UNIVERSITY OF BIRMINGHAM

ESSAYS ON THE ECONOMICS OF WILDFIRES

by

SARAH MEIER

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Birmingham Business School

School of Geography, Earth and Environmental Sciences

University of Birmingham

June 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This doctoral thesis presents three empirical essays on the economics of wildfires. The first essay models and compares the risk of extreme wildfires in Mediterranean Europe at the country level. We link geospatial data on burn perimeters, meteorological conditions, population density, and land cover from 2006 to 2019. Employing extreme value analysis the study identifies the highest risk of extreme wildfires for Portugal, followed by Greece, Spain, and Italy. We provide 10–, 20–, and 50–year return levels of burned area and associated economic losses. The second essay examines the regional economic impact of wildfires on the growth rate of gross domestic product (GDP) and employment for 233 regions in Southern Europe from 2011 to 2018. Through panel fixed effects instrumental variable estimation, the study finds a contemporaneous decrease in the annual GDP growth rate for regions affected by wildfires. The effect on employment growth is heterogeneous across sectors with a negative effect on retail and tourism offset by a positive effect on insurance and real estate activities. The third essay studies the Great Fire of 1910 in the Northwestern United States, investigating the effects of wildfire-sourced smoke pollution on excess mortality and later-life socioeconomic outcomes of children under the age of five. Utilising historical burn perime-

ters, smoke emission and dispersion modelling, as well as mortality and census data spanning from 1900 to 1940, the study finds a negative effect of smoke exposure on excess mortality in the week of the fire. Furthermore, being exposed in early childhood may lead to a decrease in some later-life socioeconomic status outcomes 20 years after the event. Collectively, these essays contribute to our understanding of wildfire risk, economic consequences, and health implications, providing valuable insights for wildfire management and policy-making efforts.

Declaration

I certify that the thesis presented for examination for the PhD degree of the University of Birmingham is solely my own work other than where collaboration is specifically indicated and that I have fully acknowledged all sources of information that have been used in the thesis. I declare that this work has not been submitted for a degree at any other university or institution.

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This doctoral thesis may not be reproduced without the author's written consent. I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I wrote this thesis as an integral part of my appointment as a research fellow and have received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement Innovative Training Networks No. 860787.

Statement of Prior Publication

A version of Chapter 1 is publicly available as: Meier, S., Strobl, E., Elliott, R. J., & Kettridge, N. (2022). Cross-country risk quantification of extreme wildfires in Mediterranean Europe. *Risk Analysis*.

A version of Chapter 2 is publicly available as: Meier, S., Elliott, R. J., & Strobl, E. (2023). The regional economic impact of wildfires: Evidence from Southern Europe. *Journal of Environmental Economics and Management*, 102787.

Acknowledgements

I am deeply indebted to my outstanding supervisors, Professors Robert Elliott and Eric Strobl, for their exceptional guidance, wisdom, and support throughout my doctoral journey at the University of Birmingham. Their invaluable insights, constructive feedback, and unwavering dedication to my research have been instrumental in shaping my work and helping me to achieve this significant milestone in my academic career. Additionally, I would like to express my genuine appreciation to Professor Nicholas Kettridge, whose advice and encouragement have been invaluable throughout my journey.

I would also like to thank the Department of Economics at the University of Birmingham for providing an intellectually stimulating and supportive environment in which to conduct my research.

To my friends, who have provided moral support, a listening ear, and a healthy dose of humour, I offer my heartfelt thanks. Your encouragement, advice, and wit have kept me grounded and motivated throughout this challenging process, and I am grateful for your unwavering friendship.

Finally, I would like to thank my family, who have always been my rock and my foundation. Your unconditional love, support, and sacrifice have sustained me through the ups and downs of life, and I am deeply grateful for everything you have done for me.

To all those who have contributed to my journey, I offer my sincere appreciation and deepest respect. Thank you for being part of my story and helping me to reach this momentous milestone in my academic career.

Contents

List of Tables

List of Figures

Introduction

Forests cover more than 30% of the Earth's surface and play a crucial role in providing ecosystems for biodiversity (FAO, 2022). Along with plant and animal species, humans have forged a deep alliance with forests relying on them for survival (Roberts, 2019). One dynamic ecological force that has shaped forests for millennia is wildfires, and besides their natural occurrence, fire has been intentionally used for landscape modifications and agricultural practices (Santín and Doerr, 2016). The strong interconnection between humans and flammable landscapes makes wildfire a natural hazard like no other as there is some capacity to either intensify or dampen the fire regime patterns by means of deliberate ignition, fire prevention, and suppression efforts (Bowman, 2018). Nonetheless, this capacity is limited, and coupled with the combined effects of climate change, landscape management choices, and land-use changes, wildfires now frequently encounter weather and fuel conditions conducive to becoming increasingly destructive. Consequently, wildfires impose a significant burden on human, economic, and environmental systems both in the directly affected as well as in surrounding and downwind areas (UNEP, 2022).

Particularly in recent years, individual wildfire events and wildfire seasons have exhibited

a noticeable trend in duration and intensity in many parts of the world (Jolly et al., 2015; Bowman et al., 2020; UNEP, 2022). For instance, the annual burned area has increased by five times in California from 1972 to 2018 (Williams et al., 2019) with an unparalleled wildfire season in 2020 (Williams et al., 2019; Safford et al., 2022). Canada has witnessed an almost three-fold yearly increase since 1959 (Hanes et al., 2019) and is living through a record-breaking wildfire season in 2023, and the Australian "Black Summer" of 2019-2020 is unprecedented with over 19 million hectares (ha) burned (Wen et al., 2022). The Mediterranean region is the most fire-prone in Europe, which was bleakly illustrated by the pronounced 2017 wildfire season in Portugal (NATURE, 2017) or the fatal 2018 Mati wildfire in Greece.

There are a number of ways in which wildfires can affect assets, societies, and ecosystems such as through direct destruction (e.g., infrastructure, homes lost, business downtime, aesthetic alteration of landscapes) or indirect impacts (e.g., health effects, traumatic experiences, loss of revenue due to decreasing tourism). This thesis puts three spotlights on specific aspects of economic wildfire impacts and by no means aims at addressing all of the aforementioned potential effects. The primary geographical focus of the first two chapters is on Southern Europe. The analysis specifically encompasses Portugal, Spain, Italy and Greece. Not only do these countries combined account for about 85% of the total burned area in Europe, but the region also marks a biodiversity hot-spot driven by species richness and a high percentage of endemic species (Batllori et al., 2013; Myers et al., 2000). Furthermore, the Mediterranean region is characterised by an extensive wildland-urban interface and many

coastal areas are highly populated. In contrast, the third chapter studies the historical event of the Big Burn in 1910, one of the largest wildfires on record in the United States, and thus studies a region where wildfires unfold on a significantly grander scale than their European counterparts. The subsequent paragraphs provide a concise overview of each chapter.

The first chapter delves into the realm of extreme wildfires in Southern Europe. Fire has formed an integral part of landscapes for centuries contributing to beneficial ecosystem functions (Holmes et al., 2008) and has been used by communities in order to modify landscapes or for agricultural practices (Doerr and Santín, 2016). However, although the affected people and ecosystems may adapt to near-normal conditions, evidence shows that it is arguably the extreme events that induce the most detrimental social, economic, and environmental impacts (Evin et al., 2018; Tedim et al., 2020). The primary objective of this chapter is to model and analyse the distribution of extremely large wildfires, evaluate their risk probabilities across different countries, and estimate associated monetary losses. Our data set is compiled using geospatial burned area data from the European Forest Fire Information System, a Fire Weather Index reanalysis product, population data, and land cover maps. We employ Extreme Value Theory, or more specifically a Point-Process characterisation of extremes using maximum likelihood estimation. Our results suggest the highest risk of extreme wildfires for Portugal, followed by Greece, Spain, and Italy with a 10-year BA return level of 50,338 ha, 33,242 ha, 25,165 ha, and 8,966 ha, respectively. Coupling these estimates with existing per hectare loss estimates leads to expected economic losses of 162–439 million € for Portugal, 81–219 million € for Spain, 41–290 million € for Greece, and 18–78 million € for Italy for

such 10-year return period extreme events.

In the second chapter, we study the regional economic effects of the aggregate wildfire occurrence and thus, instead of only looking at the extremes we include all wildfires. This is important because wildfires, like most natural hazards, are predominantly local phenomena and potential regional impacts could remain unidentified when evaluating effects on the national level (Horwich, 2000; Botzen et al., 2019). In this study, we set out to quantify the effects of wildfire occurrence, which is measured as fire numbers and percentage burned per region, on the growth of the regional gross domestic product (GDP) and employment. In our two-stage least squares panel fixed-effects estimation, employing the Fire Weather Index for the predominantly forested area as our instrument, we find an average contemporary decrease in a region's yearly GDP growth rate of 0.11-0.18% for wildfire-affected regions. For an average wildfire year over our study period (2011-2018) this suggests a production loss of 1.3-2.1 billion Euros for Southern Europe. Furthermore, our results suggest a heterogeneous impact of wildfire occurrence on the growth of the regional employment rate with negative effects on activities related to retail and tourism offset by an increase in employment in activities related to finance, insurance, and real estate.

Finally, the third chapter assesses the effect of wildfire-sourced air pollution on excess mortality and later-life socioeconomic status outcomes focusing on children under the age of five. Recent evidence shows how wildfire smoke is stalling or even reversing the efforts of improved air quality over the past decades in certain regions of the world (Ford et al., 2018; Burke et al., 2023). This is of particular concern as wildfire smoke exposure has been

linked to a number of negative health outcomes (Chen et al., 2021; Gao et al., 2023). In this study, we scrutinise these potential social costs using the Great Fire of 1910 which engulfed more than 1.2 million hectares in a mere two days in the Northwestern United States. Our data set is assimilated by utilising the historical burn perimeters and modelling the induced smoke dispersion, digitised mortality data, as well as full-count census data for the decades 1900 to 1940. For the effect on the excess mortality rate in young children, we assemble a week-county panel data set and employ a Two-Way Fixed-Effects model. We find a negative effect of smoke exposure on excess mortality in the week of the wildfire with an increase in the mortality rate of $\sim$56% compared to the average observed mortality rate in 1910 in the study area. Moreover, we track about 9,000 boys who resided in smoke-affected counties in 1910 over time and find weak evidence that they rank lower in some socioeconomic status indicators based on income and education 20 years after the event than boys who were not smoke-affected in their early childhood. We find no evidence of a persisting negative effect 30 years after the event.

Overall, this thesis makes some major contributions to the current body on the economics of wildfires. To highlight a few, the first chapter merges high-quality and homogenised data sets, and to the best of our knowledge is the first to conduct a cross-country analysis and comparison of extreme wildfires in Southern Europe. The second chapter overcomes endogeneity concerns related to wildfires introducing a novel instrument which may be used by other researchers and provides some of the first causal estimates of the effect of wildfires on regional economies. Last, the third chapter harnesses a large historic event to shed light on the health

and social costs of wildfire-induced smoke pollution on the particularly vulnerable paediatric population. This is the first study to estimate the effects of wildfire-sourced air pollution on health and socioeconomic outcomes in a historical context. Furthermore, the evaluation of later-life socioeconomic status for adults who were exposed to extreme air pollution in their childhood has neither been explored in a historical context nor within a wildfire setting in general.

The remainder of this doctoral thesis is organised as follows. The first essay titled "Cross-country risk quantification of extreme wildfires in Mediterranean Europe" is presented in Chapter 1. The second essay on "The regional economic impact of wildfires: Evidence from Southern Europe" is presented in Chapter 2, followed by the third essay, "Impacts of wildfire smoke exposure on excess mortality and later-life socioeconomic outcomes: The Great Fire of 1910" in Chapter 3. The thesis concludes with brief final remarks including potential pathways for future research.

Chapter 1

Cross-country risk quantification of extreme wildfires in Mediterranean Europe

Abstract

We estimate the country-level risk of extreme wildfires defined by burned area (BA) for Mediterranean Europe and carry out a cross-country comparison. To this end we avail of the European Forest Fire Information System (EFFIS) geospatial data from 2006-2019 to perform an extreme value analysis. More specifically, we apply a point process characterisation of wildfire extremes using maximum likelihood estimation. By modelling covariates, we also evaluate potential trends and correlations with commonly known factors that drive or affect wildfire occurrence, such as the Fire Weather Index as a proxy for meteorological conditions, population density, land cover type, and seasonality. We find that the highest risk of extreme wildfires is in Portugal (PT), followed by Greece (GR), Spain (ES), and Italy (IT) with a 10-year BA return level of 50'338 ha, 33'242 ha, 25'165 ha, and 8'966 ha, respectively. Coupling our results with existing estimates of the monetary impact of large wildfires suggests expected losses of 162-439 million € (PT), 81-219 million € (ES), 41-290 million € (GR), and 18-78 million € (IT) for such 10-year return period events.

Keywords: environmental economics, environmental hazards, extreme value statistics, risk analysis, wildfires

JEL classification codes: C6, Q5

1.1 Introduction

Wildfires affect humans, assets, and ecosystems and can lead to extensive socioeconomic and environmental impacts (Keeley et al., 2012). Within Europe, the Mediterranean region is the most fire-prone with high wildfire incidence and consequences (San-Miguel-Ayanz et al., 2020). This was bleakly illustrated by the pronounced wildfire season in 2017 with blazing fires in France and roughly 140'000 hectares (ha) burnt in Portugal (NATURE, 2017), or by the 2018 fatal fires in Greece leading to more than 100 deaths and causing major damage to the ecosystems of the susceptible *Natura 2000* protected areas (San-Miguel-Ayanz et al., 2018). Not only do Portugal (PT), Spain (ES), Italy (IT), Greece (GR), and France (FR) combined account for about 85% of the total annual burned area (BA) in Europe (De Rigo et al., 2017), the Mediterranean area is also particularly vulnerable in that it is densely populated, characterised by a large wildland urban interface (WUI) (San-Miguel-Ayanz et al., 2013), and due to the species richness as well as the high proportion of endemisms it marks a "biodiversity hotspot" (Batllori et al., 2013; Myers et al., 2000).

Notably, fire has historically played an integral role in Mediterranean Europe by performing highly beneficial ecosystem functions (Holmes et al., 2008), and has been utilised by communities for agricultural practices (e.g., to fertilise soils and control plant growth) and landscape modifications (Santín and Doerr, 2016). However, although societies and ecosystems are likely to adapt to near-normal conditions, this is arguably not the case for extreme events (Bowman et al., 2017; San-Miguel-Ayanz et al., 2013; Tedim et al., 2018). Rather, evidence shows that particularly large wildfires are linked to severe disturbances and losses

and are the cause of the bulk of social, economic, and adverse environmental impacts (Evin et al., 2018; Gill and Allan, 2008; Mendes et al., 2010).

The purpose of this study is to characterise the spatiotemporal distribution and dynamics of extremely large wildfires in Mediterranean Europe, as well as to quantify and compare their risk probabilities across countries. Since our interest lies in the risk quantification of rare or extreme events, we model the probabilistic structure of the commonly heavy-tailed right tail of the wildfire BA density distribution (Beverly and Martell, 2005; Hernandez et al., 2015; Scotto et al., 2014) applying Extreme Value Theory (EVT). A series of commonly known variables that potentially influence the production of large wildfires, such as the Fire Weather Index (FWI), population density, land cover type, and seasonality, are included as covariates to evaluate potential conditional probabilities. Employing the analytical tools provided by EVT enable to extrapolate wildfires of potentially unobserved size based on the European Forest Fire Information System (EFFIS) BA data set from 2006-2019, and thus, to quantify the risk of country-level extreme wildfires. Furthermore, we convert our estimates into rough monetary losses using figures from the existing literature to facilitate the potential application of our estimates to policy decisions.

EVT has been proven to be a suitable inferential tool for wildfire size risk quantification (Hernandez et al., 2015; Holmes et al., 2008), and has been applied globally (Jiang and Zhuang, 2011; Keyser and Westerling, 2019). For Mediterranean Europe Evin et al. (2018) evaluate the risk of large wildfires in France conditional on a new fire policy introduced in 1994. Moreover, several studies quantify and compare regional wildfire risk and regimes in

Portugal (De Zea Bermudez et al., 2009; Mendes et al., 2010; Scotto et al., 2014), which is unsurprising given the country bears the highest wildfire prevalence within Mediterranean Europe (Turco et al., 2019). However, to the best of our knowledge, ours is the first study to use EVT to perform a cross-country quantification of wildfire risk. Our contribution is three-fold. First, we merge high-quality homogenised and up-to-date geospatial data sets for the European Mediterranean region. Second, we perform a country-level analysis of extremely large wildfires in Mediterranean Europe and, third, we compare the estimated risks across the region.

The remainder of this study is organised as follows. Section 1.2 describes the data sources followed by Section 1.3 outlining the methodology underpinning the extreme value analysis. Section 1.4 summarises the results and derives monetary losses by matching our estimates with economic loss figures from the existing literature. The findings are subsequently discussed in Section 1.5, before Section 1.6 concludes.

1.2 Data and Variables

1.2.1 Burned Area (BA)

We use a high-quality BA spatial data product compiled by the Joint Research Centre (JRC) and provided by the EFFIS.[1] It is the primary source of harmonised data on wildfires in Europe, and thus enables a sound cross-country comparison. The data product is derived

[1] https://effis.jrc.ec.europa.eu.

from the semi-automatic classification of daily processing of Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery at 250-meter spatial resolution. The definite fire perimeters are refined through visual image interpretation and systematically collected fire news from various media. The data set includes fires larger than approximately 30 ha and contains information on the initial date, country, province, place, as well as the BA polygons.[2] We model the extreme BA conditional on the covariates described hereafter. For the full list of covariates refer to Table A1 in Appendix A.1.

1.2.2 Fire Weather Index (FWI)

Weather conditions are a major driver of wildfire events and are commonly applied to construct fire danger indices (Bedia et al., 2014; Krawchuk et al., 2009; Sousa et al., 2015). We employ the FWI component of the Canadian Forest Fire Weather Index System as a proxy for meteorological conditions incorporating temperature, wind speed, relative humidity, and precipitation. Providing a homogeneous numerical rating of relative fire potential resulting from the combination of the two fire behaviour indices, namely the Buildup Index and the Initial Spread Index (Van Wagner and Pickett, 1985), the FWI has become a reference index for European fire danger maps produced by the JRC (Camia et al., 2008). We use a high-resolution calculation developed by Natural Resources Canada based on the European Centre for Medium-Range Weather Forecasts ERA5-HRES[3] reanalysis product presented in McElhinny et al. (2020). To account for the effect of inter-seasonal drought, we use the FWI

[2] https://effis.jrc.ec.europa.eu/about-effis/technical-background/rapid-damage-assessment.

[3] https://cds.climate.copernicus.eu/cdsapp#!/home.

version derived from the overwintered Drought Code with a spatial resolution of 31 kilometres (0.28° on a reduced Gaussian grid).

We spatially join the centroid of every wildfire polygon to the closest grid cell of the FWI data set and extract (i) the daily FWI values one month prior to one week after the initial date of the fire, and (ii) the daily FWI values of the respective year of the fire. Using (i) we create the variables FWI on the initial date ($FWI\_InitDat$), the mean FWI of the month prior to the initial date (FWI_MP), the mean FWI of the week prior to the initial date (FWI_WP), and the mean FWI of the month prior until the week after the initial date (FWI_MP_WA). We employ (ii) to estimate the annual mean as well as the 0.5, 0.9, 0.95, 0.99 quantiles (FWI_Mean, $FWI\_q0.5$, $FWI\_q0.9$, $FWI\_q0.95$, $FWI\_q0.99$) of the FWI for the corresponding year of the fire incidence.

1.2.3 Population Density

Population density has gained widespread attention for its role as an ignition source, as a facilitator of suppression efforts, and as a factor that captures impact-related importance (Fernandes, 2019; González-Cabán, 2009; Lankoande and Yoder, 2006; Pechony and Shindell, 2010). To proxy population density near wildfires we use the Oak Ridge National Laboratory's *LandScan*[4] annual global population distribution data provided at approximately 1 kilometre (30") spatial resolution. The raster data representing the ambient population distribution is based on remote sensing imagery analysis techniques, demographic, and geographic

[4] https://landscan.ornl.gov.

data. We create approximately 4 kilometre buffers[5] around the centroid of the polygons and calculate the mean population density in counts per square kilometre of the respective *Land-Scan* year denoted by the variable *Pop_4km*.

1.2.4 Land Cover Type

The 2006, 2012, and 2018 versions of the Copernicus' CORINE land cover maps[6] are employed to categorize the EFFIS perimeters of BA to evaluate a potential correlation between land cover type and the distribution of large fires. The CORINE land cover information is derived from satellite data[7] using a minimum mapping unit of 25 ha and consists of an inventory of 44 land cover classes. We extract the dominant land cover type for each EFFIS BA polygon considering the latest version of the CORINE land cover data with respect to the initial date of the observation. We further reclassify the most prevalent land cover types for each country with regard to the extreme wildfire observations. For an overview of the dominant land cover types for each country see Table A2 of Appendix A.1. In the conducted analysis, types I to III are incorporated as indicator variables. A fourth indicator variable named *Type_Other* is created where none of the three types applies.

[5] The exact measure is 0.05 decimal degrees which at 45° N corresponds to 3'935.5 meters.

[6] https://land.copernicus.eu/pan-european/corine-land-cover.

[7] 2006: SPOT-4/5 and IRS P6 LISS III; 2012: IRS P6 LISS III and RapidEye; 2018: Sentinel-2 and Lansat-8.

1.3 Methods

1.3.1 Point Process using Maximum Likelihood Estimation

The foundation of the PP approach regarding extremal processes was originally introduced by Pickands (1971), and applied to environmental processes by Smith (1989). The PP approach is particularly suitable as it uses data efficiently and can easily be adapted to include temporal or covariate effects (Coles, 2001). We apply a non-homogeneous PP model to simulate the occurrence (i.e., frequency of exceedance) and intensity (i.e., excess) of a value of BA above a chosen threshold.

Let X_i be a series of independent and identically distributed (i.i.d.) random variables representing wildfire burned areas, and $N_n = \left\{\left(\frac{i}{n+1}, X_i\right) : i = 1, n\right\}$ be a sequence of point processes. Then, given a sufficiently large threshold u, on regions of the form $[0, 1] \times (u, \infty)$, the point process N_n is approximately a Poisson process with the intensity measure $\Lambda(A)$ shown in Equation (1.1) on a set of the form $A = [t_1, t_2] \times (u, \infty)$:

$$\Lambda(A) = n_y \left(t_2 - t_1\right) \left(1 + \xi \left(\frac{u - \mu}{\sigma}\right)\right)^{-1/\xi} \tag{1.1}$$

The interval (t_1, t_2) on the abscissa is a subset of $[0, 1]$ and n_y denotes the number of years of observations so that events in non-overlapping subsets of $[0, 1] \times (u, \infty)$ are independent and the estimated parameters ξ, μ, and σ correspond to the Generalised Extreme Value (GEV) distribution. It envelops three types of limit distributions, which are uniquely defined by the shape parameter ξ. The Fréchet distribution ($\xi > 0$) is characterised by a heavy tail, the

Gumbel distribution ($\xi = 0$) exposes an exponential decay of the tail, whereas a Weibull limit distribution ($\xi < 0$) has an upper bound. In general, a heavier tail implies that the probability of an "unexpected" event is larger, while the location μ and the scale σ parameters relate to the mean and spread of the distribution, respectively. For greater detail on the GEV see Appendix A.2.

Following Coles (2001), the model parameters are estimated by maximising the likelihood function

$$L(\mu, \sigma, \xi) = \exp\left\{-n_y \left[1 + \xi \frac{u-\mu}{\sigma}\right]^{-1/\xi}\right\} \times \prod_{i=1}^{N(A)} \sigma^{-1} \left\{\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}\right\}^{\delta_i[x_i > u]} \tag{1.2}$$

where δ_i is one if the realisation of $X_i > u$, and zero otherwise. The first part of the likelihood expression entails the contribution of the number of fire events (occurrence) characterised by the Poisson distribution with mean $\Lambda\{[0,1] \times [u, \infty)\}$. The second part shows the excess contribution of the observations (intensity) which are modelled as Generalised Pareto Distribution (GPD). σ is adjusted as $\sigma^* = \sigma(u) - \xi u$, so that the scale parameter is independent of the threshold. The cumulative GPD is given by Equation (1.3)

$$F(z; \sigma^*, \xi, u) = \begin{cases} 1 - \left[1 + \xi\left(\frac{z-u}{\sigma^*}\right)\right]^{\frac{-1}{\xi}}, & \text{for } \xi \neq 0 \\ 1 - e^{\frac{-z}{\sigma^*}}, & \text{for } \xi = 0, \end{cases} \tag{1.3}$$

where $1 + \xi(\frac{z-u}{\sigma^*}) > 0$, $z - u > 0$, and $\sigma^* > 0$.

Given that X has a GPD, the distribution of the re-scaled random variable z/σ^* is inde-

pendent of σ^* (Katz et al., 2005).

We perform the numerical optimisation using the R package *extRemes* (Gilleland and Katz, 2016) to estimate Equation (1.2).

1.3.2 Model Assumptions

The theoretical justification for using a PP characterisation of extremes is predicated on the assumptions of (i) unbiased threshold choice, (ii) stationarity, and (iii) independence of the excesses. Regarding (i), too low a threshold leads to a bias potentially violating the asymptotic basis of the model. If the threshold chosen is too high, the reduction of data points leads to high variance. We determine the individual countries' thresholds using the threshold diagnostic tools provided in the R package *extRemes*. They are based on the following rationale. Let the excesses over a threshold u be defined as $y = x - u$. Recalling from Section 1.3.1 that these excesses follow a GPD, this also holds true for all $y > 0$ of a threshold $v > u$ with

$$GPD\left(y, \sigma_v, \xi_v\right) = \frac{GPD\left((v-u)+y, \sigma_u, \xi_u\right)}{GPD\left((v-u), \sigma_u, \xi_u\right)} \tag{1.4}$$

As a consequence, Equation (1.4) can only be satisfied if $\xi_v = \xi_u$ and $\sigma_v - \xi \cdot v = \sigma_u - \xi \cdot u$. This implies that for a sufficiently high threshold, both the shape parameter ξ and the modified scale $\sigma - \xi u$ are independent of the threshold and need to be stable. Besides plotting the shape and modified scale parameters individually, the mean value of the

excesses y over u can be plotted against u which is known as the Mean Residual Life (MRL) plot (Coles, 2001). The GPD is deemed to fit the data well when a straight line starting from the selected threshold can be fitted within the confidence bands of the MRL plot, and thereby indicating a stable distribution. In practice, the visual interpretation of the MRL plot, as well as the individual parameter plots, is somewhat subjective. Thus, we additionally consider the threshold selection suggestions provided by the automated Bayesian leave-one-out cross-validation approach, which compares the extreme value predictive performance resulting from each of a set of thresholds. This approach was first introduced by Northrop et al. (2017) and is implemented in the R package *threshr* (Northrop and Attalides, 2020). As this approach is only applicable to independent observations, we compare outcomes in an iterative process with the estimation of the extremal index θ as a measure of dependence.

While Equation (1.2) implicitly assumes stationarity of the GEV parameters, we also estimate non-stationary models where ξ, μ, and σ are conditioned on various functional forms of the covariates described in Section 1.2 (as well as on seasonality variables) in order to assess assumption (ii). Equation (1.5) serves as an example of modelling a non-constant linear location parameter dependent on the mean FWI of the month prior to the initial date FWI_MP:

$$\mu(FWI\_MP) = \mu_O + \mu_1 * FWI\_MP \tag{1.5}$$

The evaluation of the non-stationary models is based on the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) and, for nested models, on the likelihood

ratio test.[8] We systematically model the location, shape, and scale parameters individually and combined starting with linear functional forms of the respective parameters. Whenever a model shows an improvement over the stationary model we explore more complex functional forms (i.e., quadratic and interactions). In cases where the parameter confidence intervals (CIs) could not be estimated via the delta method, 500 iterative bootstraps with replacement were applied to evaluate the parameter significance.

As a means to examine the independence assumption (iii), the degree of dependence is explored using the extremal index $\theta \in (0, 1]$ suggested by Ferro and Segers (2003) which is defined as:

$$\theta = \begin{cases} \min\left\{1, \frac{2\left(\sum^{N-1} i{=}1 T_i\right)^2}{(N-1)\sum_{i=1}^{N-1} T_i^2}\right\}, & \text{if } \max\{T_i : 1 \leq i \leq N-1\} \leq 2 \\ \min\left\{1, \frac{2\left(\sum_{i=1}^{N-1}(T_i-1)\right)^2}{(N-1)\sum_{i=1}^{N-1}(T_i-1)(T_i-2)}\right\}, & \text{if } \max\{T_i : 1 \leq i \leq N-1\} > 2, \end{cases} \tag{1.6}$$

where T_i denotes the length between excesses (interexceedance time). A value of the extremal index $\theta = 1$ implies complete independence, whereas $\theta \to 0$ indicates perfect dependence. In case the extremal index suggests a violation of the independence assumption the data can be declustered to filter the dependent observations.

[8] Suppose that the negative likelihood is x for the stationary base model and y for the restricted model, the deviance statistic $D = -2(y - x)$ then follows the χ^2_k distribution where k indicates the difference in the number of estimated parameters (Coles, 2001).

1.3.3 Model Fit

In order to assess the fit of the selected model, we implement two common diagnostic plots incorporated in the *extRemes* package. First, we avail of the Z-plot following Smith and Shively (1995). Let Zk be the Poisson intensity parameter integrated from exceedance time $k-1$ to exceedance time k (starting the series with $k=1$). The Z-plot then determines whether this random variable Zk is independent exponentially distributed with mean one, which corresponds to the observations lying on the diagonal. Second, we plot the kernel density functions of the observed data vs. the modelled distribution. For the particular case of the PP characterisation of extremes, the density of the calculated data block maxima is compared to the PP model with respect to the equivalent GEV.

1.3.4 Return Levels (Quantiles)

Harnessing the estimated probabilities associated with extremes, the interest is typically focused on providing estimates of the upper quantiles of the modelled distribution functions. Specifically, the return level of an extremely large fire, defined as z_p, which is associated with a return period of $1/p$ embodies a tangible outcome. It is equivalent to the $(1-p)th$ quantile of the corresponding modelled distribution by the PP representation of extremes. As the PP approach combines the Poisson distribution parameter with the GPD, the return level z_p is obtained by setting the cumulative distribution function Equation (1.3) equal to the desired

quantile $1-p$. Solving for z (for a probability p) leads to Equation (1.7) (Coles, 2001):

$$z_p = F^{-1}\left(1-p;\sigma^*,\xi,u\right) = \begin{cases} u + \left(\sigma^*/\xi\right)\left(p^{-\xi}-1\right), & \text{for } \xi \neq 0 \\ u + \sigma^* \ln(1/p), & \text{for } \xi = 0 \end{cases} \tag{1.7}$$

where the return level z_p denotes the BA level that is expected to be exceeded in any given year with probability p.

1.3.5 Economic Valuation

A transformation of the informational content of the BA return-level estimates into economic values could arguably be beneficial in supporting policy decisions. Our approach in this regard is to multiply associated per ha monetary losses with the expected BA, as suggested by Holmes et al. (2008). To this end we resort to existing studies either providing explicit per ha loss estimates or calculating per ha values by combining information on total BA with total loss estimates. To facilitate comparison over space and time, spatial values are harmonised in hectares, and monetary values are inflation-adjusted and expressed in 2020 Euros using the 2020 monthly average exchange rate (US$ = 0.87€). Values in US Dollars are deflated based on the not seasonally adjusted urban Consumer Price Index CPI.[9] Table 1.1 provides a summary of the economic impact aspects that have been included in each study as well as of the inflation-adjusted € per ha monetary values for the five papers included in the table. Note that the calculated €/ha losses are highly dependent on the estimation method

[9] https://www.bls.gov/cpi.

used, the type of damage and losses included, and the specific situation of the fire (season) that is studied in the paper. Therefore, the figures derived from the multiplication of our return levels with estimates from existing publications need to be interpreted with caution.

More specifically, two studies have a European context. The first estimates we derive are from a comprehensive report for Mediterranean forests by Merlo and Croitoru (2005) who provide figures in 2001 prices that encompass country-specific estimates of 884€/ha (GR), 1'480€/ha (IT), and 3'420€/ha (PT). This translates into inflation-adjusted monetary values expressed in 2020 Euros of 1'228€/ha (GR), 2'004€/ha (IT), and 4'728€/ha (PT). We also include the economic impact estimates from a study of Galicia, Spain by Barrio et al. (2007), which implements an ecosystem service approach based on assessing services that are affected due to wildfire existence. The reported monetary losses range from 2'249-3'162 €/ha in 2006 values. We apply the mean of this range, i.e., 3'304€/ha in 2020 Euros.

Table 1.1: Included economic impacts and €/ha economic loss estimates in 2020 Euros.

Barrio et al. (2007)	Butry et al. (2001)	Merlo and Croitoru (2005)	Rahn et al. (2014)	Safford et al. (2022)
		Country (region)		
Spain (Galicia)	US (Florida)	Portugal, Italy, Greece	US (San Diego)	US (California)
		Economic impacts considered		
ecosystem service approach • goods and services[h] • negative externalities[i] • property damage • suppression costs	7 categories • insured property damage • pine timber market long-run damage • presuppression costs[a] • tourism-related costs • suppression costs • disaster relief expenditures • human health effects[b]	direct losses due to forest fires • wood losses • restoration costs	5 categories • state/agency[c] • infrastructure[d] • natural areas[e] • business[f] • community[g]	ecological & socio-economic evaluation • structures lost • property damage • lives lost • suppression cost • ecosystem impact[j]
		€/ha loss estimates		
3'304€	3'801€	PT: 4'728€ ES: - IT: 2'004€ GR: 1'228€	3'230€	8'717€

Notes: [a] including prescribed burning; [b] short-run asthma-related healthcare costs; [c] i.e., insurance implications, short and long-term budget impacts, future resource impacts, [d] i.e., transportation repairs, water quality impacts, utility replacement, communication repairs; [e] i.e., species and habitats, erosion and flood control, watershed restoration, cultural and historic resources; [f] i.e., economic activity, employment impacts, building and property loss, tourism impacts; [g] i.e., recreation impacts, health and human services, building and property loss, public assistance; [h] i.e., tourism, timber losses, CO_2 losses derived from biomass burned; [i] i.e., losses derived from CO_2 emissions; [j] defined in terms of tree mortality.

As suitable research on Southern Europe is limited, we also include three loss estimates derived from studies of US wildfires. The Butry et al. (2001) case study assesses the Florida 1998 summer wildfires that burned a total of around 500'000 acres (202'343 ha). We apply the conservative lower bound total cost estimate of 600 million US$ (in 2001 values). Dividing the total cost by the total BA leads to an inflation-adjusted estimate of 3'801€/ha. A second study by Rahn et al. (2014) evaluates the 2003 wildfires in San Diego (US) and reports a cost of 6'500 US$ per acre (2'630 US$/ha in 2014 values), which is equivalent to 3'230€/ha in 2020 Euros. Finally, we include the recent publication on California wildfires by Safford et al. (2022) who investigate the extraordinary 2020 fire season. The authors estimate losses of 19 billion US$ for a historical record of 1.74 million ha BA which is equivalent to 8'717€/ha in 2020 Euros.[10]

1.4 Results

1.4.1 Summary Statistics

Country-level BA summary statistics of the EFFIS BA data product are presented in Table 1.2. The single largest fire in the data set that burnt 67'521 ha occurred in October 2017 in Portugal. Greece exhibits a comparably lower frequency of fires but has the largest mean,

[10] Extensive research conducted by Wang et al. (2021) estimate the economic losses for the 2018 wildfire season in California that include indirect losses and suggests total wildfire damages were in the region of 148.5 billion US$ for a total BA of 7'700 km^2. This leads to a per ha loss estimate of 165'467 €/ha in 2020 Euros, which is around 19 times larger than the Safford et al. (2022) estimate for the 2020 season. Given this estimate is far beyond all the other estimates, we do not use it in this analysis and only present the more conservative estimates. However, Wang et al. (2021) gives some indication of how far-reaching the costs of extreme wildfires are when we include indirect health costs as well as costs arising outside the affected region assuming extreme fires in Mediterranean Europe are comparable to those in California.

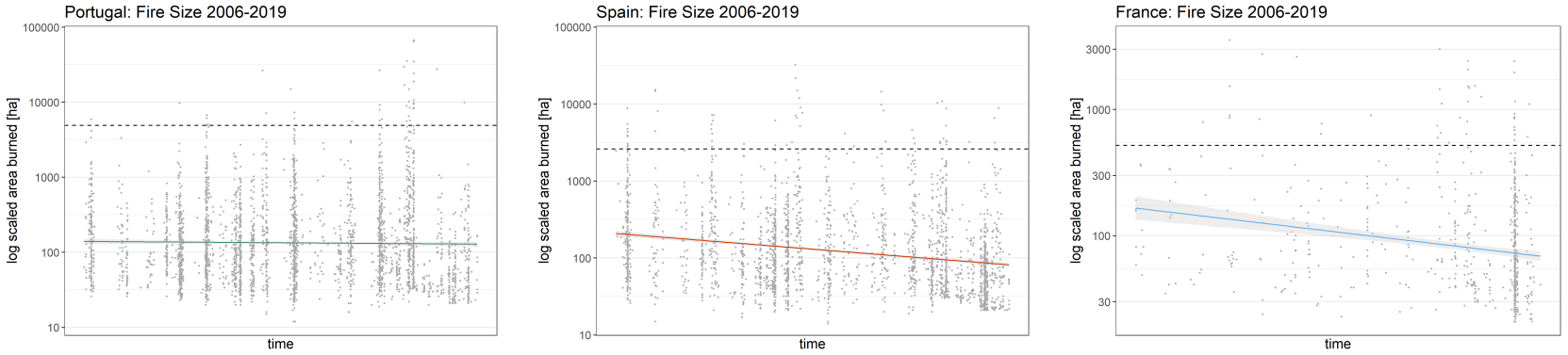
median, 75-percentile, and 90-percentile BA values. The highest annual wildfire incidence is recorded in Italy.

Table 1.2: BA summary statistics (2006 - 2019).

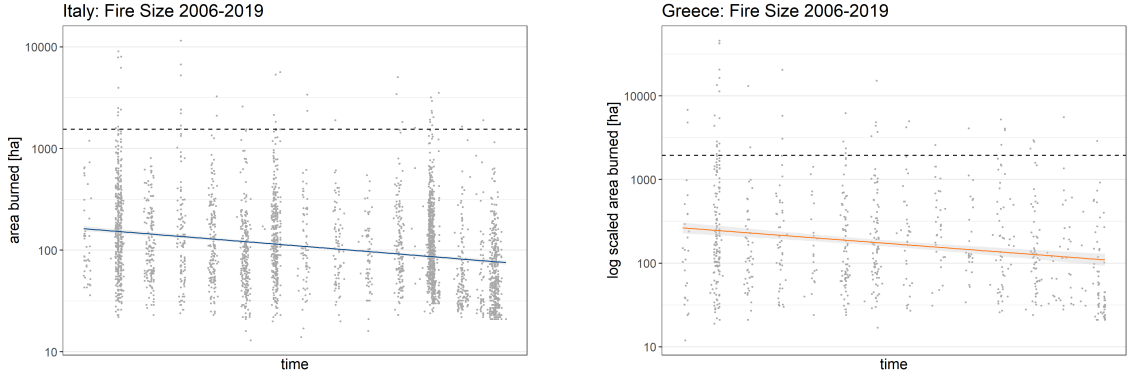
Country	n	events per year (n_a)	Mean [ha]	Median [ha]	Pctl 75 [ha]	Pctl 90 [ha]	Max [ha]
Portugal	3'084	220.3	474	106	259	768	67'521
Spain	2'412	173.3	386	95	240	678	32'424
France	668	47.7	171	65	150	340	3'555
Italy	3'260	232.9	204	91	188	372	11'550
Greece	748	53.4	761	138	412	1'325	45'809

Notes: n: number of observations; Pctl: percentile.

Figure 1.1: BA (log_{10} scaled) with indicated threshold choice (black dashed line) over the study period (2006-2019) with a generalised linear model smoothed conditional mean with CIs on the 90% level.



(a) Portugal **(b)** Spain **(c)** France



(d) Italy **(e)** Greece

Figure 1.1 presents the log-transformed wildfire BA observations from 2006 to 2019 at the country level. The data shows a slightly decreasing tendency in BA for Spain, France, Italy, and Greece, and no clear trend in Portugal. However, a slightly different picture emerges from Figure A1 of Appendix A.1 when we focus on BA extremes, herein defined as the wildfires that exceed the selected country-level threshold. While, once again, we observe decreasing BA trends for France, Italy, and Greece, no trend is evident for Spain. In contrast, the extreme BA values for Portugal exhibit an increasing trend, largely driven by the 2017 fire season.

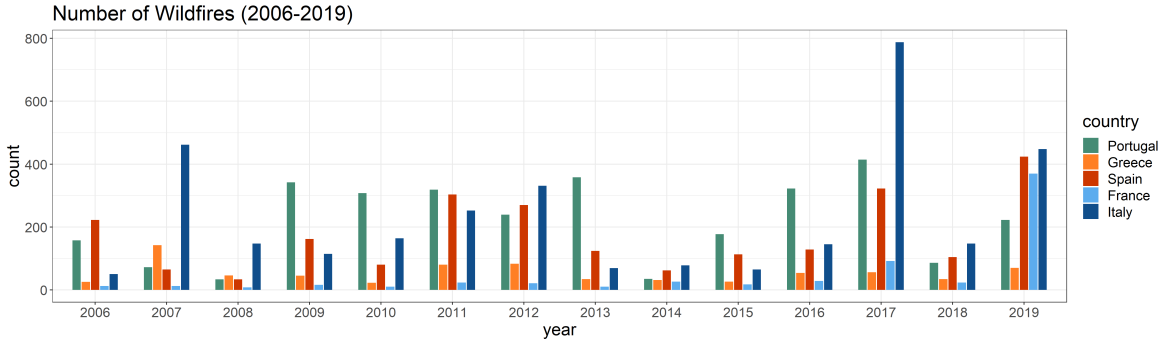
Figure 1.2 displays the annual number of wildfires and total BA over the study period enabling a direct country comparison. There are few observations and little variation over the years for Greece, while the opposite is the case for Italy. The year 2017 particularly stands out for Portugal with many fire records as well as a large total BA. For France 2019 accounts for more than half of all the observations in the study period.

Regarding the correlation between BA and the covariates, there is a general tendency of a positive correlation between the BA and the mean FWI of the week prior to the fire initial date (FWI_WP). No conclusive relationship is observable between the BA and population density. Correlation plots are provided in the supplementary material in Appendix A.1 (Figure A2 and Figure A3 for the association of the FWI and the population density, respectively).

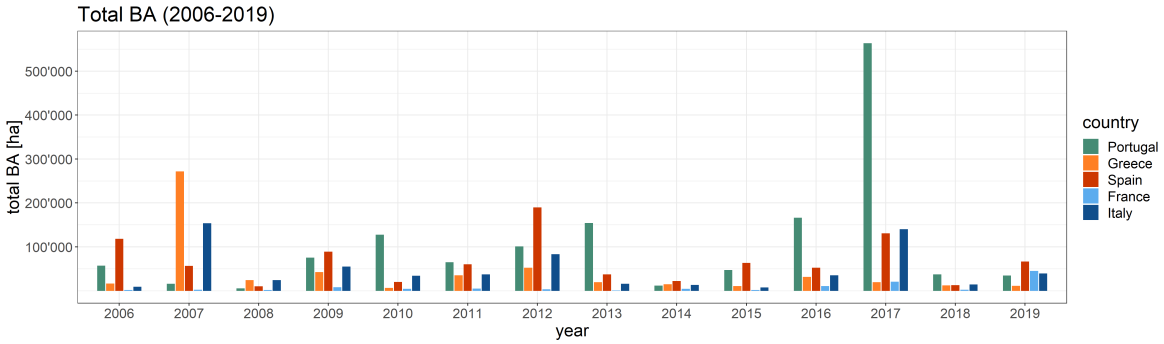
1.4.2 Threshold Selection/Dependence Test

The MRL plots with the final threshold choice (after considering all decision-supporting tools outlined in this section) are shown in Figure 1.3. Complementing the MRL plots, the

Figure 1.2: Annual number of wildfires and annual total BA in the EFFIS BA product.



