acres)	0.83	13.35	0.00	872.75	9,216
Wildfire size within 25-30 miles (thousand acres)	1.25	22.57	0.00	816.36	9,216
Wildfire size within 30-35 miles (thousand acres)	3.23	50.87	0.00	1,628.18	9,216
Wildfire size within 35-40 miles (thousand acres)	2.09	34.69	0.00	1,248.98	9,216
Wildfire size within 40-45 miles (thousand acres)	1.48	25.20	0.00	1,153.04	9,216
Wildfire size within 45-50 miles (thousand acres)	3.91	117.33	0.00	4,999.47	9,216
Days of fire closures	0.07	0.70	0.00	12.64	1,560
Days of fire closures caused by lightning fires	0.05	0.60	0.00	12.64	1,560
Days of local fires	0.66	2.48	0.00	29.00	5,376
Days of smoke	2.40	4.79	0.00	31.00	5,376
Days of smoke without local fires	1.96	3.94	0.00	31.00	5,376
Days of smoke adjusted by ozone	0.83	2.19	0.00	23.00	3,389
Controls					
Mean temperature (°C)	9.66	9.38	-12.70	32.25	9,216
Total Precipitation (mm)	68.88	103.93	0.00	1,093.07	9,216
Mean vapor pressure deficit (hPa)	10.25	8.12	0.20	45.78	9,216
Income-adjusted gasoline price index (price/thousand \$)	0.06	0.02	0.03	0.12	8,736
Wildfire size on nearby NPS units within 80 miles (thousand acres)	0.48	6.62	0.00	305.26	9,216

Note: This table reports the summary statistics at the park-by-month level for all 32 western national parks in my sample. All wildfire size variables are for the period 1992 - 2015. The fire closure variables are available from 2010 to 2019. The smoke analysis uses data available from 2006 to 2019.

Table A.2 Robustness check to alternative specifications

	(1)	(2)	(3)	(4)	(5)	(6)
	Log-linear model	Inverse hyperbolic sine (arcsinh)	PPML, Two-way clustered	PPML, Clustered by park (Bootstrap)	Log-linear, Spatial HAC	arcsinh, Spatial HAC
Fire size (thousand ac)	-0.0498** (0.0204)	-0.0480** (0.0211)	-0.0636* (0.0343)	-0.0636* (0.0342)	-0.0498** (0.0236)	-0.0480** (0.0229)
Observations	8,736	8,736	8,736	8,736	8,736	8,736
Number of Parks	32	32	32	32	32	32
Mean fire size	1.687	1.687	1.687	1.687	1.687	1.687
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Park by Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Park by Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Month-of-Sample	Yes	Yes	Yes	Yes	Yes	Yes
Lightning fires only	Yes	Yes	Yes	Yes	Yes	Yes
Nearby Parks	Yes	Yes	Yes	Yes	Yes	Yes

Note: Each column presents the coefficient estimate from re-estimating eq. (2.1) with alternative specifications or alternative ways of calculating standard errors. All regressions include park-by-year, park-by-month and month-of-sample fixed effects, as well as controls for weather and income-adjusted gasoline price. ***p<0.01, **p<0.05, *p<0.1.

Table A.3 Robustness check of the main results

	(1)	(2)	(3)	(4)	(5)
	Burned acres inside park (only)	Burned acres within 5-mile buffer around park	Piecewise linear	Quadratic	Subsample of fire season only
Fire size (thousand ac)	-0.491* (0.264)	-0.205* (0.123)	-0.0634** (0.0307)	-0.0652** (0.0320)	-0.0701** (0.0357)
Observations	8,736	8,736	8,736	8,736	2,944
Number of Parks	32	32	32	32	32
Mean fire size	0.179	0.380	1.687	1.687	1.687
Controls	Yes	Yes	Yes	Yes	Yes
Park by Year FE	Yes	Yes	Yes	Yes	Yes
Park by Month FE	Yes	Yes	Yes	Yes	Yes
Month-of-Sample	Yes	Yes	Yes	Yes	Yes
Lightning fires only	Yes	Yes	Yes	Yes	Yes
Nearby Parks	Yes	Yes	Yes	Yes	Yes

Note: Each column presents the coefficient estimate from re-estimating eq. (2.1) with an alternative definition of wildfire size, alternative functional forms for weather controls, or using the subsample of the data. All regressions include park-by-year, park-by-month and month-of-sample fixed effects, as well as controls for weather and income-adjusted gasoline price. ***p<0.01, **p<0.05, *p<0.1.

Table A.4 Effect of wildfire smoke: other adjusted smoke measures

	(1)	(2)	(3)	(4)
Panel A: Fully covered by smoke				
Fire days	-0.370*** (0.0876)		-0.366*** (0.0876)	-0.354*** (0.0845)
Smoke days		-0.135 (0.114)	-0.0722 (0.114)	0.112 (0.126)
Observations	5,376	5,376	5,376	5,376
Number of Parks	32	32	32	32
Mean fire days	0.656		0.656	0.656
Mean smoke days		1	1	0.914
Panel B: Medium density smoke				
Fire days	-0.370*** (0.0876)		-0.342*** (0.111)	-0.369*** (0.0942)
Smoke days		-0.306*** (0.0872)	-0.0588 (0.0950)	0.0104 (0.120)
Observations	5,376	3,381	3,381	3,381
Number of Parks	32	32	32	32
Mean fire days	0.656		0.656	0.656
Mean smoke days		0.955	0.955	0.653
Panel C: Overlap any developed area				
Fire days	-0.370*** (0.0876)		-0.430*** (0.104)	-0.338*** (0.0844)
Smoke days		-0.0916 (0.0625)	0.106 (0.0751)	0.115 (0.0821)
Observations	5,376	3,381	3,381	3,381
Number of Parks	32	32	32	32
Mean fire days	0.656		0.656	0.656
Mean smoke days		2.283	2.283	1.868
Controls	Yes	Yes	Yes	Yes
Park × Year FE	Yes	Yes	Yes	Yes
Park × Month FE	Yes	Yes	Yes	Yes
Month-of-Sample	Yes	Yes	Yes	Yes
Non-fire smoke days only				Yes

Note: This table presents the robustness checks for regression results of eq. (2.3) in the main text. Three panels reflect three different specifications with different measurements of smoke days. The layout of each regression table is the same as Table 2.3 in the main text. ***p<0.01, **p<0.05, *p<0.1.

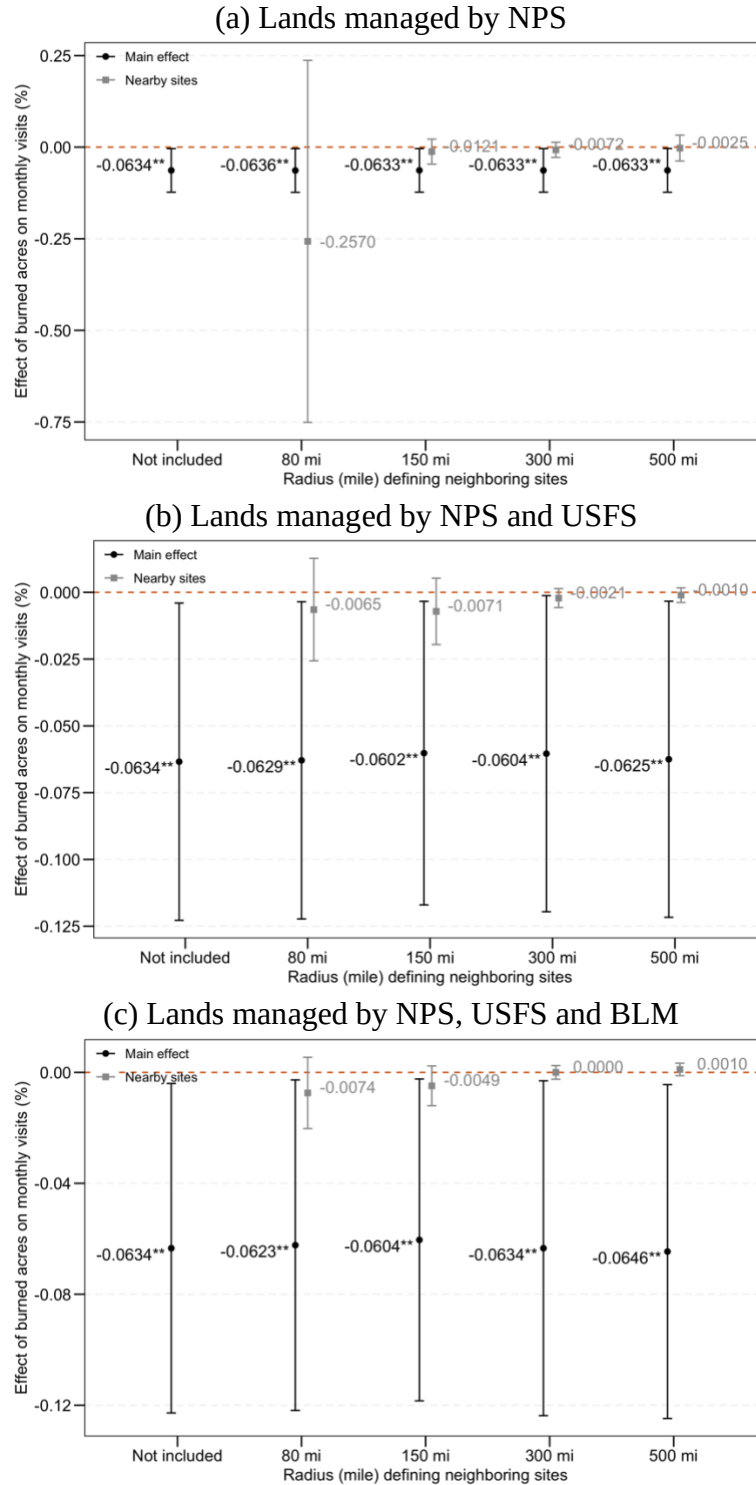


Figure A.1 Comparison of main effects with the spillover effects

Note: Each panel plots the estimated effect of wildfire size within a 50-mile radius around the national park of interest (black bars) together with the estimated effect of aggregate wildfire size at all nearby recreational areas within a specified radius (grey bars). The radius cutoff used to define “nearby” is shown on the x-axis. The panel title reports the type of the recreation units that are considered as “nearby parks” in proximity to the national park of interest.

