I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16988

On the Effects of Wildfires on Poverty in Bolivia

Gustavo Canavire-Bacarreza
Alejandro Puerta-Cuartas
Andrey Ramos

MAY 2024

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16988

On the Effects of Wildfires on Poverty in Bolivia

Gustavo Canavire-Bacarreza
World Bank, Universidad Privada Boliviana and IZA

Alejandro Puerta-Cuartas
Universidad Carlos III de Madrid

Andrey Ramos
Universidad Carlos III de Madrid

MAY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

On the Effects of Wildfires on Poverty in Bolivia*

This paper examines the impact of severe wildfire events on Bolivia's poverty and labor market outcomes. We use a panel from 2005 to 2020 utilizing NASA's MODIS Collection-6 MCD64A1 burned area product and household surveys. To attain survey representativeness at a lower geographical level, we aggregate neighboring municipalities using the max-p-region algorithm. Using the Interactive Fixed Effects Counterfactual Estimator, we estimate the causal effects of severe wildfire events on poverty, household per-capita income, and the agricultural sector. We find a significant short-term increase in poverty explained by a temporary decline in household per capita and, specifically, agricultural labor income.

JEL Classification: I32, Q54, J43
Keywords: poverty, counterfactual estimators, natural disasters

Corresponding author:
Gustavo Canavire Bacarreza
The World Bank
1818 H Street
NW Washington
DC 20433
USA
E-mail: gcanavire@worldbank.org

* The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the World Bank, its Board of Directors, or the countries it represents. We thank Juan Carlos Duque for his advice on implementing the max-p region algorithm and Jennifer Alix-Garcia, Alejandro Del Valle and Fernando Landa for their comments and suggestions. All remaining errors are our own.

1 Introduction

Wildfires are becoming more frequent and severe globally. Factors such as warmer temperatures, an increasing frequency of drought periods, and the growing rate of deforestation make forests and other vegetation more prone to ignition, increasing the risk of wildfires. Bolivia is among the countries most affected by wildfires, with an estimated 4.52 million hectares (Mha) of forested and savanna areas having burned in 2020 (Bustillo et al., 2021; Singh et al., 2022). Fire ignition in Bolivia is largely caused by slash-and-burn practices implemented by local communities to make way for agricultural and urban expansion (Carmenta et al., 2011; Devisscher et al., 2019). These fires often expand and become uncontrollable, leading to large fire events that cause significant damage to ecosystems and communities (Bustillo et al., 2021), specially to those locations with low capacity to fight against fires or challenging geographical conditions.

In addition to their negative effects on the composition, structure, and functioning of ecosystems (Jones and McDermott, 2021), wildfires can have significant labor market impacts on affected areas. On the one hand, wildfires destroy infrastructure, crops, and businesses, leading to job losses, migration, and reduced economic activity. On the other hand, wildfires can lead to increased economic activity through public spending in recovery efforts, creating jobs and stimulating economic growth (Nielsen-Pincus et al., 2014). While the literature is inconclusive on the overall impact of wildfires on labor markets, a deeper comprehension on this subject is essential for welfare improvement and fostering resilience in areas affected by this type of natural disaster. This is particularly important for Bolivia, a country with high rates of poverty and informal labor and an economy largely dependent on agriculture, forestry, and mining sectors, all of which can be significantly affected by wildfires.

In this paper, we investigate the poverty and labor market effects of severe wildfires in Bolivia from 2005 to 2020. A panel of annual burned areas by geographic unit is constructed using the NASA's Moderate Resolution Imaging Spectro-radiometer (MODIS) Collection-6 MCD64A1 burned area product. We combine this dataset with individual-level socioeconomic information obtained from the Encuesta Nacional de Hogares (ENH) conducted annually by the Bolivian Instituto Nacional de Estadística (INE). A key limitation of our socioeconomic data is its lack of representativeness at the municipality level, making it unreliable to compute aggregates at such level. To tackle this issue, we opt for an intermediate level of geographical aggregation by clustering municipalities with similar burned areas and socioeconomic characteristics using the max-p-region algorithm developed by Duque et al. (2012). Empirical evidence is provided for the validity of the regional clusterization.