(a) Number of observations



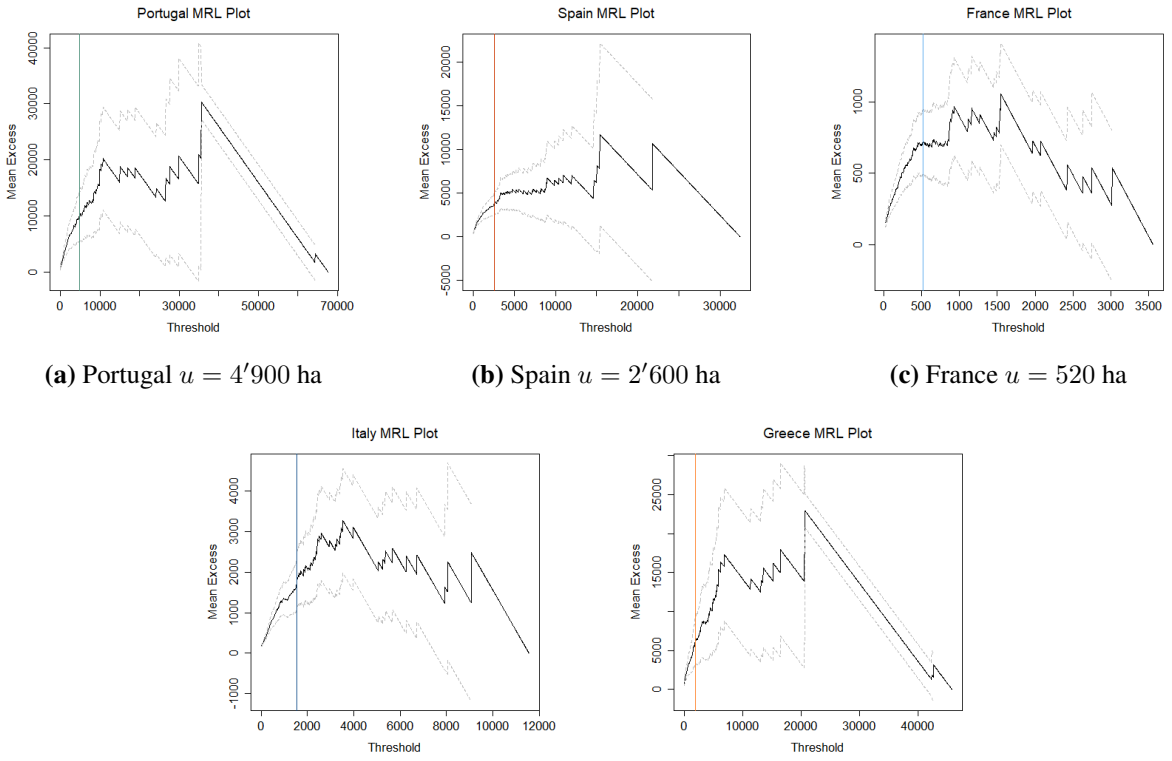
(b) Total BA

individual behaviour of the shape parameter ξ and the modified scale parameter $\sigma - \xi u$ are analysed but not shown.

The Bayesian leave-one-out cross-validation plots are presented in Figure 1.4. These show a single-run output and vary across different executions. The best threshold evaluated by this approach, denoted as u_b, is provided below the plot whenever it proved stable over 10 consecutive runs.

Table 1.3 provides a summary of the final threshold choices with the corresponding extremal indices θ, and the number of observations above the selected threshold. The excesses of Spain and Greece indicate perfect independence, while Portugal and Italy show a very high

Figure 1.3: Mean Residual Life (MRL) plots with indicated final threshold choice.



(a) Portugal $u = 4'900$ ha **(b)** Spain $u = 2'600$ ha **(c)** France $u = 520$ ha

(d) Italy $u = 1'550$ ha **(e)** Greece $u = 1'943$ ha

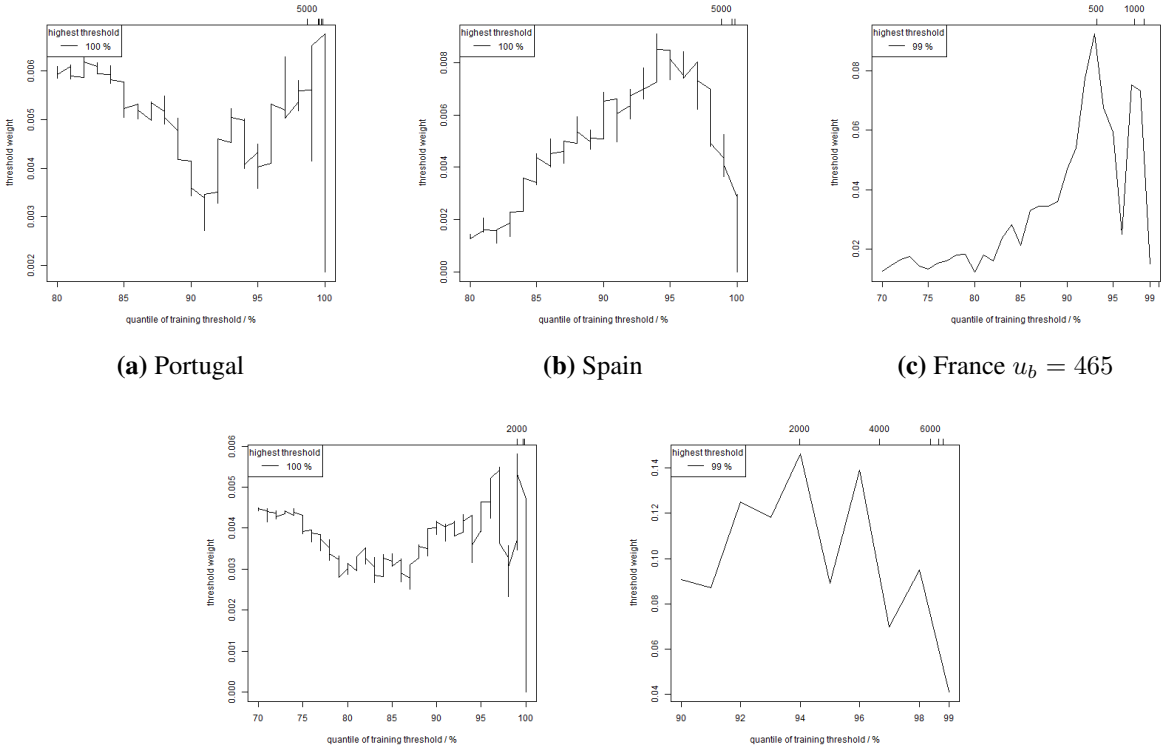
θ value. The lowest extremal index value is found for France, which did not improve after declustering.[11] The number of observations above the respective country-specific thresholds ranges from 42 to 62 and corresponds to 1.4-6.3% of the total country-level data.

1.4.3 Non-Stationarity

Since the stationary models are embedded in potential non-stationary models the results of the latter are reported first. Table 1.4 lists all the models with an improvement of the BIC > 10

[11] The specific case of modelling the extremes with the data available for France is addressed in Section 1.4.3 and Section 1.4.4.

Figure 1.4: Bayesian leave-one-out cross-validation threshold selection approach.



(a) Portugal **(b)** Spain **(c)** France $u_b = 465$

(d) Italy **(e)** Greece $u_b = 1'943$

Table 1.3: Summary statistics thresholds and extremal indices.

country	threshold u	extremal index θ	$n > u$ (% of total)
Portugal	4'900 ha	0.9	42 (1.4)
Spain	2'600 ha	1	62 (2.6)
France	520 ha	0.84	42 (6.3)
Italy	1'550 ha	0.94	45 (1.4)
Greece	1'943 ha	1	45 (6)

over the stationary model following Neath and Cavanaugh (2012) suggesting this threshold as "very strong" evidence to favour the model with the lower BIC over the competing model.

For Portugal, letting the location parameter μ depend on the mean FWI for the month prior to the initial date of the fire (FWI_MP) leads to the best model fit. The evaluation of conditional effects for the historical excesses in Spain and Italy shows that modelling the location and the shape parameter dependent on the FWI on the reported initial date ($FWI\_InitDat$) improves the model fit the most. None of the non-stationary models leads to any improvements in the model fit for Greece. For France, land cover type is found to be most influential in modelling the observed data. More specifically, modelling the location parameter conditional on land cover $Type\_I$ (Sclerophyllous vegetation) in a linear functional form not only proves to capture the empirical data best but also leads to a significant positive shift of the distributional mean. However, even though modelling non-stationarity leads to an increased model fit in specific cases, with the exception of France, results do not indicate a significant modification of the modelled GEV parameters. Consequently, based on the covariates considered, the assumption of stationarity holds true in the data sample for all countries except for France. Therefore, the reported probabilities of the stationary model are valid and comparable for Portugal, Spain, Italy, and Greece.

1.4.4 Model Selection/Model Fit

In light of no significant parameter changes modelling the extremes conditional on the implemented covariates for Portugal, Spain, Italy, and Greece, we base the subsequent model evaluations and estimations on the stationary model. Not only does the distribution of the historical extremes for France show dependence and therefore violate the stationarity assumption, but the extremal index θ in Table 1.2 also indicates higher dependence of the excesses

Table 1.4: Non-stationary models with a BIC decrease > 10 sorted by decreasing AIC.

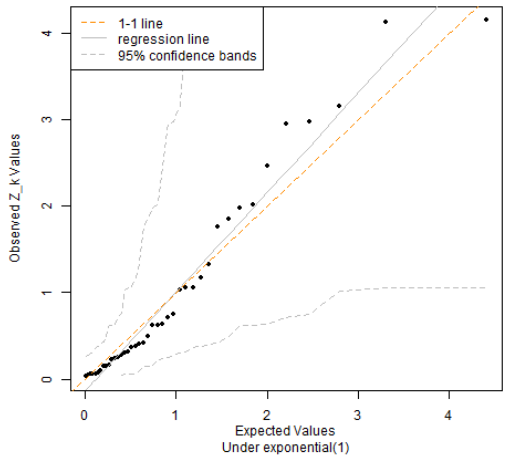
country	modeled parameter(s)	modeled covariate(s)	functional form	AIC/BIC improvement	parameter significance
Portugal	μ	FWI_MP	linear	-21.4/-19.7	μ_0 insig., $\mu_1 > 0$ *
	μ	FWI_MP	quadratic	-16.2/-12.7	$\mu_0 > 0$, μ_1 insig., μ_2 insig.
Spain	μ, σ, ξ	FWI_InitDat	μ, ξ quadratic, σ linear	-40.0/-29.7	$\mu_0 > 0$, $\mu_1 < 0$, $\mu_2 > 0$, $\sigma_0 > 0$, σ_1 insig., ξ_0 insig., $\xi_1 > 0$, $\xi_2 < 0$ *
	μ, ξ	FWI_InitDat	quadratic	-38.5/-30.3	$\mu_0 > 0$, $\mu_1 < 0$, $\mu_2 > 0$, ξ_0 insig., $\xi_1 > 0$, $\xi_2 < 0$ *
	μ	FWI_InitDat	quadratic	-28.6/-24.5	μ_0 insig., μ_1 insig., μ_2 insig.
	ξ	FWI_InitDat	quadratic	-24.6/-20.5	ξ_0 insig., $\xi_1 > 0$, $\xi_2 < 0$
France	μ	Type I	linear	-12.6/-11.3	$\mu_0 > 0$, $\mu_1 > 0$
Italy	μ, ξ	μ FWI_InitDat, ξ FWI_InitDat/DJF	μ quadratic, ξ interaction	-41.2/-32.3	$\mu_0 > 0$, μ_1 insig., μ_2 insig., ξ_0 insig., ξ_1 insig., $\xi_2 < 0$, ξ_3 insig. *
	μ, ξ	FWI_InitDat	quadratic	-36.4/-29.3	$\mu_0 > 0$, μ_1 insig., μ_2 insig., ξ_0 insig., ξ_1 insig., ξ_2 insig. *
	μ	FWI_InitDat	quadratic	-26.5/-22.9	μ_0 insig., μ_1 insig., μ_2 insig.
Greece	no model improvements				

Notes: (i) Whenever the functional form is indicated as "quadratic" the linear term is included as well. (ii) In cases where the parameter 95% CIs could not be estimated via the delta method, 500 iterative bootstraps with replacement were applied to evaluate the parameter significance (indicated with *).

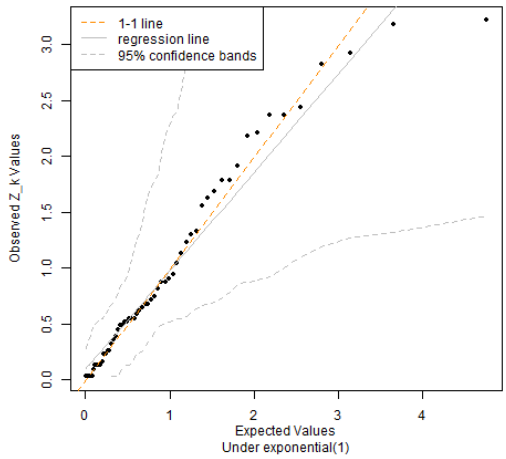
than is the case for the other countries. Thus, we exclude France from subsequent analysis.

Evaluating the model fit using Z-plots depicted in Figure 1.5 it is evident that all observations lie well within the 95% confidence bands and that there is arguably a good model fit for Portugal, Spain, and Italy, and a moderately good fit for Greece. A similar conclusion can be drawn from Figure 1.6 plotting the kernel density functions of the empirical against the modeled data. Once again, the observed data are very well modelled for Italy, and fairly well for Portugal and Spain. For Greece the modelled data, in contrast, captures the empirical data relatively less well.

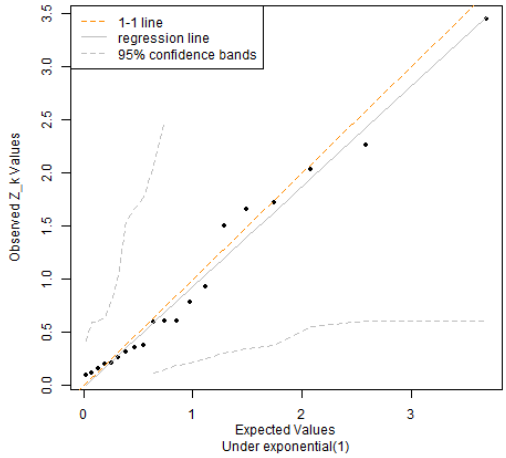
Figure 1.5: Model fit diagnostics: Z-plots.



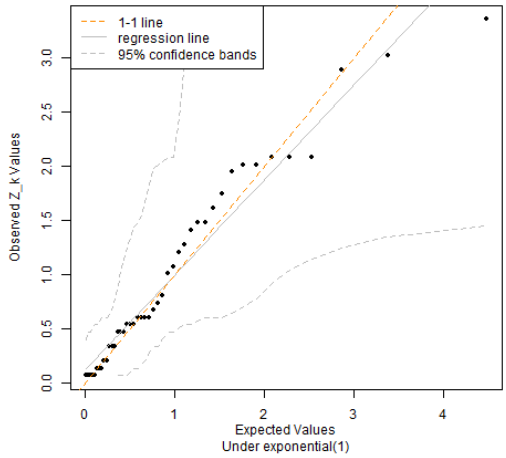
(a) Portugal



(b) Spain



(c) Italy

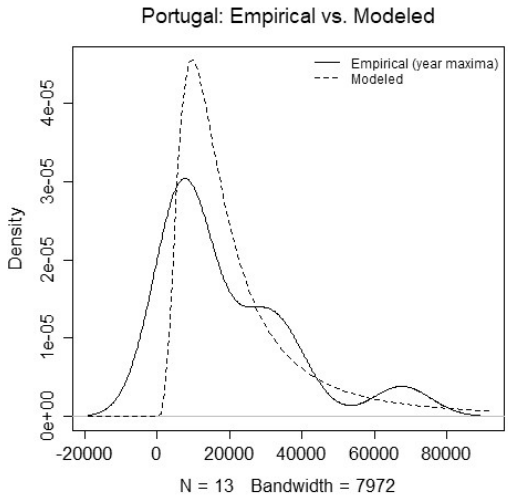


(d) Greece

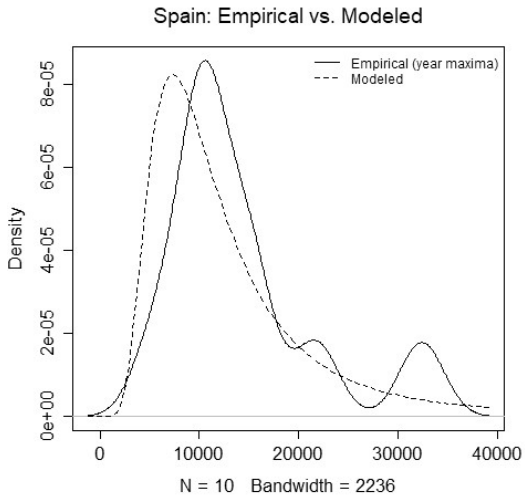
1.4.5 Parameter Estimates

The three GEV distribution parameter estimates by country are shown in Table 1.5. The largest point estimate $\hat{\mu}$ is found for Portugal, followed by Spain, Greece and Italy, respectively. Although the centre of the distribution is larger for Spain than for Greece, $\hat{\sigma}$ indicates

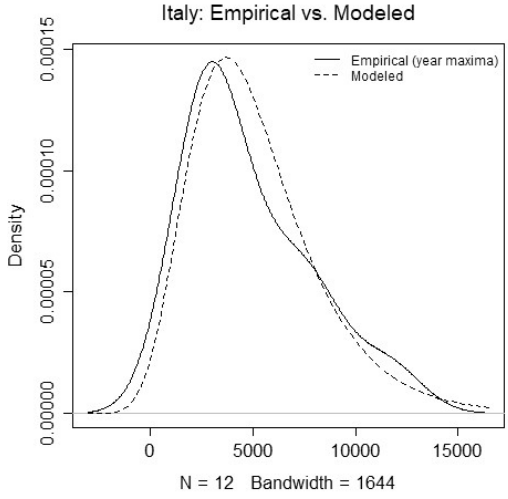
Figure 1.6: Model fit diagnostics: Density plots.



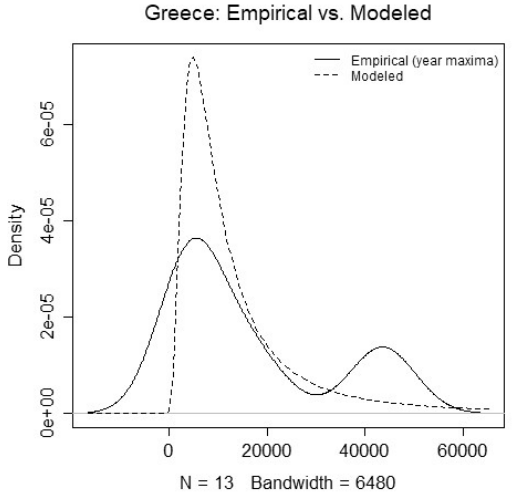
(a) Portugal



(b) Spain



(c) Italy



(d) Greece

that the spread of the distribution is wider for Greece than for Spain. In general, we observe extremely small CIs for Portugal for the location and scale parameters.

The largest shape parameter value $\hat{\xi}$, and thus the heaviest tail, is estimated for Greece followed by Portugal and is larger than 0.5 indicating that although the mean is finite, the

Table 1.5: Country-level maximum likelihood GEV parameter estimates with confidence intervals (CIs) on the 95% level.

country	location $\hat{\mu}$ [ha]	scale $\hat{\sigma}$ [ha]	shape $\hat{\xi}$	limit distribution
Portugal	13'017 (13'017,13'017)	9'061 (9'061, 9'061)	0.52 (0.28, 0.64)	Fréchet
Spain	8'673 (8'518, 8'690)	4'719 (4'700, 4'934)	0.36 (0.28, 0.46)	Fréchet
Italy	3'483 (3'397, 3'649)	1'874 (1'539, 2'031)	0.37 (-0.06, 0.55)	Gumbel
Greece	7'206 (7'084, 7'106)	5'743 (5'742, 5'767)	0.58 (0.38, 0.69)	Fréchet

Note: The CIs are estimated employing a parametric bootstrap simulating data from the fitted model.

variance is infinite (Katz et al., 2005).[12] The point estimates of the shape parameter for Spain and Italy are fairly similar. However, $\hat{\xi}$ is insignificant for Italy. On that account, the main difference in the distributions of the extremes comparing the individual countries is that Portugal, Greece, and Spain have a significantly positive shape parameter ξ indicating a Fréchet type limit distribution, while the excesses for Italy follow the Gumbel type limit distribution.

1.4.6 Return Levels and Probabilities of Exceedance

Table 1.6 displays the numerical estimates of the T-year (here with T = 5, 10, 20, 50) BA return levels, where the BA values given in ha are exceeded in one year with probability 1/T. The return levels are found to be highest in Portugal in any given return period followed by Greece, Spain, and Italy. For example, the probability that a single wildfire burns more than 50'338 ha in any given year is 10% in Portugal, while for Spain the probability of a fire

[12] A statistical moment is infinite if it converges too slowly to be integrated, and thus does not exist (Holmes et al., 2008).

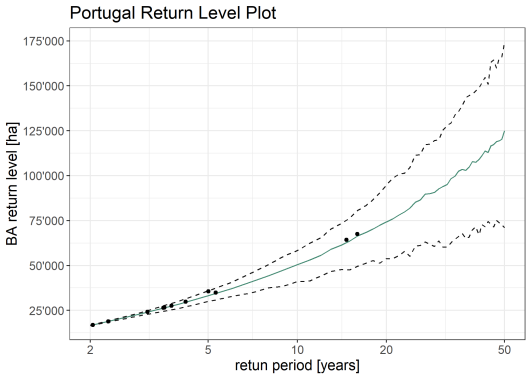
exceeding approximately this size (49’452 ha) is about 2%.

Table 1.6: Individual country return levels in ha for specific return periods.

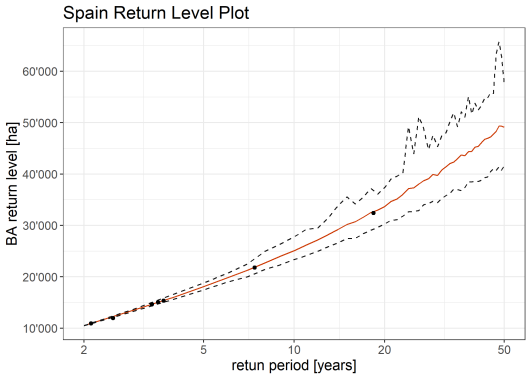
country	5-year (CI)	10-year (CI)	20-year (CI)	50-year (CI)
Portugal	33’279 (30’062, 35’832)	50’338 (39’924, 58’557)	75’256 (53’038, 94’587)	123’719 (73’838, 172’805)
Spain	18’080 (17’376, 18’822)	25’165 (23’391, 26’905)	34’017 (30’277, 39’079)	49’452 (41’636, 61’197)
Italy	7’325 (6’149, 9’025)	8’966 (6’531, 12’842)	10’890 (6’944, 12’842)	14’627 (7’053, 26’704)
Greece	20’687 (18’370, 22’372)	33’242 (28’298, 37’876)	51’764 (39’636, 64’261)	91’037 (58’694, 124’476)

Note: The return levels and CIs on the 95% level are estimated employing a parametric bootstrap simulating data from the fitted model.

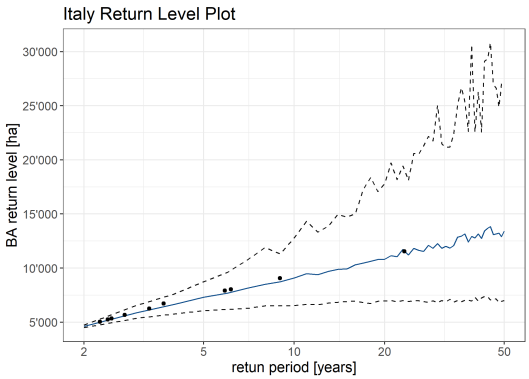
Figure 1.7: Return level plots with bootstrapped CIs on the 95% level.



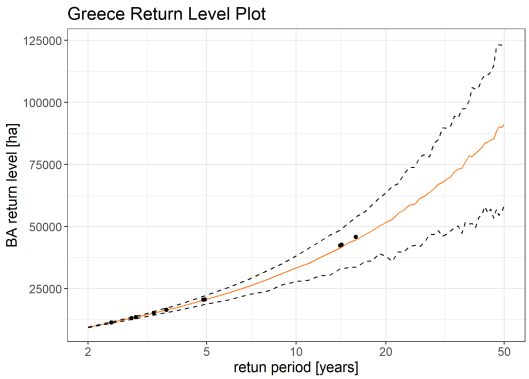
(a) Portugal



(b) Spain



(c) Italy



(d) Greece

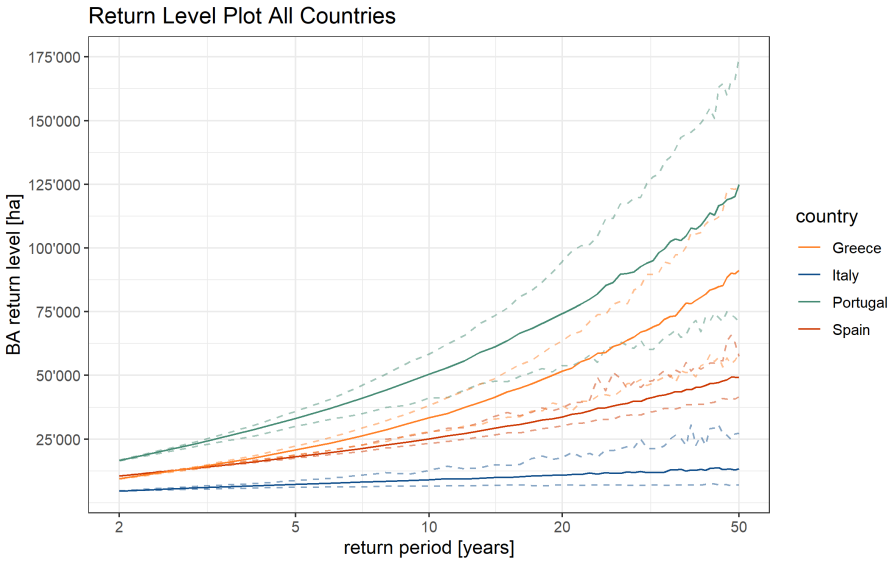
The individual country return level plots in Figure 1.7 show the distribution of the observations within the tail. Essentially, the limit distributions found for all the Mediterranean countries have no upper bound (i.e., the extremes are not converging to a specific value). Furthermore, the return level plots enable a better understanding of the different limit type distributions in a graphical fashion. In particular, the distinction between the Gumbel-type distribution found for the extremes in Italy versus the Fréchet type distributions for the other countries is distinctly visible. As the x-axis is log-transformed, the return level plot reflects Gumbel-type distributions characterised by an exponential decay of the tail as a straight line, while the Fréchet type distributions manifest as convex shapes.

We find that all the observed events lie within the bootstrapped CIs. Furthermore, the smallest confidence bands at the 95% level are observed for Spain indicating high certainty of the point estimates. In contrast, the largest CIs are apparent for Italy suggesting a wide range of potential outcomes within the 95% CI.

Looking at the largest wildfire for each country (Max value in Table 1.2), an event of such size or larger is expected to occur, on average, once every 16 years for Portugal with an annual occurrence probability of 1.9%. The calculated yearly probability for the largest observed fire in Spain is 1% and has a return period of about 18 years. In Italy, the maximum BA value is expected to be exceeded once in every 23 years with an annual probability of 2.6%, and the largest BA value for Greece is estimated as an approximately 16-year event with a yearly probability of occurrence of 1.8%.

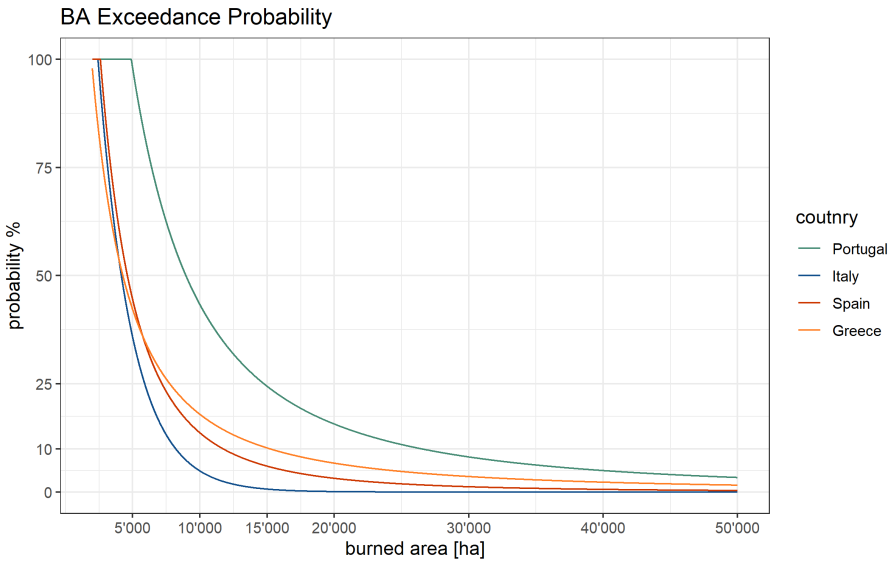
Figure 1.8 overlays the individual country return level plots to facilitate a cross-country

Figure 1.8: All country return level plot.



comparison of the extremal BA distribution. We find the highest risk for extremely large fires for any given return period for Portugal and the lowest for Italy. Comparing Greece and Spain, a higher risk for large BA's emerges for Spain for low return periods (approximately < 3 years) but above this threshold the return levels are distinctively larger for Greece.

Figure 1.9: Country-level BA exceedance probabilities in any given year.



A similar picture emerges when overlying the individual country-level BA thresholds that are exceeded in any given year with corresponding probabilities shown in Figure 1.9. The annual probability for extremely large fires decreases fastest for Italy and slowest for Portugal. The rate of the yearly probability decrease is comparably close for Spain and Greece with the ξ parameter point estimates only differing marginally.

1.4.7 Economic Valuation

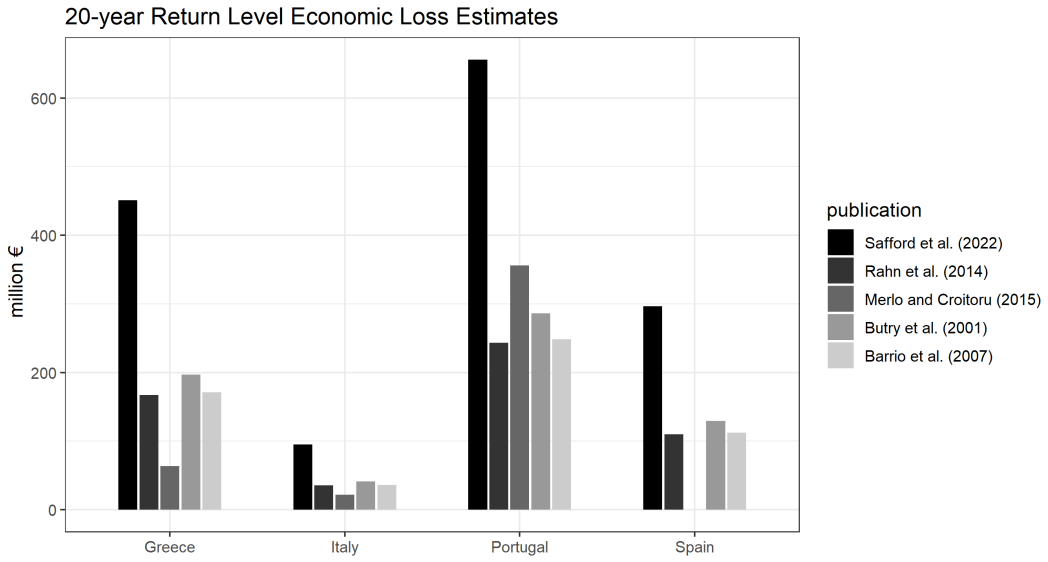
Combining our results with the economic loss figures in €/ha leads to expected return period-specific economic losses presented in Table 1.7. Allowing a comparison of the individual publications' loss calculations, Figure 1.10 graphically displays the economic loss estimates for wildfires that are expected to occur, on average once every 20 years.

Table 1.7: Range of country-level economic loss estimates for specific return levels (rl) in million € (in 2020€).

country	5-year rl	10-year rl	20-year rl	50-year rl
Portugal	107-290	162-439	243-656	400-1'078
Spain	58-158	81-219	110-297	160-431
Italy	15-64	18-78	22-95	29-128
Greece	25-180	41-290	64-451	11-794

Recall from Table 1.1, that while the estimates by Butry et al. (2001), Rahn et al. (2014), and Barrio et al. (2007) are relatively close, the country-specific €/ha estimate based on the figures in Merlo and Croitoru (2005) is lower than the other three for Italy and Greece, and

Figure 1.10: Country-level economic loss estimates for the 20-year return period.



higher for Portugal. The latest study conducted by Safford et al. (2022) clearly stands out with a distinctively larger loss estimate value.

In addition to providing economic loss estimates for specific return periods resulting from the extreme value modelling, we also show cost estimates based on the single largest observed wildfires in the study period for each country. Hence, the cost estimates come from multiplying the maximum events in Table 1.2 by the corresponding €/ha estimates derived from the existing literature. The largest wildfire event leads to an economic loss estimate of 218-589 million € for Portugal, 105-283 million € for Spain, 23-101 million € for Italy, and 56-399 million € for Greece for a specific event of that magnitude.

1.5 Discussion

1.5.1 Implications

With the quantification of the country-level risk of extreme wildfires, we are able to contribute to the empirical evidence to information-based decision-making regarding forest management for various stakeholders. Providing reliable estimates of return periods arguably has important implications for government agencies looking to adjust budget planning for fire prevention measures and suppression spending. Furthermore, the quantification of large fire risk through return levels can provide useful information for landowners regarding long-term investment and forest management choices, or for other institutions such as reinsurance companies. Moreover, the knowledge of the wildfire risk could also be used to increase awareness and thus may affect decision-making at the individual level (i.e., location choices, property protection measures, and investment in insurance associated with wildfire damage). Converting the return level estimates of extreme wildfires in Mediterranean Europe to monetary values, as we did here, arguably provides an important tool for policy-related cost-benefit analyses. For example, the associated monetary values with a return period event can assist a government in the budget allocation of both fire prevention and suppression spending by comparing their expenditures with the expected losses, particularly for extremely large wildfires over a specific time period.

Examining the specific results it is insightful to first reflect on the implications of the different distributions of extreme wildfires estimated for the individual countries in our analysis.

Most importantly, we find that these rare events follow a Fréchet type distribution for Portugal, Spain, and Greece. This is in line with regional estimates within Portugal by De Zea Bermudez et al. (2009) and Scotto et al. (2014). Out of the three limit type distributions, the Fréchet distribution has the heaviest tail indicating that the probability of rare events is much higher than commonly perceived (i.e., "extreme" wildfires are not as surprising). However, although the point estimate of the shape parameter is very similar for Spain and Italy, it is not found to be significantly positive for Italy, implying that the respective extremes follow a Gumbel distribution characterised by a lighter tail than for the other Mediterranean countries. Thus, extremely large wildfires are expected to occur less often in Italy than in Spain. Overall, we find the largest point estimate of the shape parameter of 0.58 for Greece, followed by a value of 0.52 for Portugal. This indicates that the probability of extremely large wildfires is highest in Greece when only the shape parameter of the extreme event distribution is considered (i.e., excluding the mean and the spread of the distribution). Notably, both the Fréchet and the Gumbel distributions do not converge to an upper limit but are unbounded. As a matter of fact, the extremes, and thus the associated losses, characterised by the Fréchet distribution, are limitless.

The return level results derived from the inclusion of all three parameter estimates (location, scale, and shape) indicate the highest risk of extremely large wildfires across all evaluated return periods in Portugal and the lowest risk in Italy. Comparing Greece and Portugal, the return level for up to about 3-year events is higher in Spain, but for any return period above it is found to be higher for Greece. For instance, the individual country return levels

for 10-year return period events are 50'338 ha (PT), 33'242 ha (GR), 25'165 ha (ES), and 8'966 ha (IT). For wildfires which are expected to occur on average once in 20 years, the return levels are estimated at 75'256 ha (PT), 51'764 ha (GR), 34'017 ha (ES), and 10'890 ha (IT).

Our data do not suggest that the FWI, which captures relative fire danger, affects the distribution of wildfire occurrence or their magnitude. Ideally, we would have a much longer time series which would make it possible to detect climatological changes. This means our results should be interpreted with care. In this regard, it must be pointed out that while many climate change projections suggest that Southern Europe faces an increasing risk of extreme wildfires (Bowman et al., 2017; De Rigo et al., 2017; Turco et al., 2018), Batllori et al. (2013) indicate that fire activity predictions can be highly divergent, particularly regarding precipitation-related variables. Notwithstanding the wildfire risk driven by future climate conditions, evidence also suggests that the risk associated with human exposure may increase especially with projected population growth in fire-prone regions (Knorr et al., 2016; Turco et al., 2019). Even though we did not find evidence supporting a time trend in our study, it is crucial to continue efforts to better understand the risk associated with wildfires in Mediterranean Europe. Going forward, more comprehensive and harmonised data are needed to evaluate future extreme wildfire risk scenarios incorporating climatic and demographic components as well as more detailed information on the individual fires (e.g., duration, severity, ignition point, cause) in order to distinguish which factors have the potential to influence the extremely large fires.

1.5.2 Limitations

Although predictions of events not actually observed in the historical data are common with the use of extreme value theory methods, we need to emphasise that our estimates are based on data for a particularly short time period of 14 years. In this regard encounter probability[13] suggests that the probability of observing a 5-year event given our data is approximately 96%, a 10-year event 77%, a 20-year event 51% and a 50-year event 25%, and only our 10 and 20-year event estimates are based on events witnessed in the sample period. The short time period of data may also play a role in the non-stationarity results. Although modelling the threshold excesses conditional on factors potentially influencing the distribution of extreme wildfires does lead to improved capture of the empirical data for all countries but Greece, none of the models significantly changed the extremal distribution. Whereas it is not given that the included variables would lead to a change in the distribution of extremely large wildfires over a prolonged time period, a potential underlying dependence on factors driving or affecting extreme BA is more difficult to detect in shorter periods of analysis.

The coupling of our estimated BA return levels with existing economic loss figures also comes with strong caveats, particularly with regard to regional and temporal transfers of monetary estimates, as well as through the distinct study designs incorporating disparate economic variables in the respective loss calculations. For example, only Barrio et al. (2007) and Butry et al. (2001) include any estimation of wildfire-related health costs which are of significant magnitude and thus, of rising concern as pointed out in Black et al. (2017). Fur-

[13] The encounter probability $P_e = 1 - (1 - \frac{1}{T})^n$ is the likelihood of observing a T-return period event within a specific time period denoted by n.

thermore, even though Merlo and Croitoru (2005) address country-level estimates of indirect use, option, bequest, and existence values of forests in general, they are not applied to the BA scenario and only the estimate provided by Safford et al. (2022) includes ecological (vegetation and wildlife) damage. Moreover, as the monetary valuation of indirect costs poses great challenges, besides the impediment imposed by oftentimes limited data availability particularly driven by methodological restraints, many loss calculations focus on direct impacts. However, the indirect costs are likely to exceed the reported costs as argued in CCST (2020). Thus, our calculations are conservative and the considered losses are likely to represent only some fraction of the actual economic impact. Nevertheless, our results still provide some indication of the serious implications wildfires have for many other sectors that they can reach far beyond the commonly assessed impacts.

In terms of examining the role of covariates in potential changes in the distribution of extreme wildfires, the FWI may not be the most suitable variable to capture those. Jiménez-Ruano et al. (2019) conclude that although the FWI provides useful information regarding seasonal variability and near-future trends, it is not necessarily the most advisable index to detect long-term trends. In this regard however, Pérez-Sánchez et al. (2017) do identify the FWI as the most suitable index for fire-risk ignition and spreading in semiarid areas such as the Iberian Peninsula. Likewise, De Rigo et al. (2017) point out that the FWI is well-suited as a harmonised index over different regions for weather-driven fire danger, and Fernandes et al. (2016) observe that particularly large fires exhibit stronger responses to the severity of the fire weather.

With respect to the population density covariate, Bowman et al. (2017) demonstrate that large destructive wildfires are most likely to occur in a two-sided bounded area that excludes either very sparsely or densely populated areas, and thus highlights the underlying complexity of this interdependence. On this account, our study design calculating the mean population density around the centre of a burned area might arguably be an unsatisfactory way to capture this intricate relationship, given there exists one for the extremely large fires in the studied geographical area. Additionally, we do see that many of the population density values for the extreme wildfires in our data set lie within a narrow range leading to the small explanatory power of the variable. Having information on the exact ignition location of a fire might contribute to the evaluation of the potential association between population density and extreme wildfire occurrence.

Regarding the land cover type, we face a slightly different problem as it is implemented as a categorical variable. As we look at the extremes, we focus on solely 42-62 observations for each country, and thus, we need to strictly limit the number of categories to forego having only very few wildfires for each of those. Therefore, we categorise the extreme wildfires into four main country-specific land cover type classes and thereby sacrifice some of the specificity. Comparable drawbacks arise from the categorical covariates capturing seasonality as the extremely large wildfires are assigned to one of the four seasons. However, in this case it is less a problem of simplification but rather one of unequally distributed observations per category, particularly as "off-wildfire season" categories arguably contain very few observations.

For the specific case of France we note that compared to the other countries the data are more challenging to work with. Although it is typical for all the Mediterranean countries that certain years stand out with more severe fire seasons, this is particularly pronounced for the wildfire records in France with more than half of the observations coming from 2019. This in turn leads to a comparably high dependence on extreme observations as many of the largest wildfires are recorded in a single year. Furthermore, in contrast to the other European Mediterranean countries, France geographically expands much further North and is thus characterised by more diverse land cover types. Hence, the finding that a specific land cover type, namely Sclerophyllous vegetation, leads to a positive shift in the distribution of the extremes might indicate that this is the vegetation type most dominant at the Mediterranean coastline and may be correlated with extreme wildfires.[14]

1.6 Conclusion

In this paper, we assemble a high-quality homogeneous up-to-date geospatial data set for Mediterranean Europe and perform a cross-country risk analysis of extreme wildfires defined by BA. Although modelling a variety of covariates with the potential to affect the extremal distributions, we find no evidence for non-stationarity in the observed study period. Furthermore, the threshold excesses for France in the data set do not fulfil underlying assumptions to carry out a sound EVT analysis and are thus only included in the descriptive part. In

[14] Although the aim of this paper is to model and compare country-level data, future research may benefit from regional modelling which may be particularly useful for the case of France where there is considerable heterogeneity in wildfire occurrence primarily between the north and the south of the country (and Corsica).

our results, we find the highest risk for extremely large wildfires in Portugal, followed by Greece, Spain, and Italy. We estimate the return levels for 5, 10, 20, and 50-year return period events and combine our outcomes with the existing literature on economic costs. The robust estimation of extreme wildfire events underlying an evidence-based risk assessment is arguably beneficial for governmental bodies, reinsurance institutions, landowners and residents in wildfire-prone areas providing support in information-based decision-making processes.

We emphasise the need to build international homogeneous comprehensive databases with a high spatial and temporal resolution regarding wildfire occurrence (ideally including point of origin, duration, and cause) but also dedicated to associated measures such as prevention and suppression spending, as well as individual fire event impact on ecosystems, infrastructures, properties, and people. Accompanying the extensive WUI with exposed communities, particularly in the highly populated coastal areas of Southern Europe and vulnerable ecosystems across the region, extreme wildfire events continue to pose a substantial environmental hazard for Mediterranean Europe in the future.

Chapter 2

The regional economic impact of wildfires: Evidence from Southern Europe

Abstract

We estimate the impact of wildfires on the growth rate of gross domestic product (GDP) and employment of regional economies in Southern Europe from 2011 to 2018. To this end we match Eurostat economic data with geospatial burned area perimeters based on satellite imagery for 233 Nomenclature of Territorial Units for Statistics (NUTS) 3 level regions in Portugal, Spain, Italy, and Greece. Our panel fixed effects instrumental variable estimation results suggest an average contemporary decrease in a region's annual GDP growth rate of 0.11-0.18% conditional on having experienced at least one wildfire. For an average wildfire season this leads to a yearly production loss of 1.3-2.1 billion euros for Southern Europe. The impact on the employment growth rate is heterogeneous across economic activity types in that there is a decrease in the average annual employment growth rate for activities related to retail and tourism (e.g., transport, accommodation, food service activities) of 0.09-0.15%, offset by employment growth in insurance, real estate, administrative, and support service related activities of 0.13-0.22%.

Keywords: GDP growth, employment growth, natural disasters, wildfire

JEL classification codes: O4, Q5, R1

2.1 Introduction

In recent years news coverage of orange coloured skies, evacuations, and devastation caused by wildfires has become all too familiar. Even though one tends to only hear about the most calamitous and tragic of fires, every summer Southern European countries experience a large number of fires of varying degrees of seriousness (San-Miguel-Ayanz et al., 2021, 2022). These events can be highly disruptive and destructive, affecting different sectors of the economy, such as forestry and agriculture (Butry et al., 2001; Rego et al., 2013), industry and construction (Kramer et al., 2021; Wang et al., 2021), and recreation and tourism (Kim and Jakus, 2019; Molina et al., 2019; Gellman et al., 2022; Otrachshenko and Nunes, 2022). Importantly, natural disasters, including wildfires, are for the most part localised events that are likely to induce predominantly local effects that could potentially be disguised if one only considers aggregated data at the national level (Horwich, 2000).[1] Given increasing European regional inequality particularly in Southern Europe (Iammarino et al., 2019) and the possibility that the region faces an increased risk of wildfires due to climate change (Dupuy et al., 2020; UNEP, 2022), being able to identify and quantify the potential economic impact of wildfires has important implications for regional policy making. In this paper we explicitly set out to examine the regional gross domestic product (GDP) and employment impacts of wildfires in Southern Europe since 2010.

There is now a sizeable theoretical and empirical literature focusing on the impacts of natural disasters other than wildfires on GDP growth. For example, negative effects are found

[1] Wildfires may also have more wide reaching effects through drifting smoke pollution although this aspect is not specifically considered in this study.

after hurricanes (Strobl, 2011), cyclones (Naguib et al., 2022), and floods (Parida et al., 2021). Furthermore, Barone and Mocetti (2014) show a short-term negative effect on GDP growth from a study of two earthquakes in Italy, but report a positive long-term effect for one of them. While a majority of studies do report predominantly negative effects, the literature does not offer conclusive evidence and impacts depend on a variety of dimensions, such as on severity, disaster type, and country of occurrence (Loayza et al., 2012; Fomby et al., 2013). Nevertheless, conducting a meta-analysis using more than 750 estimates from publications studying the relationship between natural disasters, Klomp and Valckx (2014) conclude that there is a genuine negative effect that is increasing over time. Similarly, Felbermayr and Gröschl (2014) construct a comprehensive disaster data set from geophysical and meteorological information as opposed to using insurance data, and also find a robust negative effect of natural disasters on GDP growth.