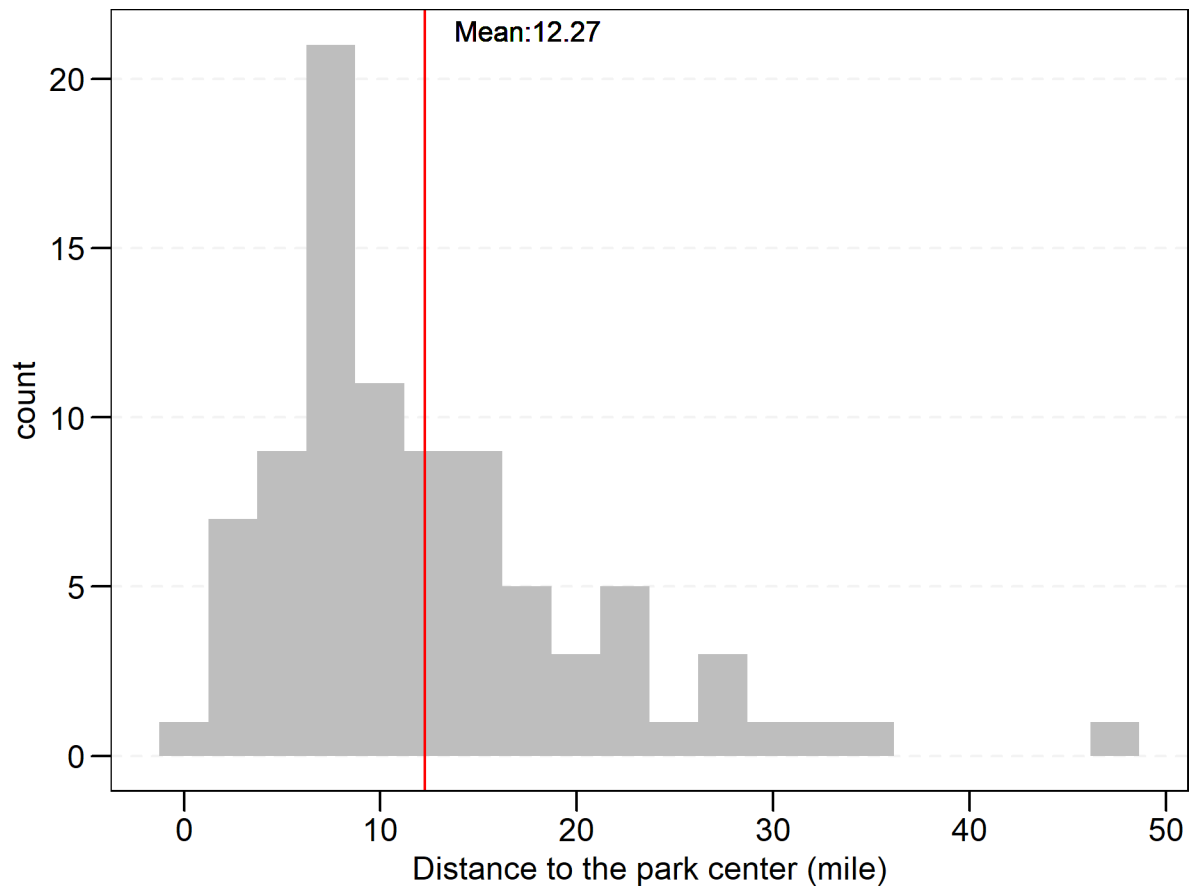


Figure A.2 The distance from park visitor centers the park centroid

Note: This figure shows the distribution of the distances from visitor centers (n = 88) to their associated parks' geographic centroids

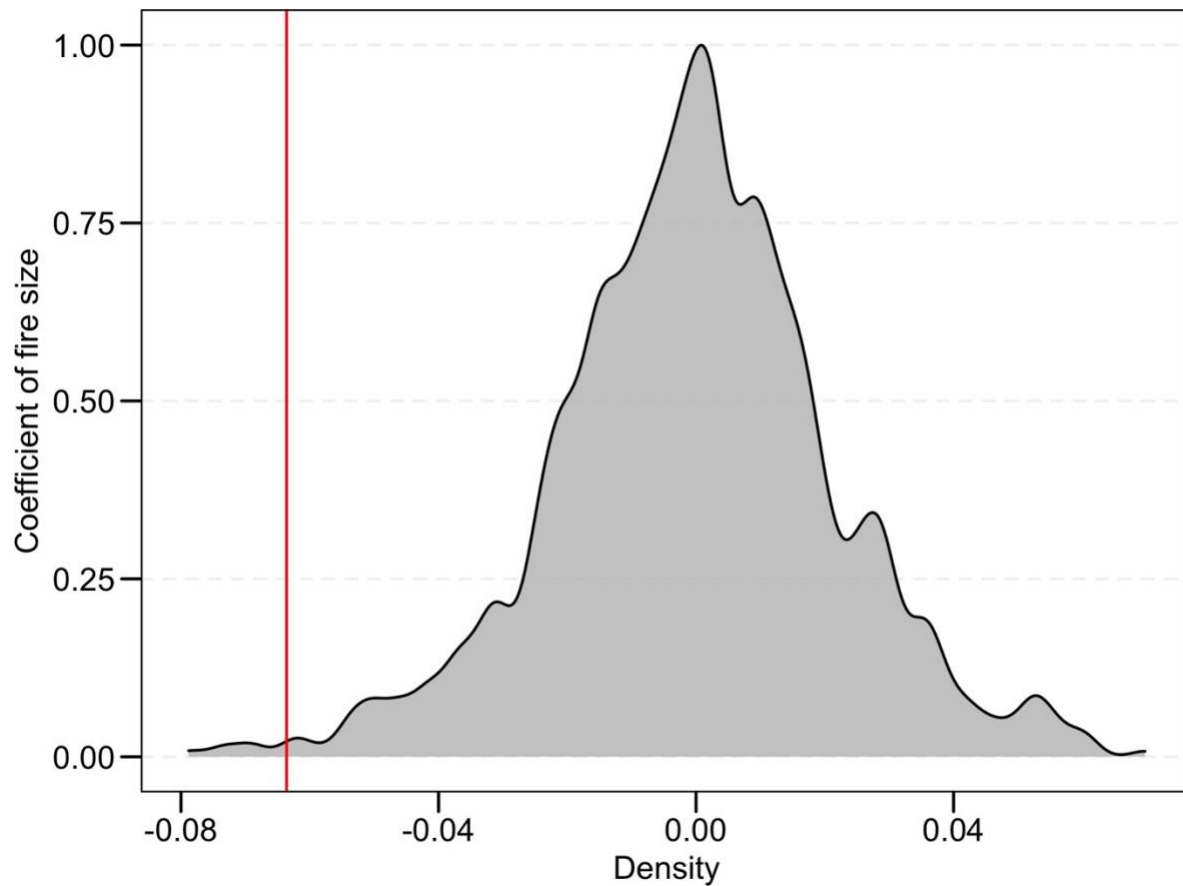


Figure A.3 Exact distribution for test of the null effect of wildfire size

Note: This figure shows the distribution for the permutation test of the null effect of wildfire size (β). I permute the dependent variables $fireSize_{iym}$ 1000 times and then plot the distribution of its coefficient estimates. The red line shows where the actual estimate reported in column (4) of Table 2.1 lies. The one-sided p-value from Wald t-statistics is 0.00, i.e., rejecting the null $\beta = 0$.

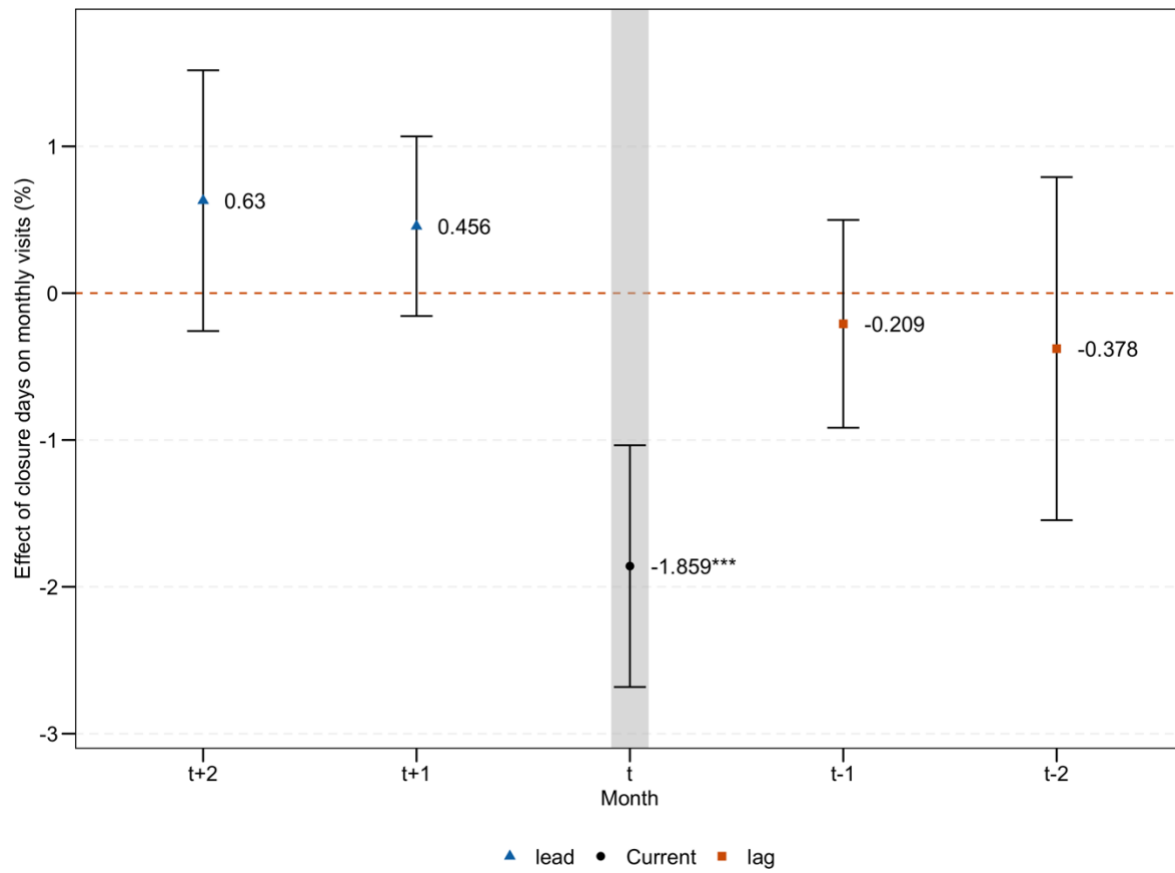


Figure A.4 Effects of leads and lags of fire closure days

Note: This figure plots estimated coefficients from a lead-and-lag test of eq. (2.2), where the independent variables are instead the fire closure days in month $t + k$ ($k = 3, 2, 1, 0, -1, -2, -3$). The whiskers represent the 95% confidence intervals based on standard errors clustered at the park-by-year level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