A number of studies have also examined the employment impact of natural disasters, although the evidence is scarcer and much more mixed. As Deryugina (2022) notes, natural disasters can affect the labour market equilibrium through a number of different channels. For example, if areas that heavily rely on tourism are impacted, employment in the hospitality sector is likely to fall. For example, Barattieri et al. (2021) show short-term negative employment and wage impacts for hurricane affected counties in Puerto Rico between 1995 and 2017. Similarly, Deryugina et al. (2018) show a short-run decline in labour market outcomes following hurricane Katrina. However, labour demand in other sectors could arguably increase through an element of "creative destruction", whereby damaged sub-optimal infras-

tructure is replaced with superior technology in the rebuilding phase. In this regard, Groen et al. (2020) find an increase in regional employment for those industries that are reconstruction related following hurricanes Katrina and Rita in 2005.

While assessing the economic impact of wildfires from a general natural hazards perspective can provide considerable insights, as pointed out by McCaffrey (2004), wildfires are also characterised by features that make them unique compared to other natural disasters. For instance, wildfires can perform beneficial functions for ecosystems under certain scenarios (Holmes et al., 2008). Moreover, wildfires are often human induced in that socioeconomic factors, such as poverty, education, or illegal activity, can contribute to the probability of wildfire occurrence (Michetti and Pinar, 2019), resulting in potential damage that is more easily mitigated or exacerbated by policy measures (e.g., land management, fire prevention) compared to other environmental hazards (Borgschulte et al., 2020).

Importantly, wildfires are particularly atypical among natural hazards since property damage can oftentimes be substantially reduced if there is large investment in manpower and equipment as described in Baylis and Boomhower (2019). Hence, central to understanding the potential economic impact of wildfires is the response during the hazard event itself. More specifically, during relatively short-duration hazards (e.g., earthquakes, hurricanes, floods) the mitigating response choice set for the direct effects is limited temporally, while wildfires can be actively "fought" and often last for several days or even weeks. Hence, an abundance of resources, including direct suppression spending and contracted services, are often made available during the wildfire event (Davis et al., 2014). If a substantial part of the employed

services and goods are provided locally, these measures can also have major indirect impacts on regional economies. From an econometric perspective these aspects that are peculiar to wildfires raise important endogeneity concerns when trying to causally identify the economic impact of wildfires compared to other environmental disaster settings.

While a considerable body of literature has studied the conceivably detrimental and immediate impact of wildfires (Morton et al., 2003; Stephenson et al., 2013; CCST, 2020), a small number of studies scrutinise their effect on traditional economic indicators, such as GDP and employment growth. The most relevant research in this area was conducted by Nielsen-Pincus et al. (2013) who examine large wildfire events in the Western United States (US) and find an increase in county-level employment growth of 1% during the quarters where fire suppression efforts took place, although the effect is heterogeneous with regard to county characteristics and economic sectors (Nielsen-Pincus et al., 2014). Furthermore, Borgschulte et al. (2020) report reduced earnings of approximately 0.04% over two years per additional smoke exposure day for the US.

Although the impact of wildfires on the labour market or on GDP growth has to date drawn little attention, two other research areas evaluating economic impacts of wildfires are better understood. On the one hand, the hedonic pricing literature demonstrates a predominantly negative effect on house prices of up to 20% following wildfires in the US (Nicholls, 2019). Furthermore, Mueller and Loomis (2014) document that although property values are negatively affected by wildfires, there is large variation across the distribution of house prices, while McCoy and Walsh (2018) find a short-lived negative effect on property values if a burn

scar can be viewed from the house. On the other hand, negative economic effects related to fire induced smoke pollution suggest that there are substantial health costs as demonstrated by Kochi et al. (2012), Richardson et al. (2012), Burke et al. (2020), Johnston et al. (2021), and Tarín-Carrasco et al. (2021). However, even for these relatively well researched aspects of wildfires, the majority of studies focus on the US and Australia, and not on Europe.

The current study makes three main contributions to the literature. First, we examine the economic implications of wildfires on regional employment and GDP growth in Europe, which to the best of our knowledge has not been explored. Since wildfires in Europe are perceived as a growing risk that predominantly affects Southern Europe, our study provides some of the first evidence on economic impacts for this fire-prone geographical region. Second, we focus on small-scale regional effects, which Horwich (2000) argues are important because natural disasters are for the most part localised events, and potential impacts are often imperceptible when studied at more aggregated geopolitical levels. As a matter of fact, neglecting potential regional economic impacts has already been identified as a major shortcoming of most previous studies addressing the impacts of natural hazards (Botzen et al., 2019). Third, in order to overcome potential endogeneity concerns when empirically estimating the economic effect of wildfires, we employ a novel causal identification strategy creating an instrumental variable (IV) by isolating climatic features for predominantly forested areas that are particularly relevant for capturing the probability of wildfire occurrence while also controlling for general and related climate conditions that might affect regional economic outcomes directly.

The empirical analysis in this paper relies on the construction of a panel data set matching annual regional economic data on employment and GDP growth from 2010 to 2018 with burned area (BA) polygons based on satellite imagery for regions in Portugal, Spain, Italy, and Greece. These data are combined with general climatic data, land cover maps, and a time-varying Fire Weather Index (FWI). Employing two-stage least squares (2SLS) instrumental variables regressions arguably allows us to causally quantify any potential effects of wildfires on annual regional employment and GDP growth in Southern Europe over our sample period.

To briefly summarise our results, we find an annual decrease in the rate of GDP growth of 0.11-0.18% for wildfire affected regions. Given that 102 regions are affected by wildfires every year on average, our findings indicate rough annual economic losses for Southern Europe in the range of 1.3-2.1 billion euros. There is also a heterogeneous impact on employment growth across economic activities where annual employment growth in tourism-related activities (e.g., accommodation, transportation, food service) decreases by 0.09-0.15%, while the sectors that include financial, insurance, real estate, and administrative activities experiences an average increase in the employment growth rate of 0.13-0.22%.

The remainder of the paper is organised as follows. Section 2.2 describes the data sources, how we constructed variables, and provides some descriptive statistics. Section 2.3 describes our identification strategy, the instrument construction, and econometric specification. Finally, the results are presented and discussed in Section 2.4 while Section 2.5 concludes.

2.2 Data and Descriptive Statistics

2.2.1 Regional Unit of Analysis and Sample Composition

The Nomenclature of Territorial Units for Statistics (NUTS) classification provides harmonised regional statistics for the European Union member and partner states. The hierarchical system divides the economic territory into major socioeconomic regions (NUTS 1), basic regions for the application of regional policies (NUTS 2), and small regions for specific diagnoses (NUTS 3). Our countries of interest include a total of 243 NUTS 3 regions, namely 25 *Entidades Intermunicipais* for Portugal, 59 *Provincias* for Spain, 107 *Provincie* for Italy, and 52 *Omades Periferiakon Enotition* for Greece (Eurostat, 2020). For data availability and comparability reasons the following regions are excluded from our analysis: The Azores and Madeira for Portugal (2 regions), the Canary Islands for Spain (7 regions), and Sud Sardegna for Italy (1 region) leaving 233 NUTS 3 regions that are used in our analysis.[2]

Table 2.1: Sample composition and descriptive statistics showing the size of NUTS 3 regions by country.

	N	Proportion (%)	Mean (km^2)	sd (km^2)	Median (km^2)
Portugal	23	10	3,860	1,948	3,345
Spain	52	22	9,588	5,251	9,317
Italy	106	46	2,774	1,679	2,454
Greece	52	22	2,534	1,706	2,339

Notes: (i) Proportion (%) = the number of regions per country as a share of all sample regions; (ii) sd = standard deviation.

Table 2.1 shows the sample composition and the disaggregated mean and median size of

[2] The omissions for Portugal and Spain are due to missing meteorological and Fire Weather Index data, and the excluded Italian region is due to rearranged regional boundaries during the study period.

the NUTS 3 regions by countries. Italy accounts for almost half of the regions (46%), Spain and Greece each add up to about one fifth, and one in ten regions is in Portugal. One can observe variation in the mean size of a unit per country, with the largest regions in Spain, and the smallest in Italy and Greece, on average.

2.2.2 Economic Data

We use data on regional level employment and per capita GDP as provided by the regional economic accounts of the Statistical Office of the European Union (Eurostat).[3] Regional accounts are derived from the corresponding national accounts, and thus, are generally defined using the concepts applied to national accounting procedures.[4] The estimation of regional GDP can follow either the production or the income approach.[5] The production approach measures regional GDP as the sum of gross value added (GVA), which is defined as the difference between output and intermediate consumption, plus taxes minus product subsidies. For the income approach, regional GDP at basic prices is derived from measuring and aggregating the regional generation of income of the economy, i.e., wages and salaries, the sum of other taxes minus subsidies on production, employers' social contributions, gross operating surplus, and consumption of fixed capital. In practice, gross operating surplus is generally

[3] https://ec.europa.eu/eurostat/web/rural-development/data (accessed in August 2021).

[4] It is noteworthy that a series of conceptual and practical difficulties arise when breaking down national data or compiling regional data directly. Challenges in accurate regional estimations involve how to account for enterprises with several regional establishments, extra-regio territory, major construction projects, cross regional boundary pipelines and cable distribution networks, or commuter flows, to name a few. For a detailed discussion with accompanying guidelines, see Chapter 13 in Eurostat (2013a) and Eurostat (2013b).

[5] Unlike for national data, the expenditure method can not be applied given the absence of data on imports and exports on the regional level.

not available by region and industry which poses a barrier to using the income approach. In general, countries are free to choose their preferred estimation approach. Hence per capita figures can be calculated for all regions excluding extra-regio measures (Eurostat, 2013b).

Our measure of regional employment is from the European Union Labour Force Survey that is based on a household sample survey of people aged 15 years and over. Persons are categorised as "employed" if any work has been performed during the survey reference week (e.g., for pay or family gain) or if they had a job at the time but were temporarily absent due to illness, holidays or educational training. The aggregated annual average of employed persons makes allowance for the fact that some people are not employed over the entire year but do casual or seasonal work (Eurostat, 2013b). Using the population data provided by Eurostat we calculate the share of employed persons in the total population.[6] We are also able to disaggregate employment growth by sections based on the Statistical classification of economic activities in the European Community (NACE) Rev.2, which is a revised classification implemented in 2007.[7] More specifically, we use Eurostat data for six categories that combine and classify a total of 21 individual economic activity sections as shown in Table 2.2 (Eurostat, 2008).

The regional economic variables are available from 2010 to 2018 and we use first differences of their logged values (i.e., growth rates) in our analysis. The geographical distribution of the average yearly employment and GDP growth rates across the NUTS 3 regions is shown

[6] https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=nama_10r_3popgdp&lang=en (accessed in August 2021).

[7] Derived from French, NACE translates as Nomenclature statistique des activités économiques dans la Communauté européenne.

Table 2.2: Statistical classification of economic activities in the European community (NACE).

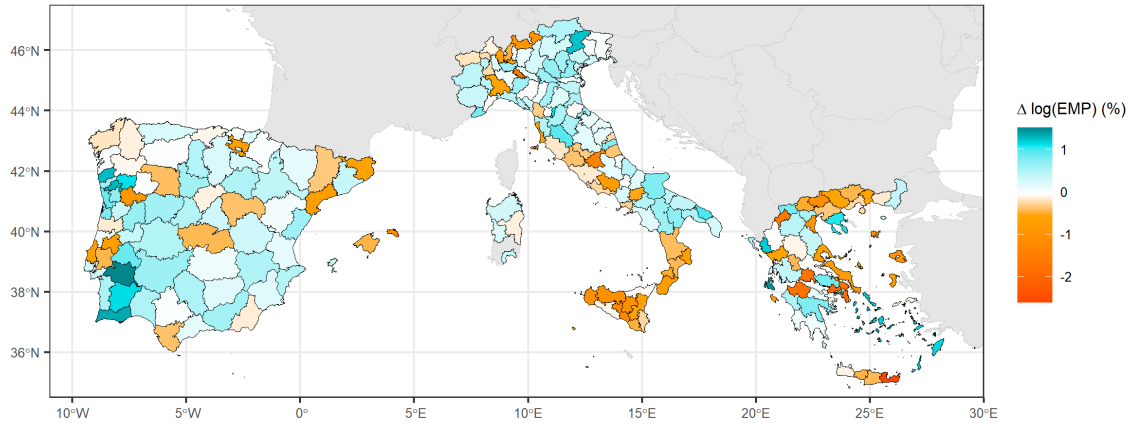
Category	Section	Description
A	A	Agriculture, forestry and fishing
B-E	B	Mining and quarrying
	C	Manufacturing
	D	Electricity, gas, steam and air conditioning supply
	E	Water supply, sewerage, waste management and remediation activities
F	F	Construction
G-J	G	Wholesale and retail trade; repair of motor vehicles and motorcycles
	H	Transportation and storage
	I	Accommodation and food service activities
	J	Information and communication
K-N	K	Financial and insurance activities
	L	Real estate activities
	M	Professional, scientific and technical activities
	N	Administrative and support service activities
O-U	O	Public administration and defence; compulsory social security
	P	Education
	Q	Human health and social work activities
	R	Arts, entertainment and recreation
	S	Other service activities
	T	Activities of households as employers
	U	Activities of extraterritorial organisations and bodies

in Figure 2.1. The maps demonstrate that while the distribution of the employment growth rate (Figure 2.1a) is fairly heterogeneous across countries, with intra-country regions experiencing both positive and negative employment growth rates over the study period, the emerging image for the GDP growth rates is strikingly different (Figure 2.1b). Rather we observe clear differences at the country-level, indicating predominantly positive GDP growth in Portugal, Spain, and Italy, and negative growth in Greece.[8]

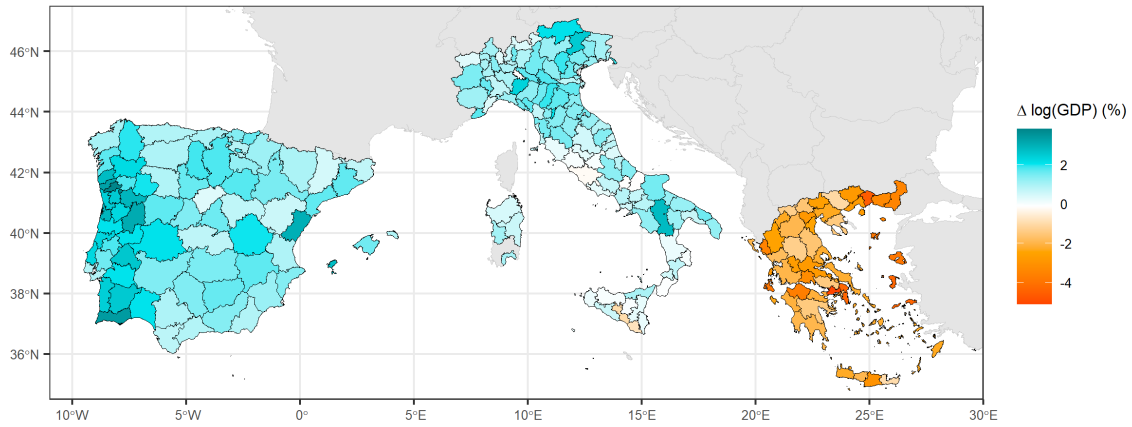
Descriptive statistics for the economic variables are summarised in Table 2.3. On average,

[8] Note, that our study period starts shortly after the financial crisis where Greece was particularly hard hit.

Figure 2.1: Average annual employment and GDP growth rates (2011-2018).



(a) Employment Growth



(b) GDP Growth

Notes: (i) economic data are from the Statistical Office of the European Union (Eurostat); (ii) Δlog(EMP) denotes the growth of the employment rate and Δlog(GDP) is the per capita GDP growth rate.

39.6% of the population is employed, and the employment growth rate is centred around zero with a slight tendency towards being positive. The smallest and largest values are within five standard deviations of the mean. The average per capita GDP is 21,184 euros, and on average positive GDP growth rates over the time period are observed. The GDP growth rates are within around six standard deviations of the mean.

Table 2.3: Descriptive statistics of economic variables (2010-2018).

	Min	Mean	sd	Median	Max	N
Employed/total population (%)	24.6	39.6	6.3	39.7	67.6	2,097
Δlog(EMP) (%)	-12.3	0.0	2.6	0.3	12.8	1,864
GDP/capita (€)	9,500	21,184	7,268	19,600	55,900	2,097
Δlog(GDP) (%)	-20.0	0.4	3.8	1.0	22.5	1,864

Notes: (i) economic data are from the Statistical Office of the European Union (Eurostat); (ii) Δlog(EMP) denotes the growth of the employment rate and Δlog(GDP) is the per capita GDP growth rate; (iii) sd = standard deviation.

2.2.3 Wildfire Impact Variables

The impact of wildfires is proxied by fire numbers as an absolute measure, and BA as a share of a region's total area. The primary data set for the construction of these variables is the high-resolution harmonised spatial BA data product provided by the European Forest Fire Information System (EFFIS).[9] This data product is based on a semi-automatic approach that combines Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery with two bands (red and near-infrared) at a 250 meter spatial resolution,[10] ancillary spatial data sets, and refinement of the perimeters through visual inspection backed up by news coverage. The burn perimeters are updated up to two times a day capturing fires larger than around 30 hectares.[11] In order to analyse potential lagged effects on the economic outcome variables our study includes all fires from 2001 to 2018.

For cross-border wildfires that affect several regions, the burn perimeters are split accord-

[9] https://effis.jrc.ec.europa.eu (accessed in July 2021).

[10] There are five bands (blue, green, as well as three short-wave infrared bands) with spatial resolution of 500 meters that help to improve BA discrimination by providing complementary information.

[11] https://effis.jrc.ec.europa.eu/about-effis/technical-background/rapid-damage-assessment (accessed in July 2021).

ing to the NUTS 3 regional boundaries.[12] We excluded all fires that burned less than five hectares after the splitting process from the fire count variable, while all burned area counts towards the BA variable. Between 2010 and 2018 the number of wildfires in the dataset is 6,709, whereby less than 1% resulted from the splitting process by regions. The total area burned over this period is approximately 2.4 million hectares.

Table 2.4: Descriptive statistics of wildfire impact variables and the Fire Weather Index (2010-2018).

	Min	Mean	sd	Median	Max	N
All observations						
FIRE	0	3	9	0	129	2,097
BA (%)	0	0.34	1.53	0	33.82	2,097
Wildfire affected observations						
FIRE	1	7	13	2	129	920
BA (%)	0.001	0.77	2.24	0.13	33.82	920
Instrument						
FWI forest	0.0	15.9	11.6	14.9	64.8	2,097

Notes: (i) FIRE indicates the annual number of wildfires per region; BA in % denotes the annual burned area relative to the total area per region; FWI forest indicates the daily mean Fire Weather Index over the summer months for predominantly forested areas; (ii) "Wildfire affected observations" includes all observations where at least one wildfire occurred in a given year; (iii) sd = standard deviation.

The wildfire impact proxy variables are summarised in Table 2.4 for 2010 to 2018.[13] The mean fire number for all regions is three, and an annual average of 0.3% of a region is burned. Only considering the observations that experienced at least one fire denoted as "wildfire affected observations", the mean fire number is seven with an average of approximately 0.8%

[12] The NUTS 2016 version of the shapefile scaled 1:1 million is provided by Eurostat at https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/nuts is used (accessed in July 2021).

[13] Although we have data from 2001 to 2018 we display the descriptive statistics for the period that matches the economic outcome variables since these figures are later used to interpret the regression coefficients.

of total area burned. In our sample, one or more wildfires occurred in about 44% of the observations, and 82% of the regions were affected over the study period. On average about a third of the regions (30%) experience a wildfire each year.

Regarding the spatial distribution of the average annual wildfire numbers shown in Figure 2.2a, most fires are observed in Southern Italy and in the Northwest of the Iberian Peninsula. Focusing on the BA (proportional to the total area of a region) displayed in Figure 2.2b, the highest values are in Central and Northern Portugal. Figure 2.2c shows the average wildfire size in hectares for each region over the study period. In contrast to Figure 2.2a and Figure 2.2b, there are comparably low values for Italy and large values for the average fire size in Greece.

2.2.4 Fire Weather Index (FWI)

The FWI is a component of the Canadian Forest Fire Weather Index System initially introduced by Van Wagner and Pickett (1985). Figure 2.3 presents a schematic of the FWI structure. The FWI captures relative fire potential, and serves as primary reference index to the Joint Research Centre (the European Commission's science and knowledge service) in the production of fire danger maps (Camia et al., 2008). The FWI is based on the combination of the two fire behaviour indices (1) Initial Spread Index (ISI) and (2) Buildup Index (BUI). The ISI estimates fire spread potential by integrating the Fine Fuel Moisture Code (FFMC), which is intended to represent fuel moisture conditions for litter fuels shaded by the forest

canopy, and surface wind speed (u and v components).[14] The BUI provides information on potential heat release incorporating fuel moisture information from deeper soil layers. More specifically, it combines the Duff Moisture Code (DMC) capturing decomposed organic material below the litter fuels, and the Drought Code (DC) representing the moisture content of the deep compact layer assessing seasonal drought effects on heavy fuels. Both the DMC and the DC are adjusted for day-length of the month.

As can be seen from Figure 2.3, the basic climatic inputs underlying the construction of the FWI are temperature, relative humidity, wind speed, and precipitation. All variables are measured at solar noon standard time. Precipitation is an accumulated measure over 24 hours. The details of the construction from the initial meteorological observations to the derivation of the fire behaviour indices ISI and BUI are beyond the scope of this paper, but are described in Van Wagner and Pickett (1985). However, we do provide an outline of the calculations for the fire behaviour indices for the FWI in Appendix B.1.

We use the daily FWI calculated by Natural Resources Canada presented in McElhinny et al. (2020). The primary meteorological inputs are from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 HRES reanalysis product with a spatial resolution of 0.25° (approximately 27-28 kilometres in our latitudes of interest).[15] We work with the FWI version using the overwintered DC, which captures inter-seasonal drought. This is preferable to using a default start value as the overwintered DC is more precise accounting

[14] u is the component of the horizontal wind towards east (zonal velocity) and v denotes its counterpart towards north (meridional velocity).

[15] Note, the primary resolution of ERA5 is 0.28125° on a reduced Gaussian grid, but the output on a regular geographical grid is 0.25°.

for precipitation in the winter months.

2.2.5 Land Cover Data

We resort to the CORINE land cover (CLC) data provided by the Copernicus Land Monitoring Service in order to distinguish between forested and non-forested areas within the studied regions.[16] CLC is specified to standardise land cover data collection in order to support environmental policy development. It was initialised in 1990 and is updated every six years. While orthocorrected satellite images provide the basis for the land cover mapping, ancillary information such as in-situ and ground survey data enhance accuracy.[17] The minimum mapping unit/width is 100 meters (25 hectares) with a thematic accuracy exceeding 85%. The CLC inventory comprises 44 land cover types (European Environment Agency, 2021).

We use the raster files of the years 2006 and 2012 and reclassify the 44 land cover types into four suitable categories for our study purposes, namely urban areas (i.e., artificial surfaces), agricultural areas, forested areas (including forests as well as shrub and/or herbaceous vegetation), as well as wetlands and water bodies (also including open spaces with little or no vegetation). See Table B1 in Appendix B.2 for the exact reclassification.

[16] https://land.copernicus.eu/pan-european/corine-land-cover (accessed in August 2022).

[17] Both the 2006 and the 2012 versions are based on the Indian Remote-Sensing Satellite P6 LISS III and on the dual date satellites (Sentinel-2 and Landsat 8). SPOT-4/5 is additionally used for year 2006 and RapidEye Earth-imaging Systems for year 2012.

2.2.6 Climatological Data

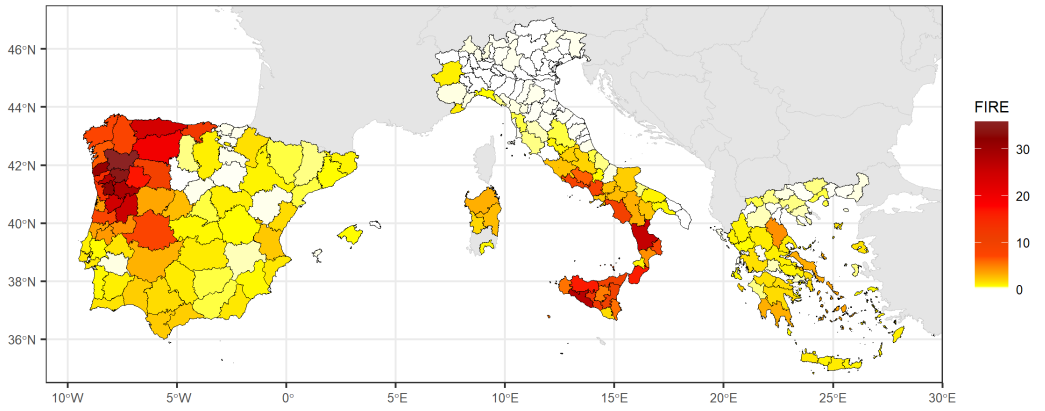
Temperature, precipitation, and relative humidity data are taken from the E-OBS, a daily gridded meteorological data set for Europe with a spatial resolution of 0.25° and is derived from in-situ observations based on the station network of the European Climate Assessment & Dataset (ECA&D) project.[18] Temperature is measured in degrees Celsius at a height of two metres and daily precipitation consists of the total amount of rain, snow, or hail (equivalent to the height in liquid water per square meter) in millimetres. Daily averaged relative humidity in percentage is based on the observational station time series from ECA&D. In order to remove data skewness, the relative humidity values are transformed by $\sqrt{100 - \textit{relative humidity}}$ before the fitting process to ensure all interpolated values are equal or smaller than 100 when converted to percentages.

For every NUTS 3 region, we take the following approach to calculate seasonal and annual average temperature, precipitation, and relative humidity for each of the four land cover type categories (i.e., urban, agriculture, forest, wetland and water bodies). First, an E-OBS gridcell for a specific land cover type is matched with a region if, in the overlapping part of the region and the E-OBS gridcell, a majority of the area is of that specific land cover type. Second, we average all the E-OBS gridcell values that are matched for a specific region and land cover type. Third, we use the daily meteorological data to calculate seasonal, i.e., summer months, and annual average temperature, precipitation and relative humidity for each of the four land cover types.

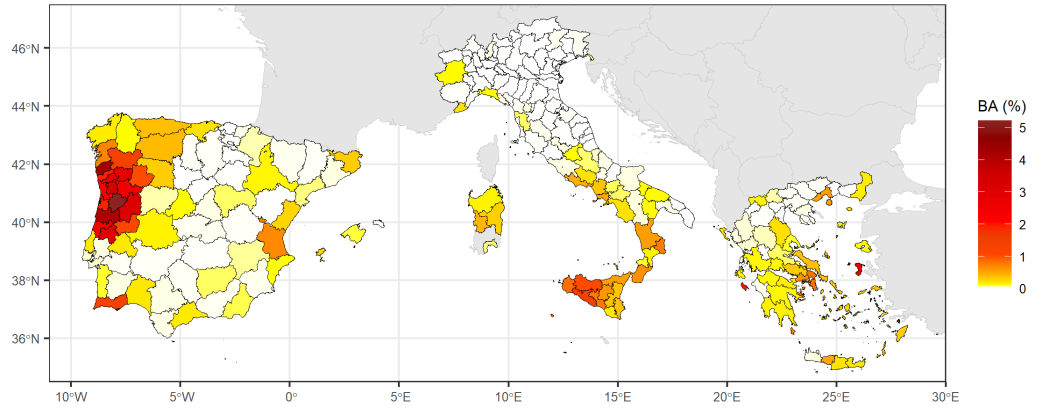
[18] For details see Cornes et al. (2018).

Even though the E-OBS data set provides information on wind speed, there are many missing values, particularly for Southern Greece and Sicily. Therefore, we instead use the 10 meter u and v wind components from the fifth generation of the ECMWF atmospheric reanalyses of the global climate (ERA5) data product (Hersbach et al., 2020). The spatial resolution of 0.25° is similar to the E-OBS data set and the temporal resolution is hourly. We extract daily values at 12 pm and match the ERA5 gridcells with the NUTS 3 regions for each land cover type in a similar manner as described for the other climatic variables. Once processed to seasonal and annual average values, wind speed is calculated from the u and v component where *wind speed* $= \sqrt{u^2 + v^2}$.

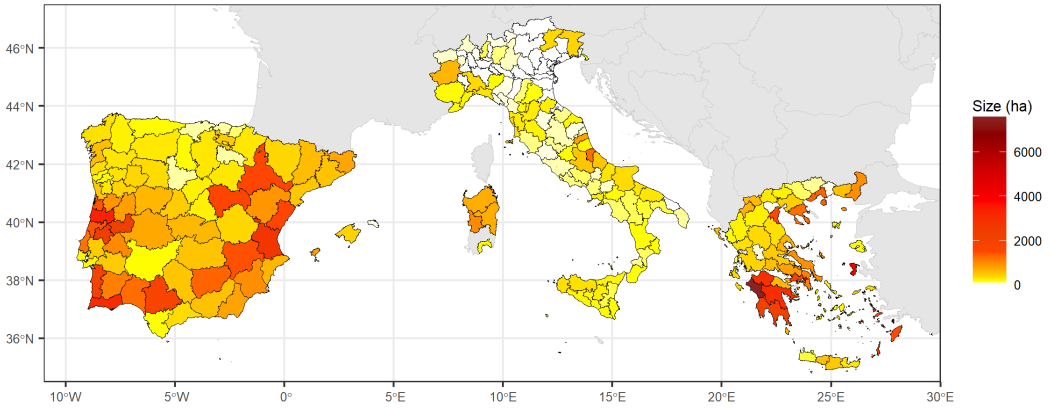
Figure 2.2: Average annual wildfire occurrence (2010-2018).



(a) Fire Number



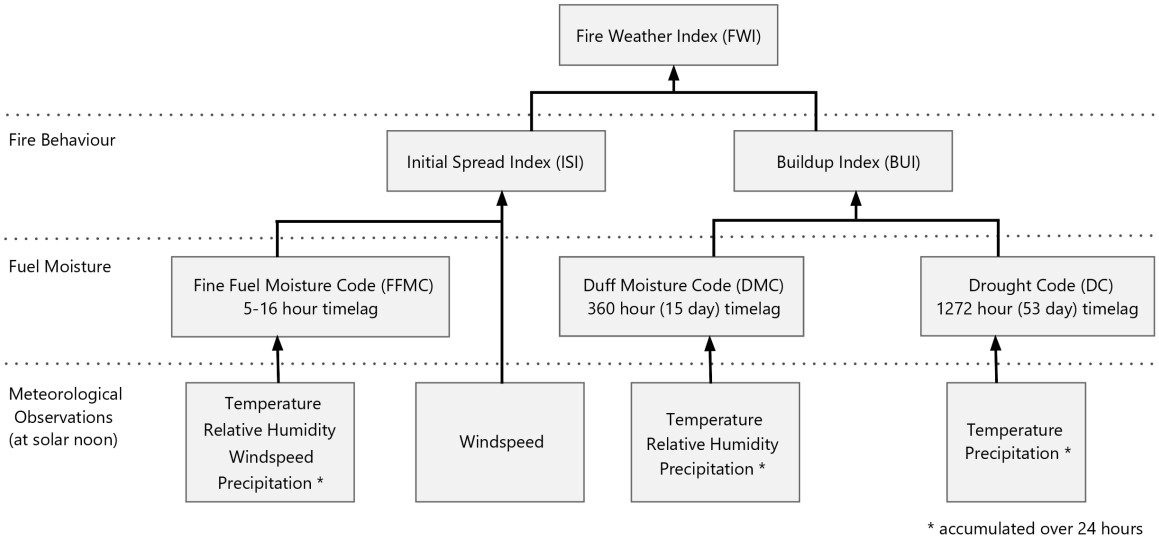
(b) Proportion Burned



(c) Average Fire Size

Notes: (i) wildfire data is taken from a high-resolution burned area product provided by the European Forest Fire Information System (EFFIS); (ii) FIRE indicates the annual number of wildfires per region; BA in % denotes the annual burned area relative to the total area per region; Size indicates the average size of a fire in hectares (ha).

Figure 2.3: Structure of the Canadian Fire Weather Index System based on Van Wagner and Pickett (1985).



2.3 Empirical Framework

2.3.1 Identification Strategy

The identification of the causal effect of wildfires on regional economies is complicated by the potential endogeneity of the wildfire proxy impact variables. As we employ geophysical measures, i.e., remotely sensed imagery defining the BA, in conjunction with regional level fixed effects, the source of endogeneity typically induced by using reported loss or damage data (e.g., through insurance claims that are likely to correlate with GDP/capita) is avoided, as outlined in Felbermayr and Gröschl (2014). However, using ordinary least squares (OLS) with regional level fixed effects, even with the geophysical based measures of the fires, might still produce biased estimates due to a number of time varying unobserved factors. Regional unobservables include, inter alia, fire and land management policies (e.g., fire prevention and suppression regimes), rural exodus/urbanisation rates, land-use changes (e.g., deforestation), land-use regulations, political instability, and local government corruption. These endogeneity concerns are particularly important for wildfires as opposed to other natural disasters because wildfires are often due to human induced activities (under the right climatic conditions) and are generally not instantaneous events.

The likely direction of bias for the aforementioned factors will differ depending on which unobservable one is considering. If the unobservable is negatively correlated with wildfire occurrence and positively correlated with the economic outcome variables or vice versa, we expect a downward bias of the OLS estimate. For example, one might expect wildfires to be

reduced if a region implements effective fire prevention measures (e.g., mechanical clearing of land, fire breaks, grazing, educational campaigns), but at the same time employment may be increased as workers are needed to carry out these interventions. Likewise, urbanisation may coincide with a larger demand for fire services, which may also increase employment and hence reduce the BA as there are more locally available suppression resources. Turning to land-use changes, if the change is from forested to agricultural land (deforestation), this is likely to increase economic activity as it might be a more profitable use of land and would also decrease wildfires which generally occur in forested areas. A downward bias would also be observed for the case of political instability or corruption levels that lead to a potential increase in BA, e.g., around elections (Skouras and Christodoulakis, 2014) which could at the same time plausibly have a negative impact on economic activity.

One might also expect upward biased estimates if certain unobservables are taken into account that are positively correlated with both wildfire occurrence and employment/output. For example, a rural exodus would result in abandoned and unmanaged forests which potentially increases wildfire occurrence, but could also increase GDP growth if there are better job opportunities in regional economic centres. Furthermore, certain environmental regulations might create perverse incentives for arson (e.g., if the burned land can subsequently be used for cultivation or construction) so that the fire numbers potentially increase, quite possibly accompanied by the creation of employment and GDP growth if the land is repurposed towards more productive activities.

Finally, OLS estimates may suffer from classical measurement error, which leads to a bias

towards zero introduced by measuring the BA using satellite data that is arguably imperfect. More specifically, the data is based on a multi-step process, which means that the data is heavily reliant on working instruments on board the satellites at all times, but also on visual inspection and manual processing. Both aspects could thus lead to attenuation bias. Moreover, as described in Alix-Garcia and Millimet (2022) and Garcia and Heilmayr (2022) using satellite data can induce non-classical (systematic) measurement error. More specifically, as noted in Section 2.2.3, not all fires smaller than 30 hectares are detected by the satellite. Therefore, the true BA might be marginally larger than in our data set. While Alix-Garcia and Millimet (2022) provide a solution if the data derived by remotely sensed imagery is used as the dependent variable in a regression, we are not aware of an applicable strategy if the independent variable is based on satellite data as, it is in our case. This thus constitutes a limitation of our study.

Our empirical strategy is to isolate time-varying fire danger for the predominantly forested areas in the summer months which is arguably a good predictor for wildfire occurrence (first stage). As described in Section 2.2.4, the FWI is based on meteorological inputs and can thus be considered as good as randomly assigned. Moreover, by construction the FWI arguably only picks up the distinctive part of the meteorological factors indicative of wildfire danger. In order to ensure the satisfaction of the exclusion restriction we implement two vectors of control variables. First, we also include the FWI in predominantly urban, agricultural, and wetlands and water body areas of a region. Second, we control for a battery of other climatic variables i.e., temperature, precipitation, relative humidity and wind speed within

the region. Every climatic variable is created separately for each of the four land cover types and is included both as summer month averages and as annual averages. By including these additional control factors we are thus ensuring that our instrument is not capturing climatic factors affecting the non-predominantly-forested areas within a region that might affect economic activity other than through wildfire occurrence, such as, for example, through their impact on the agricultural sector (Damania et al., 2020).

Since the inputs into the FWI are also temperature, precipitation, relative humidity and wind speed, the FWI captures the remaining variation through the joint occurrence of specific threshold values and/or their non-linear transformations of these in its construction.[19] Moreover, some inputs into the FWI have a different time dimension. For example, the Drought Code input into the FWI has a 53-day time lag and thus differs temporally from the precipitation and temperature in the general climate controls employed. Furthermore, the precipitation amount is adjusted to slope effects of the landscape. The Duff Moisture Code and the Drought Code are also adjusted for the day-length of the month (e.g., to account for the dry rate and potential evapotranspiration from the soil following rainfall), and thus go beyond using pure meteorological inputs in their elaborated construction designed for capturing fire danger. Our identifying assumption is thus that the instrument isolates the specific meteorological aspects leading to a substantially higher wildfire occurrence probability for

[19] For example, the Buildup Index, which forms part of the fire behaviour indices capturing heat release, is constructed of non-linear functions depending on whether today's Duff Moisture Code (denoted as P) is below are above $0.4 *$ today's Drought Code (denoted as D) as shown in Equation (9) in Appendix B.1. Subsequently the Buildup Index is used to calculate an intermediate form of the FWI, namely the duff moisture function, $f(D)$, which is once again derived from a non-linear function. More specifically, the duff moisture function is calculated as $0.626 * \text{BUI}^{0.809} + 2$ if the Buildup Index is smaller or equal to 80, and as $1000(25 + 108.64e^{-0.023*\text{BUI}}$ if the Buildup Index is larger than 80 as shown in Equation (10) in Appendix B.1.

predominantly forested areas conditional on controlling for fire danger in other areas that are arguably much less flammable, as well as for general climatological conditions within each of these area types.

2.3.2 Instrument Construction

The FWI variable for forested area implemented as the instrument is created as follows. For the intersection of each region and FWI gridcell we tabulate the share of the four re-classified land cover types. To this end, we use the latest CLC version, i.e., the FWI years 2010-2012 are matched with CLC 2006 and the FWI years 2013-2018 are matched with CLC 2012. Each NUTS 3 region is then spatially joined with all the FWI gridcells that intersect with the region under the condition that the overlapping area is predominantly forested ($>$ 50%). Subsequently, the average daily value of all matching FWI gridcells for each region is calculated. Finally, the daily mean FWI value for June, July, and August is calculated for each region as wildfires are most common in the summer months.[20] In our sample, the average FWI value for predominantly forested areas in the summer months is approximately 16 (see Table 2.4), which is described as "Moderate Fire Danger" according to the EFFIS classification[21] based on Van Wagner and Pickett (1985).[22]

[20] https://climate.copernicus.eu/esotc/2020/wildfires (accessed in December 2021).

[21] https://gwis.jrc.ec.europa.eu/about-gwis/technical-background/fire-danger-forecast (accessed in August 2021).

[22] See Table B2 in Appendix B.2 for the complete classification of the FWI ranges.

2.3.3 Econometric Specification

We evaluate the potential impact of wildfires on two economic variables in first differences, namely on the growth of the employment rate Δlog(EMP) over $t-1$ to t defined as log(employed/total pop)$_{i,t}$ $-$ log(employed/total pop)$_{i,t-1}$, and on the GDP growth rate Δlog(GDP) over $t-1$ to t defined as log(GDP/capita)$_{i,t}$ $-$ log(GDP/capita)$_{i,t-1}$, where i represents a NUTS 3 region and $t = [2011, ..., 2018]$.

We estimate the following fixed effects 2SLS linear panel model instrumenting the fire impact variables with the FWI for predominantly forested areas:

$$\text{IMPACT}_{i,t-1\to t} = \beta_1 \text{FWI forest}_{i,t-1} + \mathbf{OFWI}_{i,t-1}\boldsymbol{\gamma}_1 + \mathbf{C}_{i,t-1}\boldsymbol{\delta}_1 + \pi_t + \mu_i + \varepsilon_{i,t} \tag{2.1}$$

$$\Delta\text{ECON}_{i,t-1\to t} = \beta_2 \widehat{\text{IMPACT}}_{i,t-1} + \mathbf{OFWI}_{i,t-1}\boldsymbol{\gamma}_2 + \mathbf{C}_{i,t-1}\boldsymbol{\delta}_2 + \pi_t + \mu_i + \varepsilon_{i,t}, \tag{2.2}$$

where IMPACT$_{i,t-1\to t}$ is a placeholder for the wildfire impact proxy variables, fire numbers or BA, FWI forest$_{i,t-1}$ is the daily mean Fire Weather Index in the summer months for predominantly forested areas, $\mathbf{OFWI}_{i,t-1}$ is a vector of the FWI for the other areas i.e., predominantly urban areas, rural areas, as well as for wetlands and water bodies in the summer months. Thereby, the FWI for the other land cover types are constructed similarly to the FWI for predominantly forested areas explained in Section 2.3.2. Moreover, $\mathbf{C}_{i,t-1}$ represents a vector of climatological controls including average summer and annual temperatures, precipitation, relative humidity, and wind speed. All the climatic controls are also implemented for each of the four land cover types separately and are created similarly to the land cover

specific FWI variables explained in Section 2.3.2. π_t and μ_i account for unobserved year and regional fixed effects, respectively, and $\epsilon_{i,t}$ are idiosyncratic errors. $\Delta\text{ECON}_{i,t-1\rightarrow t}$ is alternatively defined by either $\Delta\log(\text{EMP})_{i,t-1\rightarrow t}$ or $\Delta\log(\text{GDP})_{i,t-1\rightarrow t}$, and $\widehat{\text{IMPACT}}_{i,t-1}$ is the predicted value of the wildfire impact variables in Equation (2.1).

As our unit of analysis is at the regional level, one may worry about spatial correlation across regions. More specifically, the degree of economic integration between regions is likely to increase with geographical proximity and thus economic shocks may be spatially correlated. Hence we estimate our regression models with heteroskedasticity and autocorrelation consistent (HAC) standard errors that are robust to spatial correlation.[23] The necessary geospatial inputs for the estimation of spatial HAC standard errors are created using the longitude and latitude of the region's centroids. We choose a distance threshold for spatial correlation that corresponds to the radius of the NUTS 3 region's median area presuming a circular form of a unit, and add approximately 10% to this value which results in 33 km, in order to ensure that we include adjacent regions in the spatial correlation matrix.

We also explore whether there is a lagged impact of wildfires on regional economic outcomes by including up to $t-1-z$ lagged values of the IMPACT variable in Equation (2.2). Note that in terms of instrumenting for these lagged values we do not use the complete set of lagged FWI variables in a joint 2SLS estimation framework because it would not be appropriate to expect $t-1-z$ values of FWI to be predictors for $t-1-z+n, n=1,..N$ values

[23] To test spatial correlation, we use Moran's I introduced by Moran (1950) and proposed by Cliff and Ord (1972). We implement a row-standardised inverse distance weight matrix. The null hypothesis of uncorrelated residuals is rejected for all combinations for both dependent variables and years. Hence, we implement spatial HAC standard errors.

of IMPACT. Instead we estimate Equation (2.1), generate the predicted values $\widehat{\text{IMPACT}}_{i,t-1}$, and include these and lagged values thereof in Equation (2.2). However, as the contemporary and lagged values of the predicted $\widehat{\text{IMPACT}}_{i,t-1}$ will have their own distribution, using spatial HAC standard errors would no longer be appropriate. Were spatial correlation not an issue one could instead simply generate bootstrapped standard errors. Unfortunately, there is of date no accepted method to incorporate spatial correlation into standard bootstrapping procedures. We did experiment with 1,000 re-sampled data sets using 2, 3, and 10-fold cross validation which preserved the spatial error structure. Yet, this resulted in unreasonably small standard errors, as upholding the spatial structure led to limited variation among the data sets.[24] Our solution is thus instead to implement HAC bootstrapped standard errors (1,000 replications) in the lagged estimations without being able to take account of spatial correlation. Therefore, the lagged impact findings should be interpreted cautiously.

2.4 Results and Discussion

2.4.1 Contemporary Impact

In the first two columns of Table 2.5 we present the non-instrumented impact of our two wildfire proxies on employment growth, i.e., Equation (2.2) but with direct measures of IMPACT rather than their instrumented counterparts. The results suggest there has been no significant impact of wildfires on aggregate employment growth during our sample period.

[24] The standard errors for the 2-fold cross validation are about one quarter of the spatial HAC standard errors of the contemporary time period estimations.

Columns (3) and (4) indicate that the estimations of Equation (2.1) yield a strong first stage showing that the FWI for predominantly forested areas is a positive and statistically significant determinant of the wildfire impact variables at the 0.1 percent level for fire numbers and at the 5 percent level for BA. The positive effect meets a priori expectations since a higher fire danger index value arguably leads to more favourable conditions for both the outbreak and spread of wildfires. The effect size indicates an average increase of 0.4 fires per unit increase of the FWI (Table 2.5 Column (3)). An F-statistic of 37 for fire numbers indicates that a weak instrument problem can be excluded (Stock et al., 2002). Moreover, a unit increase of the FWI is associated with an increased share of BA of 0.007 percentage points (Table 2.5 Column (4)). The F test of joint significance in the first stage is 23 for BA, also indicating no weak instrument problem. Furthermore, the reduced form estimates displayed in Column (5) does not suggest an effect of the FWI for predominantly forested area on employment growth (Table 2.5).