APPENDIX B:

SUPPLEMENTAL MATERIALS FOR Chapter 3:

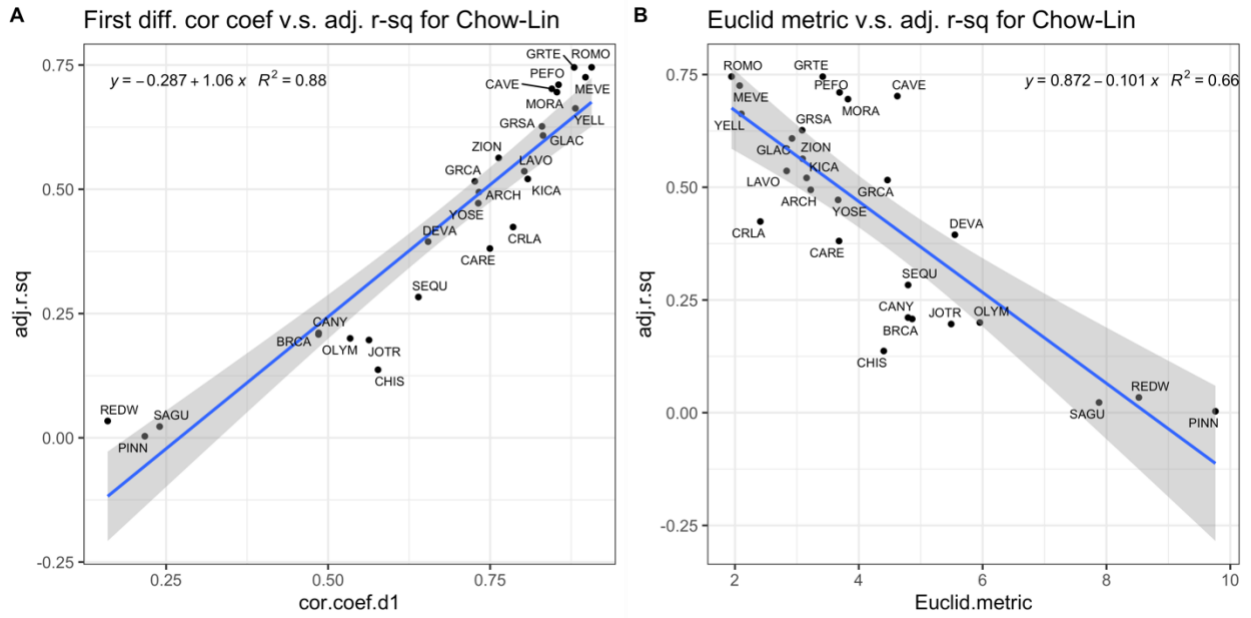


Figure B.1 A comparison of the Pearson correlation and Euclidean distance with Adjusted R-sq

Note: This figure compares two similarity metrics in terms of how they relate to the model performance of Temporal Disaggregation (TD). The Y-axis in both panels is the same, adjusted R squared of low-frequency regressions in the Chow-Lin TD method. On the X-axis are two similarity metrics. The metric in panel A is the correlation coefficient between the first-differenced series of SafeGraph visits and NPS visitation. The metric in panel B is the Euclidean distance between the z-normalized series of SafeGraph visits and NPS visitation. The linear regression lines are plotted in blue. The regression results and R squared are also illustrated on the top left of each panel.

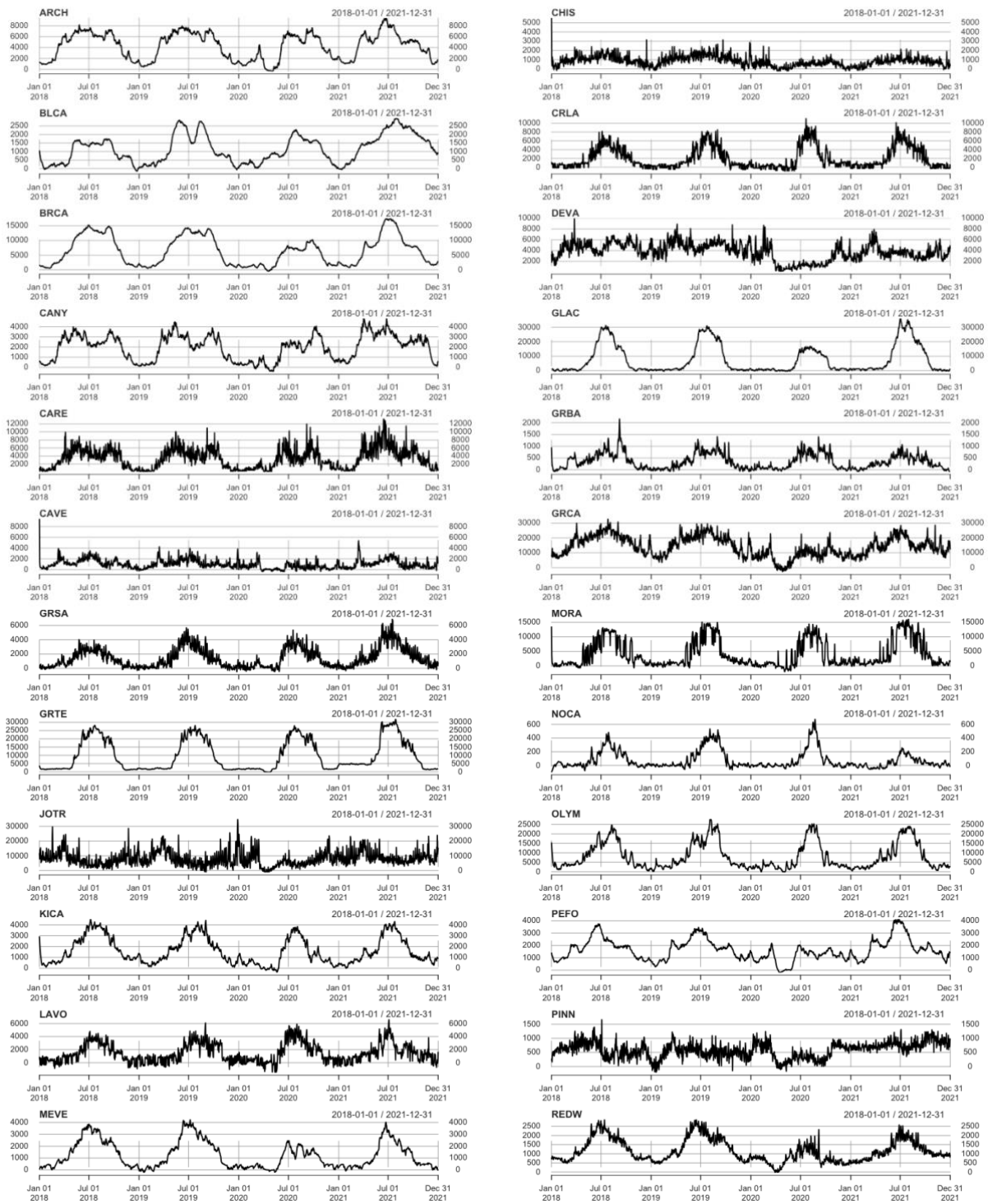


Figure B.2 Estimated daily visitation times series from Temporal Disaggregation by park

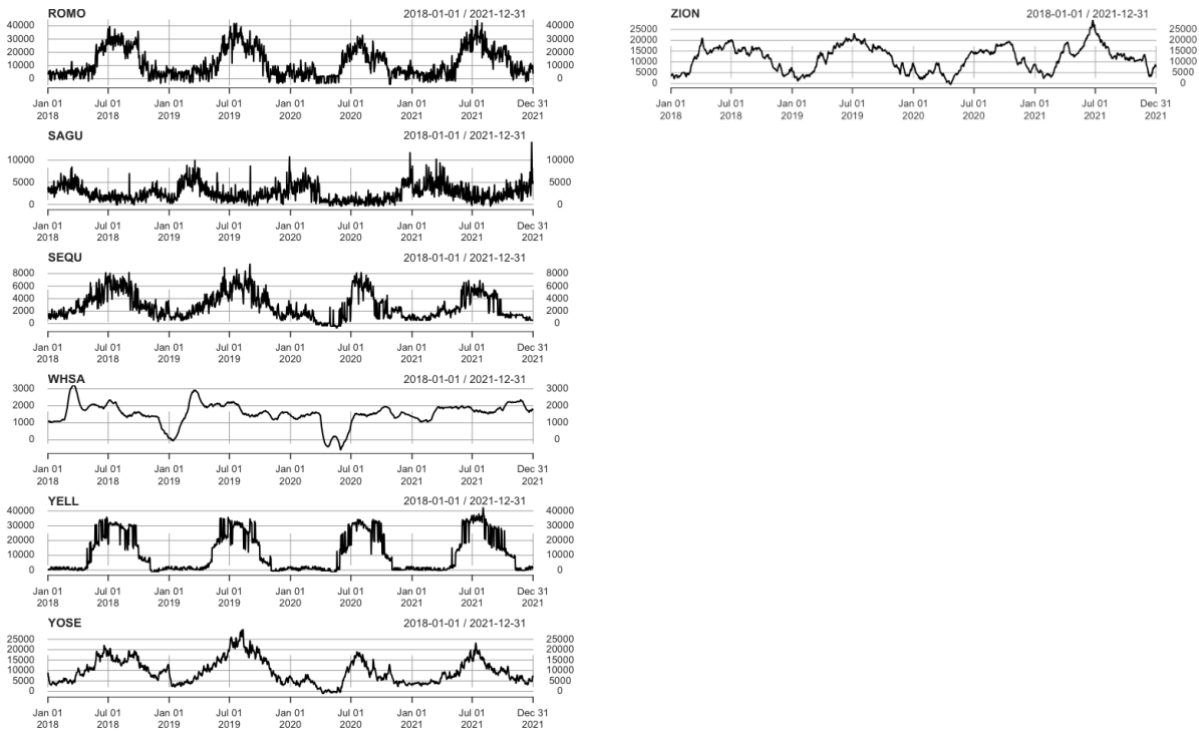


Figure B.2 (cont.)

APPENDIX C:

SUPPLEMENTAL MATERIALS FOR Chapter 4:

Table C.1 Key features of the US-focused literature

Summary of variable and sample used in these studies					
Paper	Outcome variables (1)	Weather variables (2)	Sample (3)	Functional Form (4)	Grouping (5)
Schlenker <i>et al.</i> (2005) (SHF)	Farmland values	Seasonal temperature, precipitation	Cross-section of US counties, 1982	Lin-lin	Irrigated ratio > 0.2
Deschênes and Greenstone (2007) (DG)	Agricultural profits	Growing season degree-days, total precipitation	Panel of US counties, 1982 – 2002	Lin-lin	Irrigated ratio > 0.1 State
Dall’erba and Domnguez (2016) (DD)	Farmland values	Seasonal (max) temperature, precipitation	Cross-section of Southwestern US counties, 2007	Log-lin	Elevation > median

Note: This table summarizes the key features of the three papers reviewed in this study. The five columns list their choices of dependent variables, weather/climate variables, dataset, functional form, and rules of grouping.

Table C.2 Key features of the international literature

Summary of variable and sample used in these studies					
Paper	Outcome variables (1)	Weather variables (2)	Sample (3)	Functional Form (4)	Grouping (5)
Kurukulasuriya and Mendelsohn (2007)	Annual net income	Seasonal temperature, precipitation	Cross-section of farm in Africa	Lin-lin	Irrigated farms
Garcia and Viladrich (2009)	Farmland prices	Growing season degree-days, total precipitation	Panel of Spain provinces, 1983 – 1999	Log-lin	Receives irrigation at least once
Mendelsohn et al. (2010)	Farmland values	Seasonal temperature, precipitation	Cross-section of Mexican farms, 2002	Log-lin	Irrigated farms
Kurukulasuriya et al. (2011)	Annual net income	Seasonal temperature, precipitation	Cross-section of plots in Africa	Lin-lin	Irrigated plots
Bozzola et al. (2018)	Farmland values	Seasonal temperature, precipitation	Pooled cross-section of Italian farms, 2008 – 2011	Log-lin	North, Centre, South

Note: This table summarizes the key features of the non-US studies. The five columns list their choices of dependent variables, weather/climate variables, dataset, functional form, and rules of grouping.

Table C.3 Descriptive statistics and sources

Variable	Description (Unit)	Min	Max	Mean	Sd.	Source
<i>Dependent Variable</i>						
land value	land value per acres of land in farm (\$/acre)	117.65	37623.9	3167.5	2278.6	USDA
profit	Agricultural profits per acre of farmland (\$/acre)	-513.77	1643.6	54.92	112.99	USDA
<i>Climate and Weather Variables</i>						
temp_win	Monthly average temperature (Dec. – Feb.), average of 1993-2012 (Celsius °C)	-14.22	21.58	1.68	6.8	NARR
temp_spr	Monthly average temperature (Mar. –May), average of 1993-2012 (Celsius °C)	0.16	28.29	13.54	5.07	NARR
temp_sum	Monthly average temperature (Jun.– Aug.), average of 1993-2012 (Celsius °C)	12.58	35.64	25.68	4.04	NARR
temp_aut	Monthly average temperature (Sep.– Nov.), average of 1993-2012 (Celsius °C)	2.28	27.09	14.63	4.59	NARR
prec_win	Monthly average precipitation (Dec. – Feb.), average of 1993-2012 (mm)	0.23	10.26	2.05	1.3	NARR
prec_spr	Monthly average precipitation (Mar. –May), average of 1993-2012 (mm)	0.15	5.69	2.71	0.92	NARR
prec_sum	Monthly average precipitation (Jun.– Aug.), average of 1993-2012 (mm)	0.05	5.87	2.8	0.94	NARR
prec_aut	Monthly average precipitation (Sep.– Nov.), average of 1993-2012 (mm)	0.2	6.95	2.24	0.81	NARR
gdd	Growing season degree-day (Celsius °C)	460.03	4283.18	2591.0	704.7	NARR
gsp	Growing season precipitation (mm)	0.15	40.63	15.19	6.03	NARR
<i>Socioeconomic Variables</i>						
density	Population Density per sq. mile (person/ mile ²)	0.08	787.02	65.5	79.64	Census
income	Personal Income / total population (1000\$/person)	7254	124088	28076	9018.8	BEA
<i>Soil Quality Variables</i>						
Lon	Longitude (degrees °)	-124.16	-67.64	-92.29	11.01	USGS
Lat	Latitude (degrees °)	26.12	48.82	38.29	4.91	USGS
Ec_ratio	Salinity (mmhos/cm)	0	7.19	0.2	0.56	USGS
F-ratio	Flood Frequency ratio (%)	0	0.73	0.08	0.12	USGS
Wet_ratio	Wetland ratio (%)	0	1	0.55	0.38	USGS
Slope-length	Slope steepness (%)	0.54	68.41	11.46	11.75	USGS
K-ratio	Erosion Factor (0.01 inch)	0	0.49	0.27	0.08	USGS
Sand_ratio	Sand contents (%)	0.98	96.33	38.37	19.68	USGS
Ksat_ratio	Permeability (cm/s)	0.91	113.63	20.46	18.67	USGS
Awc_ratio	Moisture Capacity (cm/cm)	0.01	0.38	0.16	0.04	USGS
Clay_ratio	Clay content (%)	1.12	55.12	18.54	8.15	USGS

Note1: The summary statistics are calculated for a balanced panel of 2,767 counties for the four most recent agricultural census years starting in 1997. All dollar values are in 2012 constant dollars.

Note2: Source abbreviation: USDA (USDA Census of Agriculture), NARR (North America Regional Reanalysis), Census (US Census Bureau), BEA (Bureau of Economic Analysis), USGS (US Geological Survey)

Table C.4 National average of group-specific projections across classifications

	Schlenker et al. (2005): Irrigated ratio > 0.2	Deschênes and Greenstone (2007): Irrigated ratio > 0.1	Deschênes and Greenstone (2007): By state boundary	Dall’erba and Domínguez (2016): Elevation > median
Cross-sectional model	6,695.147	7,938.85	-91,920.46	-420.363
Panel model	8,346.626	7,789.214	-43,956.04	11,400.11

Note: All dollar figures are converted to profits and measured in constant 2012 million US dollars.

Table C.5 Predictive accuracy of our alternative groupings based on a log-linear model

Grouping	No grouping = benchmark model	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate Region
# of groups	1	2	2	2	9
MSPE	3,086,703	2,936,164	3,039,525	3,072,950	1,853,583
MAPE	31.95%	30.90%	31.78%	31.83%	23.90%
Adjusted R^2	0.8075	0.8363	0.8205	0.8082	0.8895
F test (H_0 : benchmark model)	-	0.000	0.000	0.000	0.000

Note: This table presents the results of cross-validation from estimating a fixed coefficient model with a log-linear functional form in the cross-sectional setting. The classifications to be evaluated are in the row headings. The first row reports the number of groups for each grouping. The second row reports mean squared prediction error (MSPE). The third row reports the mean absolute percentage error (MAPE).

Table C.6 Results of the non-nested J Test based on a log-linear model

Grouping pairs (H_0 by column vs. H_1 by row)	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate region
Irrigated ratio > 0.2	-	28.7472 (0.000)	37.3303 (0.000)	17.318 (0.000)
Irrigated ratio > 0.1	3.0604 (0.002)	-	25.0477 (0.000)	8.0772 (0.000)
Elevation > median	7.0093 (0.000)	7.7623 (0.000)	-	7.4392 (0.000)
Climate region	86.781 (0.000)	91.6803 (0.742)	95.9352 (0.000)	-

Note: This table reports the results from non-nested J tests conducted using a log-linear model. Both the row and column headings list the competing classifications. The null hypothesis (H_0) is that grouping in the row outperforms the grouping in the column. Each entry lists the t-test statistic and the associated p-value.

Table C.7 Predictive accuracy of our alternative groupings based on a log-linear model and with temperature-precipitation interactions

Grouping	No grouping = benchmark model	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate Region
# of groups	1	2	2	2	9
MSPE	3,067,255	2,929,711	3,293,502	3,072,950	1,876,437
MAPE	30.91%	29.93%	30.59 %	31.83%	24.02%
Adjusted R^2	0.8185	0.8445	0.8315	0.8195	0.8917
F test (H_0 : benchmark model)	-	0.000	0.000	0.000	0.000

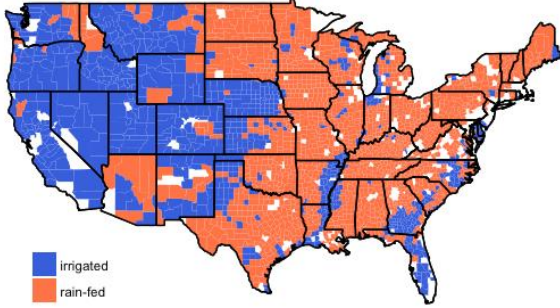
Note: This table presents the results of cross-validation from estimating a fixed coefficient model with a log-linear model functional form and temperature-precipitation interactions. The classifications to be evaluated are in the row headings. The first row reports the number of groups for each grouping. The second row reports mean squared prediction error (MSPE). The third row reports the mean absolute percentage error (MAPE).

Table C.8 Results of the non-nested J Test based on a log-linear model and with temperature-precipitation interactions

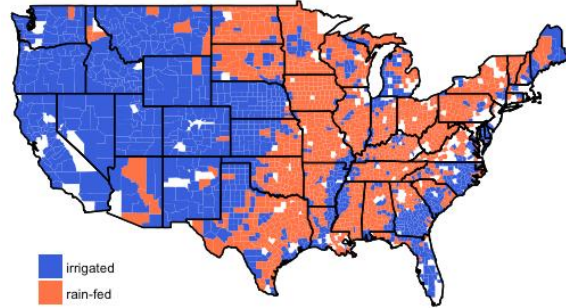
Grouping pairs (H_0 by column vs. H_1 by row)	Irrigated ratio > 0.2	Irrigated ratio > 0.1	Elevation > median	Climate region
Irrigated ratio > 0.2	-	26.481 (0.000)	35.9424 (0.000)	19.005 (0.000)
Irrigated ratio > 0.1	5.399 (0.000)	-	25.7979 (0.000)	11.335 (0.000)
Elevation > median	8.5458 (0.000)	9.7883 (0.000)	-	7.1699 (0.000)
Climate region	84.483 (0.000)	88.290 (0.742)	92.4540 (0.000)	-

Note: This table reports the results from non-nested J tests conducted using a log-linear model with temperature-precipitation interactions. Both the row and column headings list the competing classifications. The null hypothesis (H_0) is that grouping in the row outperforms the grouping in the column. Each entry lists the t-test statistic and the associated p-value.