The results of the IV estimations stated in Equation (2.2) show a positive insignificant effect of the wildfire impact proxy variables on the growth of the employment rate. Thus, like previous research conducted in the US by Nielsen-Pincus et al. (2013) who report a general positive effect of large wildfires on employment growth of approximately 1% during the quarter of the fire at the regional (county) level, we find a positive effect for Southern Europe. However, it is not significant possibly because our study differs in that (i) we look at annual vs. their quarterly data, and thus a potential seasonal effect would not be detected, (ii) we include all fires, and therefore evaluate an aggregate effect, while Nielsen-Pincus et al.

Table 2.5: Wildfires and employment growth (2011-2018).

	Δlog(EMP)		FIRE	BA	Δlog(EMP)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FIRE $\times 10^{-2}$	-0.005					0.044	
	(0.005)					(0.055)	
BA		0.040					0.240
		(0.046)					(0.307)
FWI forest			0.370***	0.067*	0.000		
			(0.107)	(0.027)	(0.000)		
Climate Ctrl	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FWI Ctrl	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.30	0.30	0.60	0.31	0.40	0.39	0.39
Model	OLS	OLS	1st stage	1st stage	Reduced	IV	IV

Notes: (i) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) spatial HAC standard errors in parentheses as implemented by Foreman (2020) with a distance cutoff value of 33 km; (iii) N = 1,864; (iv) the first stage F-statistics of 37 and 23 for fire numbers (Column (3)), and burned area (Column (4)), respectively, indicate that a weak instrument problem can be excluded (Stock et al., 2002); (v) Δlog(EMP) denotes the growth of the employment rate; FIRE is the annual number of fires (stated in 100 fires); BA is the proportion of the annual burned area per region; FWI forest denotes the mean of the daily Fire Weather Index in the summer months for predominantly forested areas; (vi) FWI controls include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; climate controls include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; fixed effects include regional and time fixed effects.

(2013) evaluate only large wildfire events,[25] and (iii) wildfires are on average much larger in the US than in Europe, and the resulting effect might thus be different.

Table 2.6 shows the effect of the wildfire proxy variables on GDP growth. The OLS results displayed in Columns (1) and (2) show a negative insignificant impact of wildfires on this economic activity indicator.[26] Unlike for aggregate employment, we find a negative significant effect of the FWI for predominantly forested areas at the 0.1 percent significance level in the reduced form (Table 2.6 Column (5)). Both wildfire impact variables show a

[25] Thereby, a wildfire is defined as large when suppression spending exceeds one million US$.

[26] Columns (3) and (4) which show the first stage are identical to Table 2.5 and are reported for completeness.

significant negative impact on GDP growth in the IV estimations (Columns (6) and (7)). One should note that the short-term negative GDP growth effects for wildfire affected regions found here are in line with the majority of the general natural disaster studies discussed in the introduction.

Table 2.6: Wildfires and GDP growth (2011-2018).

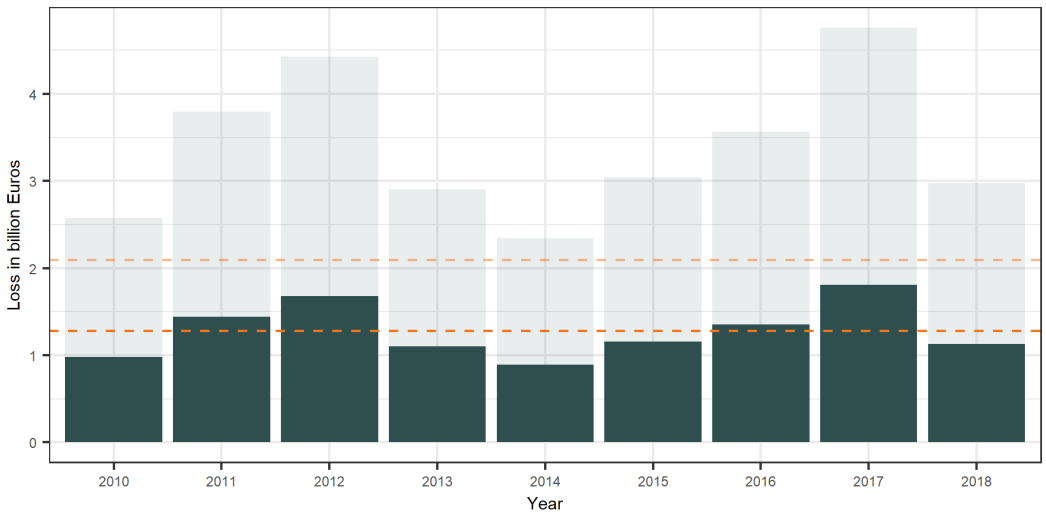
	Δlog(GDP)		FIRE	BA	Δlog(GDP)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FIRE $\times 10^{-2}$	-0.005					-0.259*	
	(0.006)					(0.107)	
BA		-0.001					-1.425*
		(0.036)					(0.687)
FWI forest			0.370***	0.067*	-0.001***		
			(0.107)	(0.027)	(0.000)		
Climate Ctrl	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FWI Ctrl	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.45	0.45	0.60	0.31	0.54	0.38	0.28
Model	OLS	OLS	1st stage	1st stage	Reduced	IV	IV

Notes: (i) * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) spatial HAC standard errors in parentheses as implemented by Foreman (2020) with a distance cutoff value of 33 km; (iii) N = 1,864; (iv) the first stage F-statistics of 37 and 23 for fire numbers (Column (3)), and BA (Column (4)), respectively, indicate that a weak instrument problem can be excluded (Stock et al., 2002); (v) Δlog(EMP) denotes the growth of the employment rate; FIRE is the annual number of fires (stated in 100 fires); BA is the proportion of the annual burned area per region; FWI forest denotes the mean of the daily Fire Weather Index in the summer months for predominantly forested areas; (vi) FWI controls include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; climate controls include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; fixed effects include regional and time fixed effects.

The point estimates on IMPACT of the IV specification in Table 2.6 indicate that, on average, an additional fire leads to a decrease in the regional annual GDP growth rate of 0.026% (Column (6)). As shown in Table 2.4, the mean wildfire number of the affected observations is 7 and thus the average wildfire affected region experiences a yearly decrease

in the GDP growth rate of 0.18% ($-0.00259 * 7 = -0.018$). The largest number of annual wildfire events in a region observed is 129. Therefore, for the most severely hit region in the "worst" observed year over our sample period this would lead to a decrease of the annual GDP growth rate of 3.3% ($-0.00259 * 129 = -0.33$). The wildfire proxy variable BA (Column (7)) is also positive and significant and suggests a decrease in a region's yearly GDP growth rate, on average, of 0.11% ($-1.425 * 0.0077 = -0.011$) conditional on having experienced at least one wildfire. Table 2.4 shows that for the most heavily affected region in the data set, the aggregated annual BA was 33.82%. In such an extreme year, the regional GDP growth rate is predicted to decrease by 4.8% ($-1.425 * 0.3382 = -0.48$).

Figure 2.4: Annual loss estimates based on a decrease in GDP growth for Southern Europe in billion euros (2010-2018).



Notes: (i) the solid bars indicate the lower bound and the transparent bars the upper bound of the estimate; (ii) the orange lines indicate the annual average losses (lower and upper bound) for the entire time period with an average of 102 wildfire affected regions per fire season.

To get a better understanding of what these changes in growth rates mean in monetary values we calculate annual average losses for Southern Europe. To this end we multiply our

estimated GDP growth effects with the mean GDP/capita value of 21,184 euros (as shown in Table 2.3) which implies an average loss of production of 23.3-38.1 euros/capita ($21,184 * 0.0011 = 23.3$ using BA and $21,184 * 0.0018 = 38.1$ using fire numbers). Subsequently, we multiply the average GDP/capita losses with the mean regional population of 538,000 to calculate the average loss in production for one affected region, which is 12.5-20.5 million euros. Figure 2.4 shows the monetary losses due to lost production for each year derived by the multiplication of the per region estimate with the number of wildfire affected regions, and therefore shows the variation related to the severity and intensity of the fire seasons. The average number of affected regions from 2010 to 2018 is 102, which suggests that losses are in the region of 1.3-2.1 billion euros for Southern Europe in a given year.

2.4.2 Employment Growth by Economic Activities

We next scrutinise the aggregate positive insignificant effect of wildfires on the growth of the employment rate in different economic activity categories to explore potential heterogeneous effects. To this end, the NACE economic activity sections are combined into six main categories as shown in Table 2.2.[27] The effects of categories A, B-E, and F are shown in Table 2.7, and the results of categories G-J, K-N, O-U are given in Table 2.8. Furthermore, the heterogeneous effects of wildfires on the growth of the employment rate by economic activity categories are visualised in Figure 2.5, showing the point estimates and the 95% confidence intervals for each category. The impact of wildfires on the growth of the employment

[27] One should note that there is no information for category A for 1 region.

rate in agriculture, forestry and fishing (Table 2.7 Columns (1) and (2)), on industries other than construction (category B-E) (Table 2.7 Columns (3) and (4)), as well as on construction (Table 2.7 Columns (5) and (6)) is positive but insignificant. Furthermore, the results show an insignificant negative effect on sector O-U, that is public administration and defence, compulsory social security, education, human health and social work activities, arts, entertainment and recreation, and repair of household goods and other services (Table 2.8 Columns (5) and (6)).

Table 2.7: Wildfires and employment growth for NACE activity categories A, BE, and F (2011-2018).

	Δ log(EMP^{A})		Δ log(EMP$^{B\text{-}E}$)		Δ log(EMP^{F})	
	(1)	(2)	(3)	(4)	(5)	(6)
FIRE $\times 10^{-2}$	0.069		0.050		0.002	
	(0.158)		(0.074)		(0.228)	
BA		0.380		0.273		0.010
		(0.866)		(0.422)		(1.257)
FWI Ctrl	Yes	Yes	Yes	Yes	Yes	Yes
Climate Ctrl	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,856	1,856	1,864	1,864	1,864	1,864

Notes: (i) * $p < 0.05$; (ii) spatial HAC standard errors in parentheses as implemented by Foreman (2020) with a distance cutoff value of 33 km; (iii) Δlog(EMP) denotes the growth in the employment rate; (iv) the superscript refers to the NACE activity where A includes agriculture, forestry and fishing, B-E is industry except construction, and F indicates construction; FIRE indicates the annual number of fires (stated in 100 fires); BA is the proportion of the annual burned area per region; (v) FWI controls include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; climate controls include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; fixed effects include regional and time fixed effects.

We find that two employment categories are significantly affected by wildfires. First, there is a negative effect of wildfires on the employment growth rate in sector G-J, which includes wholesale and retail trade, transport, accommodation and food service activities,

Table 2.8: Wildfires and employment growth for NACE activity categories G-J, K-N, and O-U (2011-2018).

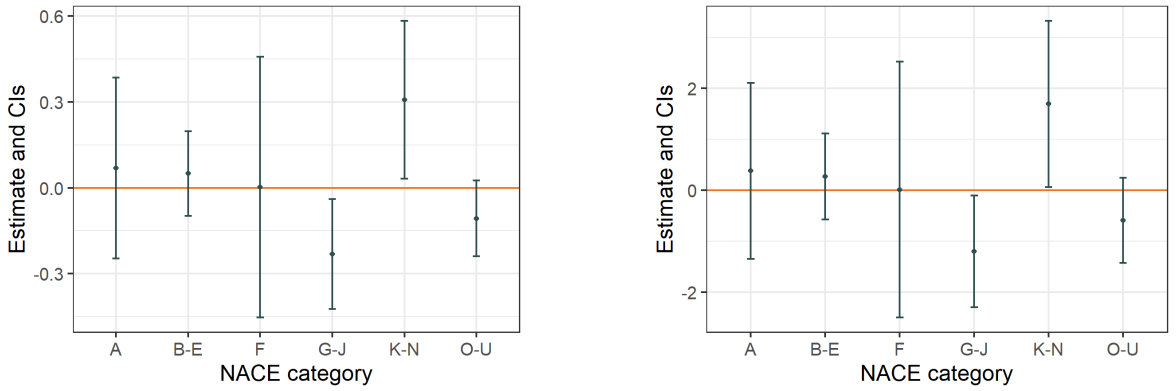
	Δ log(EMP$^{\text{G-J}}$)		Δ log(EMP$^{\text{K-N}}$)		Δ log(EMP$^{\text{O-U}}$)	
	(1)	(2)	(3)	(4)	(5)	(6)
FIRE $\times 10^{-2}$	-0.213*		0.308*		-0.107	
	(0.089)		(0.138)		(0.066)	
BA		-1.174*		1.695*		-0.591
		(0.519)		(0.817)		(0.417)
FWI Ctrl	Yes	Yes	Yes	Yes	Yes	Yes
Climate Ctrl	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	1,864	1,864	1,864	1,864	1,864	1,864

Notes: (i) * $p < 0.05$; (ii) spatial HAC standard errors in parentheses as implemented by Foreman (2020) with a distance cutoff value of 33 km; (iii) Δlog(EMP) denotes the growth in the employment rate; (iv) the superscript refers to the NACE activity where G-J includes wholesale and retail trade, transport, accommodation and food service activities, information and communication, K-N is contains financial and insurance activities, real estate activities, professional, scientific and technical activities, administrative and support service activities, and O-U includes public administration and defence, compulsory social security, education, human health and social work activities, arts, entertainment and recreation, and repair of household goods and other services; FIRE indicates the annual number of fires (stated in 100 fires); BA is the proportion of the annual burned area per region; (v) FWI controls include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; climate controls include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; fixed effects include regional and time fixed effects.

information and communication (Table 2.8 Columns (1) and (2)). This could indicate that employment activities related to retail and tourism (e.g., wholesale and retail trade; land, air, and water passenger transport; hotels, campgrounds, restaurants) are negatively affected. Once again, we multiply our estimates with the average fire numbers and BA which leads to a regional annual decrease in the rate of employment growth in category G-J of 0.09-0.15% ($-0.00213 * 7 = -0.015$ using fire numbers (Column (1)) and $-1.174 * 0.0077 = -0.009$ using BA (Column (2))) for wildfire affected regions.

We quantify the estimated results of wildfires on employment growth for the specific

Figure 2.5: Wildfires and employment growth by economic activity category (2011-2018).



(a) FIRE **(b)** BA

Notes: (i) the economic activity categories are defined following the Statistical classification of economic activities in the European Community abbreviated as NACE (see Table 2.2 for the full NACE economic activities classification); (ii) FIRE indicates the annual number of wildfires per region; BA in % denotes the annual burned area relative to the total area per region; (iii) CIs = confidence intervals.

activity categories in terms of job numbers to enhance the understanding of this magnitude. On average, 62,519 people (28.8% of the working population) are employed in the retail and tourism sections G-J per region as shown in Table B3 in Appendix B.2. Our estimates translate into 56-94 jobs lost per affected region annually ($62,519 * 0.0009 = 56$ using BA and $62,519 * 0.0015 = 94$ using fire numbers). With 102 regions that experience a wildfire in an average year this leads to a loss of 5,712-9,588 jobs for Southern Europe in the employment activity sectors including retail, transportation, as well as accommodation and food service activities.

Our findings concur with previous studies looking at recreational activities and tourism related to wildfires. For example, Kim and Jakus (2019) evaluate tourist flows in response to wildfires studying national park visits in Utah. The authors find a decrease in tourism in four out of five national parks and suggest an annual loss of 31-53 jobs based on the estimated loss

in labour income. Furthermore, Gellman et al. (2022) study the effect of wildfires and smoke exposure on more than 1,000 campgrounds in the western US showing that 1 million visitors per year are affected and estimate a decline in campground use. Evidence of wildfires affecting tourism-related industries in Southern Europe is provided by Molina et al. (2019) who estimate the economic susceptibility of recreation activities due to wildfires for the "Aracena y Picos de Aroche Natural Park" in Spain and show a susceptibility increase of 58 million euros due to travel and incidental costs. Moreover, Otrachshenko and Nunes (2022) estimate the effect of forest fires on tourist arrivals for 278 municipalities in Portugal and show that a 1% increase in BA in a given year reduces the tourist arrivals in that year by 3.5%.

The second significantly affected category, is the employment growth in NACE activity sections K-N, which include financial and insurance activities (e.g., risk and damage evaluation, financial leasing, reinsurance), real estate activities, professional, scientific and technical activities (e.g., legal and accounting services, architectural and engineering activities) as well as administrative and support service activities (e.g., renting and leasing of motor vehicles and construction machinery, temporary employment agencies activities, security and investigation activities, services to buildings and landscape activities). The magnitude of the effects indicate that wildfires lead to an increase in the regional annual employment growth in these sectors of 0.13-0.22% ($0.00308 * 7 = 0.022$) using fire numbers (Column (3)) and $1.695 * 0.0077 = 0.013$ using BA (Column (4))) conditional on a region having experienced at least one wildfire. The estimated positive employment effect in this NACE category seems sensible in response to wildfires, particularly as it incorporates insurance and damage eval-

uation, real estate activities, temporary employment activities (i.e., this includes short-term contracting possibly in the labour intensive construction sector or for additional firefighters), as well as services to buildings and landscapes activities which comprises cleaning of affected buildings and landscapes in the aftermaths of a wildfire.

On average 32,137 people (11.1%) work in sections K-N (see Table B3 in Appendix B.2) and thus a wildfire affected region experiences an annual increase of 42-71 jobs in this sector ($32,137 * 0.0013 = 42$ using BA and $32,137 * 0.0022 = 71$ using fire numbers). For an average fire season (i.e., 102 wildfire affected regions), this leads to an additional 4,284-7,242 jobs related to financial, insurance, real estate, as well as administrative and support service activities in Southern Europe.

2.4.3 Lagged Impact

Wildfires might have a more sustained effect on regional economies. Therefore, we explore whether there are lagged effects on regional employment and GDP growth of wildfires by including two lags of the wildfire impact variables in Equation (2.2). As noted earlier, the reported standard errors of all lagged estimations are not robust to spatial correlation, and thus must be interpreted relatively cautiously.

The results shown in Table 2.9 suggest consistently that it is only in the contemporary year that the GDP growth rate (Columns 4-6) is affected for both wildfire impact variables. In contrast, the results regarding the effect of BA on aggregate employment growth (and fire numbers for lag 1) indicate that there is a positive effect of the prior year. This would

Table 2.9: Wildfires and employment and GDP growth with lags (2011-2018).

	Δlog(EMP)			Δlog(GDP)		
	(1)	(2)	(3)	(4)	(5)	(6)
$FIRE_t \times 10^{-2}$	0.044	0.041	0.013	-0.259**	-0.259**	-0.222**
	(0.064)	(0.069)	(0.061)	(0.085)	(0.088)	(0.084)
$FIRE_{t-1} \times 10^{-2}$		0.089**	0.065		-0.009	-0.009
		(0.034)	(0.036)		(0.044)	(0.037)
$FIRE_{t-2} \times 10^{-2}$			0.057			-0.004
			(0.037)			(0.040)
BA_t	0.240	0.135	-0.040	-1.425**	-1.435**	-1.241**
	(0.354)	(0.382)	(0.347)	(0.468)	(0.477)	(0.444)
BA_{t-1}		0.524**	0.543**		0.052	0.171
		(0.193)	(0.207)		(0.260)	(0.210)
BA_{t-2}			0.208			0.130
			(0.208)			(0.191)

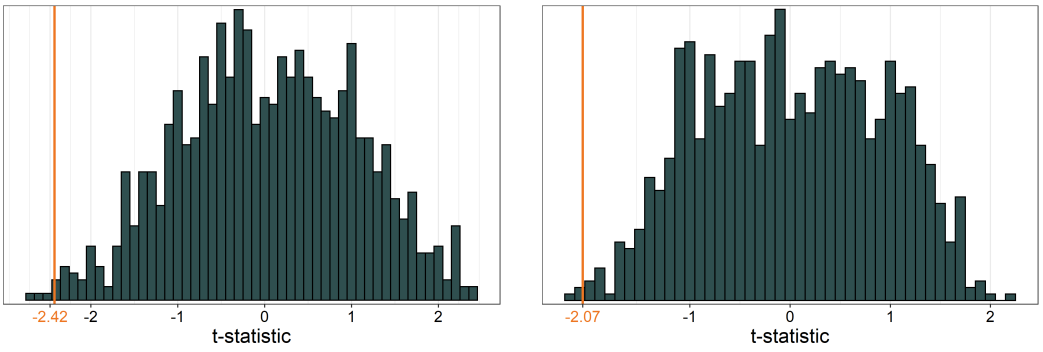
Notes: (i) * p $<$ 0.05, ** p $<$ 0.01; (ii) standard errors in parentheses are bootstrapped (1,000 replications) and clustered on the regional level; (iii) N = 1,864; (d) Δlog(EMP) denotes the growth in the employment rate and Δlog(GDP) is the GDP growth rate; FIRE$_t$ indicates the number of fires in year t (stated in 100 fires); BA$_t$ is the proportion of the annual burned area per region in year t; (iv) all estimations are run with FWI controls that include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; climate controls that include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; as well as fixed effects including regional and time fixed effects.

imply, subject to the concerns regarding the lack of spatial correlation taken account of in the standard errors, that a region's annual aggregate employment growth increases on average by 0.04-0.06% ($0.00089 * 7 = 0.006$ using fire numbers in Column (2), $0.524 * 0.0077 = 0.004$ using BA in Column (2), or $0.543 * 0.0077 = 0.004$ using BA in Column (3)) conditional on having experienced at least one wildfire. This aligns with recent research on the economic effects of natural disasters presented in Deryugina (2022) showing long-term labour market resilience for wealthy countries.

2.4.4 Robustness Checks

To test the robustness of our baseline estimations, we conduct the Fisher randomisation test introduced by Fisher (1937) for the estimates of the wildfire impact variables on GDP growth. We randomly reshuffle the fire numbers and BA across space and time (keeping the instrument and the control variables fixed) and run the IV regressions performing 1,000 iterations. The results displayed in Figure 2.6 show the high level of significance of our results (indicated by the t-statistic of the actual estimate) with a p-value of 0.004 for both fire numbers and for BA. This demonstrates that our results are not driven by chance.

Figure 2.6: Fisher randomisation test of wildfire impact variables and GDP growth with 1,000 iterations.

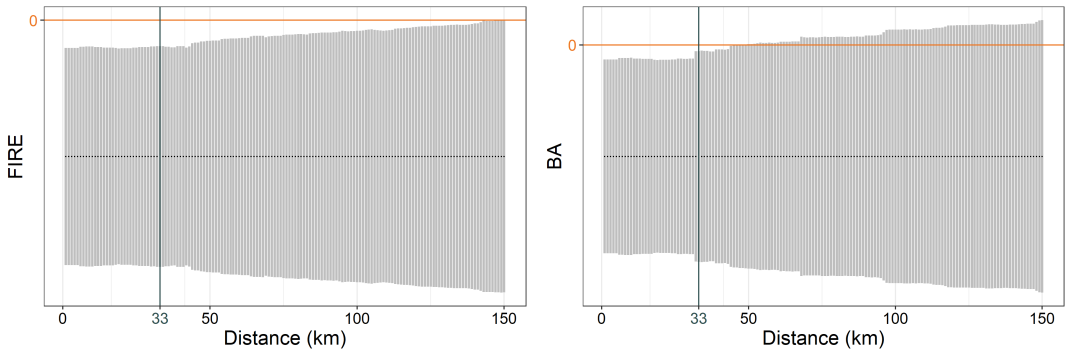


(a) FIRE **(b)** BA

Notes: (i) the vertical line indicates the t-statistic of our actual estimate; (ii) FIRE indicates the annual number of wildfires per region; BA in % denotes the annual burned area relative to the total area per region.

As described in Section 2.3, we choose 33km as the distance cutoff for the spatially robust HAC standard errors since this reflected the median distance between regions' centroids. To explore how sensitive our results are to this choice we incrementally increase the threshold and re-estimate Equation (2.2). Figure 2.7 shows that the spatial standard errors increase in

Figure 2.7: Spatial HAC standard errors with varying distance cutoff estimating the wildfire impact on GDP growth.



(a) FIRE **(b)** BA

Notes: (i) the vertical line indicates the selected cutoff value of 33 km; (ii) FIRE indicates the annual number of wildfires per region; BA in % denotes the annual burned area relative to the total area per region.

distance and become insignificant after we choose values of approximately 40 to 140 kilometres for the BA and fire numbers, respectively. Thus, our findings are only robust to assuming that potential regional economic shocks or spillover effects are limited to mostly adjacent regions.

To explore a potential economic impact beyond the directly affected BA we create buffers of one and five kilometres around each wildfire BA polygon. The underlying idea is to evaluate whether the effects extend to surrounding areas given those arguably suffer from indirect impacts (e.g., road closures, business downtime, decrease in tourism). Similar to the baseline estimations, we find significant negative effects of the wildfire impact variables on the GDP growth rate and insignificant positive effects on employment growth for the buffered estimations. The magnitude of the coefficient decreases with increasing buffer size as shown in Table 2.10.

Finally, one might be concerned that migration potentially impacts our findings. As ex-

Table 2.10: Wildfires and employment and GDP growth with buffered estimations (2011-2018).

	Direct	Buffer 1km	Buffer 5km
Δlog(EMP)			
FIRE $\times 10^{-2}$	0.044	0.016	0.001
BA	0.240	0.088	0.012
Δlog(GDP)			
FIRE $\times 10^{-2}$	-0.259*	-0.093*	-0.007*
BA	-1.425*	-0.522*	-0.070*

Notes: (i) * $p < 0.05$; (ii) direct effects incorporate the actual burned area; for potential effects beyond the burned area, buffers of size 1 and 5 km area created around each polygon; (iii) Δlog(EMP) denotes the growth of the employment rate; Δlog(GDP) is the per capita GDP growth rate; FIRE is annual number of fires per region (in 100 fires); BA is the proportion of the annual burned area per region; (iv) climate controls include summer and annual means of the variables temperature, precipitation, relative humidity, and wind speed for each of the four land cover type categories separately; FWI controls include the Fire Weather Index for predominantly urban and agricultural areas as well as for wetlands and water bodies; fixed effects include regional and time fixed effects.

plained in Section 2.3 we implement spatial standard errors, robust to various cutoffs between 40 to 140 kilometres, in our main estimations. This would take account of migration into neighbouring regions as long as these act through shocks captured in the error term. We additionally run our specification in Equation (2.2) but using regional population growth as the dependent variable. More precisely, as long as births and deaths are not directly affected by wildfires, or their effects cancel each other out, or in net are less than any effect on migration, then any impact on population growth can be considered to be due to net migration. However, the results of this exercise showed that neither the reduced form (coefficient 0.00; standard error 0.00) nor the IV estimates for fire numbers (coefficient 0.007; standard error 0.008) and BA (coefficient 0.039; standard error 0.045) were significant. Thus, either there is no effect on net migration or the effect is cancelled out by impacts of the birth net of the death rate.

Nevertheless, we need to emphasise that our study estimates the aggregated effects of all

wildfire occurrences per region and year, most of which are arguably not disastrous. This is in strong contrast to the existing migration literature in the natural hazards context that sets a focus on large-scale devastating events (Karácsonyi et al., 2021).[28] For example, Sheldon and Zhan (2022) find county out-migration following hurricanes and floods using Federal Emergency Management Agency (FEMA) data. Regarding wildfires, Winkler and Rouleau (2021) suggest that extreme wildfires (as declared by FEMA) in the US may be associated with out-migration. Similarly, Boustan et al. (2020) report increased out-migration for severe fire events using FEMA data from 1920 to 2010 in the US, although it needs to be pointed out that in their specification a severe fire event is associated with at least 25 mortalities. Out of the more than 6,000 wildfires in our dataset, less than a hand full would be categorised accordingly.

There are reasons why residents are unlikely to migrate after being affected by a wildfire. For instance, areas that are fire-prone often simultaneously draw people due to their intrinsic environmental amenities, as extensively outlined in McConnell et al. (2021). More precisely, even for large and devastating wildfires, the negative impact may not be big enough to outweigh the amenity-draw to that very place. Evidence in line with this notion shows that even for disastrous wildfire events, such as the California's 2017 North Bay fires which resulted in more than 6,000 structures damaged or destroyed, a small minority of affected households moved out of the county (Sharygin, 2021). Finally, a recent study by McConnell et al. (2021) investigates fires that are known to have destroyed at least one building, i.e., 16% of fires in

[28] We hereby focus on studies looking at revealed preferences and do not discuss publications studying stated preferences (e.g., surveys quantifying the intention to move after severe wildfire events) such as Nawrotzki et al. (2014) or Berlin Rubin and Wong-Parodi (2022).

their sample, and find in- as well as out-migration for the entire sample and an increase in out-migration for fires that destroyed more than 17 buildings. The authors state as a broader conclusion that residents largely remain in fire-prone regions after less destructive events, which per definition are arguably more destructive than the majority of fires in our data set.

2.5 Conclusions

In this paper we link high resolution satellite data of aggregate wildfire burned areas with regional economic data for Southern Europe, enabling us to contribute to a deeper understanding of how these events impact local economic outcomes. Given that wildfire incidents are likely correlated with various unobservable factors, such as land management policies, wildfire prevention strategies, and land-use changes, and can be set intentionally, the events are treated as endogenous in our analysis. To overcome this concern we use a measure of wildfire occurrence probability for predominantly forested areas based on relevant climatic features as an instrumental variable, while controlling for fire danger in non-forested area as well as for general climatic conditions that might directly affect regional economies. Importantly, the analysis indicates that not taking account of the endogeneity of wildfires is likely to lead to biased estimates on economic impacts. The proposed instrumental variable strategy might thus also prove to be a useful approach for other researchers interested in the economic implications of wildfires.

Our causally identified results for Southern Europe show a consistent negative contemporary effect of wildfires on the annual regional GDP growth rate ranging from 0.11 to 0.18%.

For the most severe wildfire years, the effect can lead to a decrease in the GDP growth rate of approximately 3.3-4.8% using fire numbers and burned area, respectively. The disaggregated employment analysis by economic activity categories reveals heterogeneous impacts, where industries such as wholesale and retail trade, transport, accommodation and food service activities are experiencing a negative employment effect of 0.09-0.15%, plausibly as a result of disruptions related to tourism. In contrast, our results show a positive effect of wildfires on regional employment growth of 0.13-0.22% in sectors including financial, insurance, and real estate activities, as well as short-term contracting activities.

Overall our study provides novel evidence that wildfires lead to a significant decrease in the regional GDP growth rate for Southern Europe. Although wildfires have formed an integral part of the Mediterranean landscapes for centuries, the public institutional response could benefit from an extensive evaluation of mitigation and prevention mechanisms (e.g., mechanical clearing, prescribed burning, grazing, land management activities) to reduce the negative impacts on local economies. As illustrated in Bayham et al. (2022), economic interdependencies and inefficiencies in fire-prone landscapes render wildfire management highly complex and large research gaps remain. European wide data collection efforts on these aspects at the regional level would allow researchers to further investigate the possible role of these interventionist factors. Such insights would importantly allow regional policy makers to explicitly evaluate strategies to strengthen the resilience of regional economies, particularly since the potential damage of wildfires is predicted to become more pronounced in the future (Dupuy et al., 2020).

Chapter 3

Impacts of wildfire smoke exposure on excess mortality and later-life socioeconomic outcomes: The Great Fire of 1910

Abstract

The Great Fire of 1910 in the Northwestern United States remains one of the largest wildfires on record burning more than 1.2 million hectares in just over two days and had a pivotal role in shaping the fire suppression regime that dominated fire management policy for decades. This paper studies the impact of fire-sourced smoke pollution on excess mortality and later-life socioeconomic outcomes of children under the age of five for this event. To this end, using historical burn perimeters we employ wildfire smoke emission and dispersion tools to model pollution exposure and assemble a county-level weekly panel data set combining mortality data with full-count census data from 1900 to 1940. Our Two-Way Fixed Effects estimates suggest a positive effect of smoke exposure on excess mortality in the week of the fire. Furthermore, we find weaker evidence that smoke exposure in early childhood is associated with a decrease in some later-life socioeconomic status indices based on education and income in 1930 although the impact had disappeared by 1940.

Keywords: wildfire, air pollution, health effects, socioeconomic outcomes

JEL classification codes: I1, N3, N5, Q5

3.1 Introduction

Over the past few decades, wildfire seasons have exhibited a discernible pattern of increasing duration and intensity in many regions across the world (Bowman et al., 2020; UNEP, 2022). The impacts of these fires extend far beyond the immediate burn zone, as released pollutants can travel thousands of kilometres (Sapkota et al., 2005; Kollanus et al., 2017), imposing a substantial burden on human, economic, and environmental systems in the surrounding and downwind areas. Burke et al. (2023) show that wildfire smoke has eroded 23% of the gains in improved air quality in much of the United States (US) since 2016 and increasing wildfire-sourced particulate matter (PM) 2.5 are projected to offset air quality improvements due to decreasing anthropogenic emissions in the Southwestern US by the mid 21st century (Ford et al., 2018).[1] These recent developments and future projections in air quality degradation ultimately pose substantial harm to human health, particularly as recent toxicological studies suggest that the chemical composition of wildfire-specific PM2.5 emissions may be more harmful than equal doses of ambient PM2.5 emissions (Aguilera et al., 2021).[2]

This study exploits the Great Fire of 1910 in the Northwestern US to empirically evaluate the causal effect of smoke exposure on health and human capital outcomes. This extreme

[1] PM2.5 refers to particles that are 2.5 microns or less in diameter.

[2] Wildfire PM primarily consists of carbonaceous material, with at least 50% organic carbon. It exhibits a higher oxidative potential compared to ambient urban PM due to the presence of these polar organic compounds. Oxidative potential refers to a substance's ability to induce oxidative stress, which can lead to inflammation within the body (Verma et al., 2009). In addition, the smoke generated by wildfires is characterised by a greater concentration of ultra-fine particles compared to typical urban air pollution, increasing the likelihood of deeper deposition within the respiratory system (Schöllnberger et al., 2002).

wildfire event not only remains one of the largest fires in US history engulfing more than 1.2 million hectares in just over two days, but also had a pivotal role in shaping fire management policy for decades initiating a rigorous wildfire prevention and suppression regime that lasted for much of the 20th century (Van Wagtendonk, 2007) and unknowingly paved the path for more extreme wildfires by creating "fire deficits" (Marlon et al., 2012; Steel et al., 2015). Importantly, at the time of the Great Fire, the potential short and long term health effects of wildfire smoke exposure appear to have been partially undervalued and considered only of secondary concern compared to the direct economic consequences, thus likely minimising avoidance behaviour compared to that which is found in modern settings (Burke et al., 2022).[3] We use this large-scale historical event to not only assess the immediate health impact but also evaluate whether potential adverse effects persist over time.

Fire-sourced air pollution and health outcomes have been at the centre of a rich body of literature. Smoke exposure has previously been associated with negative physiological (e.g., respiratory and cardiovascular morbidity and mortality) and psychological (e.g. post-traumatic stress disorder, depression) health effects in the general population; for extensive reviews see Liu et al. (2015); Chen et al. (2021); Grant and Runkle (2022); Gao et al. (2023). We focus here on young children, defined as those under the age of five, as they have been shown to be particularly vulnerable to adverse health effects due to their ongoing development, higher respiratory intake relative to body weight, and incomplete physiological barriers

[3] Fowler et al. (2020) describe the chronology of global air quality stating that air pollution has been recognised as a threat to human health since the year 400 before the common era. The authors mention that people would have felt discomfort by emissions close to fires inside shelters and that many societies preferred open fires or chimneys in the 1800s. Yet, besides that example of the use of fire for living purposes, the concerns around air pollution are related to coal burning, urban centres and industry.

(Holm et al., 2021). These barriers, including the nasal, gut, and lung epithelial barriers, are not fully developed in children, allowing a larger proportion of harmful particles to penetrate deep into their lungs (Bennett et al., 2007). In addition, the incomplete blood-brain barrier can lead to extensive neuroinflammation and subsequent cell loss in the central nervous system, potentially resulting in cognitive deficits (Brockmeyer and D'Angiulli, 2016). An extensive review on the association between wildfire smoke and health effects in children specifically is provided by Holm et al. (2021). The authors point out that the most robust literature revolves around respiratory effects, oftentimes measured through emergency department visits or hospital admissions (Stowell et al., 2019; Pratt et al., 2019; Ye et al., 2021).

The first part of our study assesses the effect of smoke exposure on mortality and thus focuses on a health outcome that can be measured with higher certainty compared to morbidity measures, given its unambiguous definition (Acosta and Irizarry, 2022). In this regard, previous work by Doubleday et al. (2020) suggest a possible effect of wildfire smoke exposure on respiratory deaths in young children in Washington, while Xue et al. (2021) find that each $\mu g/m^3$ increment of fire-sourced PM2.5 is associated with a 2.31% increased risk of child mortality for low-income and middle-income countries. Furthermore, evidence of the 1997 Indonesian wildfires on early-life mortality suggests 15,600 "missing children" due to decreased air quality, driven by prenatal exposure (Jayachandran, 2009).

The second part of our study aims to evaluate the potential persistent effects of early-life smoke exposure on later-life socioeconomic outcomes. Evaluating later-life physical and physiological outcomes, Rosales-Rueda and Triyana (2019) analyse the pollution induced by

the 1997 Indonesian forest fires and report that on average pollution-exposed children have a shorter stature three years after exposure, and exhibit a lower lung capacity 10 years post-exposure. However, only children who were smoke-exposed in-utero are found to be shorter 10 and 17 years post-exposure. While the literature on the effects of urban and industrial-sourced air pollution on cognitive performance or human capital formation in children is growing, wildfire-smoke-specific research is scarce. One exception is Wen and Burke (2022) who identify that a higher cumulative annual wildfire smoke exposure is associated with lower test scores for children in the US.[4]

A study on the impact of wildfires is subject to a number of endogeneity concerns. For example, unlike other natural hazards such as earthquakes or tornadoes, it is hard to argue that wildfires are completely exogenous shocks as their occurrence and intensity depends on land management choices and are oftentimes human-induced which means wildfire events may be correlated with local socioeconomic characteristics. To overcome these endogeneity concerns, we use a historical burned area map and model the smoke emission and dispersion to reconstruct the wildfire smoke exposure at the county level using meteorological inputs that are arguably exogenous to our outcome variables of interest. For the first part of the analysis, we combine the air quality exposure measure at the county level with digitised death records obtained from Ancestry.com and the Integrated Public Use Microdata Series (IPUMS) complete count 1900 and 1910 censuses to assemble a weekly county-level panel data set and employ a Difference-in-Differences design to study short-term weekly excess

[4] Cleland et al. (2022) show that cognitive performance in a brain-training game is negatively associated with short-term wildfire-sourced smoke exposure. However, this study is on adults only.

mortality. In the second part of this study, we link boys who were under the age of five in 1910 over time using IPUMS 1910 to 1940 full-count individual and household data in order to evaluate their socioeconomic status outcomes 20-30 years later and compare men who were smoke-exposed in their early childhood to men who were not conditional on initial endowments and characteristics.

The results of the first part of this study suggest an immediate impact of smoke exposure on excess mortality for children under the age of five in the week of the wildfire event. More specifically, the excess mortality rate is 23.5 per 100,000 which signifies an increase of $\sim$56% relative to the average observed mortality rate in 1910 across the entire study area. In the second part of the analysis, we find weaker evidence of a negative effect of wildfire-sourced smoke exposure in early childhood on some later-life socioeconomic status indicators that are based on occupational and educational performance 20 years after the Great Fire of 1910. However, this effect seems to disappear 30 years after the event.

This paper provides a number of contributions to the current literature on the health impacts of air pollution. First, existing studies have almost exclusively been limited to modern settings despite polluted air being considered a problem from medieval times (Brimblecombe, 2011). Research that has an historical setting includes Heblich et al. (2021) who show that the impact of air pollution during the industrial revolution still persist today.[5] In other studies, Beach and Hanlon (2018) use information on local coal use and wind patterns to show that exposure to air pollution in the 1850s increased infant mortality rates across England and

[5] Heblich et al. (2021) demonstrate that the persistent east-west divide in neighbourhood sorting in 2011 is partly a result of being exposed to different levels of air pollution during the industrial revolution.

Wales by between 6 and 8 per cent while Bailey et al. (2018) show that local coal intensity exposure of WWI enlisted English and Welsh men during childhood has longer term health effects by reducing adulthood height. We add to this nascent literature by exploring for the first time the impact of another source of air pollution, namely wildfires, in a historical context, but also looking jointly at both short and long term impacts. For the latter, in contrast to Bailey et al. (2018), we specifically look at the socioeconomic implications for adults of early life air pollution exposure which has not yet been explored in a historical context nor within the setting of wildfires in general.

The remainder of this paper is structured as follows. Section 3.2 presents the historical background and the unfolding of the Great Fire of 1910. Section 3.3 provides details of the data sets used followed by Section 3.4 which explains the empirical framework. In Section 3.5 we present the results and robustness checks and Section 3.6 concludes.

3.2 Historical Background

The Great Fire of 1910, commonly known as the Big Burn or the Big Blowup, was a catastrophic event that occurred in a rugged and remote geographical region characterised by towering mountains which form part of the Northern Rocky Mountain chain, dense forests, pristine alpine lakes, and thriving wildlife. For millennia, indigenous peoples had inhabited the area spanning Idaho, Montana, and eastern Washington, while in the late 19th century, European settlers began to migrate westward, working in mining, logging, and the construction of the Northern Pacific Railway. Conflict was commonplace including a struggle for

land and resources among the various groups of settlers and with the native nations. In addition, societal upheaval such as segregation and internal disputes among European immigrants contributed to the turbulent atmosphere of the time.[6]

In the years leading up to the Big Burn, Theodore Roosevelt served as the President of the United States and was a strong advocate for nature conservation. Under his leadership, 150 national forests, 51 federal bird reserves, and 5 national parks were established. The US Forest Service, which was founded in 1905, employed mostly young Yale college graduates in forestry to work as rangers in the newly established national forests. However, their presence was met with resistance from some local communities who feared the loss of profitable land. The responsibilities of forest rangers included constructing trails, installing telephone lines, building cabins, and extinguishing small fires. Despite the US Forest Service's controversial reputation, some men, such as Ed Pulaski, who joined the service as a ranger in 1908, transitioned from mining and railroad jobs to working as a ranger. Pulaski later gained fame for his heroism during the Big Burn by leading his 45-man crew to safety in an old mining tunnel.

Throughout this period, fires were a common occurrence in the region, but the summer of 1910 proved particularly disastrous. While the native population had adapted to the regularity of wildfires, the new settlers were largely unfamiliar with the potentially devastating consequences of forest fires. Fires that year were reported as early as April and were often started by mine operators, coal-burning trains, or lightning strikes. In the spring and summer of 1910 warm weather conditions were highly anomalous as shown in Diaz and Swetnam (2013), ex-

[6] The information presented in this section is primarily derived from Egan (2009), unless otherwise specified.

acerbating concerns over safety. In the town of Wallace, Idaho, forest supervisor William Weigl faced mounting pressure to enhance fire control vigilance. However, the understaffed Forest Service struggled to maintain operations. While the position of a forest ranger offered better pay than mining work (approximately $15 as opposed to $13 per 60-hour week), the job as a ranger was unpopular due to its perceived dangers and ideological conflicts over conservation versus resource extraction. Rumours even circulated that the government had passed legislation to pay rangers only for completed work, making it even less attractive to work as a ranger, in an attempt to dismantle the organisation. This meant that due to the lack of funds for the Forest Service, rangers oftentimes had to wait weeks to be paid. In a desperate attempt to gain control over the situation, Weigl called a town meeting to recruit additional help, but unfortunately, no volunteers came forward.

On 26 July 1910, a lightning storm ignited numerous fires and prompted the Forest Service to request additional firefighters. As the situation worsened, all available men were dispatched to the fires, even extending the search to Taft, Idaho, a notorious, lawless mining town referred to as the "wickedest city in America", infamous for its high murder rate, heavy drinking, and brothels. In early August, 60 prisoners, including a murderer and a bank robber who remained handcuffed, were released and sent to the fire lines. Shortly thereafter, President William Taft authorised the deployment of 2,500 military troops, including the 25th Infantry, also known as "Buffalo soldiers", based in Spokane, Washington. These African-American troops, primarily from the South, had been tasked with a variety of duties including battling Native Americans, quelling civil unrest, catching thieves, and protecting

settlers while also building roads. The 25th infantry were dispatched to the towns of Wallace and Avery, Idaho, and to Missoula, Montana to fight on the fire lines along many others with no experience or training.

While it is challenging to find evidence on the beliefs or awareness of potential negative effects of wildfire smoke on health at the time, historical records suggest that the main concerns around wildfires were related to the supply of timber as well as the role of forests in flood prevention. A leaflet by James Wilson, the United States Secretary of Agriculture, printed in 1900 states the following: "The great annual destruction of forests by fire is an injury to all persons and industries. The welfare of every community is dependent upon a cheap and plentiful supply of timber, and a forest cover is the most effective means of preventing floods and maintaining a regular flow of streams used for irrigation and other useful purposes (Miller and Cohen, 2021, p. 8)." Although Egan (2009, p. 141) writes about the symptoms that the residents in the town of Wallace, Idaho experienced due to the heavy smoke in the buildup to the fire, they are mentioned as "low-grade tortures." The passage reads: "People could tolerate the ever-present smoke, though it wasn't good for children and the elderly, made eyes redden and throats scratchy and brought on a ragged cough ... They put up with these low-grade tortures because shorter days told them summer was almost over, and they had lived through a humdinger, and soon the rains would come and wash the town clean."