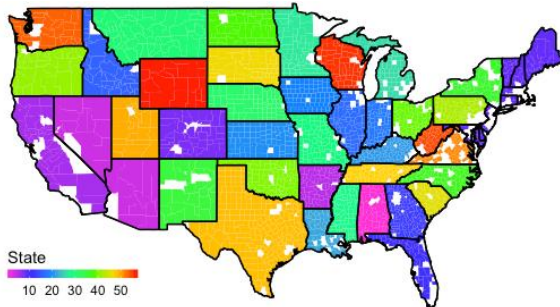
Panel A: Grouping in SFH: Irrigated ratio > 0.2



Panel B: Grouping in DG: Irrigated ratio > 0.1



Panel C: Grouping in DG: By state boundary



Panel D: Grouping in DD: Elevation > median

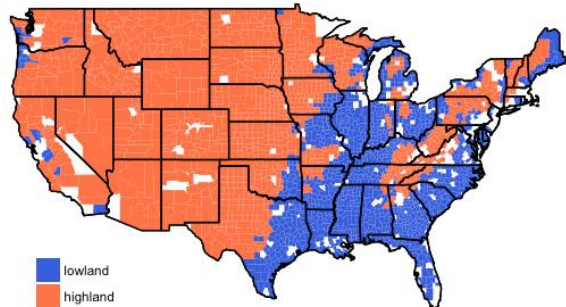


Figure C.1 Mapping the groupings currently used in the literature

Note: This figure plots the group boundary of: A. grouping by the 20% irrigation level cutoff ratio as in Schlenker et al. (2005); B. grouping by the 10% irrigation level cutoff ratio as in Deschênes and Greenstone (2007); C, grouping based on state boundaries as in Deschênes and Greenstone (2007); D, grouping by the median elevation level as in Dall’erba and Dominguez (2016).

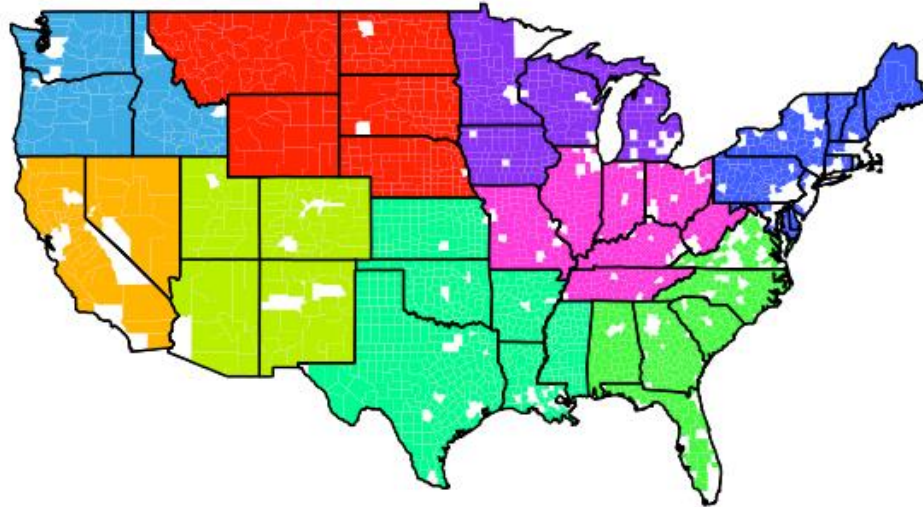


Figure C.2 Climate Regions

Note: This figure plots the boundary of US climate regions as defined by the National Oceanic and Atmospheric Administration (NOAA).

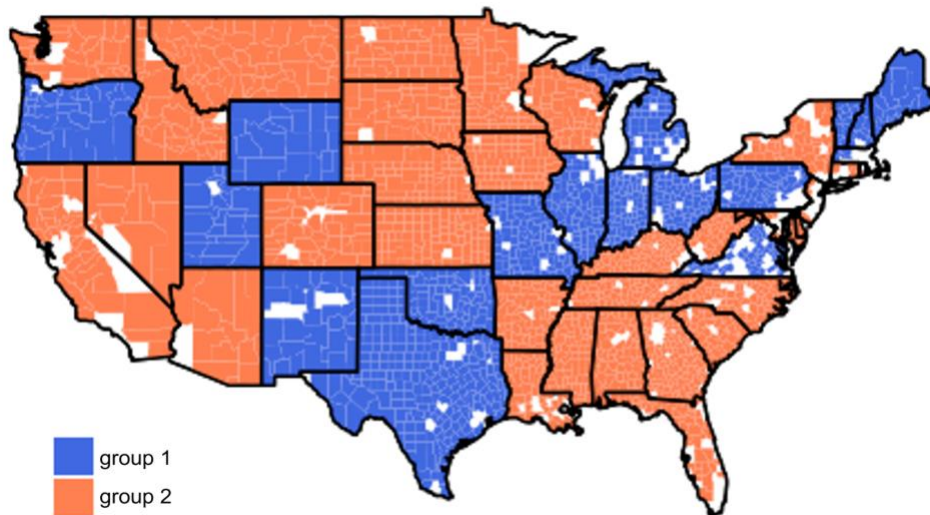


Figure C.3 Grouping generated by C-Lasso

Note: This figure graphically presents a data-driven classification generated by C-Lasso. The data used is a state-level balanced panel from 1993 to 2002.

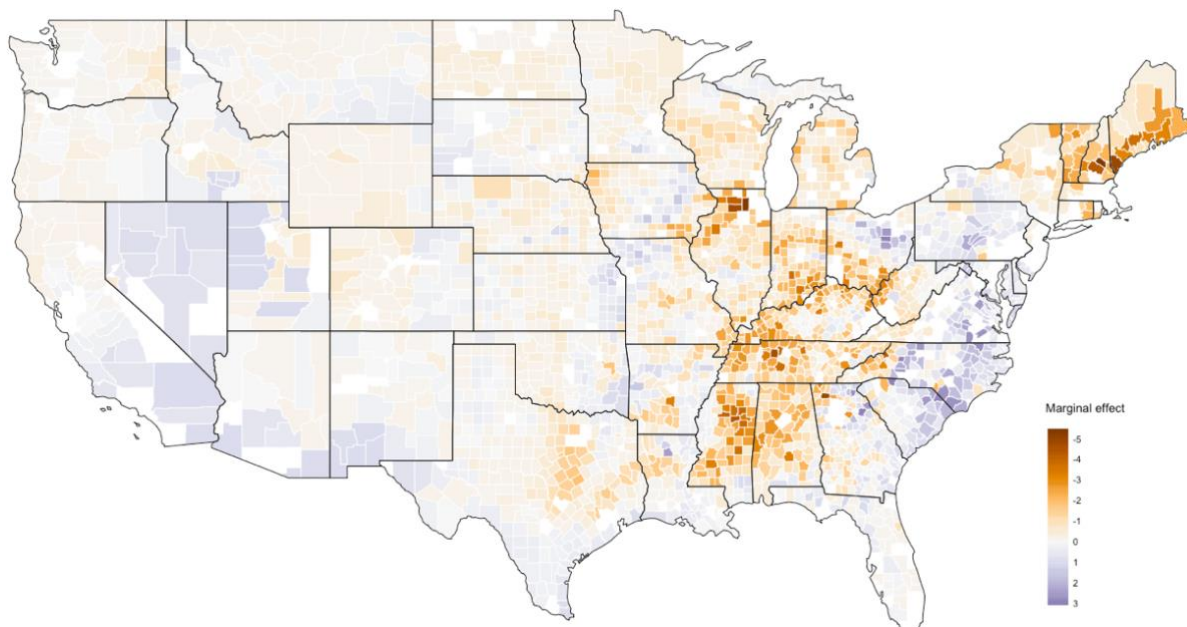


Figure C.4 The point estimates of marginal effect from training causal forests

Note: This figure maps the marginal effect of summer temperature in standard deviation units.

Supplementary Text

C-Lasso

We experiment with the C-Lasso⁴⁰ (or classifier-Lasso) suggested by (Su et al., 2016b) to determine the number of groups and to identify the group membership of each county. Su et al. (2016) demonstrate that C-Lasso can yield a consistent classification which classifies each individual into the correct group with probability approaching one. This method provides a mechanism to estimate the latent groupings that best accommodate the data, without assuming a priori knowledge of the group structure.

The starting point of C-Lasso algorithm is a panel data model with complete slope heterogeneity⁴¹:

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_i + \mu_i + \eta_t + \varepsilon_{it} \quad (\text{C.1})$$

where y_{it} denotes dependent variable. Independent variables are $\mathbf{x}_{it} = (x_{it,1}, \dots, x_{it,p}, \dots, x_{it,P})$, of which each element represents p th variable for the i th individual at time t . $\boldsymbol{\beta}_i$ is a $p \times 1$ vector of individual slope parameters. μ_i is individual fixed effect and η_t is time fixed effect. The following optimization describes how C-Lasso is designed to select the optimal numbers of groups:

$$\min_{\boldsymbol{\beta}_i, \boldsymbol{\alpha}_j} Q_{J,\lambda}(\boldsymbol{\beta}_i, \boldsymbol{\alpha}_k) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta}_i - \mu_i - \eta_t)^2 + \frac{\lambda}{N} \sum_{i=1}^N \prod_{j=1}^{J=1} \|\boldsymbol{\beta}_i - \boldsymbol{\alpha}_j\| \quad (\text{C.2})$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_J)'$ is a matrix of group-specific slope parameters with j th vector being the same slope shared by the individuals belonging to group j . J denotes the unknown number of groups, which is bounded by the maximum number of groups, J_{max} . In eq. (C.2), $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_N)'$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_J)'$ are chosen by minimizing the C-Lasso objective function $Q_{J,\lambda}$,

⁴⁰ Lasso stands for least absolute shrinkage and selection operator.

⁴¹ Under this assumption, slope coefficients are treated as completely different across different individual units and each β_i is estimated by separate regression. Traditional panel model frequently assumes complete slope homogeneity.

given the number of groups J and a tuning parameter λ . For each combination of J and λ , We obtain a set of estimates of $\{\hat{\beta}_i(J, \lambda), \hat{\alpha}_j(J, \lambda)\}$ by minimizing $Q_{J, \lambda}$. λ serves as a weight to the penalty term⁴², which will be jointly determined with unknown J by minimizing the following information criterion (IC)⁴³:

$$\min_{1 \leq j \leq J_{max}, \lambda} IC(J, \lambda) = \phi(y_{it}, x_{it}, \hat{\beta}_i(J, \lambda), \hat{\alpha}_j(J, \lambda)) \quad (C.3)$$

Intuitively, the correct numbers of groups J^* is determined by a measure of goodness fit of eq. (C.1) based on estimates of $\{\hat{\beta}_i(J, \lambda), \hat{\alpha}_j(J, \lambda)\}$. With such J^* and λ^* , an individual i is assigned to group j^* , if $\hat{\beta}_i(J^*, \lambda^*)$ is close to some $\hat{\alpha}_{j^*}(J^*, \lambda^*)$, i.e. $\hat{\beta}_i = \hat{\alpha}_{j^*}$, where $j^* = \operatorname{argmin}_{1 \leq j \leq J^*} \|\hat{\beta}_i - \hat{\alpha}_j\|$.⁴⁴

C-Lasso produces grouping that enjoys the classification consistency in a panel data model with unknown numbers of groups and latent group structure. In contrast to regular clustering algorithms like K-means clustering that relies on the similarity between observations of explanatory variables to assign group membership, C-Lasso classifies individuals so that the dissimilarity of response relationship of outcome variables to explanatory variables within a group is smaller than those in a different group.

Causal Forests

We implement Wager and Athey (2018)'s causal forest approach which extends random forest algorithm to the problem of estimating heterogeneous causal effects in a randomized study. Causal

⁴² Following (Su et al., 2016b), $\lambda = cS_Y^2 T^{-1/3}$. c is a geometrically increasing sequence of 10 point in $(0.2, \dots, 2)$, i.e. $c = 0.2 \times 10^{\frac{r-1}{9}}$ where $r = 1, \dots, 10$; S_Y^2 is the sample variance of $\tilde{y}_{it}$, where $\tilde{y}_{it} = y_{it} - \bar{y}_i$.

⁴³ See Page 2227 of (Su et al., 2016b)) for the explicit form of IC.

⁴⁴ For example, suppose $p = 1$, $N = 3$, $J^* = 2$, and estimates of $\hat{\beta}_1 = 10$, $\hat{\beta}_2 = 5$, $\hat{\beta}_3 = 1$ and $\hat{\alpha}_1 = 4$, $\hat{\alpha}_2 = 8$. Then grouping structure is selected as: individual 1 belongs to group 2, while individuals 2 and 3 belong to group 1.

forests can also be applied in observational studies with continuous treatment. In our case, we want to estimate the impact of a climate variable W_i on the outcome Y_i for individual county i given covariates X s. Namely, we compute an average partial effect $\tau(x) = \frac{Cov(Y_i, W_i | X_i = x)}{Var(W_i | X_i = x)}$ assuming unconfoundedness, that is, the treatment assignment W_i is as good as random conditionally on covariates. Thus, we are no longer primarily interested in predicting the outcome Y_i . We are instead interested in predicting this function $\tau(x)$, the treatment effect at x . Compared to a standard decision tree, a causal tree in the causal forest divides the data and finds splits where treatment effects differ the most. As a result, the data in each terminal leaf are as similar as possible and the average effect is the predicted treatment effect $\hat{\tau}_l$ for observations with X s belonging to the same terminal leaf l . To ensure valid statistical inference, a causal forest is made up of “honest trees”: the algorithm partitions the data in two, one for determining the splits in the tree and another one for estimating the treatment effect in each leaf. As such, the variance of causal forests can be consistently estimated.

REFERENCES:

- Abatzoglou, J.T., Williams, A.P., 2016. Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences* 113, 11770–11775.
- Abeysinghe, T., Lee, C., 1998. Best linear unbiased disaggregation of annual GDP to quarterly figures: the case of Malaysia. *Journal of Forecasting* 17, 527–537.
- Adams, R.M., 1989. Global Climate Change and Agriculture: An Economic Perspective. *American Journal of Agricultural Economics* 71, 1272.
- Aghion, P., Howitt, P., Mayer-Foulkes, D., 2005. The effect of financial development on convergence: Theory and evidence. *The quarterly journal of economics* 120, 173–222.
- Alba, C., Pan, B., Yin, J., Rice, W.L., Mitra, P., Lin, M., Liang, Y., 2022. COVID-19's Impact on Visitation Behavior to US National Parks From Communities of Color - Evidence From Mobile Phone Data.
- Alonso-Almeida, M.-M., Borrajo-Millán, F., Yi, L., 2019. Are Social Media Data Pushing Overtourism? The Case of Barcelona and Chinese Tourists. *Sustainability* 11, 3356.
- Altangerel, K., Kull, C.A., 2013. The prescribed burning debate in Australia: conflicts and compatibilities. *Journal of Environmental Planning and Management* 56, 103–120.
- Balch, J.K., Bradley, B.A., Abatzoglou, J.T., Nagy, R.C., Fusco, E.J., Mahood, A.L., 2017. Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences* 114, 2946–2951.
- Baltagi, B.H., Fingleton, B., Pirotte, A., 2014. Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices. *Journal of Urban Economics* 80, 76–86.
- Barreca, A., Clay, K., Deschênes, O., Greenstone, M., Shapiro, J., 2016a. Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy* 124, 105–159.
- Barreca, A., Clay, K., Deschenes, O., Greenstone, M., Shapiro, J.S., 2016b. Adapting to Climate Change : The Remarkable Decline in the US Temperature-Mortality Relationship over the Twentieth Century. *Journal of Political Economy* 124, 105–159.
- Bawa, R.S., 2017. Effects of wildfire on the value of recreation in western North America. *Journal of sustainable forestry* 36, 1–17.
- Bellemare, M.F., Wichman, C.J., 2020. Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics* 82, 50–61.

- Bergstrom, J.C., Stowers, M., Shonkwiler, J.S., 2020. What Does the Future Hold for U.S. National Park Visitation? Estimation and Assessment of Demand Determinants and New Projections. *Journal of Agricultural and Resource Economics* 45.
- Black, C., Tesfaigzi, Y., Bassein, J.A., Miller, L.A., 2017. Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. *Environmental Toxicology and Pharmacology* 55, 186–195.
- Blanc, E., Schlenker, W., 2017. The Use of Panel Models in Assessments of Climate Impacts on Agriculture. *Review of Environmental Economics and Policy* 11, 258–279.
- Borgschulte, M., Molitor, D., Zou, E., 2022. Air Pollution and the Labor Market: Evidence from Wildfire Smoke (NBER Working Papers No. 29952). National Bureau of Economic Research, Inc.
- Brey, S.J., Fischer, E.V., 2016. Smoke in the City: How Often and Where Does Smoke Impact Summertime Ozone in the United States? *Environmental Science & Technology* 50, 1288–1294.
- Brey, S.J., Ruminski, M., Atwood, S.A., Fischer, E.V., 2018. Connecting smoke plumes to sources using Hazard Mapping System (HMS) smoke and fire location data over North America. *Atmospheric Chemistry and Physics* 18, 1745–1761.
- Brücker, H., Siliverstovs, B., 2006. On the estimation and forecasting of international migration: how relevant is heterogeneity across countries? *Empirical Economics* 31, 735–754.
- Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* 28, 281–298.
- Bun, M.J., Harrison, T.D., 2019. OLS and IV estimation of regression models including endogenous interaction terms. *Econometric Reviews* 38, 814–827.
- Bureau of Economic Analysis (BEA), U.S., 2020. Outdoor Recreation Satellite Account, U.S. and States, 2019. [WWW Document]. URL <https://www.bea.gov/news/2020/outdoor-recreation-satellite-account-us-and-states-2019>
- Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., Wara, M., 2021. The changing risk and burden of wildfire in the United States. *Proceedings of the National Academy of Sciences* 118.
- Burke, M., Dykema, J., Lobell, D.B., Miguel, E., Satyanath, S., 2015. Incorporating Climate Uncertainty into Estimates of Climate Change Impacts. *Review of Economics and Statistics* 97, 461–471.
- Burkhardt, J., Bayham, J., Wilson, A., Carter, E., Berman, J.D., O'Dell, K., Ford, B., Fischer, E.V., Pierce, J.R., 2019. The effect of pollution on crime: Evidence from data on particulate matter and ozone. *Journal of Environmental Economics and Management* 98, 102267.

- Cahill, K., Collins, R., McPartland, S., Pitt, A., Verbos, R., 2018. Overview of the Interagency Visitor Use Management Framework and the uses of social science in its implementation in the National Park Service, in: The George Wright Forum. George Wright Society, pp. 32–41.
- Callaway, B., Goodman-Bacon, A., Sant’Anna, P.H.C., 2021. Difference-in-Differences with a Continuous Treatment.
- Callaway, B., Sant’Anna, P.H.C., 2021. Difference-in-Differences with multiple time periods. *Journal of Econometrics* 225, 200–230.
- Chamberlin, G., 2009. Methods Explained. *Economic & Labour Market Review* 3, 48–57.
- Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D., Leskovec, J., 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589, 82–87.
- Chen, B., 2007. An Empirical Comparison of Methods for Temporal Disaggregation at the National Accounts. Office of Directors Bureau of Economic Analysis, Washington, DC 86.
- Cheng, H., 2014. Analysis of Panel Data, Econometric Society Monograph. Cambridge University Press.
- Chestnut, L.G., Rowe, R.D., 1990. Preservation values for visibility protection at the national parks. Cooperative Agreement# CR-813-686. Research Triangle Park, NC: US Environmental Protection Agency.
- Chiou, L., Tucker, C., 2020. Social Distancing, Internet Access and Inequality.
- Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society* 591–605.
- Chow, G.C., Lin, A.-L., 1976. Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association* 71, 719–721.
- Chow, G.C., Lin, A.-L., 1971. Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics* 53, 372.
- Ciammola, A., Palma, F.D., Marini, M., 2005. Temporal disaggregation techniques of time series by related series: a comparison by a Monte Carlo experiment 44.
- Clark, M., Wilkins, E.J., Dagan, D.T., Powell, R., Sharp, R.L., Hillis, V., 2019. Bringing forecasting into the future: Using Google to predict visitation in U.S. national parks. *Journal of Environmental Management* 243, 88–94.