On the 19th of August some embers hit Wallace, and although they were extinguished before setting any structures on fire, the first people started to pack their belongings (Krainz, 2012). On the 20th of August, strong hurricane-force Palouser winds from the southwest

picked up and merged a number of small fires into one large, all-consuming conflagration. The Wallace town Major ordered an evacuation by train by around midnight that day. One should note, however, that the evacuations were carried out fairly late and only moved people out of the immediately threatened burn zone so many evacuees were transported to regions that were within the heavily smoke-affected area (e.g., Missoula), and were arguably exposed before and during the evacuation. Moreover, although the population rushed to the trains only women and children were given space with the Buffalo soldiers tasked with throwing men off the trains. The population of the mining town Taft, Idaho, had refused to evacuate and were intoxicated by the time the fire arrived, which resulted in panic, and the entire town was destroyed. Meanwhile, Ed Pulaski's 45-man firefighting crew was surrounded and trapped by the flames, and sought shelter in an old mine tunnel with 40 of the men surviving and making it back to Wallace before being immediately sent to the hospital that had not burned down. The fire was finally extinguished on the 22nd of August when a cold front swept over the Northern Rocky Mountains, bringing steady rain and some early snowfall. Overall, the fire killed 87 people, most of whom were firefighters, burned five towns to the ground, partly destroyed many others, and burned over 3 million acres ($\sim$1.2 million hectares) of forest.[7]

One can use newspaper reports during the course of the event to get an idea of how the fire was perceived at the time.[8] For instance, The Missoulian (Missoula, Montana) reported on the 21st of August in a short notice that "The pall of smoke overhanging the town was so

[7] There were no children among the recorded casualties.

[8] All newspaper articles referenced within this study are obtained from the website https://www.newspapers.com. We specify the newspaper as well as the date of publication for each source individually within the text.

dense that the electric lights were turned on at 3 o'clock in the afternoon." One day later, the Great Falls Tribune (Great Falls, Montana) mentions the wildfire smoke a number of times. On a general note, it is stated that "A dense pall of smoke hangs all over western Montana. In Missoula it was as dark as midnight at 5 o'clock, the dense smoke being given a lurid hue, which had all the semblance of the glow of fire, but which was probably due to the sun." For Great Falls, a town that was, similar to Missoula, spared by the flames, the column reads "The atmosphere in this vicinity has been heavily charged with smoke for the past twenty-four hours and at a late hour this morning it was almost impossible to distinguish objects in any portion of the city for a distance of three blocks." Even the cause of a derailed train was partly attributed to poor visibility mentioning that "Train No.2 on the Great Northern was wrecked one and a half miles west of Rudyard ... The cause of the wreck was attributed partly to the dense smoke which prevailed last night and today. The sun being entirely obscured and it being only possible to see a few hundred feet ahead." However, we find no reference to possible health concerns related to the event.

The catastrophic Great Fire of 1910 served as a critical turning point in the history of the US Forest Service. The Service faced considerable criticism for its emphasis on conservation rather than logging, with many believing that more active forest management practices would have prevented such a calamitous event. Nevertheless, Theodore Roosevelt used the tragedy as a call to action, resulting in the provision of much-needed funding and resources to the Forest Service to prevent similar disasters in the future. This event was thus instrumental in solidifying support for the Forest Service's fire management mission and led to the

establishment of a vigorous fire suppression regime. In 1935, the 10 am rule was introduced, which mandated that every fire must be extinguished by 10 am the following morning. The mission of relentless fire prevention and suppression was consolidated by the invention of "Smokey Bear" in 1944, a symbol for the joint effort to promote forest fire prevention. According to Van Wagtendonk (2007) fire suppression was the only fire policy implemented by the federal land management agencies and it was not until 1974 that the US Forest Service transitioned to a fire management regime allowing some fires ignited by lightning to burn in specific wilderness areas.

The legacy of the Big Burn bore the imprint of an exceedingly stringent fire suppression policy which proved effective in preventing major wildfire disasters for much of the 20th century. However, it also arguably unintentionally led to the development of denser and less diverse forests as well as a "fire deficit" as discussed by Marlon et al. (2012), which ultimately and unknowingly set the stage for even more devastating fires at the end of the 20th and the beginning of the 21st century (Steel et al., 2015).

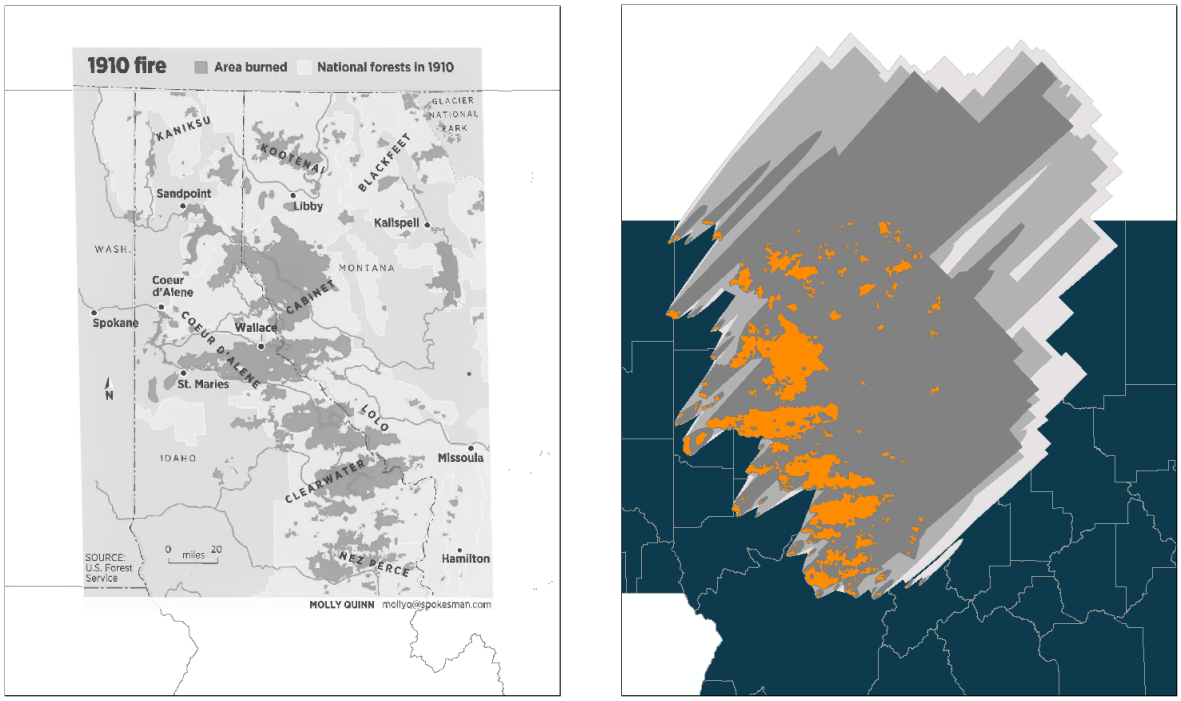
3.3 Data and Descriptive Statistics

3.3.1 Burn Perimeters

A historical map of the Great Fire of 1910's burn perimeters created by the US Forest Service is accessible from a number of historical sources. We obtain the required data from

the Spokesman Review.[9] The digital raster image is georeferenced using the shapefile of the US states provided by the US Census bureau.[10] Figure 3.1a shows how the image of the historical burned area is matched to a spatial reference system using the state boundaries. We then delineate each fire scar individually and create a shapefile consisting of 176 burned area polygons.

Figure 3.1: Historical map of the burn perimeters and the resulting modelled smoke plumes.



(a) Historical burn perimeters

(b) Modelled smoke plumes

Notes: (i) the historical map shown in panel Figure 3.1a created by the US Forest Service shows the burn perimeters of the 1910 fires and is georeferenced using ArcGIS; (ii) the county shape file is provided by the Big Ten Academic Alliance Geoportal and shows the historical county boundaries in 1910; (iii) the merged smoke plumes shown in Figure 3.1b are modelled using BlueSky Playground version 3.5.1; (iv) in Figure 3.1b the orange area shows the burn perimeters. The darkest grey shaded area indicates the area where the hourly peak PM2.5 pollution is hazardous (PM2.5>526 μg/m^3). The medium grey area shows unhealthy hourly peak PM2.5 pollution of the values (PM2.5>130 μg/m^3) and the lightest grey scale denotes moderate hourly peak pollution (PM2.5>38 μg/m^3).

[9] https://www.spokesman.com/stories/2010/aug/15/1910-fire-region-consumed/.

[10] https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html.

3.3.2 Smoke Modelling

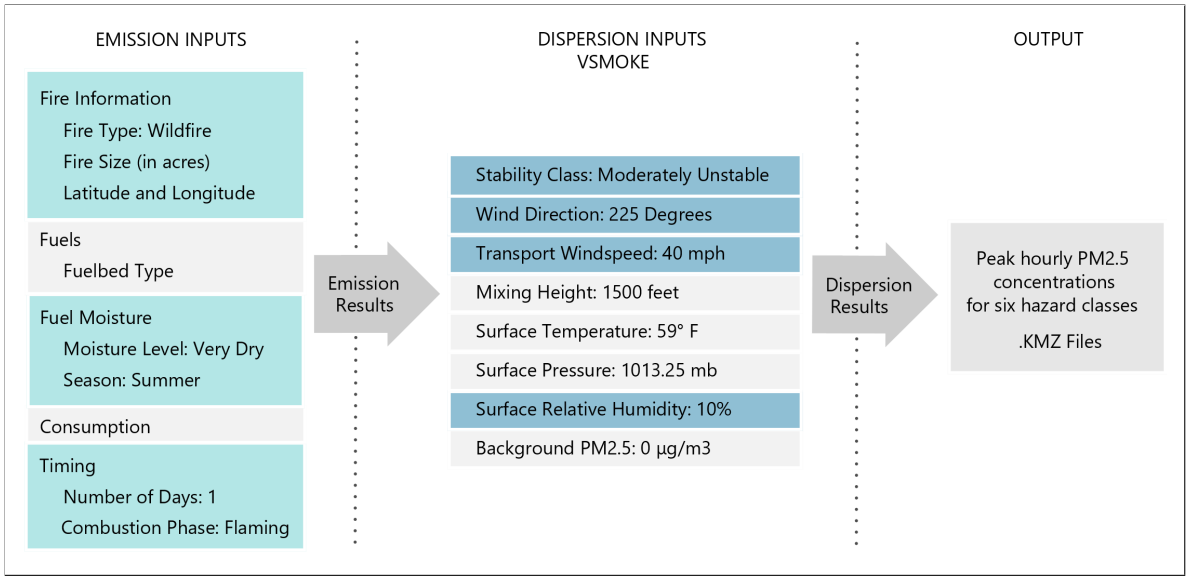
The BlueSky modelling framework is used to simulate smoke emission and dispersion of the fire. The framework is specifically designed to facilitate the modelling of the impact of cumulative smoke resulting from different types of fires, such as wildland, agricultural, and prescribed fires. Known for its high modularity, the BlueSky modelling framework connects advanced models in the fields of fuels, consumption, emission, meteorology, and air quality, and allows for multiple options for each stage of the modelling process.[11] A series of sequential processing steps, either containing individual component models or data sets are interconnected starting with information on the fire and fuel loads, advancing to fuel consumption, and concluding with smoke emissions and transport. Therefore, unlike in the sense of traditional models, BlueSky operates as a modular framework integrating existing models and data sets within a cohesive structure (Larkin et al., 2009). Modelling simulations are conducted using version 3.5.1 of the interactive online platform BlueSky Playground.[12]

The smoke modelling has to be carried out individually for each of the 176 burned area polygons which makes the process computationally demanding. Figure 3.2 provides a schematic overview of the Bluesky modelling framework. The highlighted parts indicate where a value

[11] The fuels information datasets combines the Fuels Characteristic Classification System (FCCS) by the U.S. Forest Service Fire and Environmental Research Applications (FERA) team under the lead of Don McKenzie and LANDFIRE by the U.S. Forest Service Missoula Fire Lab. The consumption model CONSUME is developed by the U.S. Forest Service FERA team including Roger Ottmar, Susan Prichart, and Gary Anderson. The emission factor model embedded in the BlueSky framework is based on the work by Prichard et al. (2020) and the VSMOKE-GIS dispersion model was originally introduced by Harms and Lavdas (1997) and further developed by Scott Godrick.

[12] BlueSky Playground has been developed by Sonoma Technology Inc. and is publicly accessible under https://tools.airfire.org/playground/v3.5/emissionsinputs.php. It has been supported by the US Forest Service, the National Fire Plan, the Joint Fire Sciences Program, NASA, and the Department of the Interior.

Figure 3.2: Schematic overview of the smoke emission and dispersion process using the BlueSky modelling framework.



Notes: (i) the above model shows the conceptual framework of the BlueSky modelling tool divided into the emission and dispersion inputs; (ii) the highlighted parts of the framework indicate where value supplied by the user is strictly required for the framework to run or where our input choices deviate from the default settings; (iii) F = Fahrenheit, PM = particulate matter, mb = millibars, $\mu g/m^3$ = micro-grams per cubic meter, KMZ = Keyhole Markup Language Zipped.

provided by the user is necessary for the models to run or where our input choice deviates from the default settings. The first part, which is shown in the left-hand Column in Figure 3.2, comprises information on the emissions of the fire. To create the necessary values in the fire information section, we calculate the size in acres as well as the longitude and latitude of the centroid for each of the burned area polygons. In addition, the fire type is specified as "wildfire" as opposed to prescribed or agricultural fire. The fuelbed type is automatically selected based on the location of the fire on a modern map (we are implicitly assuming that the type of forest has not changed between 1910 and now).

Within the fuel moisture inputs section, the season is specified as "summer." Regarding the fuel moisture section, the moisture level is set to the most extreme value "very dry"

informed by the work of Diaz and Swetnam (2013) which noted exceptional hot and dry climatic conditions in 1910 that were not matched again until 2012. Furthermore, there are some indications solidifying this choice in the book by Egan (2009). For example, the author recalls that "People had endured a summer without moisture, the driest in a generation ... (Egan, 2009, p 227)" and writes regarding fuel moisture on the 19th of August that "when Emma Pulaski walked outside of town [Wallace] to greet her husband, and all the vegetation crunched underfoot, brown and crisp to the touch, she knew that everything and everybody in this pocket of people in the mountains had been reduced to fuel (Egan, 2009, p 143)" or "the wind sucked moisture from the forest, blowing hard in the afternoons, leaving all that standing timber as if it was just hung out to dry (Egan, 2009, p 108)."

The default settings regarding consumption (percentage of shrub and canopy) remain unchanged. Although the Big Blowup lasted for approximately 48 hours, the simulation is run for one day as each polygon is modelled individually and likely burned for no more than 24 hours.[13] Finally, guided by historical records recollecting that: "... single wall of yellow and orange ... and it burned at the crowns, the highest tips of the trees exploding into the air, flying off to light the crowns of other tall trees (Egan, 2009, p 115)", it was deemed best to select the flaming combustion phase over the smouldering combustion phase within the timing section.

The output from the smoke emission input modelling generates an emission report that is subsequently used in the second modelling step, the smoke dispersion modelling shown in the

[13] The historical records as described in Section 3.2 provide a good timeline for the unfolding of the event. References to the duration, such as "For forty-eight hours, no one knew whether they would see another day or recognize their homes again" in Egan (2009, p 227) help to emphasise the brevity of the individual events.

middle Column of Figure 3.2. For the dispersion modelling part, BlueSky Playground allows for two options using either VSMOKE, available since 2021, where the user can choose the meteorological impact factors, or the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model which uses observed meteorological data. Even though HYSPLIT would be preferable from an accuracy standpoint it is only available for more recent years. Thus, the VSMOKE model is used which, based on historical records, most closely matches the meteorological conditions at the time.

We select the stability class "Moderately Unstable", and thus deviate from the default setting of "Near Neutral" for two reasons. First, there is evidence of turbulent atmospheric conditions due to the Palouse wind event triggering the wildfire escalation. One passage in Egan (2009) reads "... the Palouse is one of those curious places in the West where a weather system can form benign and transform into something ferocious long after it has left the cradle of its creation. ... the air over the Palouse can be volatile, or violent. So it was on the Saturday afternoon of August 20th, when atmospheric conditions gave birth to a Palouser that lifted the red dirt of the hills and slammed into the forests – not as a gust or an episodic blow, but as a battering ram of forced air (Egan, 2009, p 154)." Second, the Great Fire of 1910 is likely to have created its own weather system that influences atmospheric conditions. Although research on strong convective processes associated with extreme wildfire events, known as pyroconvection, has only evolved recently as noted in Dowdy and Pepler (2018), and therefore the specific identification as such may not have been used in historic reports, there is strong evidence of pyroconvection during the Big Burn. For instance, passages read:

"The chain reaction of a wildfire had begun. Heated plant matter released hydrogen and carbon while drawing in oxygen, and the whole of it was on the run, a weather system of its own (Egan, 2009, p 155)" or "Looking up, he saw towering columns of black smoke above the treeline, and then – the strangest of things – the columns themselves would explode into flame, sending off fire hundreds of feet above the treetops, like the towering spigot of a refinery. What he saw – known as fire whirls – can reach temperatures of 2,000 degrees, with a downdraft in the center and violent updrafts on the outer rings (Egan, 2009, p 159)."

The wind direction is defined as coming from the southwest i.e., 225 degrees, as it appears to be have been the predominant wind direction of the Palouser winds during the event: "Everyone knew about Palousers, the warm winds from the southwest (Egan, 2009, p 4)" or " ... a lumbering advance of flame pushed along by the prevailing winds from the south and west (Egan, 2009, p 145)." As for the transport wind speed, the value of 40 mph is selected. The peak windspeed was reported to be well beyond that: "By now, the conscripted air was no longer a Palouser but a firestorm of hurricane-force winds, in excess of eighty miles an hour (Egan, 2009, p 156)." Further passages indicate windspeeds of 50-60 mph (Egan, 2009, p 155,158). However, we want to select a value that is arguably more appropriate to capture an average over the 24-hour model simulation. Finally, with a value of 10% for relative humidity, we deviate from the default value of 25% based on climatological records by Diaz and Swetnam (2013) who point out extremely low relative humidity with values around 20% or lower for the affected areas. Moreover, historical records read: "He [Billy Greeley, Forest Service's Region 1 forester] had never seen the woods so ready to explode. Low humidity,

always a key indicator, had dropped to the level of the Mojave Desert (Egan, 2009, p 114)."

After running the VSMOKE model, the dispersion result can be exported as a Keyhole Markup Language Zipped (.kmz) file indicating the peak hourly PM2.5 concentration plume in μg/m^3 in six subgroups based on the hazard level. We import the individual files to ArcGIS and merge the smoke plumes from all individual fire polygons considering three levels of pollution: (i) "hazardous", where peak hourly PM2.5 concentrations are above 526μg/m^3, (ii) "unhealthy", with peak hourly PM2.5 concentrations exceeding 130μg/m^3, and (iii) "moderate", where the hourly peak PM2.5 pollution was at least 38μg/m^3. The merged smoke plumes with the three different concentration levels are shown in Figure 3.1b. Note, that the different hazard classes do not make a difference when it comes to categorising which counties are smoke affected.[14]

3.3.3 Population Data and County Boundary Changes

We use the anonymised full-count census population data provided by the Integrated Public Use Microdata Series (IPUMS) USA for the decades 1900 and 1910 (Ruggles et al., 2021). To match the census data with the smoke exposure modelling at the county level, the corresponding county-level Federal Information Processing Standard (FIPS) code for the IPUMS data is generated by combining information on the Inter-University Consortium for Political and Social Research (ICPRS) with the state code for all the census years.

To account for boundary changes from 1900 to 1910 we use (i) the Big Ten Academic Al-

[14] Excluded from the treatment group are the two counties that had less than 1% smoke coverage.

liance (BTAA) Geoportal which provides historical shape files for the years 1900 and 1910[15] and (ii) the Atlas of Historical County Boundaries which documents county-level boundary changes for all US states.[16] For counties affected by boundary changes between 1900 and 1910 we approximate the population in 1900 based on the county boundaries of 1910 taking account of the week of the enforcement of the change. For example, on 21st February 1907 (International Organization for Standardization (ISO) week 8), county Kootenai was split into counties Kootenai and Bonner. Thus, a theoretical 1900 Bonner County would have consisted of 61% of 1910 county Kootenai and the 1900 Kootenai County would only be 39% of the 1910 Kootenai County. Hence, for all weeks before the boundary change, the 1900 population of county Kootenai and a hypothetical 1900 Bonner County is adjusted accordingly. The summary of the calculations of these boundary changes is shown in Table C1 in Appendix C.1.[17]

The boundary change adjustments are necessary to interpolate weekly population over time which is needed to calculate time-varying weekly mortality rates from 1905 to 1910. The 1900 census data was collected on the 1st of June 1990 (ISO week 22) and the 1910 census data on the 15th April 1910 (ISO week 15).[18] After adjusting the 1900 census data to the 1910 county areas we linearly interpolate the population numbers using the two censuses

[15] https://geo.btaa.org/ e.g., https://geo.btaa.org/catalog/harvard-nhgis-pop1910 for the year 1910.

[16] https://digital.newberry.org/ahcb/project.html.

[17] Note, that we assume that the population is equally distributed across the area which is of course only an approximation of the actual distribution. Ideally, one would know which share of the population was affected by the county boundary changes between 1900 and 1910. However, we have no information regarding this.

[18] https://www.census.gov/history/www/through_the_decades/overview/.

and create a weekly county-level population panel data set from 1905 to 1910.[19]

3.3.4 Mortality Data

We obtain the county-level mortality data for the years 1905 to 1910 from the private genealogy company Ancestry.com LLC using their online products of digitised death records including Ancestry.com and Find a Grave. We run separate searches for deaths for each of the years per county for the three fire-affected states of Idaho, Montana, and the eastern part of Washington.[20] The search results on Ancestry.com draw on a number of digitised data sources, such as state and county records, newspapers' obituary sections, church records, and gravestones. Table C2 in Appendix C.1 provides an overview of the data sources on Ancestry.com used for our specific search.

The data cleaning process involved several steps including the cleaning of the individual variables (e.g., date of birth, date of death, age at death, name), inferring missing values from other variables (e.g., inferring age at death from date of birth and date of death), and removing duplicates. The detailed steps of the cleaning procedure are explained in Table C3 in Appendix C.1. Table 3.1 shows how the observation numbers evolve in our sample after each step. Our final data comprises the deaths from 1905 to 1910 for the 70 sample counties and includes 12,876 individual mortality records for children under the age of five.

[19] Note, that the adjustment may introduce measurement error. However, since the boundary changes are arguably unrelated to smoke exposure this could only lead to a bias towards zero of our estimates and would not invalidate our results.

[20] We only look at the eastern part of Washington as the fire occurred at the very eastern border to Montana and the wind direction is reported to be from the Southwest to the Northeast. Thus, the western part of Washington was unlikely to be affected.

Table 3.1: Mortality sample restrictions through the data cleaning process.

Data cleaning steps	Observations
All death records 1905-1910	113,862
1. Cleaning date of death	100,250
2. Cleaning names	100,050
3. Cleaning date of birth and age at death	100,047
4. Sub-setting children under the age of five	28,981
5. Removing duplicates	12,876

Notes: (i) the above table shows the steps of the cleaning process of the digitised death data obtained from Ancestry.com; (ii) a more detailed explanation of the individual steps is provided in Table C3 in Appendix C.1.

We aggregate the individual death records from 1905 to 1910 to weekly death counts using the International Organization for Standardization (ISO) week date system as each week consists of 7 days. Moreover, we adjust the weekly death counts to county boundary changes during this period in a similar manner to that previously described for the population data.

3.3.5 Construction of the Excess Mortality Rate

Estimating excess mortality has emerged as a progressively effective method for quantifying the impact of an event (Acosta and Irizarry, 2022). Although the concept of excess mortality has gained considerable attention during the COVID-19 pandemic to, for example, monitor the progression of COVID-19 (Msemburi et al., 2023) or to compare its impact across countries (Karlinsky and Kobak, 2021), it can and has been applied to a wide range of different events. In the realm of natural disasters, for example, Santos-Burgoa et al. (2018) estimate the excess deaths related to Hurricane Maria in Puerto Rico, and Morita et al. (2017) study the indirect excess mortality risk associated with the 2011 triple disaster in Fukushima, Japan. For wildfires specifically, excess mortality risk has been studied by Hänninen et al.

(2009) in Finland related to East European wildfires and by Kochi et al. (2012) for the 2003 southern Californian wildfires.

Excess mortality can be defined as the additional deaths that occur in a given period of time due to a health event compared to the deaths that would normally have occurred in the absence of that event. Since the counterfactual i.e., the number of deaths in the absence of the health event can not be observed, a common approach to estimate excess mortality is to subtract the baseline mortality from the observed mortality. This baseline is oftentimes based on observations for the same region prior to the event. As implemented by many health monitoring institutions worldwide (e.g., European mortality monitoring activity (EuroMOMO) and the United Kingdom Office for Health Improvement and Disparities) we use a reference period of five years prior to the event i.e., we calculate weekly mortality for children under five years from 1905-1909 for each county to give us our baseline.

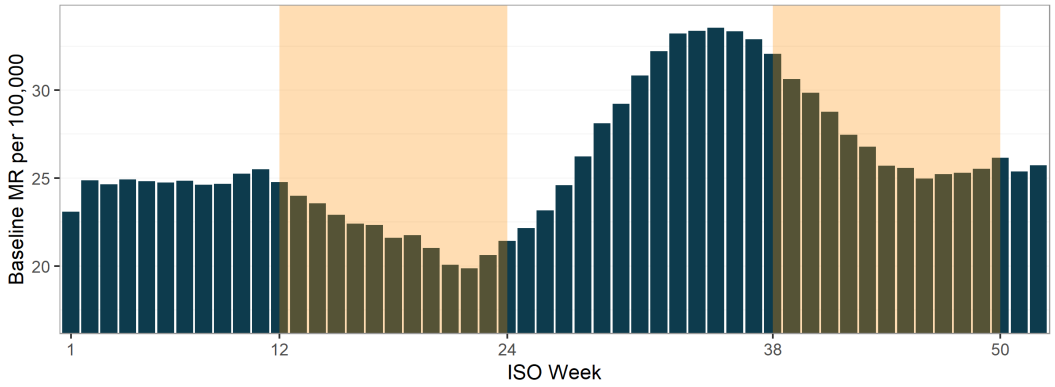
Our context is the impact of the Great Fire of 1910 measuring excess mortality at the regional level. Since the study area is rural and sparsely populated, and the temporal resolution is relatively low i.e., weekly, it is important to take the population size into account. Therefore, we use death rates, calculated as the number of deaths per 100,000 of the population. In addition, since mortality rates are affected by seasonality throughout the year, we calculate the week-specific mortality rate for each ISO week of the year. Weekly excess mortality rates are derived from a three step procedure. First, we calculate the baseline mortality rate for our reference period. As death counts are volatile our baseline mortality rate is smoothed using

each specific ISO week $\pm$ 1 week as shown in Equation (3.1):[21]

$$BMR_w = \frac{1}{15} \sum_{t=1905}^{1909} \sum_{w=-1}^{1} MR_{wt}, \tag{3.1}$$

where BMR represents the baseline mortality rate for children under the age of five in ISO week w, t denotes the year, and MR_{wt} is the mortality rate in week w in year t.

Figure 3.3: Seasonal course of the baseline mortality rates for children under the age of five in the study area.



Notes: (i) Baseline MR indicates the baseline mortality rate for children under the age of five and ISO stands for International Organization for Standardization; (ii) the mortality data is obtained from the genealogy company Ancestry.com and the population data is retrieved from the 1900 and 1910 US censuses; (iii) the orange-shaded areas show the meteorological spring (ISO weeks 12-24) and autumn (ISO weeks 38-50).

The baseline mortality rate estimations for children under the age of five range from approximately 20 to 34 weekly deaths per 100,000 as an average for all the counties in our sample. Figure 3.3 shows the seasonal trend over a 52-week year. The highest mortality rates are observed in the summer and mortality tends to decrease in spring and autumn which are shaded in orange. Table 3.2 presents some descriptives. The first row shows the average

[21] Using a smoothing approach is standard in widely applied excess mortality algorithms used by public health offices such as the Farrington and Noufaily algorithms described in Noufaily et al. (2012).

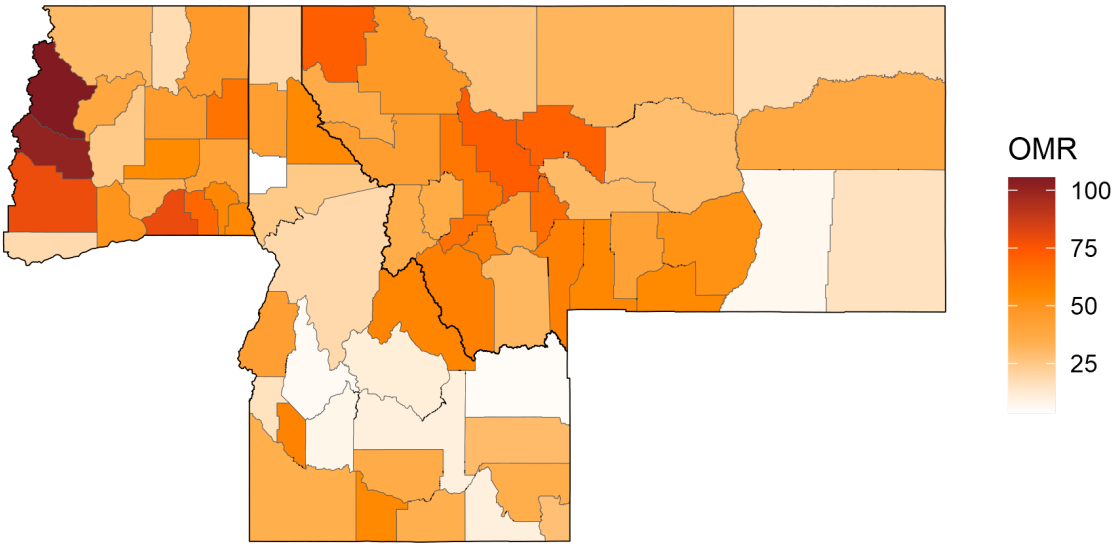
weekly baseline mortality rates. Although there are no specific figures on the death rates of the time for children under the age of five, we can compare our estimates to the period mortality statistics reports by the Department of Commerce and Labour at the Bureau of the Census for the entire population. Reported all age annual mortality rates for the period are equivalent to approximately 25 to 27 weekly deaths per 100,000 for the 1905-1909 period for rural regions in the registration area (Department of Commerce and Labor, 1912).

Second, we calculate the county-level observed mortality rate of each ISO week in 1910. Figure 3.4 shows the geographical distribution of the average mortality rate over all 52 weeks and the second row of Table 3.2 contains the descriptive statistics of the mortality rate. The observed average weekly mortality rate for children under the age of five in 1910 is 42 per 100,000 and is thus, on average, higher than the baseline mortality rate of 26 that is derived from the reference period. This might be because the mortality rate was actually higher in 1910 compared to the reference period or that we are capturing more digitised death records in 1910 than in the previous years. If the latter is the case this could be driven by the fact that the state of Washington is part of the mortality statistics registration area from 1908 and Montana from 1910.[22]

Third, we calculate the county-level weekly excess mortality rates for all ISO weeks of 1910 by subtracting the baseline mortality rates derived from the reference period 1905-1909 from the observed weekly mortality rates in 1910. Figure 3.5 shows the geographical distribution of the excess mortality rate in 1910 and the third row of Table 3.2 shows the excess

[22] Idaho was not part of the census bureau mortality statistics until the 1920s.

Figure 3.4: Average weekly observed mortality rate of children under the age of 5 (number of deaths per 100,000) in 1910.



Notes: (i) OMR denotes the observed mortality rate for children under the age of five in 1910; (ii) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iii) the mortality data is retrieved from the genealogy company Ancestry.com; (iv) the county shape file is provided by The Big Ten Academic Alliance Geoportal and shows the historical county boundaries.

mortality rate descriptive statistics for all our observations i.e., 52 weeks for 70 counties.

3.3.6 Census Crosswalks

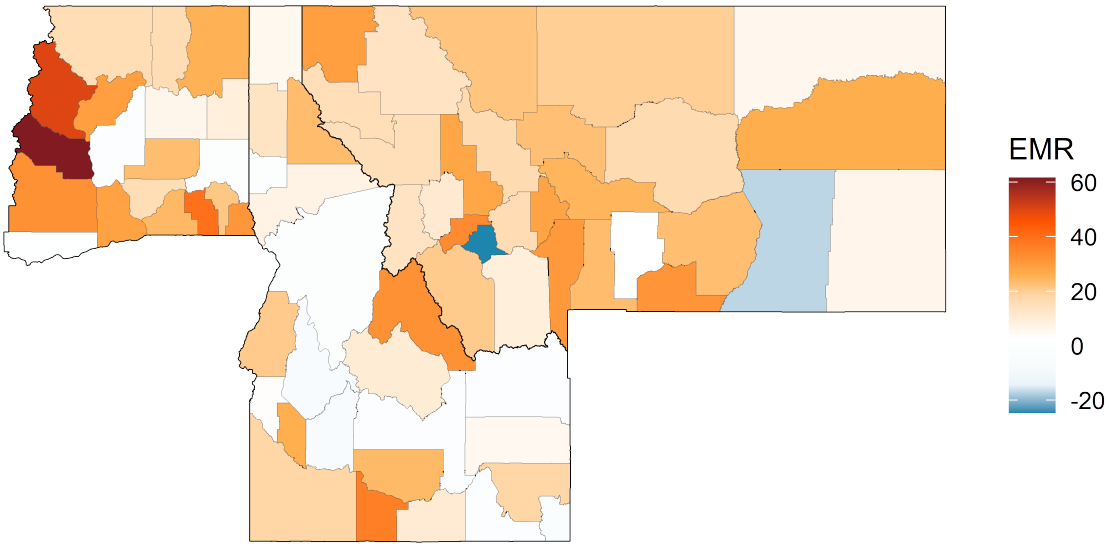
For the analysis of the long-term effect, we use the IPUMS full-count household and individual data for the years 1910, 1930, and 1940 (Ruggles et al., 2021). In order to link in-

Table 3.2: Descriptive statistics of weekly baseline, observed, and excess mortality rates for children per 100,000 under the age of five in 1910.

	Min	Mean	SD	Median	Max	N
BMR Age < 5	0	26	21	22	154	3,640
OMR Age < 5	0	42	72	0	761	3,640
EMR Age < 5	-154	16	72	-6	761	3,640

Notes: (i) SD = standard deviation; (ii) BMR indicates the baseline mortality rate from 1905-1909, OMR denotes the observed mortality rate in 1910, and EMR is the excess mortality rate in 1910; (iii) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iv) the mortality data is retrieved from the genealogy company Ancestry.com.

Figure 3.5: Average weekly excess mortality rate of children under the age of five (number of deaths per 100,000) in 1910.



Notes: (i) EMR denotes the excess mortality rate for children under the age of five in 1910; (ii) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iii) the mortality data is retrieved from the genealogy company Ancestry.com; (iv) the county shape file is provided by The Big Ten Academic Alliance Geoportal and shows the historical county boundaries.

dividuals over time we use the crosswalk files provided by the Census Linking Project implementing the Abramitzky, Boustan and Eriksson (ABE) exact standard algorithm (Abramitzky et al., 2022a,b).[23] The automated approaches create a very low rate of false positives as shown in Abramitzky et al. (2021). The links are undertaken based on variables that are expected to remain constant over time, typically the birth year, name, gender, and county or state of birth; see Abramitzky et al. (2021) for the step-wise procedure. Note that in the crosswalk files only males can be linked because females are harder to track given many change their last names after marriage.

For 1910 we use all individuals of the full-count census to create a number of control

[23] Section 3.5 provides a robustness check using the conservative version of the ABE exact algorithm.

variables at the county level, including average age, percentage of farm households, and the economic structure captured as the percentage of workers in different industries. The full list and descriptive statistics of the county-level variables are shown in Table C4 in Appendix C.1. Furthermore, we extract the boys who were under the age of five in 1910, including their corresponding parental information. In addition, we link the boys with their household characteristics in 1910 using the household serial number. Finally, we merge our variables of interest with the crosswalk files for 1930 and 1940.

At the individual level we construct a race indicator where zero indicates that a person is white and one indicates they are non-white. Nativity is defined as zero if both parents are born in the US and one if at least one parent is born outside the US. At the household level, we include family size which denotes the number of family members, the number of families in one household, and indicator variables for whether it is a family-, farm-, urban-, or mortgage-paying household. The parental characteristics on the mothers' side are limited as women are usually not part of the labour force although we do include the age and nativity of the mother. For the child's father, we obtain numeric information regarding their age, education, and earnings scores. Indicator variables on nativity, employment as well as the industries where they work at are created, where industry codes are classified into 12 categories based on the 1950 Census Bureau industrial classification system. Table 3.3 shows the descriptive statistics for the 9,029 boys that we were able to link from 1910 to 1930.[24]

[24] The descriptive table is extracted for all individuals with a 1930 value for the variables Occupational Income Score and for the Duncan Socioeconomic Index. The sample is slightly smaller i.e., 8,954 individuals when the outcome variables are Occupational Earnings Score, Occupational Siegel Prestige Score, and the Nam-Powers-Boyd Occupational Status Score. The descriptive statistics of the linked men in 1940 closely mirror the numbers from 1930 and are not shown in the interest of conciseness.

The outcome variables of interest are the socioeconomic status measures in 1930 and 1940.[25] More specifically we look at three indicators that are solely based on income: (i) the Occupational Income Score which is the annual income in hundreds US$, (ii) the Occupational Earnings Score denoting the percentage of persons in occupations having lower median earnings than the respondent's occupation, and (iii) the Occupational Siegel Prestige Score which is based on survey methods asking about "general and social standings" of specific occupations. Moreover, we study two occupational standings measures that are derived from a combination of income and education. The Nam-Powers-Boyd Occupational Status Score on a scale from 0 to 100 and the Duncan Socioeconomic Index on a scale from 0 to 96 are alternative measures of socioeconomic status which are based on the median income level and educational attainment associated with each occupational category.

[25] Note, that in this study socioeconomic status and occupational standings are used interchangeably.

Table 3.3: Descriptive statistics of the variables including individual, household, and parental characteristics of 9,029 boys linked to 1930.

	Min	Mean	SD	Median	Max
Individual					
Non-white (%)	0.0	0.7	8.3	0	100
Non-american born parent (%)	0.0	28.4	45.1	0	100
Household (1910)					
Family size	3.0	5.9	2.2	6	19
Families in household	1.0	1.3	0.7	1	8
Non-family household (%)	0.0	0.4	6.0	0	100
Urban household (%)	0.0	17.5	38.0	0	100
Farm household (%)	0.0	58.9	49.2	100	100
Paying mortgage (%)	0.0	35.4	47.8	0	100
Mother (1910)					
Age	15.0	31.4	6.8	31	66
Non-american born parent (%)	0.0	46.5	49.9	0	100
Father (1910)					
Age	18.0	37.3	8.0	37	70
Non-american born parent (%)	0.0	48.6	50.0	0	100
Education score	0.0	8.6	12.8	5	94
Earnings score	1.4	30.1	29.9	10	100
Unemployed (%)	0.0	5.1	21.9	0	100
Ind. Agriculture, forestry & fishing (%)	0.0	63.2	48.2	100	100
Ind. Mining (%)	0.0	3.7	19.0	0	100
Ind. Construction (%)	0.0	6.5	24.6	0	100
Ind. Manufacturing (%)	0.0	5.7	23.3	0	100
Ind. Transportation, communication & utilities (%)	0.0	5.3	22.3	0	100
Ind. Wholesale and retail trade (%)	0.0	7.9	27.0	0	100
Ind. Finance, insurance, and real estate (%)	0.0	1.6	12.4	0	100
Ind. Business and repair services (%)	0.0	1.1	10.3	0	100
Ind. Personal services (%)	0.0	1.8	13.2	0	100
Ind. Entertainment and related services (%)	0.0	0.1	3.6	0	100
Ind. Professional and related services (%)	0.0	1.7	13.0	0	100
Ind. Public administration (%)	0.0	1.4	11.6	0	100

Notes: (i) note that all the variables in percentages (%) are indicator variables and therefore, a median of 100 means that the majority of observations are of the value 1, and 0 means that the majority of observations are of the value 0; similarly, a maximum value of 100 for indicator variables in percentages (%) denotes that the maximum value is 1; (ii) SD = standard deviation, Ind. = Industry; (iii) the variables are obtained from the Integrated Public Use Microdata Series USA 1910 full-count individual and household censuses.

Table 3.4: Descriptive statistics of the socioeconomic status indices of the linked men in 1930 and 1940.

	Min	Mean	SD	Median	Max	N
1930						
Occupational Income Score (in 100s US$)	3.0	19.4	9.5	20.0	80.0	9,029
Occupational Earnings Score	0.6	35.7	28.0	39.7	100.0	8,954
Occupational Siegel Prestige Score	9.3	29.6	12.2	28.4	81.5	8,954
Nam-Powers-Boyd Occupational Status Score	3.6	35.3	26.4	25.1	100.0	8,954
Duncan Socioeconomic Index	3.0	22.2	19.5	15.0	96.0	9,029
1940						
Occupational Income Score (in 100s US$)	3.0	25.0	11.0	24.0	80.0	10,648
Occupational Earnings Score	0.6	49.4	29.7	52.6	100.0	10,621
Occupational Siegel Prestige Score	9.3	36.8	13.3	36.7	81.5	10,621
Nam-Powers-Boyd Occupational Status Score	3.6	48.6	28.1	48.7	100.0	10,621
Duncan Socioeconomic Index	3.0	31.6	24.1	19.0	96.0	10,493

Notes: (i) SD = standard deviation: (ii) the variables are obtained from the Integrated Public Use Microdata Series USA 1930 and 1940 full-count individual censuses.

3.4 Empirical Framework

3.4.1 Identification Strategy

Unlike other natural hazards, such as earthquakes or hurricanes, arguably wildfires do not occur completely randomly in space and can thus be not considered exogenous to various economic outcomes such as health or socioeconomic status. Reasons for this include that wildfire incidence is dependent on anthropogenic factors such as land-use changes (e.g., deforestation) or land and fire management policies that are potentially correlated with health and socioeconomic characteristics. Further endogeneity concerns arise from the fact that many fires start directly due to human activity either through negligence or intentional actions. As a matter of fact, in the context of the Great Fire of 1910, some of the ignitions occurred due to sparks flying off coal-burning trains and it is speculated that some of the fires were set deliberately for political and economic reasons.[26] Thus, wildfire occurrence may be correlated with a number of unobserved economic dimensions or behavioural patterns that potentially also affect child mortality or socioeconomic outcomes, and not accounting for such unobserved factors would lead to biased estimates of the studied outcomes.

To address the aforementioned endogeneity concerns we model the smoke plume utilising meteorological inputs such as wind direction and transport wind speed, which arguably induces exogenous variation in smoke exposure that can be leveraged to estimate the causal effect of smoke pollution on the counties' excess mortality rates and later-life socioeconomic

[26] For instance, it has been noted that "rangers were openly suspicious of these fires: they heard numerous stories that the blazes had been deliberately set – as a way to clear land, to get title, to ensure that a patch of woods not remain for long as part of Roosevelt's reserves (Egan, 2009, p.109)."

status. The identification strategy of using wind direction to estimate the impacts of fire-sourced air pollution on health has been applied in various contexts. For example, Rangel and Vogl (2019) exploit daily changes in agricultural fire location and wind direction to relate in-utero smoke exposure to health at birth outcomes in the sugar-growing region of the Brazilian state of São Paulo. Furthermore, Rocha and Sant'Anna (2022) employ an instrumental variable strategy combining the monthly variation of wind direction in surrounding municipalities to estimate the effect of deforestation-related smoke pollution on morbidity and mortality for municipalities in the Brazilian Amazon, while Pullabhotla and Souza (2022) use daily data on wind direction to study the effect of agricultural fires on hypertension risk in India.

As our modelling of the smoke due to the wildfires is based on a number of meteorological factors it can be considered strictly exogenous as these are unlikely to have been anticipated (e.g., no endogenous selection into treatment). To also exclude the possibility that the treatment and the comparison group were nevertheless on different pathways regarding their excess mortality rates before the event, we also test pre-treatment differences of the two groups in leads of the treatment variable.

Given the source of the smoke is the fire, it is in the nature of the event that the smoke emission is highly correlated with the burned area itself. As we are interested in estimating the effect of smoke pollution rather than the potential direct effects of the fire, we control for the burned area in a similar manner as for smoke exposure, namely creating an indicator variable that is one if a county comprises burned area and zero otherwise in both our short-

term and long-term estimations.

For the short-term impacts, we analyse the effect of wildfire smoke pollution on excess mortality implementing a Difference-in-Differences design. The central idea is that the excess mortality rates of a population that is smoke-affected (treatment group) would have evolved in a similar manner in the absence of the smoke as the population that was unaffected by the event (comparison group). Assuming that this key assumption holds, identification of the effect of smoke pollution on health relies upon using the comparison group for the unobservable counterfactual outcome in the absence of the event. One might worry about contamination from potential changes in mortality rates due to other period factors (e.g., epidemics). However, those are unlikely to be related to smoke exposure and are thus not of concern. Moreover, these potentially confounding shocks or common trends would be picked up by the week-fixed effects assuming they affect the entire study area.

In the estimation of long-term effects of smoke exposure on later-life occupational standings, we link boys who are under the age of five in 1910 over time and assess their socioeconomic outcomes 20 and 30 years later i.e., in 1930 and 1940, respectively.[27] Since our data are inherently cross-sectional, not allowing us to control for all possible confounding county and individual level factors, we control for a large number of individual, parental, household, and county-level characteristics that may affect later-life socioeconomic status in order to isolate the remaining variation attributable to smoke exposure. Thus, for the socioeconomic

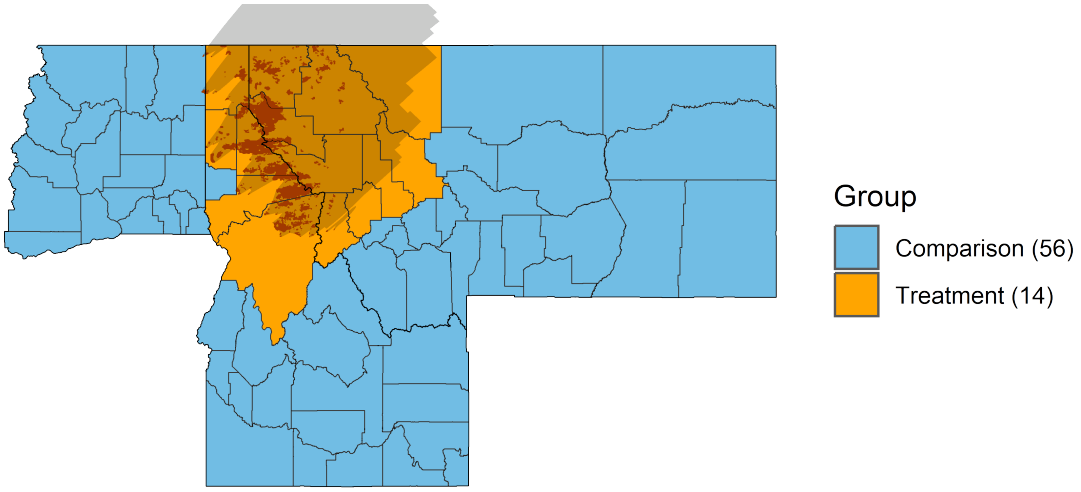
[27] Note that we exclude men who are still "in school" in 1930 and 1940, respectively so only men who have completed their education are observed. For instance, for the Duncan Socioeconomic Index, 492 men in 1930 and 60 men in 1940 are dropped from the analysis as they are enrolled in an educational institution (including college, university, or night school).

regressions we assume causal identification strategy conditional on these controls.

3.4.2 Treatment and Comparison Group

The categorisation of the counties into the treatment group consisting of 14 counties and the comparison group of 56 counties is shown in Figure 3.6, where we additionally include the burned area to jointly visualise the source of the fires. In terms of population, in 1910 about 191,000 people were living in the treatment area and 937,000 in the comparison area. Approximately 19,000 (9.9%) and 105,000 (11.2%) children are under the age of five in the treatment and comparison group, respectively.

Figure 3.6: Treatment and comparison group for smoke-affected counties.



Notes: (i) a county is classified as smoke affected if any part of the county was exposed to hazardous hourly peak pollution (PM2.5>526 μg/m^3); (ii) the county shape file shows the historical county boundaries provided by The Big Ten Academic Alliance Geoportal; (iii) the grey shaded area shows the modelled moderate hazard smoke plume employing the BlueSky smoke modelling framework and the red shaded area indicates the burn perimeters of the fire.

Table 3.5 shows the balance tests across treatment and comparison counties comparing a number of variables at the county level in terms of their eight month average before the

Great Fire of 1910. Accordingly, the average population per county is 13,533 and 16,498 in the treatment and comparison areas, respectively. Note that the higher average population for the comparison group is driven by Spokane County, Washington as the commercial centre of the inland Northwest, and the only county with a population exceeding 100,000 as shown in Figure C1 in Appendix C.1. Excluding Spokane County, the average population for the comparison group is 14,258. The number of children under the age of five is slightly higher in the comparison group per county. As for the calculated mortality rates, the baseline mortality rates and the observed mortality rates are on average slightly higher in the treatment group, and the excess mortality rates are slightly lower in the treatment group prior to the event. However, reassuringly none of the variables of interest are significantly different pre-treatment as shown by the t-statistic and corresponding p-values.[28]

Table 3.5: Balance table showing the county-level average population and mortality rates of children under the age of five for the 32 ISO calendar weeks ($\sim$8 months) before the fire.

	Comparison	Treatment	Difference	t-stat	p-value
Total population	16,498	13,533	2,966	0.55	0.59
Population < 5	1,845	1,357	488	0.95	0.34
BMR < 5 per 100,000	24.0	26.1	-2.2	-0.49	0.63
OMR < 5 per 100,000	39.7	41.3	-1.6	-0.23	0.82
EMR < 5 per 100,000	15.7	15.1	0.5	0.10	0.92

Notes: (i) stars indicate significance according to * $p < 0.0.5$, ** $p < 0.01$, *** $p < 0.001$; (ii) the population variables are obtained from the Integrated Public Use Microdata Series USA 1900 and 1910 full-count individual censuses; (iii) the mortality data is retrieved from the genealogy company Ancestry.com; (iv) BMR denotes the weekly baseline mortality rate per 100,000 and is derived by taking week-specific smoothed averages from 1905-1909; OMR stands for the average observed weekly mortality rate per 100,000 in the ISO calendar weeks 1-32; EMR indicates the weekly excess mortality rate per 100,000 in the ISO calendar weeks 1-32 and is calculated by subtracting the weekly baseline mortality rate from the weekly observed mortality rate.

[28] Similarly, we check the distribution of the average weekly excess mortality rates of all weeks in 1910 before the fire i.e., ISO week 1 to 32 by treatment and comparison group. The distribution shows no clear structural difference between the groups as shown in Figure C2 in Appendix C.1.

3.4.3 Econometric Specification

We estimate a dynamic two-way fixed effects Difference-in-Differences model for the short-term effects shown in Equation (3.2) including 16 pre– and post-event periods (∼4 months) i.e., leads and lags of the treatment variable of the wildfire event which is denoted as week 0:

$$EMR_{it} = \sum_{\substack{k=-16 \\ k\neq -1}}^{16} \beta_k \times \mathbb{1}\{t = k\} \times S_i + \sum_{\substack{k=-16 \\ k\neq -1}}^{16} \gamma_k \times \mathbb{1}\{t = k\} \times BA_i + \mu_i + \lambda_t + \varepsilon_{it}, \quad (3.2)$$

where EMR_{it} is the excess mortality rate in county i and week t, and the week before the event is omitted from the equation to normalise the estimates of β_{-1} and γ_{-1} to 0 in order to estimate the effects relative to the reference period. S_i represents the treatment group indicator and is equal to 1 if county i is smoke-affected in week t, and 0 otherwise. BA_i stands for the burned area and is modelled similarly to smoke and is equal to 1 if county i comprises some burned area in week t and 0 otherwise. County fixed effects, μ_i, account for county-specific time-invariant characteristics and week fixed effects, λ_t, captures common shocks that might potentially affect our study region at large. ε_{it} is the error component. The coefficients of interest are β_k for $k \geq 0$ (lags) which capture the effect of smoke-exposed in post-event period k relative to the pre-event week -1. The coefficients $\beta_k < -1$ (leads) can be interpreted as pre-event differences in excess mortality rates between treatment and comparison groups.

Error terms ε_{it} are clustered at the county level due to the possibility of persistent cor-

relations between idiosyncratic disturbances within counties on a weekly basis. Thus, we allow for serial correlation within the cross-sectional units over time. Given the treatment is captured at the county level this is consistent with recent work on appropriate clustering adjustment by Abadie et al. (2023).

For the estimation of the long-term effect of smoke exposure on later-life socioeconomic outcomes, we estimate an Ordinary Least Squares (OLS) cross-section regression as specified in Equation (3.3):

$$\begin{aligned} SES_{ic}^{d} = \beta_1 S_{ic} + \beta_2 BA_{ic} + IND_i \boldsymbol{\beta}_3 + PAR_{ic} \boldsymbol{\beta}_4 + HH_{ic} \boldsymbol{\beta}_5 \\ + CTY_{ic} \boldsymbol{\beta}_6 + STATE_{ic} \boldsymbol{\beta}_7 + \varepsilon_i, \quad d = [1930, 1940], \end{aligned} \tag{3.3}$$

where SES_{ic}^{d} indicates the socioeconomic status outcome variable in decade d of individual i who resided in county c in 1910. We study the decades 1930 and 1940. S_{ic} is an indicator variable that is equal to 1 if individual i's county of residence c in 1910 was smoke affected and 0 otherwise. In a similar manner, BA_{ic} represents an indicator variable that is equal to 1 if individual i's county of residence c in 1910 comprised some burned area and 0 otherwise. Furthermore, IND_i, PAR_{ic}, and HH_{ic} are vectors of individual, parental, and household characteristics of individual i who resided in county c in 1910, respectively. The variables contained in the vector are shown in Table 3.3. CTY_{ic} is a vector denoting the characteristics of county c in which individual i resided in 1910. The complete list of the variables included at the county level is shown in Table C4 in Appendix C.1. Finally, $STATE_{ic}$ is a vector of indicator variables for each state that individual i's county of residence c in 1910

belongs to and ε_i indicates the error term. Standard errors are again clustered at the county level following the similar reasoning as described for the short-term analysis.

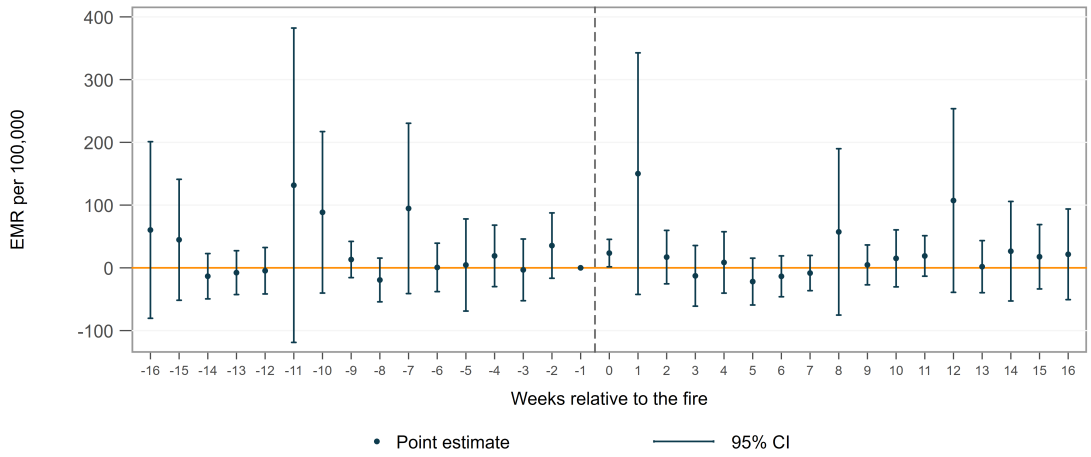
3.5 Results and Discussion

3.5.1 Short-term Excess Mortality

The point estimates and confidence intervals of Equation (3.2) for the 16 pre– and post-event weeks of the Great Fire are shown graphically in Figure 3.7. Reassuringly none of the leads are statistically significant indicating that there was no difference in the excess mortality rate between the treatment and comparison group prior to the event and hence, as expected, there were no anticipation effects. The important observation is that there is a positive effect of smoke exposure on the excess mortality rate for children under the age of five in the week of the wildfire i.e., week 0, but no such effect in the 16 weeks following the event.

Table 3.6 presents the corresponding regression table resulting from Equation (3.2). In the interest of brevity, a condensed version of the estimated lagged coefficients is presented, where the full table showing all lags for week 0 to week 16 is provided in Table C5 in Appendix C.1. In Column (1) we show the estimates of only including county-fixed effects, while Column (2) presents results of additionally controlling for week fixed effects. Accordingly, only accounting for time invariant county unobservables implies that smoke exposure had no impact on mortality of under five year olds. In contrast, also allowing for common time specific factors indicates an excess mortality rate in the treatment counties in the week

Figure 3.7: Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five.



Notes: (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data is obtained from the genealogy company Ancestry.com and the population data is retrieved from the 1910 US census.

of the fire of 49.3 per 100,000. However, additionally including fire exposure in Column (3) to also capture the direct impact of the wildfires, and are the estimates corresponding to Figure 3.7, reduces the point estimate by over 50 per cent. The estimated coefficient suggests that smoke exposure due to the wildfire increased excess mortality rate by 23.5 per 100,000 for children under the age of five in the week of the event.

Comparing our estimated excess mortality rate of 23.5 per 100,000 to the observed weekly mortality rate of 42 per 100,000 over the entire year of 1910, as taken from Table 3.2, suggests a 56% increase in excess mortality in the week of the fires due to smoke exposure. This immediate impact on excess mortality is in line with findings provided by Johnston et al. (2011) who study all-cause non-accidental mortality due to bush fires and dust storms from 1997 to 2004 in Sydney. The authors report a same-day increase in mortality controlling for temperature for all age groups. Moreover, Doubleday et al. (2020) assess non-traumatic

Table 3.6: Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 6).

	Excess Mortality Rate		
	(1)	(2)	(3)
Smoke event	32.4	49.3*	23.5*
	(16.8)	(19.8)	(11.0)
1 week after event	52.9	45.8	150.1
	(47.6)	(50.3)	(96.5)
2 weeks after event	13.6	23.6	17.1
	(19.0)	(24.1)	(21.3)
3 weeks after event	-4.2	5.4	-12.7
	(31.4)	(34.4)	(24.2)
4 weeks after event	-19.6	-11.1	8.6
	(22.5)	(25.0)	(24.5)
5 weeks after event	-14.5	-22.0	-21.9
	(27.4)	(31.9)	(18.7)
6 weeks after event	-12.4	-6.4	-13.5
	(23.0)	(26.2)	(16.3)
Controls			
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.06	0.07	0.10
N	2,310	2,310	2,310

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Equation (3.2); (iii) the sample includes 70 counties; (iv) Burned area denotes the percentage burned of a county; (v) standard errors are clustered at the county level; (vi) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data is retrieved from the genealogy company Ancestry.com.

mortality associated with wildfire smoke exposure from 2006 to 2017 in Washington State and report that previous-day smoke exposure poses the highest mortality risk and that it diminishes rapidly such that there is no evidence of an elevated mortality risk after two days.

3.5.2 Long-term Socioeconomic Status

The regression table for the 1930 census linked estimations of Equation (3.3) is shown in Table 3.7. The results in Columns (1) to (3) indicate that there was no impact of smoke

exposure during early childhood on surviving male adults 20 years after exposure when they were between 20 and 24 years old, for indicators based solely on income, i.e., the Occupational Income Score, the Occupational Earnings Score, and the Occupational Siegel Prestige Score. In contrast, there is a significant negative effect for the two occupational standings indicators that are jointly based on education and income. The point estimates suggest that the Nam-Powers-Boyd Occupational Status Score is 2.6 points lower for men who were smoke-exposed in early childhood compared to non-exposed men (Column (4)). This translates into a 7.4% decrease relative to the sample mean of 35.3 (Table 3.4). The estimation for the Duncan Socioeconomic Index shown in Column (5) indicates that smoke-exposed men also rank on average 2.6 points lower than non-exposed men, which translates to a decrease of 11.6% relative to the mean of 22.2 points shown in Table 3.4.[29]

In terms of the other controls, at the individual level non-whites have lower socioeconomic status outcomes than whites. Moreover, growing up in a larger family or on a farm is associated with lower later-life occupational standings while growing up in an urban household is linked to better performance in later-life socioeconomic status. Regarding parental characteristics, our results suggest that an increase in the age of the mother, as well as the father's Occupational Earnings Score, are positively associated with better ranking on later-life occupational standings. Reassuringly, these results are consistent with those in the traditional

[29] One may be concerned that the boys that survived the wildfire-sourced air pollution are systematically different (e.g., healthier, stronger, richer) than the non-smoke-affected boys and thus, that our sample is characterised by a "survival bias". However, assuming this was the case, our results showing a negative effect on the later-life socioeconomic status of the boys who were smoke-exposed would be an underestimation of the true effect.

labour economics literature.[30]

When we assess the linked sample for 1940, we still observe negative effects on a majority of the occupational standings indicators of early-childhood exposure to wildfire smoke, however, they are not significant. The regression results are shown in Table 3.8. Therefore, the negative effect on men at the age of 20-24 years does not seem to persist once they are 30-34 years old. A possible reason is that once the workers have spent longer in the labour-market, other factors such as experience or connections, may be more influential than the early-childhood health shock.

More generally, one should note that the adverse effect of early childhood wildfire exposure on later-life socioeconomic status indicators found at least 20 years after the wildfire may have potentially arisen through health effects that affected both physiological as well as cognitive aspects of development. More specifically, the possibility that adverse effects of wildfire smoke exposure on respiratory and cardiovascular morbidity may have longer-lasting implications for later-life health outcomes has been outlined in Section 3.1. For example, it has been found that smoke exposure in early childhood is associated with shorter stature by the age of 17 presented by Rosales-Rueda and Triyana (2019), and Tan-Soo and Pattanayak (2019) propose that decrease in height due to fires could result in an average monthly income decrease of approximately 4% in adulthood. Nevertheless, it is important to note that the

[30] For the Duncan Socioeconomic Index the coefficient of burned area is positive and significant. This may result from rebuilding efforts after the fire in the long run which may have had a positive effect on socioeconomic status. However, we do not propose a causal interpretation of the burned area coefficient because there are arguably other factors that are correlated with fire incidence and socioeconomic outcomes as pointed out in Section 3.4.

study by Rosales-Rueda and Triyana (2019) predominantly focused on prenatal exposure, while the current research examines children under the age of five, making direct comparisons between the two studies difficult. Moreover, focusing on the cognitive channel, Wen and Burke (2022) find a negative association between annual cumulative wildfire-smoke exposure and school test scores for the US using data from 2009 to 2016. However, the children in this paper are between 8-14 years old and thus older than the children in our study.

Given the literature on wildfire-smoke exposure and cognitive as well as neuropsychological development in children (particularly under the age of five) is limited, it is helpful to further draw on evidence from PM2.5 exposure due to other sources. In the literature review on air pollution and neuropsychological development in children by Suades-González et al. (2015) the authors report that there is inadequate or insufficient evidence on the association between PM2.5 and cognitive ad psycho-motor development. However, the authors suggest a positive link between postnatal PM2.5 exposure and autism spectrum disorder (ASD). For example, Talbott et al. (2015) conduct a population-based case-control study in Pennsylvania and find a significant association of PM2.5 exposure at the age of 2 and childhood ASD. Note that while many studies assess prenatal exposure to air pollution, evidence for young children is scarce (Suades-González et al., 2015).

Table 3.7: Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood.

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.6	-2.0	-0.9	-2.6*	-2.6***
	(0.5)	(1.3)	(0.5)	(1.1)	(0.7)
Burned area	-0.2	-1.4	0.2	0.8	2.7*
	(0.6)	(1.8)	(0.8)	(1.4)	(1.0)
Individual					
Non-white	-5.3***	-16.0***	-4.9**	-15.7***	-8.2***
	(0.9)	(2.7)	(1.5)	(2.4)	(1.6)
Non-american born parent	-0.2	-0.6	-0.3	-0.9	-0.6
	(0.3)	(0.9)	(0.4)	(0.8)	(0.6)
Household (1910)					
Family size	-0.2***	-0.6***	-0.5***	-0.9***	-0.8***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	-0.1	-0.2	0.2	0.1	0.3
	(0.2)	(0.5)	(0.2)	(0.4)	(0.3)
Non-family household	0.6	3.0	1.5	3.7	4.3
	(2.0)	(5.9)	(2.4)	(5.0)	(3.5)
Urban household	1.2**	3.0*	1.9***	4.6***	4.5***
	(0.4)	(1.3)	(0.5)	(1.0)	(0.7)
Farm household	-1.5***	-4.3***	-0.8	-4.4***	-2.4***
	(0.3)	(1.0)	(0.5)	(1.0)	(0.7)
Paying mortgage	0.1	0.4	0.4	0.6	0.6
	(0.2)	(0.6)	(0.3)	(0.5)	(0.4)
Parents (1910)					
Mother: Age	0.1***	0.3***	0.2***	0.4***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Mother: Non-american born parent	-0.3	-0.9	0.0	-0.4	-0.1
	(0.2)	(0.7)	(0.3)	(0.6)	(0.4)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)
Father: Non-american born parent	-0.3	-1.3	0.1	-0.5	0.6
	(0.3)	(0.7)	(0.3)	(0.6)	(0.5)
Father: Education score	0.0	0.0	0.0**	0.1	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0**	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	-0.2	-0.5	-0.5	-1.2	-1.5
	(0.3)	(1.0)	(0.5)	(1.0)	(0.7)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.12	0.13	0.06	0.14	0.11
N	9,029	8,954	8,954	8,954	9,029

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Table 3.8: Regression of socioeconomic outcomes in 1940 on smoke exposure in early childhood.

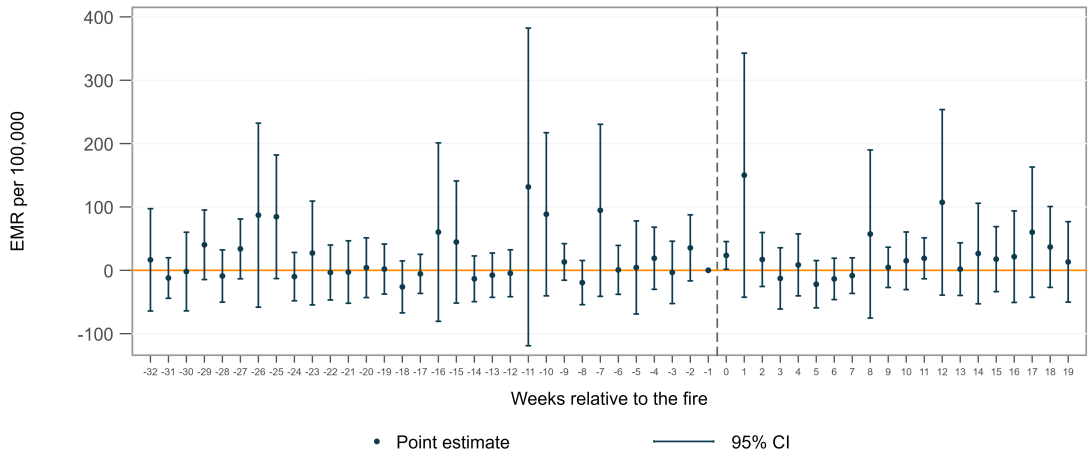
	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.8	-1.6	0.2	-0.8	0.1
	(0.4)	(1.4)	(0.6)	(1.4)	(1.0)
Burned area	0.8	1.2	0.2	0.7	0.6
	(0.6)	(2.0)	(0.6)	(1.8)	(1.2)
Individual					
Non-white	-2.9*	-8.3**	-6.0***	-12.1***	-9.7***
	(1.1)	(2.7)	(1.7)	(2.6)	(2.6)
Non-american born parent	-0.4	-1.2	-0.5	-1.4*	-1.0
	(0.3)	(0.8)	(0.3)	(0.7)	(0.6)
Household (1910)					
Family size	-0.3***	-0.8***	-0.5***	-1.0***	-1.0***
	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)
Families in household	0.2	0.5	0.5*	0.6	0.7*
	(0.1)	(0.4)	(0.2)	(0.4)	(0.3)
Non-family household	-0.3	-0.2	0.0	1.3	1.0
	(1.8)	(5.3)	(1.6)	(4.3)	(3.6)
Urban household	1.1*	1.7	0.9	2.8**	3.4**
	(0.4)	(1.1)	(0.5)	(1.1)	(1.0)
Farm household	-1.2***	-3.5***	-0.2	-3.1**	-1.5
	(0.4)	(1.0)	(0.5)	(1.0)	(0.9)
Paying mortgage	-0.1	0.1	-0.3	-0.1	-0.7
	(0.2)	(0.6)	(0.3)	(0.6)	(0.5)
Parents (1910)					
Mother: Age	0.1***	0.2***	0.2***	0.3***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	0.0	-0.6	0.3	-0.6	0.2
	(0.2)	(0.6)	(0.3)	(0.6)	(0.4)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-american born parent	-0.0	-0.9	0.8*	-0.2	0.8
	(0.3)	(0.7)	(0.3)	(0.7)	(0.5)
Father: Education score	0.1***	0.1***	0.1***	0.2***	0.2***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0**	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	0.5	1.6	0.3	1.5	0.8
	(0.5)	(1.3)	(0.6)	(1.1)	(1.0)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.09	0.09	0.06	0.11	0.11
N	10,635	10,608	10,608	10,608	10,481

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

3.5.3 Robustness Checks

To corroborate that our short-term excess mortality finding in the week of the wildfire event is not driven by chance we extend the evaluated period from 16 pre– and post-event periods to the full calendar year including all 52 ISO weeks of 1910. Our main result is unchanged in that there is only a significant impact of smoke exposure on excess mortality in the week of the fire and in none of the other 51 ISO weeks of the year. The point estimates and confidence intervals are shown in Figure 3.8.

Figure 3.8: Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five for full calendar year 1910.



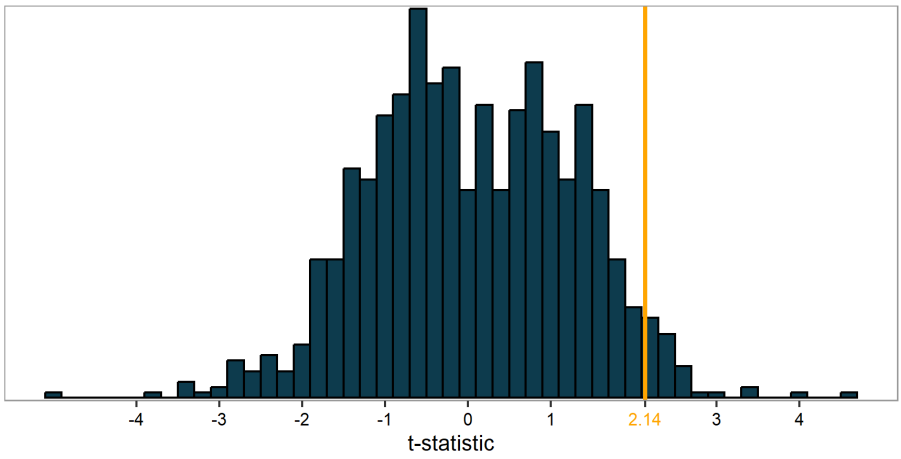
Notes: (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data is obtained from the genealogy company Ancestry.com and the population data is retrieved from the 1910 US census.

We also conduct a number of permutation tests in the spirit of Fisher (1937). More specifically, we randomise which of the 14 out of the 70 counties are smoke-affected, and then also randomise in which 10 of these 14 a fire event took place. Subsequently, we run Equation (3.2) performing 1,000 iterations and plot the distribution of the corresponding t-statistic.

The p-value is derived by the rank of the actual estimate. For the short-term excess mortality value in the week of the wildfire, this permutation test indicates a p-value of 0.034 (1-966/1,000) as shown in Figure 3.9 which demonstrates that the result is unlikely to be driven by chance. For the later-life socioeconomic outcomes, the similar test is performed for the estimates that are significant in 1930 i.e., the Nam-Powers-Boyd Occupational Status Score and the Duncan Socioeconomic Index as shown in Table 3.7. The distribution of the t-statistics derived from the 1,000 iterations for both variables is shown in Figure 3.10. While the actual t-statistic of the Duncan Socioeconomic Index (Figure 3.10b) is highly significant with a p-value of 0.018 (1-982/1,000), the p-value of the Nam-Powers-Boyd Occupational Status Score (Figure 3.10a) is 0.078 (1-922/1,000) and thus only significant at the 8% level. This indicates that although we have confidence in the negative effect of early-life smoke exposure on the Duncan Socioeconomic Index this is somewhat weaker for the case of the Nam-Powers-Boyd Occupational Status Score.

One might be worried that too many comparison counties are included in our estimation that are peripheral to the source. To investigate we run two different estimations reducing the number of comparison counties (i) visually around the smoke plume (29 comparison counties), and (ii) using a calculated buffer area around the modelled smoke plume of 250 km, and include all the counties in the comparison group that are within that buffer (40 comparison counties). The corresponding maps are shown in Figure C3 in Appendix C.2. The estimations of Equation (3.2) indicate slightly higher weekly excess mortality rate coefficients of 28.7 and 26.9 per 100,000 for the estimation with 29 and 40 adjacent comparison counties,

Figure 3.9: Permutation test of the effect of smoke exposure on excess mortality in the week of the fire with 1,000 iterations.

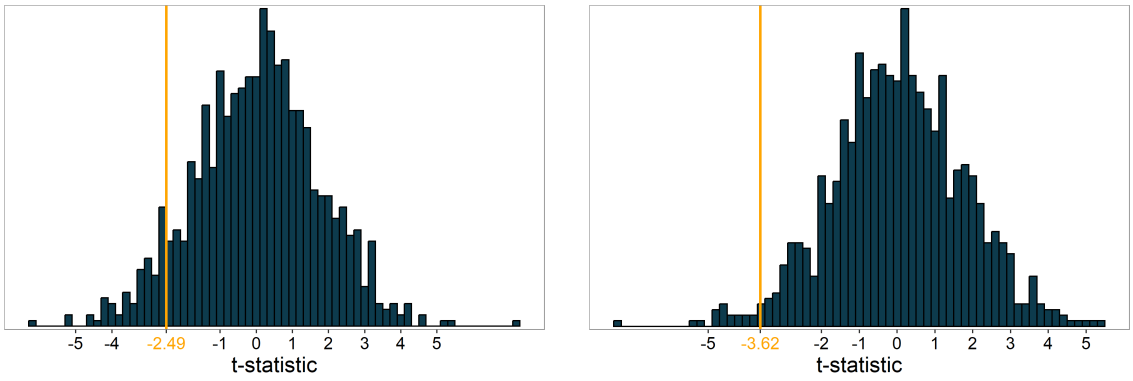


Notes: (i) this plot shows the distribution of the t-statistic for 1,000 iterations of Equation (3.2) randomly assigning treatment (smoke exposure) to 14 out of the 70 counties of which 10 are also containing burned area; (ii) the orange line indicates the t-statistic of our baseline estimate presented in Table 3.6 Column (3).

respectively. Both the Difference-in-Differences plots (Figure C4) and the regression tables (Table C6 and Table C7) are shown in Appendix C.2. As for the long-term impact on later-life socioeconomic outcomes, the point estimates are still negative but insignificant for both versions of reduced comparison counties for 1930 (Table C8 and Table C9). Note, that the sample size decreases to approximately 6,700 men for the subsample including 40 comparison counties and to approximately 4,900 men for the sample with 29 comparison counties and thus may lack statistical power.

For both the short-term and the long-term estimations, one might also want to assess whether our results are robust to the exclusion of Spokane County, Washington and Missoula County, Montana from the comparison group for two reasons. The first concern is regarding the major difference to the other counties as Spokane County is the economic centre of the

Figure 3.10: Permutation test of the effect of smoke exposure later-life socioeconomic status indices with 1,000 iterations.



(a) Nam-Powers-Boyd Occ. Status Score

(b) Duncan Socioeconomic Index

Notes: (i) this plot shows the distribution of the t-statistic for 1,000 iterations of Equation (3.3) randomly assigning treatment (smoke exposure) to 14 out of the 70 counties of which 10 are also containing burned area; (ii) the orange line indicates the t-statistic of our baseline estimates presented in Table 3.7 Columns (4) and (5).

area at the period and therefore, might be structurally different.[31] The second concern is regarding the evacuation procedures during the fire. The evacuations by train from within the immediate burn zone from small towns surrounding Wallace were carried out in two directions (Krainz, 2012). On the one hand, people were evacuated to Missoula, Montana which is within our treatment area. On the other hand, people were evacuated to Spokane which is part of our comparison group. These evacuations were performed rather late when the population was arguably already smoke-exposed.[32] At the same time one may worry that evacuees to

[31] Note, that at least for the short-term analysis this is captured by county-fixed effects. For the long-term analysis, we try to capture county characteristics with a vector of county-level controls. However, this is likely to be imperfect.

[32] This assumption is corroborated by newspaper articles reporting on evacuations. The Spokesman-Review (Spokane, Washington) writes on Monday, the 22nd of August on the recollection of an evacuee that "On our way to this city we passed through stretches of raging fire. A pall of smoke hung over everything. In places it was like early dusk. For 10 miles you could not see a coach ahead. The cars were crowded with people and smoke, and the heat from the blazing timber could be felt on the window panes." Similarly, the Idaho Daily Statesman (Boise, Idaho) reports on the same day that "The smoke became so dense in the gulch that it was impossible for the fleeing travellers to keep on the road, and Rev. Grier was compelled to lead the entire way, lighting up the gloom with a lantern."

Spokane would not be smoke exposed to the same extent as the population remaining in the smoke-affected areas.[33] We find that our excess mortality rate estimate for the fire week is robust to dropping Spokane County and Missoula County from the comparison group with a coefficient of 26.8 per 100,000 which is slightly larger than our baseline result of 23.5 per 100,000 (see Figure C5 for the DD plot and Table C10 for the regression table). In terms of the long-term estimation we still find significant negative effects for the Duncan Socioeconomic Index with a reduction of 2.1 points in 1930 (Table C11 in Appendix C.2). While the Nam-Powers-Boyd Occupational Status Score is still negative it is now insignificant. As for the full sample, we continue to find no significant effects on later-life socioeconomic status indicators in 1940 (Table C12 in Appendix C.2).

Finally, For the long-term analysis, we re-run our estimations using a different matching algorithm i.e., the ABE exact conservative matching algorithm which employs stricter matching rules than the ABE standard matching approach. More specifically, while in the standard matching approach, the individuals are required to be unique by name within the respective year of birth, this requirement needs to hold true within two years of birth to be established as a match in the ABE exact conservative approach (Abramitzky et al., 2021). Table C13 in Appendix C.2 shows that in line with our baseline estimations, there is a negative effect on both the socioeconomic status indicators that are based on both income and education. The magnitude is similar for the Nam-Powers-Boyd Occupational Status Score (–2.7) and slightly higher for the Duncan Socioeconomic Index i.e., –3 vs. –2.6 in the baseline estimation. Sim-

[33] Also note that only very few people in the smoke-affected region were evacuated with around 1,200 persons being evacuated to Spokane.

ilar to the baseline estimation there is no effect in 1940 (see Table C14 in Appendix C.2).

3.6 Conclusions

In this study we assessed the smoke-induced short and long-term impacts of the Great Fire of 1910, one of the largest wildfires on record in the US, on excess mortality and later-life socioeconomic outcomes in children under the age of five. To this end we used historical burn perimeters within a wildfire smoke emission and dispersion model to proxy smoke exposure and combined this with mortality records and linked full-count census data from 1900 to 1940. Our econometric estimations suggest a short-term effect on the excess mortality rate for children residing in smoke-affected regions in the week of the wildfire. Furthermore, we find weaker evidence that boys who were under the age of five at the time of the Great Fire of 1910 ranked lower on some socioeconomic status indices in 1930 if they resided in smoke-affected counties compared to boys who did not.

More generally our paper provides novel evidence both in the economic assessment of smoke exposure during wildfires in terms of short-term health effects during childhood and later life socioeconomic outcomes in a historic context where avoidance behaviour is relatively limited. Thereby this empirical study contributes to a deeper understanding of the implications of major wildfire events for public health and human capital formation. Both these aspects are especially important given the prolonged and intensified wildfire seasons that have occurred over the last few decades globally and are projected to substantially deteriorate air quality in many places in the future (Ford et al., 2018; Burke et al., 2023). While

the affected area of the Great Fire of 1910 was sparsely populated, our results suggest that the implications of wildfires of a similar magnitude may be heavily amplified if the downwind regions are more densely populated (Pan et al., 2023). Nevertheless, it should be noted that while our context of a rather extreme level of fire-sourced air pollution has advantages for empirical identification, one must be cautious to extrapolate the findings to the more typical pollution levels induced by wildfire smoke.

While there is a growing body of literature studying the effects of wildfire-induced air pollution on direct and short-term health outcomes, there is arguably a lack of evidence on the potential long-term social costs that may arise due to impairments on human capital formation and economic well-being. Neglecting these potential social costs can lead to an incomplete evaluation of the true costs of wildfires on societies. The attempt to quantify fire-sourced air pollution effects, and thereby also the potential of health benefits due to its abatement or exposure avoidance, is particularly important as wildfire duration and intensity are projected to increase in many regions of the world. While our paper is a first attempt of filling this gap using a large historical event, the potential social costs warrant considerably more investigation.

Concluding Remarks

Each of the three essays sheds light on a very specific aspect of wildfires and their potential societal, economic, and health implications. The findings of the first essay indicate that extreme wildfires in Mediterranean Europe, particularly in Portugal, Greece, Spain, and Italy, occur more frequently than oftentimes perceived. This insight holds important implications for various stakeholders. Government agencies can utilise the country-level loss estimates provided for specific time horizons to adjust budget planning for fire prevention measures and suppression spending. Landowners can make informed decisions regarding long-term investment and forest management strategies based on the quantified risk of large fires. Reinsurance companies can incorporate the knowledge of wildfire risk to assess potential liabilities and set appropriate premiums. Additionally, the study emphasises the importance of converting return level estimates of extreme wildfires into monetary values for policy-related cost-benefit analyses. However, it is crucial to acknowledge the limitations of the study, such as the short data period used for analysis and the need for comprehensive and harmonised data to evaluate future wildfire risk scenarios, particularly regarding factors that can not be picked up by satellite imagery such as wildfire prevention and suppression spending.

The second essay uses high-resolution satellite data on wildfire-burned areas in Southern Europe and links those with regional economic data to understand the impact of wildfires on local economies. The analysis overcomes endogeneity concerns of wildfires by using a measure of wildfire danger for predominantly forested areas as an instrumental variable while controlling for fire danger and general climatic conditions in non-forested areas. The findings reveal a consistent negative effect of wildfires on the annual regional GDP growth rate, ranging from 0.11% to 0.18% conditional on having experienced a fire. The employment analysis shows heterogeneous impacts across different industries, with negative effects in sectors related to retail and tourism and positive effects in sectors such as finance, insurance, and real estate. The study highlights the need for a comprehensive evaluation of mitigation and prevention mechanisms to reduce the negative impacts of wildfires on local economies. Further research and data collection efforts are necessary to understand the complex factors involved in wildfire management and develop strategies to strengthen regional economic resilience.

The third essay estimates the short and long-term impacts of the Great Fire of 1910, one of the largest wildfires in US history, on excess mortality and socioeconomic outcomes in children under the age of five. By analysing historical data sources, the study finds evidence of a short-term increase in the mortality rate for children in smoke-affected regions during the week of the wildfire. Additionally, it suggests that boys who experienced smoke exposure in early childhood may suffer later-life consequences manifested by lower socioeconomic status outcomes 20 years after the event compared to men who were not smoke-exposed. While the Great Fire of 1910 affected sparsely populated areas, the study suggests that the consequences

of similar wildfires could be amplified in more densely populated downwind regions. Yet, the awareness of the harmful impacts of wildfire-sourced smoke exposure on air quality and the availability of means to avoid exposure may reduce adverse effects on human health. The study highlights the necessity of conducting additional research on the long-term social costs of wildfires, including their effects on human capital formation and economic well-being. Importantly, failing to consider these costs is crucial, as it can result in an inadequate comprehension of the holistic impact of wildfires on societies.

While these essays contribute to understanding the economic implications of wildfires, they offer only limited insights into specific aspects. Particularly as wildfire intensity and duration are projected to increase in many parts of the world, empirical evidence is warranted for a number of aspects. Extensive evidence-based research is needed to examine the effects of wildfires in developing countries as well as in geographical regions that have historically been unaffected by wildfires but are expected to face impacts in the future due to changing climatic conditions. Moreover, it would be valuable to investigate the heterogeneous effects of wildfires on different population groups, considering factors such as demographic characteristics, wealth, gender, and race. This research would facilitate the development of solutions to better protect and support disadvantaged or vulnerable populations. Additionally, further exploration of the long-term effects of wildfire-sourced smoke exposure is essential, as studies in this area are still scarce. A deeper understanding of these effects would support investments aimed at minimising exposure and increasing awareness. Lastly, although not explicitly addressed in this thesis, future research should focus on exploring the impacts

of wildfires on biodiversity (e.g., through ecosystems or species). This investigation is crucial for obtaining a comprehensive understanding of the potential effects that are likely to significantly impact societies in indirect and less conspicuous ways compared to the more immediately visible impacts.

Bibliography

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., and Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3):865–918.

Abramitzky, R., Boustan, L., Eriksson, K., Rashid, M., and Pérez, S. (2022a). Census Linking Project: 1910-1930 Crosswalk.

Abramitzky, R., Boustan, L., Eriksson, K., Rashid, M., and Pérez, S. (2022b). Census Linking Project: 1910-1940 Crosswalk.

Acosta, R. J. and Irizarry, R. A. (2022). A flexible statistical framework for estimating excess mortality. *Epidemiology*, 33(3):346–353.

Aguilera, R., Corringham, T., Gershunov, A., and Benmarhnia, T. (2021). Wildfire smoke impacts respiratory health more than fine particles from other sources: Observational evidence from Southern California. *Nature Communications*, 12(1).

Alix-Garcia, J. and Millimet, D. (2022). Remotely Incorrect? Accounting for Nonclassi-

cal Measurement Error in Satellite Data on Deforestation. *Journal of the Association of Environmental and Resource Economists*.

Bailey, R. E., Hatton, T. J., and Inwood, K. (2018). Atmospheric pollution, health, and height in late nineteenth century britain. *The Journal of Economic History*, 78(4):1210–1247.

Barattieri, A. et al. (2021). The short-run, dynamic employment effects of natural disasters: New insights. *Working Paper ISSN 2292-0838*.

Barone, G. and Mocetti, S. (2014). Natural disasters, growth and institutions: A tale of two earthquakes. *Journal of Urban Economics*, 84:52–66.

Barrio, M., Loureiro, M., and Chas, M. L. (2007). Aproximación a las pérdidas económicas ocasionadas a corto plazo por los incendios forestales en Galicia en 2006. *Economía Agraria y Recursos Naturales*, 7(14):45–64.

Batllori, E., Parisien, M. A., Krawchuk, M. A., and Moritz, M. A. (2013). Climate change-induced shifts in fire for Mediterranean ecosystems. *Global Ecology and Biogeography*, 22(10):1118–1129.

Bayham, J., Yoder, J. K., Champ, P. A., and Calkin, D. E. (2022). The economics of wildfire in the United States. *Annual Review of Resource Economics*, 14.

Baylis, P. and Boomhower, J. (2019). Moral hazard, wildfires, and the economic incidence of natural disasters (No. w26550). *National Bureau of Economic Research*.

Beach, B. and Hanlon, W. W. (2018). Coal smoke and mortality in an early industrial economy. *The Economic Journal*, 128(615):2652–2675.

Bedia, J., Herrera, S., Camia, A., Moreno, J. M., and Gutiérrez, J. M. (2014). Forest fire danger projections in the Mediterranean using ENSEMBLES regional climate change scenarios. *Climatic Change*, 122(1-2):185–199.

Bennett, W. D., Zeman, K. L., and Jarabek, A. M. (2007). Nasal contribution to breathing and fine particle deposition in children versus adults. *Journal of Toxicology and Environmental Health - Part A: Current Issues*, 71(3):227–237.

Berlin Rubin, N. and Wong-Parodi, G. (2022). *As California burns: The psychology of wildfire- and wildfire smoke-related migration intentions*. Springer Netherlands.

Beverly, J. L. and Martell, D. L. (2005). Characterizing extreme fire and weather events in the Boreal Shield ecozone of Ontario. *Agricultural and Forest Meteorology*, 133(1-4):5–16.

Black, C., Tesfaigzi, Y., Bassein, J. A., and 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.

Borgschulte, M., Molitor, D., and Zou, Z. Y. (2020). Air pollution and the labor market: Evidence from wildfire smoke. *NBER Working Paper*, pages 1–47.

Botzen, W. J., Deschenes, O., and Sanders, M. (2019). The economic impacts of natural disasters: A review of models and empirical studies. *Review of Environmental Economics and Policy*, 13(2):167–188.

Boustan, L. P., Kahn, M. E., Rhode, P. W., and Yanguas, M. L. (2020). The effect of natural disasters on economic activity in US counties: A century of data. *Journal of Urban*

Economics, 118:103257.

Bowman, D. (2018). Wildfire science is at a loss for comprehensive data. *Nature*, 560(7716):7.

Bowman, D. M., Kolden, C. A., Abatzoglou, J. T., Johnston, F. H., van der Werf, G. R., and Flannigan, M. (2020). Vegetation fires in the Anthropocene. *Nature Reviews Earth and Environment*, 1(10):500–515.

Bowman, D. M., Williamson, G. J., Abatzoglou, J. T., Kolden, C. A., Cochrane, M. A., and Smith, A. M. (2017). Human exposure and sensitivity to globally extreme wildfire events. *Nature Ecology and Evolution*, 1(3):1–6.

Brimblecombe, P. (2011). *The big smoke: A history of air pollution in London since medieval times*. Routledge.

Brockmeyer, S. and D'Angiulli, A. (2016). How air pollution alters brain development: The role of neuroinflammation. *Translational Neuroscience*, 7(1):24–30.

Burke, M. et al. (2020). The changing risk and burden of wildfire in the US. *NBER Working Paper Series*, No. 27423.

Burke, M. et al. (2022). Exposures and behavioural responses to wildfire smoke. *Nature Human Behaviour*, 6(10):1351–1361.

Burke, M. et al. (2023). Wildfire influence on recent US pollution trends. *NBER Working Paper Series*, January(30882):1–41.