- Cochrane, M.A., Moran, C.J., Wimberly, M.C., Baer, A.D., Finney, M.A., Beckendorf, K.L., Eidenshink, J., Zhu, Z., 2012. Estimation of wildfire size and risk changes due to fuels treatments. *Int. J. Wildland Fire* 21, 357.
- Creany, N.E., Monz, C.A., D'Antonio, A., Sisneros-Kidd, A., Wilkins, E.J., Nesbitt, J., Mitrovich, M., 2021. Estimating trail use and visitor spatial distribution using mobile device data: An example from the nature reserve of orange county, California USA. *Environmental Challenges* 4, 100171.
- Dall'Erba, S., Chen, Z., Nava, N.J., 2021. US interstate trade will mitigate the negative impact of climate change on crop profit. *American Journal of Agricultural Economics* 103, 1720–1741.
- Dall'erba, S., Domínguez, F., 2016a. The Impact of Climate Change on Agriculture in the Southwestern United States: The Ricardian Approach Revisited. *Spatial Economic Analysis* 11, 46–66.
- Dall'erba, S., Domínguez, F., 2016b. The Impact of Climate Change on Agriculture in the Southwestern United States: The Ricardian Approach Revisited. *Spatial Economic Analysis* 11, 46–66.
- Darwin, R., 1999. The impact of global warming on agriculture: A Ricardian analysis: Comment. *American Economic Review* 89, 1049–1052.
- Davidson, R., MacKinnon, J.G., 1981. Several Tests for Model Specification in the Presence of Alternative Hypotheses. *Econometrica: Journal of the Econometric Society* 49, 781–793.
- Davis, J., Heller, S.B., 2017. Using causal forests to predict treatment heterogeneity: An application to summer jobs. *American Economic Review* 107, 546–50.
- De Chaisemartin, C., d'Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review* 110, 2964–96.
- Denton, F.T., 1971. Adjustment of monthly or quarterly series to annual totals: an approach based on quadratic minimization. *Journal of the American Statistical Association* 66, 99–102.
- Deschênes, O., Greenstone, M., 2007a. The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather. *American Economic Review* 97, 354–385.
- Deschênes, O., Greenstone, M., 2007b. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *American Economic Review* 97, 354–385.
- Duffield, J.W., Neher, C.J., Patterson, D.A., Deskins, A.M., 2013. Effects of wildfire on national park visitation and the regional economy: a natural experiment in the Northern Rockies. *International Journal of Wildland Fire* 22, 1155.

- EIA, U., 2020. Weekly retail gasoline and diesel prices. US Energy Information Administration (EIA).
- Englin, J., Boxall, P.C., Chakraborty, K., 1996. Valuing the Impacts of Forest Fires on Backcountry Forest Recreation. *Forest Science* 42, 6.
- Englin, J., Holmes, T.P., Lutz, J., 2008. Wildfire and the Economic Value of Wilderness Recreation, in: Holmes, T.P., Prestemon, J.P., Abt, K.L. (Eds.), *The Economics of Forest Disturbances*. Springer Netherlands, Dordrecht, pp. 191–208.
- Fernandez, R.B., 1981. A Methodological Note on the Estimation of Time Series. *The Review of Economics and Statistics* 63, 471.
- Fisher, A.C., Hanemann, W.M., Roberts, M.J., Schlenker, W., 2012. The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather: Comment. *American Economic Review* 102, 3749–3760.
- Fisher, D.M., Wood, S.A., Roh, Y.-H., Kim, C.-K., 2019. The Geographic Spread and Preferences of Tourists Revealed by User-Generated Information on Jeju Island, South Korea. *Land* 8, 73.
- Fisichelli, N.A., Schuurman, G.W., Monahan, W.B., Ziesler, P.S., 2015. Protected Area Tourism in a Changing Climate: Will Visitation at US National Parks Warm Up or Overheat? *PLOS ONE* 10, e0128226.
- Ford, B., Burke, M., Lassman, W., Pfister, G., Pierce, J.R., 2017. Status update: is smoke on your mind? Using social media to assess smoke exposure. *Atmospheric Chemistry and Physics* 17, 7541–7554.
- Fotheringham, A.S., Brunson, C., Charlton, M., 2003. Geographically weighted regression: the analysis of spatially varying relationships. John Wiley & Sons.
- Freeman III, A.M., Herriges, J.A., Kling, C.L., 2014. *The Measurement of Environmental and Resource Values*, 3rd Edition. ed. Routledge.
- Gellman, J., Walls, M., Wibbenmeyer, M., 2022. Wildfire, smoke, and outdoor recreation in the western United States. *Forest Policy and Economics* 134, 102619.
- Goodman, A.C., Dubin, R.A., 1990. Sample Stratification with Non-Nested Alternatives: Theory and a Hedonic Example. *The Review of Economics and Statistics* 72, 168.
- Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225, 254–277.
- Goolsbee, A., Syverson, C., 2021. Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020. *Journal of Public Economics* 193, 104311.

- Gordon, J.S., Gruver, J.B., Flint, C.G., Luloff, A., 2013. Perceptions of wildfire and landscape change in the Kenai Peninsula, Alaska. *Environmental management* 52, 807–820.
- Greene, W., 2018. *Econometric Analysis*, 8th edition. ed. Pearson.
- Grossman, G., Kim, S., Rexer, J.M., Thirumurthy, H., 2020. Political partisanship influences behavioral responses to governors' recommendations for COVID-19 prevention in the United States. *Proceedings of the National Academy of Sciences* 117, 24144–24153.
- Haider, W., Knowler, D., Trenholm, R., Moore, J., Bradshaw, P., Lertzman, K., 2019. Climate change, increasing forest fire incidence, and the value of visibility: evidence from British Columbia, Canada. *Canadian Journal of Forest Research* 49, 1242–1255.
- Herwartz, H., Xu, F., 2009. Panel data model comparison for empirical saving-investment relations. *Applied Economics Letters* 16, 803–807.
- Hesseln, H., Loomis, J.B., González-Cabán, A., 2003. The effects of fire on hiking demand: A travel cost study of Colorado and Montana, in: *USDA Forest Service Proceedings RMRS-P-29*. pp. 177–186.
- Heutel, G., Miller, N.H., Molitor, D., 2021. Adaptation and the mortality effects of temperature across US climate regions. *Review of Economics and Statistics* 103, 740–753.
- Hewer, M., Scott, D., Fenech, A., 2016. Seasonal weather sensitivity, temperature thresholds, and climate change impacts for park visitation. *Tourism Geographies* 18, 297–321.
- Hoerling, M., Eischeid, J., Perlwitz, J., Quan, X.-W., Wolter, K., Cheng, L., 2016. Characterizing recent trends in US heavy precipitation. *Journal of Climate* 29, 2313–2332.
- Hsiang, S., 2016. Climate Econometrics. *Annual Review of Resource Economics* 8, 43–75.
- Islam, M.R., 2009. Evaluation of Different Temporal Disaggregation Techniques and an Application to Italian GDP 12.
- Jay, J., Heykoop, F., Hwang, L., de Jong, J., Kondo, M., 2021. Effects of the COVID-19 Pandemic on Park Use in U.S. Cities.
- Juhász, L., Hochmair, H., 2020. Studying Spatial and Temporal Visitation Patterns of Points of Interest Using SafeGraph Data in Florida. *GI_Forum* 1, 119–136.
- Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., Kruse, J., 2020. Multiscale dynamic human mobility flow dataset in the US during the COVID-19 epidemic. *Scientific data* 7, 1–13.
- Karl, T.R., Koss, W.J., 1984. Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983.

- Keiser, D., Lade, G., Rudik, I., 2018. Air pollution and visitation at U.S. national parks. *Science Advances* 4, eaat1613.
- Kim, M.-K., Jakus, P.M., 2019. Wildfire, national park visitation, and changes in regional economic activity. *Journal of Outdoor Recreation and Tourism* 26, 34–42.
- Klein, L.R., 1958. The Estimation of Distributed Lags. *Econometrica* 26, 553.
- Knittel, C.R., Stolper, S., 2019. Using machine learning to target treatment: The case of household energy use. National Bureau of Economic Research.
- Kubo, T., Uryu, S., Yamano, H., Tsuge, T., Yamakita, T., Shirayama, Y., 2020. Mobile phone network data reveal nationwide economic value of coastal tourism under climate change. *Tourism Management* 77, 104010.
- Kupfer, J.A., Li, Z., Ning, H., Huang, X., 2021. Using Mobile Device Data to Track the Effects of the COVID-19 Pandemic on Spatiotemporal Patterns of National Park Visitation. *Sustainability* 13, 9366.
- Lancaster, T., 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391–413.
- Larsen, L.N.D., Howe, P.D., Brunson, M., Yocom, L., McAvoy, D., Berry, E.H., Smith, J.W., 2021. Risk perceptions and mitigation behaviors of residents following a near-miss wildfire. *Landscape and Urban Planning* 207, 104005.
- Lawson, M., 2021. Innovative new ways to count outdoor recreation: Using data from cell phones, fitness trackers, social media, and other novel data sources. *Headwaters Economics*.
- LeSage, J., Pace, R.K., 2009. Introduction to spatial econometrics. Chapman and Hall/CRC.
- Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: A literature review. *Tourism Management* 68, 301–323.
- Li, X., Law, R., Xie, G., Wang, S., 2021. Review of tourism forecasting research with internet data. *Tourism Management* 83, 104245.
- Liang, Y., Yin, J., Pan, B., Lin, M., Chi, G., 2021. Assessing the validity of SafeGraph data for visitor monitoring in Yellowstone National Park. *Travel and Tourism Research Association: Advancing Tourism Research Globally* 13.
- Litterman, R.B., 1983. A Random Walk, Markov Model for the Distribution of Time Series. *Journal of Business & Economic Statistics* 1, 169–173.
- Loomis, J., González-Cabán, A., Englin, J., 2001. Testing for Differential Effects of Forest Fires on Hiking and Mountain Biking Demand and Benefits. *Journal of Agricultural and Resource Economics* 16.

- Loomis, J.B., 2000. Counting on Recreation Use Data: A Call for Long-Term Monitoring. *Journal of Leisure Research* 32, 93–96.
- Manning, R., Valliere, W., Anderson, L., Pettengill, P., Reigner, N., Budruk, M., Laven, D., Hallo, J., Park, L., Bacon, J., Abbe, D., Goonan, K., 2011. Defining, Measuring, Monitoring, and Managing the Sustainability of Parks for Outdoor Recreation. *Journal of Park and Recreation Administration* 29, 14.
- Manning, R.E., 2011. *Studies in Outdoor Recreation: Search and Research for Satisfaction*. Oregon State University Press.
- Manning, R.E., 1985. Crowding norms in backcountry settings: A review and synthesis. *Journal of Leisure Research* 17, 75–89.
- Martin, W.E., Martin, I.M., Kent, B., 2009. The role of risk perceptions in the risk mitigation process: The case of wildfire in high risk communities. *Journal of Environmental Management* 91, 489–498.
- Mendelsohn, R., Dinar, A., 2003a. Climate, Water, and Agriculture. *Land Economics* 79, 328–341.
- Mendelsohn, R., Dinar, A., 2003b. Climate , Water , and Agriculture. *Land Economics* 79, 328–341.
- Mendelsohn, R., Nordhaus, W.D., Shaw, D., 1994. The Impact of Global Warming on Agriculture: A Ricardian Analysis. *American Economic Review* 84, 753–771.
- Mendelsohn, R.O., Massetti, E., 2017. The Use of Cross-Sectional Analysis to Measure Climate Impacts on Agriculture: Theory and Evidence. *Review of Environmental Economics and Policy* 11, 280–298.
- Merrill, N.H., Atkinson, S.F., Mulvaney, K.K., Mazzotta, M.J., Bousquin, J., 2020. Using data derived from cellular phone locations to estimate visitation to natural areas: An application to water recreation in New England, USA. *PLOS ONE* 15, e0231863.
- Miller, N., Molitor, D., Zou, E., 2017. Blowing Smoke: Health Impacts of Wildfire Plume Dynamics.
- Miller, S., 2020. Causal forest estimation of heterogeneous and time-varying environmental policy effects. *Journal of Environmental Economics and Management* 103, 102337.
- Moeltner, K., Englin, J., 2004. Choice Behavior Under Time-Variant Quality: State Dependence Versus “Play-It-by-Ear” in Selecting Ski Resorts. *Journal of Business & Economic Statistics* 22, 214–224.
- Monz, C., Creany, N., Nesbitt, J., Mitrovich, M., 2021. Mobile Device Data Analysis to Determine the Demographics of Park Visitors. *Journal of Park and Recreation Administration*.