Butry, D. T., Mercer, D. E., Prestemon, J. P., Pye, J. M., and Holmes, T. P. (2001). What is

the price of catastrophic wildfire? *Journal of Forestry*, 99(11):9–17.

Camia, A., Amatulli, G., and San-Miguel-Ayanz, J. (2008). Past and future trends of forest fire danger in Europe. Technical report, Joint Research Centre.

CCST (2020). The costs of wildfire in California: An independent review of scientific and technical information. Technical report, California Council on Science and Technology, Sacramento, California.

Chen, H., Samet, J. M., Bromberg, P. A., and Tong, H. (2021). Cardiovascular health impacts of wildfire smoke exposure. *Particle and Fibre Toxicology*, 18(1):1–22.

Cleland, S. E. et al. (2022). Short-term exposure to wildfire smoke and PM2.5 and cognitive performance in a brain-training game: A longitudinal study of U.S. adults. *Environmental Health Perspectives*, 130:1–12.

Cliff, A. and Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4(3):267–284.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., and Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17):9391–9409.

Damania, R., Desbureaux, S., and Zaveri, E. (2020). Does rainfall matter for economic growth? Evidence from global sub-national data (1990–2014). *Journal of Environmental*

Economics and Management, 102:102335.

Davis, E. J., Moseley, C., Nielsen-Pincus, M., and Jakes, P. J. (2014). The community economic impacts of large wildfires: A case study from Trinity county, California. *Society and Natural Resources*, 27(9):983–993.

De Rigo, D., Libertà, G., Houston Durrant, T., Artés Vivancos, T., and San-Miguel-Ayanz, J. (2017). Forest fire danger extremes in Europe under climate change: Variability and uncertainty. Technical report, Publication Office of the European Union, Luxembourg.

De Zea Bermudez, P., Mendes, J., Pereira, J. M., Turkman, K. F., and Vasconcelos, M. J. (2009). Spatial and temporal extremes of wildfire sizes in Portugal. *International Journal of Wildland Fire*, 18(8):983–991.

Department of Commerce and Labor (1912). Mortality statistics: 1910. *U.S. Census Bureau*, 109:1–191.

Deryugina, T. (2022). Economic effects of natural disasters. *IZA World of Labor*, pages 1–10.

Deryugina, T., Kawano, L., and Levitt, S. (2018). The economic impact of hurricane Katrina on its victims: Evidence from individual tax returns. *American Economic Journal: Applied Economics*, 10(2):202–33.

Diaz, H. F. and Swetnam, T. W. (2013). The wildfires of 1910: Climatology of an extreme early twentieth-century event and comparison with more recent extremes. *Bulletin of the American Meteorological Society*, 94(9):1361–1370.

Doerr, S. H. and Santín, C. (2016). Global trends in wildfire and its impacts: Perceptions

versus realities in a changing world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696).

Doubleday, A. et al. (2020). Mortality associated with wildfire smoke exposure in Washington state, 2006-2017: A case-crossover study. *Environmental Health: A Global Access Science Source*, 19(1):1–10.

Dowdy, A. J. and Pepler, A. (2018). Pyroconvection risk in Australia: Climatological changes in atmospheric stability and surface fire weather conditions. *Geophysical Research Letters*, 45(4):2005–2013.

Dupuy, J.-l. et al. (2020). Climate change impact on future wildfire danger and activity in Southern Europe: A review. *Annals of Forest Science*, 77(35).

Egan, T. (2009). *The Big Burn: Teddy Roosevelt & the fire that saved America*. Houghton Mifflin Harcourt, New York.

European Environment Agency (2021). CORINE land cover - User manual. *Copernicus Land Monitoring Service*, 1.0:128.

Eurostat (2008). NACE Rev. 2 – Statistical classification of economic activities in the European community. Technical report, European Commission, Luxembourg.

Eurostat (2013a). European system of accounts ESA 2010. Technical report, European Commission, Luxembourg.

Eurostat (2013b). Manual on regional accounts methods. Technical report, European Commission, Luxembourg.

Eurostat (2020). Statistical regions in the European Union and partner countries. Technical report, European Commission, Luxembourg.

Evin, G., Curt, T., and Eckert, N. (2018). Has fire policy decreased the return period of the largest wildfire events in France? A Bayesian assessment based on extreme value theory. *Natural Hazards and Earth System Sciences*, 18(10):2641–2651.

FAO (2022). *The state of the world's forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies*. The Food and Agriculture Organization, Rome.

Felbermayr, G. and Gröschl, J. (2014). Naturally negative: The growth effects of natural disasters. *Journal of Development Economics*, 111:92–106.

Fernandes, P. M. (2019). Variation in the Canadian fire weather index thresholds for increasingly larger fires in Portugal. *Forests*, 10(10).

Fernandes, P. M., Barros, A. M., Pinto, A., and Santos, J. A. (2016). Characteristics and controls of extremely large wildfires in the Western Mediterranean Basin. *Journal of Geophysical Research: Biogeosciences*, 121(8):2141–2157.

Ferro, C. A. and Segers, J. (2003). Inference for clusters of extreme values. *Royal Statistical Society*, 65(2):545–556.

Fisher, R. A. (1937). *The design of experiments*. Oliver & Boyd, Edinburgh & London, 2nd edition.

Fomby, T., Ikeda, Y., and Loayza, N. (2013). The growth aftermath of natural disasters.

Journal of Applied Econometrics, 28:412–434.

Ford, B. et al. (2018). Future fire impacts on smoke concentrations, visibility, and health in the contiguous United States. *GeoHealth*, 2(2):229–247.

Foreman, T. (2020). SPATIAL_HAC_IV: Stata module to estimate an instrumental variable regression, adjusting standard errors for spatial correlation, heteroskedasticity, and autocorrelation. *Statistical Software Components S458872, Boston College Department of Economics*.

Fowler, D. et al. (2020). A chronology of global air quality. *Philosophical Transactions Royal Society A*, 378(20190314).

Gao, Y. et al. (2023). Long-term impacts of non-occupational wildfire exposure on human health: A systematic review. *Environmental Pollution*, 320:121041.

Garcia, A. and Heilmayr, R. (2022). Conservation impact evaluation using remotely sensed data. *Available at SSRN 4179782*.

Gellman, J., Walls, M., and Wibbenmeyer, M. (2022). Wildfire, smoke, and outdoor recreation in the Western United States. *Forest Policy and Economics*, 134:102619.

Gill, A. M. and Allan, G. (2008). Large fires, fire effects and the fire-regime concept. *International Journal of Wildland Fire*, 17(6):688–695.

Gilleland, E. and Katz, R. W. (2016). ExtRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, 72(8):1–39.

González-Cabán, A. (2009). Proceedings of the third international symposium on fire eco-

nomics, planning, and policy: Common problems and approaches. Technical report, United States Department of Agriculture, Forest Service, Albany, California.

Grant, E. and Runkle, J. D. (2022). Long-term health effects of wildfire exposure: A scoping review. *The Journal of Climate Change and Health*, 6:100110.

Groen, J. A., Kutzbach, M. J., and Polivka, A. E. (2020). Storms and jobs: The effect of hurricanes on individuals' employment and earnings over the long term. *Journal of Labor Economics*, 38(3):653–685.

Hanes, C. C., Wang, X., Jain, P., Parisien, M. A., Little, J. M., and Flannigan, M. D. (2019). Fire-regime changes in Canada over the last half century. *Canadian Journal of Forest Research*, 49(3):256–269.

Hänninen, O. O., Salonen, R. O., Koistinen, K., Lanki, T., Barregard, L., and Jantunen, M. (2009). Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. *Journal of Exposure Science and Environmental Epidemiology*, 19(4):414–422.

Harms, M. F. and Lavdas, L. G. (1997). Users guide to VSMOKE-GIS for workstations. Technical report, United States Department of Agriculture, Forest Service, Southern Research Station.

Heblich, S., Trew, A., and Zylberberg, Y. (2021). East-side story: Historical pollution and persistent neighborhood sorting. *Journal of Political Economy*, 129(5):1508–1552.

Hernandez, C., Keribin, C., Drobinski, P., and Turquety, S. (2015). Statistical modelling of

wildfire size and intensity: A step toward meteorological forecasting of summer extreme fire risk. *Annales Geophysicae*, 33(12):1495–1506.

Hersbach, H. et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Holm, S. M., Miller, M. D., and Balmes, J. R. (2021). Health effects of wildfire smoke in children and public health tools: A narrative review. *Journal of Exposure Science and Environmental Epidemiology*, 31(1):1–20.

Holmes, T. P., Prestemon, J. P., and Abt, K. L. (2008). *An Introduction to the Economics of Forest Disturbance*. Springer.

Horwich, G. (2000). Economic lessons of the Kobe earthquake. *Economic Development and Cultural Change*, 48(3):521–542.

Iammarino, S., Rodriguez-Pose, A., and Storper, M. (2019). Regional inequality in Europe: Evidence, theory and policy implications. *Journal of Economic Geography*, 19(2):273–298.

Jayachandran, S. (2009). Air quality and early-life mortality evidence from Indonesia's wildfires. *Journal of Human Resources*, 44(4):916–954.

Jiang, Y. and Zhuang, Q. (2011). Extreme value analysis of wildfires in Canadian boreal forest ecosystems. *Canadian Journal of Forest Research*, 41(9):1836–1851.

Jiménez-Ruano, A., Rodrigues Mimbrero, M., Jolly, W. M., and de la Riva Fernández, J. (2019). The role of short-term weather conditions in temporal dynamics of fire regime

features in mainland Spain. *Journal of Environmental Management*, 241:575–586.

Johnston, F., Hanigan, I., Henderson, S., Morgan, G., and Bowman, D. (2011). Extreme air pollution events from bushfires and dust storms and their association with mortality in Sydney, Australia 1994-2007. *Environmental Research*, 111(6):811–816.

Johnston, F. H. et al. (2021). Unprecedented health costs of smoke-related PM2.5 from the 2019–20 Australian megafires. *Nature Sustainability*, 4(1):42–47.

Jolly, W. M. et al. (2015). Climate-induced variations in global wildfire danger from 1979 to 2013. *Nature Communications*, 6:1–11.

Karácsonyi, D., Taylor, A., and Bird, D. (2021). *The demography of disasters: Impacts for population and place*. Springer Nature.

Karlinsky, A. and Kobak, D. (2021). Tracking excess mortality across countries during the covid-19 pandemic with the world mortality dataset. *eLife*, 10:1–21.

Katz, R. W., Brush, G. S., and Parlange, M. B. (2005). Statistics of extremes: Modeling ecological disturbances. *Ecology*, 86(5):1124–1134.

Keeley, J. E., Bond, W. J., Bradstock, R. A., Pausas, J. G., and Rundel, P. W. (2012). *Fire in Mediterranean ecosystems: Ecology, evolution and management*. Cambridge University Press, New York.

Keyser, A. R. and Westerling, A. L. R. (2019). Predicting increasing high severity area burned for three forested regions in the Western United States using extreme value theory. *Forest Ecology and Management*, 432:694–706.

Kim, M. K. and Jakus, P. M. (2019). Wildfire, national park visitation, and changes in regional economic activity. *Journal of Outdoor Recreation and Tourism*, 26:34–42.

Klomp, J. and Valckx, K. (2014). Natural disasters and economic growth: A meta-analysis. *Global Environmental Change*, 26(1):183–195.

Knorr, W., Arneth, A., and Jiang, L. (2016). Demographic controls of future global fire risk. *Nature Climate Change*, 6(8):781–785.

Kochi, I., Champ, P. A., Loomis, J. B., and Donovan, G. H. (2012). Valuing mortality impacts of smoke exposure from major Southern California wildfires. *Journal of Forest Economics*, 18(1):61–75.

Kollanus, V. et al. (2017). Mortality due to vegetation fire – originated PM 2.5 exposure in Europe - Assessment for the years 2005 and 2008. *Environmental Health Perspectives*, 125(1):30–37.

Krainz, T. A. (2012). Fleeing the big burn: Refugees, informal assistance, and welfare practices in the progressive Era. *Journal of Policy History*, 24(3):405–431.

Kramer, H. A., Butsic, V., Mockrin, M. H., Ramirez-Reyes, C., Alexandre, P. M., and Radeloff, V. C. (2021). Post-wildfire rebuilding and new development in California indicates minimal adaptation to fire risk. *Land Use Policy*, 107:105502.

Krawchuk, M. A., Cumming, S. G., and Flannigan, M. D. (2009). Predicted changes in fire weather suggest increases in lightning fire initiation and future area burned in the mixedwood boreal forest. *Climatic Change*, 92(1-2):83–97.

Lankoande, M. and Yoder, J. (2006). An econometric model of wildfire suppression productivity. *Washington State University School of Economic Sciences*, pages 1–23.

Larkin, N. K. et al. (2009). The BlueSky smoke modeling framework. *International Journal of Wildland Fire*, 18(8):906–920.

Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., and Bell, M. L. (2015). A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environmental Research*, 136:120–132.

Loayza, N. V., Olaberría, E., Rigolini, J., and Christiaensen, L. (2012). Natural disasters and growth: Going beyond the averages. *World Development*, 40(7):1317–1336.

Marlon, J. R. et al. (2012). Long-term perspective on wildfires in the Western USA. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9):535–543.

McCaffrey, S. (2004). Thinking of wildfire as a natural hazard. *Society and Natural Resources ISSN*, 17(6):509–516.

McConnell, K. et al. (2021). Effects of wildfire destruction on migration, consumer credit, and financial distress. *Federal Reserve Bank of Cleveland Working Paper No. 21-29*.

McCoy, S. J. and Walsh, R. P. (2018). Wildfire risk, salience & housing demand. *Journal of Environmental Economics and Management*, 91:203–228.

McElhinny, M., Beckers, J. F., Hanes, C., Flannigan, M., and Jain, P. (2020). A high-resolution reanalysis of global fire weather from 1979 to 2018 - Overwintering the Drought Code. *Earth System Science Data*, 12(3):1823–1833.

Mendes, J. M., de Zea Bermudez, P. C., Pereira, J., Turkman, K. F., and Vasconcelos, M. J. (2010). Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. *Environmental and Ecological Statistics*, 17(1):1–28.

Merlo, M. and Croitoru, L. (2005). *Valuing mediterranean forests - Towards total economic value*. CABI Publishing.

Michetti, M. and Pinar, M. (2019). Forest fires across Italian regions and implications for climate change: A panel data analysis. *Environmental and Resource Economics*, 72(1):207–246.

Miller, D. and Cohen, S. (2021). *The Big Burn: The Northwest's great forest fire of 1910*. Mountain Press Publishing Company, 8th printing edition.

Molina, J. R., González-Cabán, A., and Rodríguez y Silva, F. (2019). Wildfires impact on the economic susceptibility of recreation activities: Application in a Mediterranean protected area. *Journal of Environmental Management*, 245:454–463.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1):17–23.

Morita, T. et al. (2017). Excess mortality due to indirect health effects of the 2011 triple disaster in Fukushima, Japan: A retrospective observational study. *Journal of Epidemiology and Community Health*, 71(10):974–980.

Morton, D. C., Roessing, M. E., Camp, A. E., and Tyrrell, M. L. (2003). Assessing the environmental, social, and economic impacts of wildfire. *Global Institute of Sustainable Forestry*.

Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., and Wakefield, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613(7942):130–137.

Mueller, J. M. and Loomis, J. B. (2014). Does the estimated impact of wildfires vary with the housing price distribution? A quantile regression approach. *Land Use Policy*, 41:121–127.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403:853–858.

Naguib, C., Pelli, M., Poirier, D., and Tschopp, J. (2022). The impact of cyclones on local economic growth: Evidence from local projections. *CIRANO Working Paper*, 26.

NATURE (2017). Spreading like wildfire. *Nature Climate Change*, 7(11):755.

Nawrotzki, R. J., Brenkert-Smith, H., Hunter, L. M., and Champ, P. A. (2014). Wildfire-migration dynamics: Lessons from Colorado's Fourmile Canyon fire. *Society and Natural Resources*, 27(2):215–225.

Neath, A. A. and Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):199–203.

Nicholls, S. (2019). Impacts of environmental disturbances on housing prices: A review of the hedonic pricing literature. *Journal of Environmental Management*, 246:1–10.

Nielsen-Pincus, M., Moseley, C., and Gebert, K. (2013). The effects of large wildfires on employment and wage growth and volatility in the Western United States. *Journal of*

Forestry, 111(6):404–411.

Nielsen-Pincus, M., Moseley, C., and Gebert, K. (2014). Job growth and loss across sectors and time in the Western US: The impact of large wildfires. *Forest Policy and Economics*, 38:199–206.

Northrop, P. J. and Attalides, N. (2020). threshr: Threshold selection and uncertainty for extreme value analysis.

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society*, 66(1):93–120.

Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N., and Charlett, A. (2012). An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, 32(7):1206–1222.

Otrachshenko, V. and Nunes, L. C. (2022). Fire takes no vacation: Impact of fires on tourism. *Environment and Development Economics*, 27(1):86–101.

Pan, S. et al. (2023). Quantifying the premature mortality and economic loss from wildfire-induced PM2.5 in the contiguous U.S. *Science of the Total Environment*, 875:162614.

Parida, Y., Saini, S., and Chowdhury, J. R. (2021). Economic growth in the aftermath of floods in Indian states. *Environment, Development and Sustainability*, 23(1):535–561.

Pechony, O. and Shindell, D. T. (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. *Proceedings of the National Academy of Sciences*

of the United States of America, 107(45):19167–19170.

Pérez-Sánchez, J., Senent-Aparicio, J., Díaz-Palmero, J. M., and Cabezas-Cerezo, J. d. D. (2017). A comparative study of fire weather indices in a semiarid South-eastern Europe region. Case of study: Murcia (Spain). *Science of the Total Environment*, 590-591:761–774.

Pickands, J. (1971). The two-dimensional poisson process and extremal processes. *Journal of Applied Probability*, 8(4):745–756.

Pratt, J. R. et al. (2019). A national burden assessment of estimated pediatric asthma emergency department visits that may be attributed to elevated ozone levels associated with the presence of smoke. *Environmental Monitoring and Assessment*, 191.

Prichard, S. J. et al. (2020). Wildland fire emission factors in North America: Synthesis of existing data, measurement needs and management applications. *International Journal of Wildland Fire*, 29(2):132–147.

Pullabhotla, H. K. and Souza, M. (2022). Air pollution from agricultural fires increases hypertension risk. *Journal of Environmental Economics and Management*, 115:102723.

Rahn, M., Hale, K., Brown, C., and Edwards, T. (2014). Economic impacts of wildfires: 2003 San Diego wildfires in retrospect. Technical report, San Diego State University Wildfire Research Center.

Rangel, M. A. and Vogl, T. S. (2019). Agricultural fires and health at birth. *Review of Economics and Statistics*, 101(4):616–630.

Rego, F., Louro, G., and Constantino, L. (2013). The impact of changing wildfire regimes on wood availability from Portuguese forests. *Forest Policy and Economics*, 29:56–61.

Richardson, L. A., Champ, P. A., and Loomis, J. B. (2012). The hidden cost of wildfires: Economic valuation of health effects of wildfire smoke exposure in Southern California. *Journal of Forest Economics*, 18(1):14–35.

Roberts, P. (2019). *Tropical forests in prehistory, history, and modernity*. Oxford University Press.

Rocha, R. and Sant'Anna, A. A. (2022). Winds of fire and smoke: Air pollution and health in the Brazilian Amazon. *World Development*, 151:105722.

Rosales-Rueda, M. and Triyana, M. (2019). The persistent effects of early-life exposure to air pollution: Evidence from the Indonesian forest fires. *Journal of Human Resources*, 54(4):1037–1080.

Ruggles, S. et al. (2021). IPUMS Ancestry full count data: Version 3.0 1900-1940. *IPUMS*.

Safford, H. D., Paulson, A. K., Steel, Z. L., Young, D. J., and Wayman, R. B. (2022). The 2020 California fire season: A year like no other, a return to the past or a harbinger of the future? *Global Ecology and Biogeography*, pages 1–21.

San-Miguel-Ayanz, J. et al. (2018). Forest fires in Europe, Middle East and North Africa 2017. Technical report, Joint Research Centre, Luxembourg.

San-Miguel-Ayanz, J. et al. (2020). Forest fires in Europe, Middle East and North Africa 2019. Technical report, Joint Research Center, Luxembourg.

San-Miguel-Ayanz, J. et al. (2021). Forest fires in Europe, Middle East and North Africa 2020. Technical report, Joint Research Centre, Luxembourg.

San-Miguel-Ayanz, J. et al. (2022). Forest fires in Europe, Middle East and North Africa 2021. Technical report, Joint Research Centre, Luxembourg.

San-Miguel-Ayanz, J., Moreno, J. M., and Camia, A. (2013). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management*, 294:11–22.

Santín, C. and Doerr, S. H. (2016). Fire effects on soils: The human dimension. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696):28–34.

Santos-Burgoa, C. et al. (2018). Differential and persistent risk of excess mortality from Hurricane Maria in Puerto Rico: A time-series analysis. *The Lancet Planetary Health*, 2(11):e478–e488.

Sapkota, A. et al. (2005). Impact of the 2002 Canadian forest fires on particulate matter air quality in Baltimore City. *Environmental Science and Technology*, 39(1):24–32.

Schöllnberger, H., Aden, J., and Scott, B. R. (2002). Respiratory tract deposition efficiencies: Evaluation of effects from smoke released in the Cerro Grande Forest Fire. *Journal of Aerosol Medicine*, 15(4):387–399.

Scotto, M. G. et al. (2014). Area burned in Portugal over recent decades: An extreme value analysis. *International Journal of Wildland Fire*, 23(6):812–824.

Sharygin, E. (2021). Estimating migration impacts of wildfire: California's 2017 North Bay

fires. In *The Demography of Disasters*, pages 49–70. Springer.

Sheldon, T. L. and Zhan, C. (2022). The impact of hurricanes and floods on domestic migration. *Journal of Environmental Economics and Management*, 115:102726.

Skouras, S. and Christodoulakis, N. (2014). Electoral misgovernance cycles: Evidence from wildfires and tax evasion in Greece. *Public Choice*, 159(3):533–559.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, 4(4):367–377.

Smith, R. L. and Shively, T. S. (1995). Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold. *Atmospheric Environment*, 29(23):3489–3499.

Sousa, P. M., Trigo, R. M., Pereira, M. G., Bedia, J., and Gutiérrez, J. M. (2015). Different approaches to model future burnt area in the Iberian Peninsula. *Agricultural and Forest Meteorology*, 202:11–25.

Steel, Z. L., Safford, H. D., and Viers, J. H. (2015). The fire frequency-severity relationship and the legacy of fire suppression in California forests. *Ecosphere*, 6(1).

Stephenson, C., Handmer, J., and Robyn, B. (2013). Estimating the economic, social and environmental impacts of wildfires in Australia. *Environmental Hazards*, 12(2):93–111.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, 20(4):518–529.

Stowell, J. D. et al. (2019). Associations of wildfire smoke PM2.5 exposure with cardiorespiratory events in Colorado 2011–2014. *Environment International*, 133:105151.

Strobl, E. (2011). The economic growth impact of hurricanes: Evidence from U.S. coastal counties. *Review of Economics and Statistics*, 93(2):575–589.

Suades-González, E., Gascon, M., Guxens, M., and Sunyer, J. (2015). Air pollution and neuropsychological development: A review of the latest evidence. *Endocrinology (United States)*, 156(10):3473–3482.

Talbott, E. O. et al. (2015). Fine particulate matter and the risk of autism spectrum disorder. *Environmental Research*, 140:414–420.

Tan-Soo, J. S. and Pattanayak, S. K. (2019). Seeking natural capital projects: Forest fires, haze, and early-life exposure in Indonesia. *Proceedings of the National Academy of Sciences of the United States of America*, 116(12):5239–5245.

Tarín-Carrasco, P. et al. (2021). Impact of large wildfires on PM10 levels and human mortality in Portugal. *Natural Hazards and Earth System Sciences*, 21(9):2867–2880.

Tedim, F. et al. (2018). Defining extreme wildfire events: Difficulties, challenges, and impacts. *Fire*, 1(1):9.

Tedim, F. et al. (2020). What can we do differently about the extreme wildfire problem. In *Extreme wildfire events and disasters: Root causes and new management strategies*, chapter 13, pages 233–263. Elsevier Inc.

Turco, M. et al. (2018). Exacerbated fires in Mediterranean Europe due to anthropogenic

warming projected with non-stationary climate-fire models. *Nature Communications*, 9(1):1–9.

Turco, M., Herrera, S., Tourigny, E., Chuvieco, E., and Provenzale, A. (2019). A comparison of remotely-sensed and inventory datasets for burned area in Mediterranean Europe. *International Journal of Applied Earth Observation and Geoinformation*, 82.

UNEP (2022). Spreading like wildfire: The rising threat of extraordinary landscape fires. Technical report, United Nations Environment Programme, Nairobi.

Van Wagner, C. E. and Pickett, T. L. (1985). Equations and FORTRAN program for the Canadian forest fire weather index system. Technical report, Canadian Forestry Service, Ottawa.

Van Wagtendonk, J. W. (2007). The history and evolution of wildland fire. *Fire Ecology*, 3(2):3–17.

Verma, V., Polidori, A., Schauer, J. J., Shafer, M. M., Cassee, F. R., and Sioutas, C. (2009). Physicochemical and toxicological profiles of particulate matter in Los Angeles during the October 2007 Southern California wildfires. *Environmental Science and Technology*, 43(3):954–960.

Wang, D. et al. (2021). Economic footprint of California wildfires in 2018. *Nature Sustainability*, 4(3):252–260.

Wen, B., Wu, Y., Xu, R., Guo, Y., and Li, S. (2022). Excess emergency department visits for cardiovascular and respiratory diseases during the 2019–20 bushfire period in Australia: A

two-stage interrupted time-series analysis. *Science of the Total Environment*, 809:152226.

Wen, J. and Burke, M. (2022). Lower test scores from wildfire smoke exposure. *Nature Sustainability*, 5(11):947–955.

Williams, A. P. et al. (2019). Observed impacts of anthropogenic climate change on wildfire in California. *Earth's Future*, 7(8):892–910.

Winkler, R. L. and Rouleau, M. D. (2021). Amenities or disamenities? Estimating the impacts of extreme heat and wildfire on domestic US migration. *Population and Environment*, 42(4):622–648.

Xue, T. et al. (2021). Associations between exposure to landscape fire smoke and child mortality in low-income and middle-income countries: A matched case-control study. *The Lancet Planetary Health*, 5(9):e588–e598.

Ye, T. et al. (2021). Risk and burden of hospital admissions associated with wildfire-related PM2·5 in Brazil, 2000–15: A nationwide time-series study. *The Lancet Planetary Health*, 5(9):e599–e607.

A Appendix for Chapter 1

A.1 Supplementary Material

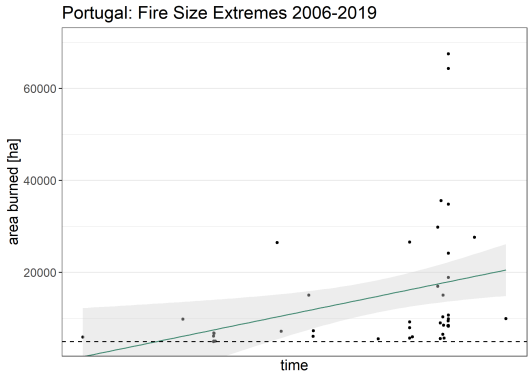
Table A1: List of covariates.

category	name	type	description
FWI	*FWI_InitDat*	numeric	FWI on the reported initial date of the fire
	FWI_MP	numeric	mean FWI of the month prior to the initial date of the fire
	FWI_WP	numeric	mean FWI of the week prior to the initial date of the fire
	FWI_MP_WA	numeric	mean FWI of the month prior until the week after the initial date
	FWI_Mean	numeric	annual FWI mean of corresponding year of the fire incidence
	FWI_q0.5	numeric	annual FWI median of corresponding year of the fire incidence
	FWI_q0.9	numeric	0.9 quantile of annual FWI of the corresponding year of the fire incidence
	FWI_q0.95	numeric	0.95 quantile of annual FWI of the corresponding year of the fire incidence
	FWI_q0.99	numeric	0.99 quantile of annual FWI of the corresponding year of the fire incidence
population density	*Pop_4km*	numeric	mean population density in approx. 4 km buffer around the perimeter centroid (counts per square kilometer)
land cover type	*Type_I*	indicator	predominant land cover type I
	Type_II	indicator	predominant land cover type II
	Type_III	indicator	predominant land cover type III
	Type_Other	indicator	1 if not of types I - III
seasonality	*DJF*	indicator	fire in winter months (December, January, February)
	MAM	indicator	fire in spring months (March, April, Mai)
	JJA	indicator	fire in summer months (June, July, August)
	SON	indicator	fire in autumn months (September, October, November)

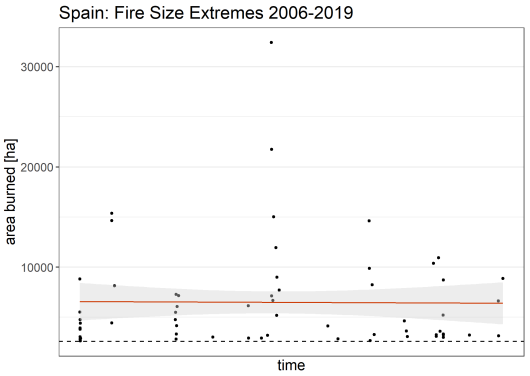
Table A2: Country-level dominant CORINE land cover types.

country	Type I	Type II	Type III
Portugal	Transitional woodland-shrub	Moors and heathland	Broad-leaved forest Coniferous forest Mixed forest
Spain	Broad-leaved forest Coniferous forest	Sclerophyllous vegetation	Moors and heathland
France	Sclerophyllous vegetation	Natural grasslands	Transitional woodland-shrub
Italy	Sclerophyllous vegetation	Non-irrigated arable land Pastures	Agro-forestry area Broad-leaved forest Coniferous forest
Greece	Sclerophyllous vegetation	Transitional woodland-shrub	Land principally occupied by agriculture

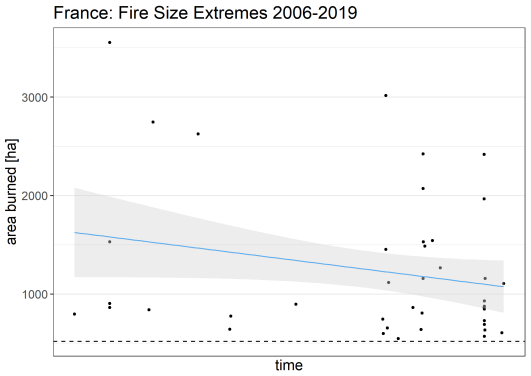
Figure A1: BA over the study period (2006-2019) for the extremes (observations exceeding indicated threshold u) with a generalised linear model smoothed conditional mean with CIs on the 90% level.



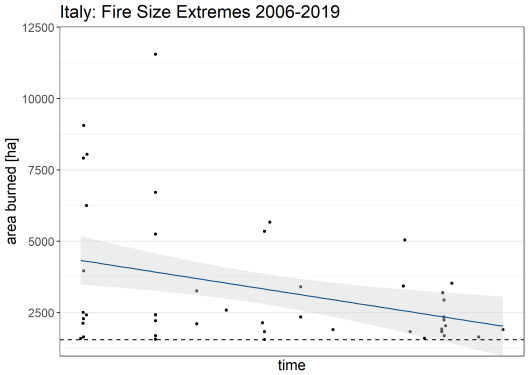
(a) Portugal extremes



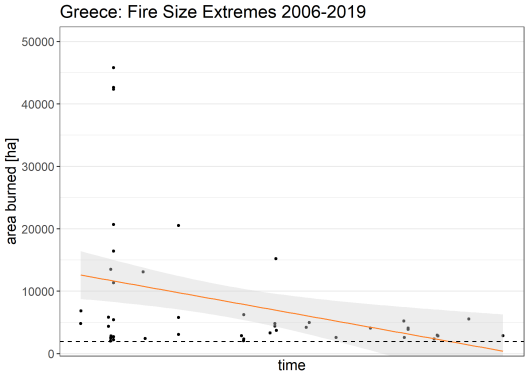
(b) Spain extremes



(c) France extremes

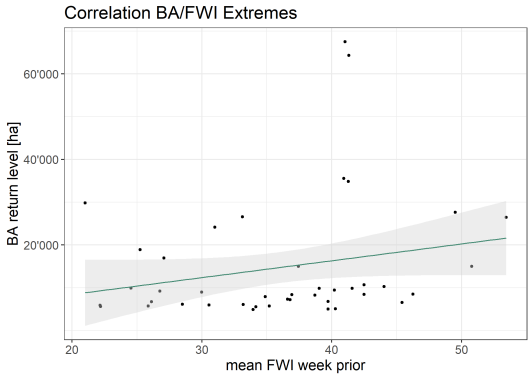


(d) Italy extremes

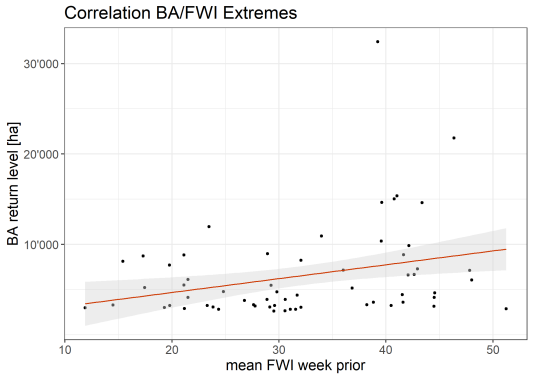


(e) Greece extremes

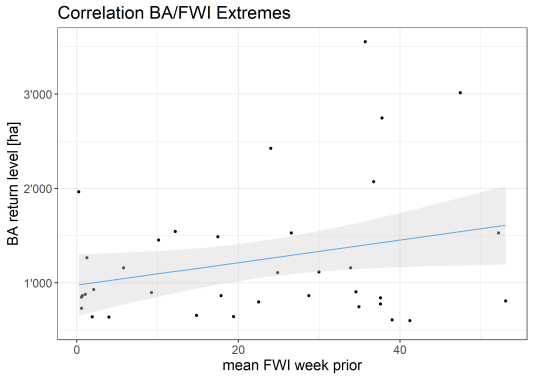
Figure A2: Correlation between BA and the mean FWI of the week prior to the initial date of the fire with a generalised linear model smoothed conditional mean with CIs on the 90% level.



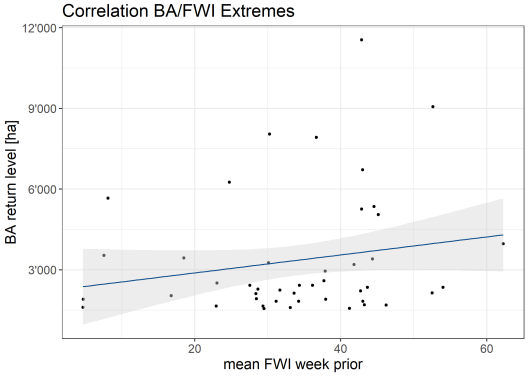
(a) Portugal



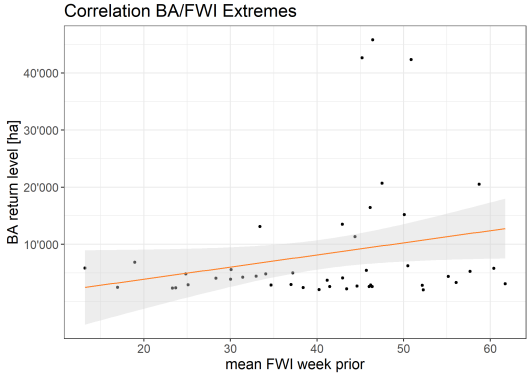
(b) Spain



(c) France

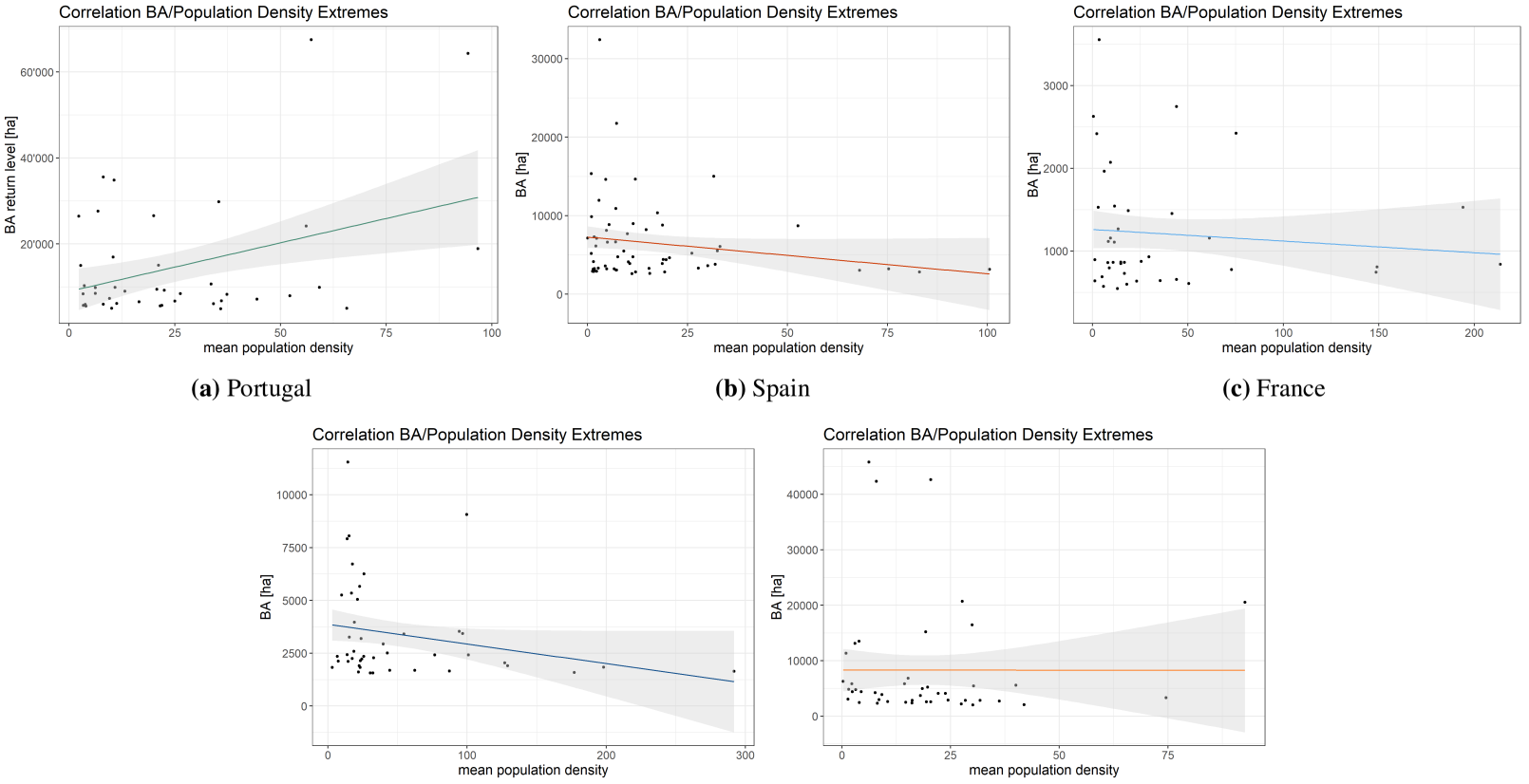


(d) Italy



(e) Greece

Figure A3: Correlation between BA and population density with a generalised linear model smoothed conditional mean with CIs on the 90% level. Outlier observations are excluded (253 for Spain, 822 for Italy, as well as 1320, and 507 for France.)



(a) Portugal **(b)** Spain **(c)** France

(d) Italy **(e)** Greece

A.2 Generalised Extreme Value (GEV) Distribution

This section is entirely based on Coles (2001) and Gilleland and Katz (2016).

Let $M_n = max\{X_1, .., X_n\}$ be the maximum of a sequence of i.i.d. random variables X_i with a common distribution function F. Then, the distribution of M_n can theoretically be derived as

$$\begin{aligned} \Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\ &= \Pr\{X_1 \leq z\} \times \cdots \times \Pr\{X_n \leq z\} \\ &= \{F(z)\}^n. \end{aligned} \tag{4}$$

However, as F is unknown, we look for approximate models for F^n. Therefore, we are allowing a linear normalisation of M_n for a sequence of real numbers $a_n > 0$ and b_n, so that $\frac{M_n - b_n}{a_n}$ converges in distribution as $n \to \infty$.

$$\Pr\{(M_n - b_n)/a_n \leq z\} \xrightarrow{n\to\infty} G(z) \tag{5}$$

$G(z)$ belongs to the family of GEV distributions[34] of the form

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \tag{6}$$

defined on a set $z : 1 + \xi(z - \mu)/\sigma > 0$ with the scale and shape parameters satisfying $-\infty < \sigma < \infty$ and $-\infty < \xi < \infty$.

[34] Given it is a non-generate distribution function.

The GEV distribution introduced in Equation (6) envelops three types of limit distributions determined by the sign of the shape parameter ξ. The *Fréchet* distribution results from $\xi > 0$ and is defined by a heavy tail (density of $G(z)$ decays polynomially). $\xi < 0$ indicates an upper bounded *Weibull* distribution function. Finally, the *Gumbel* type (density of $G(z)$ decays exponentially) results by taking the limit as $\xi \to 0$ leading to

$$G(z) = \exp\left[-\exp\left\{-\left(\frac{z-\mu}{\sigma}\right)\right\}\right], -\infty < z < \infty. \tag{7}$$

B Appendix for Chapter 2

B.1 Fire Weather Index Equations

Given the complexity of the FWI calculations, we limit ourselves to illustrate the derivations of the direct fire behaviour inputs to the FWI, which are the Initial Spread Index (ISI) and the Buildup Index (BUI). Thus, we will not elude to the underlying functions of the respective inputs, namely the function of wind $f(W)$, the fine fuel moisture function $f(F)$, today's Duff Moisture Code denoted as P, and today's Drought Code denoted as D. The exposition of the exact equations here forth strongly draws on Van Wagner and Pickett (1985), where the full set of equations based on the primary input variables is described.

The Initial Spread Index is defined by Equation (8), whereby $f(W)$ is a function of wind and $f(F)$ is the fine fuel moisture function.

$$\text{ISI} = 0.208 * f(W) * f(F) \tag{8}$$

The Buildup Index shown in Equation (9) is a function of today's Duff Moisture Code (DMC)

denoted as P and today's Drought Code (DC) denoted as D.

$$\text{BUI} = \begin{cases} 0.8 * PD/(P + 0.4 * D), & \text{if } P \leq 0.4 * D \\ \text{P} - [1 - 0.8 * D/(P + 0.4 * D)][0.92 + (0.0114 * P)^{1.7}], & \text{if } P > 0.4 * D \end{cases} \tag{9}$$

The output of Equation (9) is subsequently used as input to calculate the duff moisture function, $f(D)$, shown in Equation (10).

$$f(D) = \begin{cases} 0.626 * \text{BUI}^{0.809} + 2, & \text{if BUI} \leq 80 \\ 1000(25 + 108.64e^{-0.023 * \text{BUI}}, & \text{if BUI} > 80 \end{cases} \tag{10}$$

Equation (11) derives B, which is an intermediate form of the FWI, by scaling and multiplying today's ISI with the duff moisture function.

$$B = 0.1 * \text{ISI} * f(D) \tag{11}$$

Finally, Equation (12) shows the derivation of the FWI in it's final form.

$$\text{FWI} = \begin{cases} B, & \text{if B} \leq 1 \\ 2.72(0.434 * ln(B))^{0.647}, & \text{if } B > 1 \end{cases} \tag{12}$$

B.2 Supplementary Material

Table B1: CORINE land cover types and reclassification.

Reclassification	CORINE land cover type
Urban	**1. Artificial surfaces**
	1.1 Urban fabric
	1.2 Industrial, commercial and transport units
	1.3 Mine, dump and construction sites
	1.4 Artificial, non-agricultural vegetated areas
Agriculture	**2. Agricultural areas**
	2.1 Arable land
	2.2 Permanent crops
	2.3 Pastures
	2.4 Heterogeneous agricultural areas
Forest	**3. Forest and seminatural areas**
	3.1 Forests
	3.2 Shrub and/or herbaceous vegetation associations
Wetlands and water bodies	3.3 Open spaces with little or no vegetation
	4. Wetlands
	4.1 Inland wetlands
	4.2 Coastal wetlands
	5. Water bodies
	5.1 Inland waters
	5.2 Marine waters

Note: This overview is based on information provided by Copernicus. For greater detail on sub-classifications, see https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html (accessed in September 2022).

Table B2: Fire Weather Index classification based on Van Wagner and Pickett (1985).

Fire danger	FWI ranges
Very low	< 5.2
Low	5.2 - 11.2
Moderate	11.2 - 21.3
High	21.3 - 38
Very high	38 - 50
Extreme	> 50

Note: Upper bound excluded in the FWI ranges.

Table B3: Descriptive statistics of employment categories (2010-2018).

	Min	Mean	sd	Median	Max	Proportion (%)
Category A	0	11,276	9,943	8,400	72,900	9.7
Category B-E	350	32,598	43,850	17,700	391,500	14.7
Category F	400	14,186	19,707	8,800	219,300	6.7
Category G-J	2,750	62,519	107,080	33,200	1,154,500	28.8
Category K-N	460	32,137	73,532	13,900	776,800	11.1
Category O-U	1,620	65,144	109,268	37,100	1,092,800	29

Notes: (i) economic data are used from the Statistical Office of the European Union (Eurostat); (ii) the letter refers to the NACE activity where A includes agriculture, forestry and fishing, B-E is industry except construction, F indicates construction, G-J includes wholesale and retail trade, transport, accommodation and food service activities, information and communication, K-N contains financial and insurance activities, real estate activities, professional, scientific and technical activities, administrative and support service activities, and O-U includes public administration and defence, compulsory social security, education, human health and social work activities, arts, entertainment and recreation, and repair of household goods and other services; (iii) Proportion (%) = average of employment activity category divided by the total employment for each observation; (iv) sd = standard deviation.

C Appendix for Chapter 3

C.1 Supporting Information

Table C1: County-level boundary changes from 1900 to 1910.

FIPS	County 1910	State	Matching Area 1900	Comment
16017	Bonner	ID	61% Kootenai	21 Feb 1907 (W08): Kootenai split in Kootenai and Bonner
16031	Cassia	ID	57% Cassia	21 Feb 1907 (W08): Cassia split in Cassia and Twin Falls
16055	Kootenai	ID	39% Kootenai	21 Feb 1907 (W08): Kootenai split in Kootenai and Bonner
16069	Nez Perce	ID	100% Nez Perce & 53% Shoshone	1904: Nez Perce gained from Shoshone
16079	Shoshone	ID	47% Shoshone	1904: Shoshone lost to Nez Perce
16083	Twin Falls	ID	43% Cassia	21 Feb 1907 (W08): Cassia split in Cassia and Twin Falls
30017	Custer	MT	58% Custer	1901: Custer split in Custer and Rosebud
30023	Deer Lodge	MT	14% Deer Lodge & 32% Silver Bow	1901: Deer Lodge lost to Powell 1903 Deer Lodge gained from Silver Bow
30029	Flathead	MT	64% Flathead	1 Jul 1909 (W26): Flathead split in Flathead and Lincoln
30053	Lincoln	MT	36% Flathead	1 Jul 1909 (W26): Flathead split in Flathead and Lincoln
30063	Missoula	MT	56% Missoula	1 Mar 1906 (W09): Missoula split in Missoula and Sanders
30077	Powell	MT	84% Deer Lodge	1901: Powell created from Deer Lodge *
30087	Rosebud	MT	42% Custer	1901: Custer split in Custer and Rosebud
30089	Sanders	MT	44% Missoula	1 Mar 1906 (W09): Missoula split in Missoula and Sanders
30093	Silver Bow	MT	68% Silver Bow	1903: Silver Bow lost to Deer Lodge
53005	Benton	WA	24% Yakima & 17% Klikatat	8 Mar 1905 (W10): Benton created from Klikatat and Yakima
53017	Douglas	WA	40% Douglas	24 Feb 1909 (W08): Douglas split in Douglas and Grant
53025	Grant	WA	60% Douglas	24 Feb 1909 (W08)16: Douglas split in Douglas and Grant
53039	Klikatat	WA	80% Klikatat	8 Mar 1905 (W10): Part of Klikatat lost to creation of Benton
53077	Yakima	WA	76% Yakima	8 Mar 1905 (W10): Part of Yakima lost to creation of Benton

Notes: * 0.3% of the area was used from Lewis and Clark but we do not consider attributions lower than 1%; (i) the exact taking effect date of the boundary change is only indicated if it affects the mortality data i.e., changes from 1905 to 1909; (ii) in the comment column the ISO week is indicated in parentheses; (iii) the shape files of the county boundaries in the year 1900 and 1910 are obtained from https://geo.btaa.org and the exact information on the boundary changes is from https://digital.newberry.org/ahcb/project.html.

Table C2: Ancestry data sources for county-level deaths from 1905-1910.

General
• U.S., Find a Grave Index, 1600s-Current • U.S., Newspapers.com Obituary Index, 1800s-current • U.S., Presbyterian Church Records, 1701-1970
Montana
• Beaverhead County, Montana, U.S., County Records, 1862-2009 • Missoula and Ravalli County, Montana Cemeteries • Montana, U.S., County Births and Deaths, 1830-2011 • Montana, U.S., State Deaths, 1907-2018 • Web: Gallatin County, Montana, U.S., Death Index, 1856-2014
Idaho
• Idaho, U.S., County Birth and Death Records, 1863-1970 • Idaho, U.S., Death Index, 1890-1964 • Idaho, U.S., Death Records, 1890-1969 • Idaho, U.S., Select Deaths and Burials, 1907-1965 • Salt Lake City, Utah, U.S., Cemetery Records, 1848-1992 • Salt Lake County, Utah, U.S., Death Records, 1908-1949
Washington
• Washington, U.S., Death Index, 1940-2017 • Washington, U.S., Death Records, 1883-1960 • Washington, U.S., Select Death Index, 1907-1960

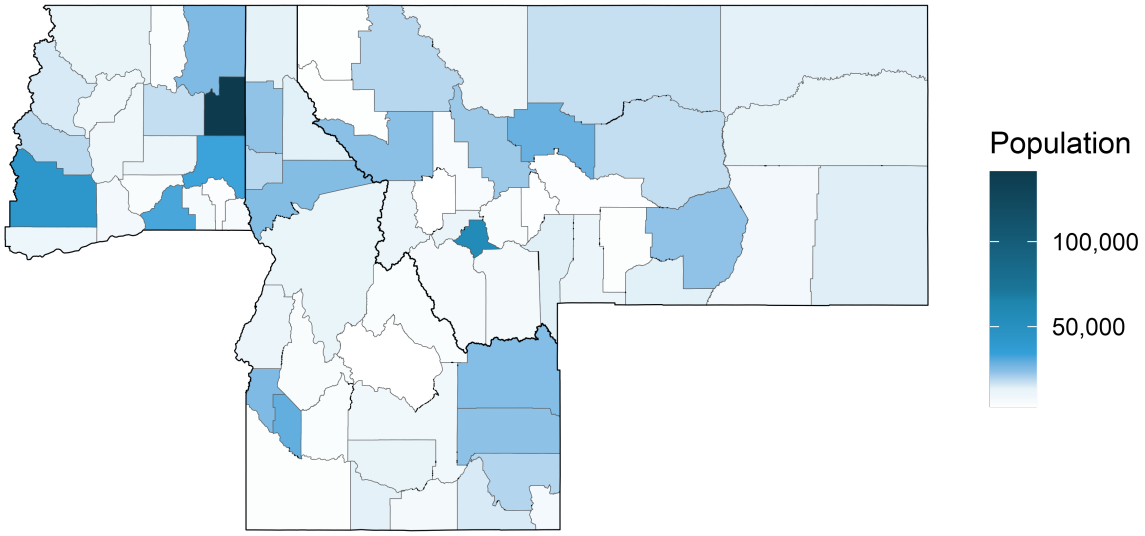
Note: The source for this table is the genealogy company Ancestry.com.

Table C3: The individual steps of the data cleaning process of the mortality data

1. Cleaning date of death
• Remove non-numeric characters e.g., “Ab”, “abt”, “about”, “unknown date”* • Drop observations if the date of death is longer than 11 characters • Drop observations with year range e.g., 1900-1913 • Drop observations only including the month and year of death but not the day
2. Cleaning names
• Drop observations if name includes “Void” • Remove special characters from names e.g., numbers, “Sgt”, “1lt”* • Drop observations if the name includes characters such as “?”, “...”, “#”, “$”
3. Cleaning date of birth and age at death
• Remove characters from date of birth e.g., “ab”, “about”, “unknown”, “not obtainable”, “Don’t know about”* • Replace the birth date with NA if no characters are left • Add 1st of January if only birth month and year are given in the date of birth • Drop observations if the birth year is missing • Adjust all two-digit years to four-digit years • Drop observations with a negative age • Infer age from date of birth and date of death if age is missing
4. Sub-setting children under the age of five
• Subset observations with the age at death under five years
5. Removing duplicates
• Drop duplicate observations if several observations are identical on the variables name, gender, date of death, state, and county FIPS • Drop duplicate observations if several observations are identical on the variables date of birth, date of death, age at death, and county FIPS • Drop duplicate observations if several observations are identical on the variables date of birth, name, age at death, and county FIPS • Drop duplicate observations if several observations are identical on the variables gender, name, birth year, death year, and county FIPS

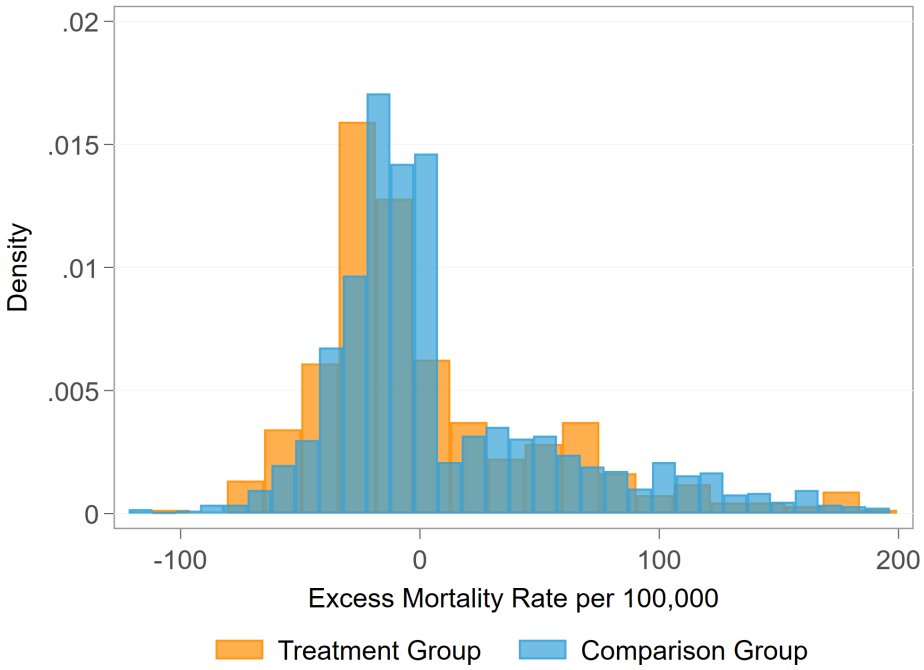
Notes: * These steps also include an extensive number of typos and misspellings; (i) The above table shows the specific steps that are undertaken to clean the digitised mortality data obtained from Ancestry.com; (ii) FIPS denotes the abbreviation for Federal Information Processing Standard and is a 5-digit county identification code.

Figure C1: County-level population 1910



Notes: (i) the map shows the distribution of population at the county level in 1910 where the highest populated county is Spokane County, Washington; (ii) the county shape file shows the historical county boundaries provided by The Big Ten Academic Alliance Geoportal; (iii) county-level population is retrieved from the Integrated Public Use Microdata Series USA 1910 full-count census.

Figure C2: Average weekly excess mortality rate for children under the age of five in the ∼8 months before the week of the fire by treatment and comparison group.



Notes: (i) Observations with and excess mortality rate exceeding 200 are excluded from this figure; (ii) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA; (iii) the mortality data is retrieved from the genealogy company Ancestry.com.

Table C4: Descriptive statistics of the control variables at the county level for the 70 sample counties in 1910.

	Min	Mean	SD	Median	Max
Average age	22.0	28.2	2.1	28.5	32.2
Average Socioeconomic indicator	15.4	20.4	2.6	19.9	30.1
Average family size	2.9	4.0	0.7	3.8	6.2
Number of families in household	1.2	2.8	0.9	2.8	5.3
Non-family household (%)	2.3	9.3	4.6	8.4	22.9
Place $>$ 1,000 habitants (%)	0.0	11.7	13.0	10.5	48.7
Farm household (%)	2.3	39.5	15.1	42.1	73.4
Rented property (%)	15.4	33.0	9.9	33.6	55.0
Paying mortgage (%)	4.4	24.8	12.1	24.3	56.6
Multigenerational household (%)	68.7	81.3	5.6	81.9	92.6
Non-white (%)	0.0	3.3	4.8	1.7	30.0
Non-american parent (%)	21.3	45.1	11.9	45.0	76.8
In school (%)	11.6	19.2	4.8	18.1	32.8
Unable to read and write (%)	0.5	4.3	3.7	3.3	20.4
Unemployed (%)	1.1	6.1	3.5	5.5	17.3
In labour force (%)	3.3	8.2	3.7	7.3	26.2
Ind. Agriculture, forestry & fishing (%)	4.2	48.3	18.1	52.4	78.4
Ind. Mining (%)	0.0	7.0	11.1	1.8	53.4
Ind. Construction (%)	1.1	6.6	4.0	5.4	23.5
Ind. Manufacturing (%)	1.5	7.5	8.5	4.5	50.1
Ind. Transportation, communication & utilities (%)	2.5	13.4	8.3	11.6	37.4
Ind. Wholesale and retail trade (%)	3.8	7.7	2.5	7.4	15.3
Ind. Finance, insurance, and real estate (%)	0.3	1.3	0.7	1.2	4.3
Ind. Business and repair services (%)	0.4	1.1	0.4	1.1	2.3
Ind. Personal services (%)	1.3	3.2	1.2	2.9	8.4
Ind. Entertainment and related services (%)	0.0	0.3	0.2	0.3	1.0
Ind. Professional and related services (%)	0.8	2.2	0.7	2.1	4.1
Ind. Public administration (%)	0.3	1.4	1.3	1.0	9.0

Notes: (i) SD = standard deviation, Ind. = Industry; (ii) the variables are obtained from the Integrated Public Use Microdata Series USA 1910 full-count individual and household censuses.

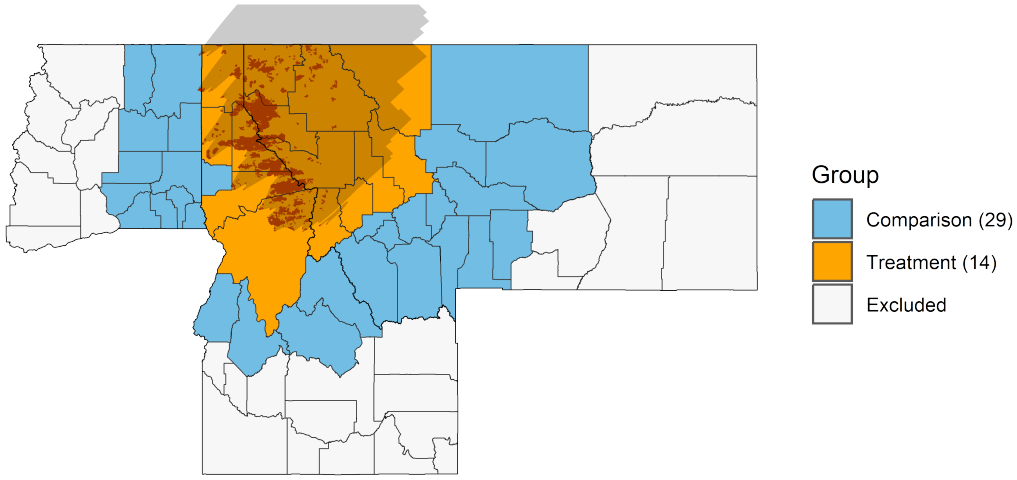
Table C5: Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 16).

	Excess Mortality Rate		
	(1)	(2)	(3)
Smoke event	32.4	49.3*	23.5*
	(16.8)	(19.8)	(11.0)
1 week after event	52.9	45.8	150.1
	(47.6)	(50.3)	(96.5)
2 weeks after event	13.6	23.6	17.1
	(19.0)	(24.1)	(21.3)
3 weeks after event	-4.2	5.4	-12.7
	(31.4)	(34.4)	(24.2)
4 weeks after event	-19.6	-11.1	8.6
	(22.5)	(25.0)	(24.5)
5 weeks after event	-14.5	-22.0	-21.9
	(27.4)	(31.9)	(18.7)
6 weeks after event	-12.4	-6.4	-13.5
	(23.0)	(26.2)	(16.3)
7 weeks after event	-15.7	-2.2	-8.3
	(19.0)	(22.9)	(14.1)
8 weeks after event	7.9	6.3	57.3
	(36.0)	(38.2)	(66.5)
9 weeks after event	-21.0	-7.3	4.7
	(18.9)	(21.5)	(15.9)
10 weeks after event	-10.5	-3.2	15.1
	(20.1)	(22.8)	(22.8)
11 weeks after event	3.1	15.0	19.0
	(25.9)	(28.4)	(16.2)
12 weeks after event	38.5	42.7	107.3
	(43.8)	(45.5)	(73.4)
13 weeks after event	-11.2	-27.9	1.9
	(27.1)	(32.0)	(20.8)
14 weeks after event	-12.3	-18.1	26.5
	(27.4)	(31.0)	(39.8)
15 weeks after event	-16.8	-16.8	17.6
	(23.1)	(26.1)	(25.7)
16 weeks after event	-11.2	-3.2	21.5
	(28.9)	(31.5)	(36.2)
Controls			
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.06	0.07	0.10
N	2,310	2,310	2,310

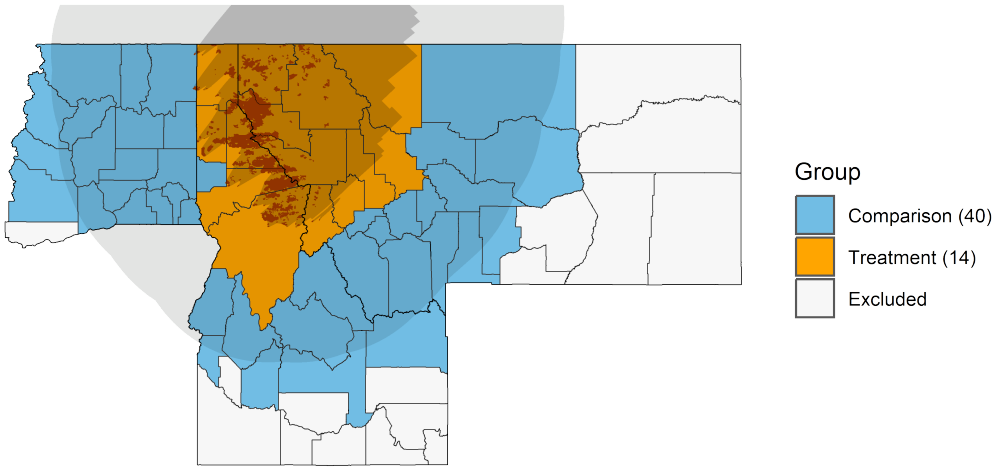
Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Equation (3.2); (ii) the sample includes 70 counties; (iii) Burned area denotes the percentage burned of a county; (iv) standard errors are clustered at the county level; (v) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data is retrieved from the genealogy company Ancestry.com.

C.2 Robustness Checks

Figure C3: Treatment and comparison group for smoke-affected counties only including comparison counties surrounding smoke-affected counties.



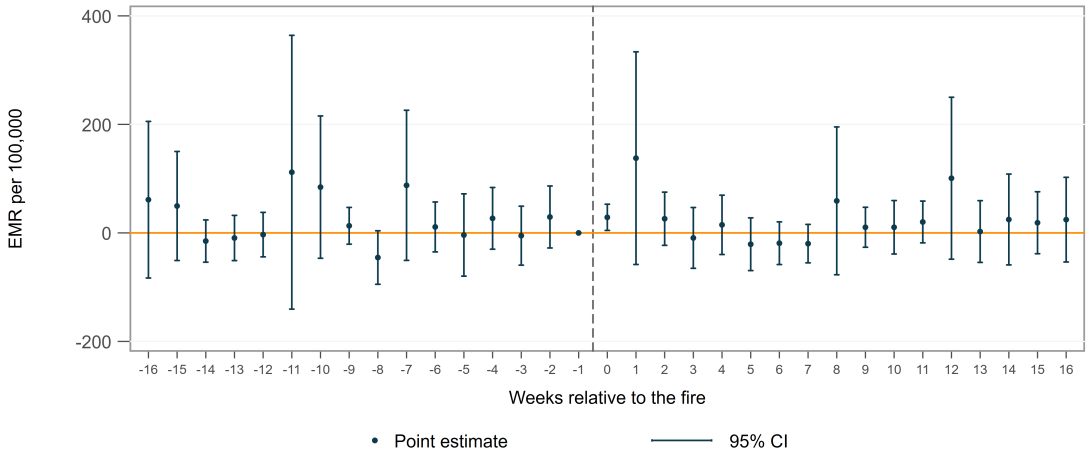
(a) Visual inspection



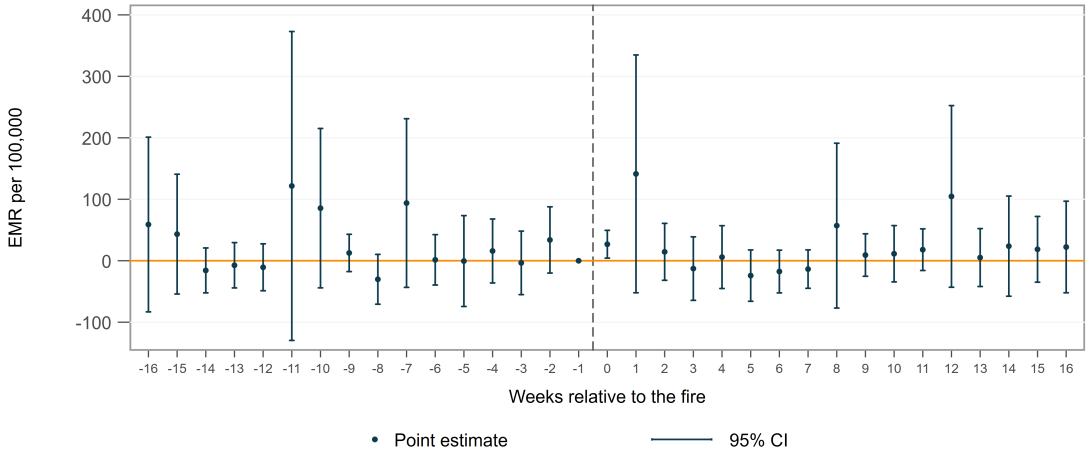
(b) Within 250-kilometre buffer

Notes: (i) a county is classified as smoke affected if any part of the county was exposed to hazardous hourly peak pollution (PM2.5>526 $\mu g/m^3$); (ii) the county shape file shows the historical county boundaries provided by The Big Ten Academic Alliance Geoportal; (iii) the lighter grey area shows the moderate hourly peak pollution area and the darker shaded grey area indicates the 250-kilometre buffer around this area.

Figure C4: Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five only including comparison counties surrounding smoke affected counties.



(a) Upon visual inspection



(b) Within 250-kilometre buffer

Notes: (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data is obtained from the genealogy company Ancestry.com and the population data is retrieved from the 1910 US census.

Table C6: Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 16) only including 29 surrounding comparison counties upon visual inspection.

	Excess Mortality Rate		
	(1)	(2)	(3)
Smoke event	32.4	55.2*	28.7*
	(17.0)	(20.7)	(12.0)
1 week after event	52.9	33.9	137.8
	(48.0)	(55.3)	(97.1)
2 weeks after event	13.6	35.7	26.1
	(19.1)	(29.3)	(24.3)
3 weeks after event	-4.2	9.3	-9.2
	(31.7)	(38.5)	(27.8)
4 weeks after event	-19.6	-1.5	14.8
	(22.7)	(28.8)	(27.1)
5 weeks after event	-14.5	-20.5	-20.9
	(27.6)	(37.6)	(24.1)
6 weeks after event	-12.4	-14.2	-19.0
	(23.2)	(29.9)	(19.5)
7 weeks after event	-15.7	-18.2	-19.8
	(19.2)	(27.0)	(17.5)
8 weeks after event	7.9	11.3	59.1
	(36.3)	(41.3)	(67.5)
9 weeks after event	-21.0	1.2	10.4
	(19.0)	(22.8)	(18.3)
10 weeks after event	-10.5	-8.8	10.4
	(20.3)	(26.1)	(24.4)
11 weeks after event	3.1	16.9	20.2
	(26.1)	(31.0)	(19.1)
12 weeks after event	38.5	36.9	100.8
	(44.2)	(48.3)	(74.0)
13 weeks after event	-11.2	-25.6	2.5
	(27.3)	(40.8)	(28.2)
14 weeks after event	-12.3	-18.3	24.7
	(27.6)	(34.9)	(41.5)
15 weeks after event	-16.8	-13.6	18.7
	(23.3)	(30.1)	(28.3)
16 weeks after event	-11.2	2.0	24.4
	(29.2)	(34.8)	(38.7)
Controls			
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.05	0.07	0.11
N	1,419	1,419	1,419

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Equation (3.2); (ii) the sample includes 70 counties; (iii) Burned area denotes the percentage burned of a county; (iv) standard errors are clustered at the county level; (v) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data is retrieved from the genealogy company Ancestry.com.

Table C7: Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 16) only including 40 surrounding comparison courtiers within a 250-kilometre buffer.

	Excess Mortality Rate		
	(1)	(2)	(3)
Smoke event	32.4	53.4*	26.9*
	(16.9)	(20.0)	(11.3)
1 week after event	52.9	35.8	141.3
	(47.8)	(52.1)	(96.5)
2 weeks after event	13.6	19.9	14.5
	(19.1)	(26.1)	(23.1)
3 weeks after event	-4.2	4.9	-12.8
	(31.6)	(36.2)	(25.8)
4 weeks after event	-19.6	-14.4	5.9
	(22.6)	(26.6)	(25.5)
5 weeks after event	-14.5	-25.1	-24.2
	(27.5)	(33.9)	(20.8)
6 weeks after event	-12.4	-12.1	-17.5
	(23.1)	(27.5)	(17.3)
7 weeks after event	-15.7	-9.6	-13.6
	(19.1)	(24.8)	(15.6)
8 weeks after event	7.9	7.1	57.1
	(36.1)	(39.4)	(66.9)
9 weeks after event	-21.0	-0.7	9.3
	(18.9)	(22.0)	(17.2)
10 weeks after event	-10.5	-8.0	11.4
	(20.2)	(23.6)	(22.8)
11 weeks after event	3.1	13.8	18.0
	(26.0)	(29.0)	(16.9)
12 weeks after event	38.5	40.4	104.7
	(44.0)	(46.6)	(73.7)
13 weeks after event	-11.2	-22.7	5.2
	(27.2)	(35.2)	(23.5)
14 weeks after event	-12.3	-20.9	23.8
	(27.5)	(33.1)	(40.6)
15 weeks after event	-16.8	-14.7	18.6
	(23.2)	(27.5)	(26.7)
16 weeks after event	-11.2	-1.5	22.4
	(29.0)	(32.8)	(37.1)
Controls			
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.06	0.07	0.11
N	1,782	1,782	1,782

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Equation (3.2); (ii) the sample includes 70 counties; (iii) Burned area denotes the percentage burned of a county; (iv) standard errors are clustered at the county level; (v) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data is retrieved from the genealogy company Ancestry.com.

Table C8: Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood only including surrounding comparison counties (upon visual inspection).

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.3	-1.8	0.2	-1.7	-1.0
	(0.7)	(1.7)	(0.7)	(1.4)	(1.1)
Burned area	0.7	0.6	2.0	4.3*	5.3***
	(1.0)	(2.5)	(1.0)	(2.1)	(1.4)
Individual					
Non-white	-4.1**	-12.8**	-3.9*	-13.4***	-7.6**
	(1.2)	(4.1)	(1.7)	(3.5)	(2.3)
Non-american born parent	-0.2	-1.0	-0.1	-1.1	-0.4
	(0.4)	(1.1)	(0.5)	(1.0)	(0.7)
Household (1910)					
Family size	-0.2***	-0.7**	-0.3**	-0.8***	-0.7***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	-0.1	-0.3	0.3	0.0	0.4
	(0.2)	(0.6)	(0.1)	(0.4)	(0.3)
Non-family household	2.8	7.6	3.1	7.6	6.9
	(2.8)	(8.9)	(3.1)	(6.8)	(4.6)
Urban household	1.0*	2.1	2.1**	4.9***	4.9***
	(0.5)	(1.6)	(0.7)	(1.3)	(1.0)
Farm household	-1.1*	-3.0*	-0.2	-3.7*	-2.3*
	(0.5)	(1.4)	(0.6)	(1.5)	(1.1)
Paying mortgage	-0.1	-0.2	-0.0	-0.0	0.2
	(0.2)	(0.6)	(0.4)	(0.6)	(0.5)
Parents (1910)					
Mother: Age	0.1**	0.2**	0.1**	0.3***	0.2***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	-0.4	-0.8	-0.3	-0.5	-0.4
	(0.4)	(1.1)	(0.4)	(0.9)	(0.6)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.0
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-american born parent	-0.5	-1.6	0.2	-0.6	0.3
	(0.4)	(1.0)	(0.5)	(1.0)	(0.8)
Father: Education score	0.0	0.0	0.1**	0.1	0.1*
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0*	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	0.1	0.6	0.1	-0.1	-1.1
	(0.5)	(1.6)	(0.9)	(1.8)	(1.3)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.14	0.14	0.07	0.16	0.12
N	4,877	4,838	4,838	4,838	4,877

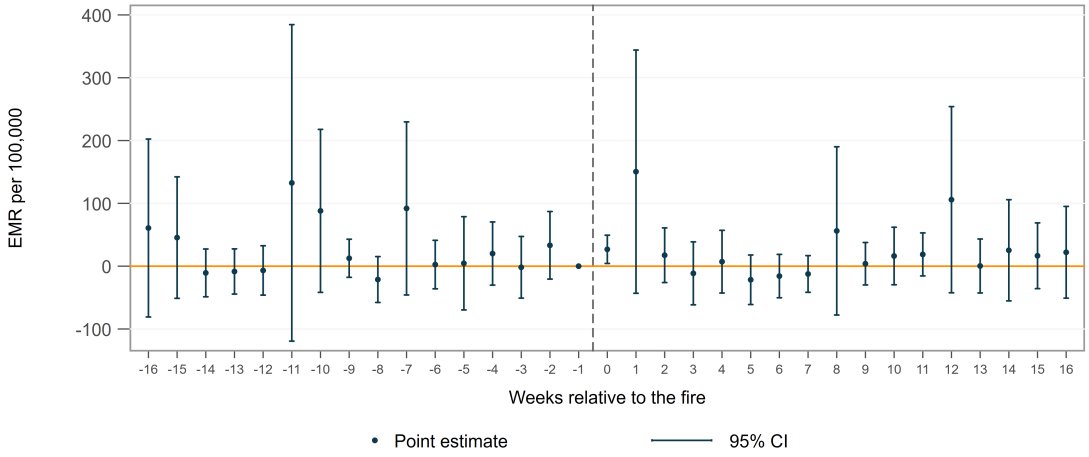
Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Table C9: Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood only including surrounding comparison counties (within 250-kilometre buffer).

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.3	-1.6	-0.4	-1.9	-1.7
	(0.6)	(1.6)	(0.6)	(1.4)	(0.9)
Burned area	0.1	-0.3	0.5	1.8	2.8*
	(0.6)	(1.7)	(0.7)	(1.5)	(1.1)
Individual					
Non-white	-4.5***	-14.1***	-5.2**	-15.2***	-8.6***
	(0.9)	(3.1)	(1.7)	(3.0)	(2.1)
Non-american born parent	-0.2	-0.8	-0.2	-1.1	-0.5
	(0.4)	(1.1)	(0.5)	(1.0)	(0.7)
Household (1910)					
Family size	-0.3***	-0.8***	-0.4***	-0.9***	-0.8***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	-0.2	-0.5	0.2	-0.2	0.2
	(0.2)	(0.5)	(0.2)	(0.4)	(0.3)
Non-family household	3.0	8.8	4.1	9.0	8.0
	(2.2)	(6.9)	(2.7)	(5.6)	(4.1)
Urban household	1.2**	2.6	2.0**	4.8***	4.8***
	(0.4)	(1.5)	(0.6)	(1.2)	(0.9)
Farm household	-1.4**	-3.8**	-0.2	-3.8**	-2.1*
	(0.4)	(1.3)	(0.5)	(1.3)	(0.9)
Paying mortgage	0.1	0.6	0.3	0.7	0.7
	(0.2)	(0.6)	(0.3)	(0.5)	(0.4)
Parents (1910)					
Mother: Age	0.1***	0.3***	0.2***	0.3***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	-0.4	-1.0	-0.2	-0.6	-0.3
	(0.3)	(0.8)	(0.3)	(0.7)	(0.5)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.0
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-american born parent	-0.4	-1.4	0.1	-0.3	0.6
	(0.3)	(0.8)	(0.4)	(0.8)	(0.6)
Father: Education score	0.0	0.0	0.1**	0.1	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0**	0.1***	0.0*	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	-0.2	-0.1	-0.3	-0.4	-1.2
	(0.4)	(1.2)	(0.7)	(1.3)	(1.0)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.13	0.13	0.06	0.14	0.11
N	6,730	6,676	6,676	6,676	6,730

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Figure C5: Difference-in-Differences point estimates and 95% confidence intervals of smoke exposure on excess mortality of children under the age of five excluding Spokane County and Missoula County.



Notes: (i) EMR denotes the excess mortality rate and CI indicates the 95% confidence interval; (ii) the mortality data is obtained from the genealogy company Ancestry.com and the population data is retrieved from the 1910 US census.

Table C10: Difference-in-Differences regression results of smoke exposure on the excess mortality rate of children under the age of five (weeks 0 to 16) excluding Spokane County and Missoula County.

	Excess Mortality Rate		
	(1)	(2)	(3)
Smoke event	38.5*	55.7**	26.8*
	(16.9)	(20.0)	(11.3)
1 week after event	56.9	50.0	150.5
	(51.1)	(53.7)	(97.0)
2 weeks after event	14.1	24.1	17.4
	(20.4)	(25.5)	(21.8)
3 weeks after event	-1.5	7.9	-11.4
	(33.8)	(36.7)	(25.1)
4 weeks after event	-22.0	-13.6	7.1
	(24.2)	(26.6)	(25.0)
5 weeks after event	-12.8	-20.9	-21.6
	(29.4)	(33.8)	(19.7)
6 weeks after event	-17.7	-11.7	-15.8
	(24.2)	(27.3)	(17.3)
7 weeks after event	-25.2	-11.6	-12.4
	(17.8)	(22.0)	(14.6)
8 weeks after event	6.6	5.3	56.2
	(38.7)	(40.9)	(67.1)
9 weeks after event	-24.0	-9.3	3.9
	(20.1)	(22.7)	(16.9)
10 weeks after event	-8.7	-0.5	16.2
	(21.5)	(24.2)	(23.0)
11 weeks after event	2.3	14.5	18.7
	(27.9)	(30.3)	(17.2)
12 weeks after event	36.1	41.1	105.8
	(47.1)	(48.8)	(74.2)
13 weeks after event	-13.3	-30.2	0.3
	(29.1)	(33.9)	(21.5)
14 weeks after event	-14.1	-19.5	25.3
	(29.4)	(33.0)	(40.4)
15 weeks after event	-18.4	-18.2	16.6
	(24.8)	(27.7)	(26.2)
16 weeks after event	-9.3	-1.1	22.1
	(31.1)	(33.7)	(36.6)
Controls			
County fixed effects	✓	✓	✓
Week fixed effects		✓	✓
Burned area			✓
R^2	0.07	0.07	0.10
N	2,244	2,244	2,244

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$; (ii) the table shows the coefficients of the Difference-in-Differences estimation as stated in Equation (3.2); (ii) Burned area denotes the percentage burned of a county; (iii) standard errors are clustered at the county level; (iv) the population data is compiled from the 1910 US full-count census provided by the Integrated Public Use Microdata Series (IPUMS) USA and the mortality data is retrieved from the genealogy company Ancestry.com.

Table C11: Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood excluding Spokane County and Missoula County.

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.4	-1.2	-0.6	-1.9	-2.1**
	(0.4)	(1.2)	(0.5)	(1.0)	(0.7)
Burned area	-0.5	-2.4	-0.1	0.2	2.6*
	(0.5)	(1.4)	(0.8)	(1.4)	(1.0)
Individual					
Non-white	-5.4***	-15.9***	-5.3**	-15.4***	-7.6***
	(1.0)	(3.0)	(1.8)	(2.6)	(1.8)
Non-american born parent	-0.2	-0.7	-0.5	-1.1	-0.9
	(0.3)	(1.0)	(0.4)	(0.9)	(0.6)
Household (1910)					
Family size	-0.2***	-0.6***	-0.5***	-0.9***	-0.8***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	-0.1	-0.3	0.2	0.0	0.2
	(0.2)	(0.5)	(0.2)	(0.4)	(0.3)
Non-family household	0.8	3.7	2.1	4.3	5.1
	(2.1)	(6.2)	(2.5)	(5.3)	(3.6)
Urban household	1.2*	2.9	1.4**	4.1**	4.1***
	(0.5)	(1.6)	(0.5)	(1.2)	(0.8)
Farm household	-1.7***	-4.9***	-1.0*	-4.9***	-2.7***
	(0.3)	(1.0)	(0.5)	(1.0)	(0.7)
Paying mortgage	0.1	0.5	0.3	0.6	0.5
	(0.2)	(0.6)	(0.3)	(0.5)	(0.4)
Parents (1910)					
Mother: Age	0.1***	0.3***	0.2***	0.3***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	-0.2	-0.7	0.2	-0.1	0.1
	(0.2)	(0.7)	(0.3)	(0.6)	(0.4)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-american born parent	-0.4	-1.5	-0.0	-0.7	0.5
	(0.3)	(0.8)	(0.3)	(0.7)	(0.5)
Father: Education score	0.0	0.0	0.1*	0.1*	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0*	0.1***	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	-0.3	-0.8	-0.5	-1.5	-1.5
	(0.4)	(1.1)	(0.6)	(1.1)	(0.8)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.11	0.12	0.05	0.12	0.09
N	8,107	8,041	8,041	8,041	8,107

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Table C12: Regression of individual socioeconomic status outcomes in 1940 on smoke exposure in early childhood excluding Spokane County and Missoula County.

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.6	-0.9	0.2	0.0	0.6
	(0.4)	(1.2)	(0.5)	(1.1)	(0.9)
Burned area	0.4	-0.1	-0.2	-0.3	0.1
	(0.5)	(1.6)	(0.6)	(1.4)	(1.1)
Individual					
Non-white	-2.5*	-8.1*	-5.6**	-11.8***	-8.9**
	(1.2)	(3.1)	(1.8)	(2.9)	(3.0)
Non-american born parent	-0.3	-1.1	-0.4	-1.3	-0.9
	(0.3)	(0.9)	(0.3)	(0.8)	(0.6)
Household (1910)					
Family size	-0.3***	-0.8***	-0.5***	-1.0***	-1.0***
	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)
Families in household	0.3	0.6	0.6**	0.7	0.7*
	(0.2)	(0.4)	(0.2)	(0.4)	(0.4)
Non-family household	-0.6	-1.5	-0.1	0.9	1.1
	(2.0)	(5.8)	(1.8)	(4.8)	(4.0)
Urban household	0.9	1.1	0.8	2.3	2.9*
	(0.5)	(1.2)	(0.6)	(1.2)	(1.2)
Farm household	-1.3***	-3.8***	-0.4	-3.4**	-2.0*
	(0.4)	(1.0)	(0.5)	(1.0)	(0.9)
Paying mortgage	-0.2	-0.1	-0.4	-0.2	-0.8
	(0.2)	(0.6)	(0.3)	(0.6)	(0.5)
Parents (1910)					
Mother: Age	0.1***	0.2***	0.2***	0.3***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	0.0	-0.6	0.3	-0.6	0.2
	(0.2)	(0.7)	(0.3)	(0.6)	(0.5)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Father: Non-american born parent	-0.2	-1.3	0.5	-0.7	0.3
	(0.2)	(0.7)	(0.3)	(0.7)	(0.5)
Father: Education score	0.1***	0.1**	0.1***	0.1***	0.2***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0***	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	0.4	1.5	-0.1	1.5	0.7
	(0.5)	(1.5)	(0.6)	(1.2)	(1.1)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.08	0.08	0.06	0.10	0.10
N	9,505	9,480	9,480	9,480	9,364

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Table C13: Regression of individual socioeconomic status outcomes in 1930 on smoke exposure in early childhood using ABE standard conservative matching algorithm.

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.7	-1.8	-0.9	-2.7*	-3.0***
	(0.5)	(1.4)	(0.5)	(1.2)	(0.8)
Burned area	-0.1	-1.6	0.4	1.0	3.1*
	(0.6)	(1.8)	(0.8)	(1.6)	(1.2)
Individual					
Non-white	-6.1***	-18.2***	-5.9***	-17.2***	-8.7***
	(0.9)	(3.2)	(1.5)	(2.6)	(1.7)
Non-american born parent	-0.2	-0.8	-0.3	-0.9	-0.6
	(0.4)	(1.1)	(0.5)	(1.0)	(0.7)
Household (1910)					
Family size	-0.2***	-0.6***	-0.5***	-0.9***	-0.8***
	(0.1)	(0.2)	(0.1)	(0.2)	(0.1)
Families in household	-0.0	-0.1	0.3	0.2	0.3
	(0.2)	(0.5)	(0.2)	(0.4)	(0.3)
Non-family household	1.0	4.7	0.9	3.1	2.1
	(2.2)	(6.7)	(2.5)	(5.6)	(3.7)
Urban household	1.3**	3.2*	1.9***	5.0***	4.9***
	(0.4)	(1.3)	(0.5)	(1.0)	(0.7)
Farm household	-1.6***	-4.8***	-0.7	-4.6***	-2.3**
	(0.4)	(1.2)	(0.5)	(1.2)	(0.8)
Paying mortgage	0.2	0.7	0.5	0.9	0.8*
	(0.2)	(0.6)	(0.3)	(0.5)	(0.4)
Parents (1910)					
Mother: Age	0.1***	0.3***	0.2***	0.4***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)
Mother: Non-american born parent	-0.4	-1.1	-0.1	-0.6	-0.3
	(0.2)	(0.7)	(0.3)	(0.6)	(0.4)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)
Father: Non-american born parent	-0.4	-1.3	0.0	-0.5	0.5
	(0.3)	(0.8)	(0.3)	(0.7)	(0.5)
Father: Education score	0.0	-0.0	0.0*	0.1	0.1**
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0**	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	-0.4	-1.2	-0.5	-1.5	-1.4
	(0.4)	(1.1)	(0.6)	(1.1)	(0.8)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.13	0.13	0.06	0.14	0.11
N	7,801	7,738	7,738	7,738	7,801

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1930; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.

Table C14: Regression of individual socioeconomic status outcomes in 1940 on smoke exposure in early childhood using ABE standard conservative matching algorithm.

	Income			Income & Education	
	(1) INCOME	(2) ERSCOR	(3) PRESGL	(4) NPBOSS	(5) DSEI
Smoke	-0.9*	-2.2	0.5	-1.1	0.1
	(0.4)	(1.3)	(0.6)	(1.3)	(1.0)
Burned area	0.9	1.8	0.1	1.2	0.6
	(0.6)	(2.1)	(0.7)	(2.0)	(1.3)
Individual					
Non-white	-3.7**	-10.4**	-7.5***	-14.4***	-11.3***
	(1.3)	(3.3)	(1.6)	(2.9)	(2.7)
Non-american born parent	-0.6	-1.8*	-0.4	-1.7*	-1.0
	(0.3)	(0.9)	(0.3)	(0.7)	(0.6)
Household (1910)					
Family size	-0.3***	-0.8***	-0.6***	-1.1***	-1.1***
	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)
Families in household	0.2	0.6	0.4	0.6	0.6
	(0.2)	(0.4)	(0.2)	(0.4)	(0.3)
Non-family household	1.1	3.2	1.0	4.8	3.7
	(2.1)	(6.1)	(1.7)	(4.8)	(4.2)
Urban household	1.2**	1.8	1.1*	3.2**	3.6***
	(0.4)	(1.0)	(0.5)	(1.0)	(1.0)
Farm household	-1.1**	-3.3**	0.1	-2.9**	-1.2
	(0.4)	(1.1)	(0.5)	(1.1)	(0.9)
Paying mortgage	-0.0	0.5	-0.3	0.4	-0.3
	(0.2)	(0.7)	(0.3)	(0.6)	(0.5)
Parents (1910)					
Mother: Age	0.1***	0.2***	0.2***	0.4***	0.3***
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Mother: Non-american born parent	0.2	-0.1	0.3	-0.3	0.4
	(0.2)	(0.6)	(0.3)	(0.6)	(0.5)
Father: Age	-0.0	-0.1	-0.0	-0.1	-0.1
	(0.0)	(0.1)	(0.0)	(0.1)	(0.1)
Father: Non-american born parent	-0.1	-1.0	0.7	-0.4	0.6
	(0.3)	(0.8)	(0.4)	(0.8)	(0.6)
Father: Education score	0.1***	0.1***	0.1***	0.2***	0.2***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Earnings score	0.0***	0.1***	0.0**	0.1***	0.1***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
Father: Unemployed	0.5	1.4	0.5	1.6	1.4
	(0.5)	(1.4)	(0.6)	(1.2)	(1.1)
Controls					
Industry father	✓	✓	✓	✓	✓
County characteristics	✓	✓	✓	✓	✓
State indicator	✓	✓	✓	✓	✓
R^2	0.10	0.09	0.07	0.12	0.12
N	9,184	9,162	9,162	9,162	9,042

Notes: (i) stars indicate significance according to * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; (ii) this table shows the results of the Ordinary Least Squares regression shown in Equation (3.3) estimation the effect of wildfire smoke exposure in early childhood on later-life socioeconomic status conditional on controls; (iii) the data is obtained from the Integrated Public Use Microdata Series full-count censuses 1910 and 1940; (iv) INCOME stands for the Occupational Income Score (in 100s US$), ERSCOR is the Occupational Earnings Score, PRESGL denotes the Occupational Siegel Prestige Score, NPBOSS is the Nam-Powers-Boyd Occupational Status Score, and DSEI represents the Duncan Socioeconomic Index.