- Monz, C., Mitrovich, M., D'Antonio, A., Sisneros-Kidd, A., 2019. Using Mobile Device Data to Estimate Visitation in Parks and Protected Areas: An Example from the Nature Reserve of Orange County, California. *Journal of Park and Recreation Administration*.
- Mosley, L., Eckley, I., Gibberd, A., 2021. Sparse Temporal Disaggregation. arXiv:2108.05783.
- Nagy, R., Fusco, E., Bradley, B., Abatzoglou, J.T., Balch, J., 2018. Human-related ignitions increase the number of large wildfires across US ecoregions. *Fire* 1, 4.
- Pavía-Miralles, J.M., 2010. A Survey of Methods to Interpolate, Distribute and Extra- polate Time Series. *Journal of Service Science and Management* 3, 449–463.
- Pesaran, M.H., 2015. *Time Series and Panel Data Econometrics*. Oxford University Press.
- Plantinga, A.J., Lubowski, R.N., Stavins, R.N., 2002. The effects of potential land development on agricultural land prices. *Journal of Urban Economics* 52, 561–581.
- Polsky, C., 2004. Putting Space and Time in Ricardian Climate Change Impact Studies: Agriculture in the U.S. Great Plains, 1969–1992. *Annals of the Association of American Geographers* 94, 549–564.
- Polsky, C., Easterling, W.E., 2001. Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities. *Agriculture, Ecosystems & Environment* 85, 133–144.
- PRISM Climate Group, 2004. PRISM Climate Data.
- Quilis, E.M., 2018. Temporal disaggregation of economic time series: The view from the trenches: Temporal disaggregation from the trenches. *Statistica Neerlandica* 72, 447–470.
- Rice, W.L., Pan, B., 2021. Understanding changes in park visitation during the COVID-19 pandemic: A spatial application of big data. *Wellbeing, Space and Society* 2, 100037.
- Rice, W.L., Rushing, J., Thomsen, J., Whitney, P., 2021. Exclusionary Effects of Campsite Allocation through Reservations in US National Parks: Evidence from Mobile Device Location Data. *Journal of Park and Recreation Administration*.
- Rice, W.L., Taff, B.D., Miller, Z.D., Newman, P., Zipp, K.Y., Pan, B., Newton, J.N., D'Antonio, A., 2020. Connecting motivations to outcomes: A study of park visitors' outcome attainment. *Journal of Outdoor Recreation and Tourism* 29, 100272.
- Riper, C.J. van, Foelske, L., Kuwayama, S.D., Keller, R., Johnson, D., 2020. Understanding the role of local knowledge in the spatial dynamics of social values expressed by stakeholders. *Applied Geography* 123, 102279.
- Rolph, G.D., Draxler, R.R., Stein, A.F., Taylor, A., Ruminski, M.G., Kondragunta, S., Zeng, J., Huang, H.-C., Manikin, G., McQueen, J.T., Davidson, P.M., 2009. Description and

- Verification of the NOAA Smoke Forecasting System: The 2007 Fire Season. *Weather and Forecasting* 24, 361–378.
- Rossi, L., Zanetti Chini, E., 2021. Temporal disaggregation of business dynamics: New evidence for U.S. economy. *Journal of Macroeconomics* 69, 103337.
- Rothman, H.K., 2005. A test of adversity and strength: wildland fire in the National Park system. US Department of the Interior, National Park Service.
- Runte, A., 1997. National parks: the American experience. U of Nebraska Press.
- SafeGraph, 2022. Top 6 Mobile Location Data Providers for Foot Traffic Insights [WWW Document]. Top 6 Mobile Location Data Providers for Foot Traffic Insights | SafeGraph. URL <https://www.safegraph.com/guides/mobile-location-data-providers> (accessed 5.5.22).
- SafeGraph, 2019. Quantifying Sampling Bias in SafeGraph Patterns [WWW Document]. Google Colab. URL <https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#scrollTo=oBADlZgHTsWg> (accessed 5.5.22).
- Sánchez, J.J., Baerenklau, K., González-Cabán, A., 2016. Valuing hypothetical wildfire impacts with a Kuhn–Tucker model of recreation demand. *Forest Policy and Economics* 71, 63–70.
- Santos Silva, J.M.C., Cardoso, F.N., 2001. The Chow-Lin method using dynamic models. *Economic Modelling* 18, 269–280.
- Schlenker, W., Hanemann, W.. M., Fisher, A.C., 2005a. Will U.S. Agriculture Really Benefit from Global Warming ? Hedonic for Irrigation Approach. *The American Economic Review* 95, 395–406.
- Schlenker, W., Hanemann, W.M., Fisher, A.C., 2006a. The impact of global warming on US agriculture: an econometric analysis of optimal growing conditions. *Review of Economics and statistics* 88, 113–125.
- Schlenker, W., Hanemann, W.M., Fisher, A.C., 2005b. Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach. *The American Economic Review* 95, 395–406.
- Schlenker, W., Hanemann, W.M.M., Fisher, A.C., 2006b. The Impact of Global Warming on the U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions. *Review of Economics and Statistics* 88, 113–125.
- Schlenker, W., Roberts, M.J., 2009a. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106, 15594–15598.

- Schlenker, W., Roberts, M.J., 2009b. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences* 106, 15594–15598.
- Schultze, W.D., Brookshire, D.S., Walther, E.G., MacFarland, K.K., 1983. The economic benefits of preserving visibility in the national parklands of the Southwest. *Nat. Resources J.* 23, 149.
- Scott, D., Jones, B., Konopek, J., 2007. Implications of climate and environmental change for nature-based tourism in the Canadian Rocky Mountains: A case study of Waterton Lakes National Park. *Tourism Management* 28, 570–579.
- Sessions, C., Wood, S.A., Rabotyagov, S., Fisher, D.M., 2016. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management* 183, 703–711.
- Short, K.C., 2017. Spatial wildfire occurrence data for the United States, 1992-2015 [FPA_FOD_20170508].
- Silva, J.S., Tenreyro, S., 2011. Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters* 112, 220–222.
- Silva, J.S., Tenreyro, S., 2006. The log of gravity. *The Review of Economics and statistics* 88, 641–658.
- Smith, A.E., Kemp, M.A., Savage, T.H., Taylor, C.L., 2005. Methods and results from a new survey of values for eastern regional haze improvements. *Journal of the Air & Waste Management Association* 55, 1767–1779.
- Smith, J.W., Seekamp, E., McCreary, A., Davenport, M., Kanazawa, M., Holmberg, K., Wilson, B., Nieber, J., 2016. Shifting demand for winter outdoor recreation along the North Shore of Lake Superior under variable rates of climate change: A finite-mixture modeling approach. *Ecological Economics* 123, 1–13.
- Sonter, L.J., Watson, K.B., Wood, S.A., Ricketts, T.H., 2016. Spatial and Temporal Dynamics and Value of Nature-Based Recreation, Estimated via Social Media. *PLOS ONE* 11, e0162372.
- Starbuck, C.M., Berrens, R.P., McKee, M., 2006. Simulating changes in forest recreation demand and associated economic impacts due to fire and fuels management activities. *Forest Policy and Economics* 8, 52–66.
- Su, L., Shi, Z., Phillips, P.C.B., 2016a. Identifying Latent Structures in Panel Data. *Econometrica* 84, 2215–2264.
- Su, L., Shi, Z., Phillips, P.C.B., 2016b. Identifying Latent Structures in Panel Data. *Econometrica* 84, 2215–2264.

- Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225, 175–199.
- Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports* 7, 17615.
- Thapa, B., Cahyanto, I., Holland, S.M., Absher, J.D., 2013. Wildfires and tourist behaviors in Florida. *Tourism Management* 36, 284–292.
- Timmins, C., 2006. Endogenous Land use and the Ricardian Valuation of Climate Change. *Environmental & Resource Economics* 33, 119–142.
- U.S. Department of the Interior, 2021. National Parks Overcrowding [WWW Document]. National Parks Overcrowding | U.S. Department of the Interior. URL <https://www.doi.gov/ocl/national-parks-overcrowding> (accessed 5.2.22).
- Vanschoenwinkel, J., Mendelsohn, R., Van Passel, S., 2016. Do Western and Eastern Europe have the same agricultural climate response? Taking adaptive capacity into account. *Global Environmental Change* 41, 74–87.
- Vaske, J.J., Lyon, K.M., 2014. Linking the 2010 Census to National Park Visitors. Natural Resource Technical Report NPS/WASO/NRTR—2014/880. National Park Service, Fort Collins, Colorado.
- Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113, 1228–1242.
- Wendling, T., Jung, K., Callahan, A., Schuler, A., Shah, N.H., Gallego, B., 2018. Comparing methods for estimation of heterogeneous treatment effects using observational data from health care databases. *Statistics in medicine* 37, 3309–3324.
- White, E., Bowker, J., Askew, A.E., Langner, L.L., Arnold, J.R., English, D.B., 2016. Federal outdoor recreation trends: effects on economic opportunities. Technical report, U.S. Department of Agriculture, Forest Service, Pacific Northwest Station. 945.
- Wood, S.A., Guerry, A.D., Silver, J.M., Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. *Scientific Reports* 3, 2976.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Yang, A., Yang, J., Yang, D., Xu, R., He, Y., Aragon, A., Qiu, H., 2021. Human Mobility to Parks Under the COVID-19 Pandemic and Wildfire Seasons in the Western and Central United States. *GeoHealth* 5.
- Ziesler, P.S., 2020. Statistical abstract: 2019. Natural Resource Data Series NPS/NRSS/EQD/NRDS—2020/1272. National Park Service, Fort Collins, Colorado.

Ziesler, P.S., Pettebone, D., 2018. Counting on Visitors: A Review of Methods and Applications for the National Park Service's Visitor Use Statistics Program. *Journal of Park and Recreation Administration* 36, 39–55.