We estimate the causal effects of severe wildfires on various socioeconomic outcomes, including poverty, income, and occupation, using the counterfactual estimators for causal inference introduced by Liu et al. (2022). This methodology allows us to compute the dynamic average treatment effects on the treated (ATT) by directly imputing counterfactual outcomes for treated units, with our units corresponding to geographical clusters. This approach is particularly useful when studying wildfire treatments, because it accommodates the ability to switch the treatment indicator on and off. However, this methodology requires the definition of a valid treatment indicator. Directly using the proportion of burned areas relative to the total region size is not suitable, due to the nonrandom occurrence of wildfires across regions. Moreover, the continuous presence of wildfires can lead to stronger adaptation and mitigation strategies in affected areas, potentially masking the true effects of wildfires on socioeconomic outcomes. To overcome this issue, we propose to identify severe wildfire events by considering the historical evolution of wildfires in each geographical unit. Therefore, a wildfire event is classified as severe if its deviation from the historical mean is sufficiently high and the treatment indicator is defined accordingly. The identifying assumption is the unexpected nature of severe wildfires: these events are considered unforeseen and could not have been predicted in advance, enabling us to treat the treatment variable as exogenous with respect to the socioeconomic trends in the affected region.

Our main finding reveals a significant increase in poverty following a severe wildfire event. Poverty rises by 8 percent in the year after a wildfire and this effect is 7.7 percent after two years. However, after three years, the effect becomes statistically insignificant, indicating that the impact of wildfires is short-lived. The mechanism behind this finding is the reduction in agricultural income, consistent with the damage to crops and agricultural infrastructure, resulting in a decrease in individuals' income following a severe wildfire.

The remainder of this paper proceeds as follows: section 2 describes the data sources. Section 3 presents the methodology used for regional clusterization, the definition of the treatment, the identifying assumptions, and the estimation of causal effects. Section 4 reports the results of our empirical analysis and robustness checks. Section 5 concludes.

2 Data

To examine how severe wildfires impact socioeconomic outcomes in Bolivia, we must gather dependable data on (i) the extent of areas affected by wildfires in specific geographic regions, and (ii) the socioeconomic and demographic factors that serve as outcomes and control variables in our empirical analysis. We describe the data collection and processing in this section.

2.1 Burned areas

We obtained data on burned areas from the MODIS Collection-6 burned area mapping product (MCD64), as developed by Giglio et al. (2018). This product provides detailed information about the spatial extent and estimated date of biomass burning on a global scale, with a spatial resolution of 500 meters. It utilizes daily 500-meter MODIS surface reflectance data and 1-kilometer MODIS active fire observations, employing a detection algorithm to establish probabilistic thresholds for categorizing individual 500-meter grid cells as either burned or unburned.[1] These data classify burned areas through the Julian day of the given month in each monthly GeoTIFF. Subsequently, we aggregated these features annually, creating yearly raster data sets that were then overlaid with a shapefile of Bolivia at the municipality level. Through this process, we constructed a panel for burned areas in Bolivian municipalities from 2005 to 2020.

2.2 Socioeconomic data

The source of socioeconomic information for this study is the ENH conducted by the INE in Bolivia. The ENH contains information on the socioeconomic indicators for individuals and households across the entire country, including rural areas. The variables contained in the ENH span the location of the households and a wide range of household members' characteristics such as sex, age, educational attainment, occupation, sector of activity, income, and poverty. In the ENH, a household is considered to be poor if its income is below the poverty line. Accordingly, we use a monetary binary measure for poverty.

To have comparable variables across time, we homogenize the surveys and draw both

[1] The MODIS burned area data are acquired from the University of Maryland's website in the form of GeoTIFF files.

the dependent variables and the controls, which allows us to identify the effect of severe wildfires on socioeconomic outcomes. After pooling the surveys, the individual-level data set comprises information on about 450,000 individuals over the study period 2005-2020. It should be noted that there are no data available for 2010, because the ENH was not conducted during that year. In our empirical analysis, we treat the year 2010 as missing.

2.3 Aggregation at the geographical level

An important limitation of the ENH data set is its cross-sectional nature, meaning that the same individuals are not followed over time. Because most methodologies employed to estimate the effects of wildfires on socioeconomic outcomes rely on having a panel of observations, it is necessary to aggregate our individual-level observations to a certain geographical extent.

While the most detailed identification geographical location provided by the ENH is the municipality of residence, the survey is representative only at the state level after 2011. Consequently, aggregating data at the municipal level and performing estimation could potentially yield sensitive and unreliable estimates. Even though it is possible to assess the effect of wildfires at the state level, this would entail a substantial loss of information. For instance, we would have to discard data before 2011, because those data are not representative. Additionally, as wildfires do not occur uniformly within a state, this approach leads to a loss of variability. To illustrate this, consider the aggregation of small municipalities with a substantial proportion of burned areas alongside larger municipalities where wildfires did not occur. This would result in a region with a small proportion of burned area, leading to a loss of crucial information on the heterogeneity in burning behavior.

To address this challenge, we opt for an intermediate level of geographical aggregation. Specifically, we construct clusters of municipalities with similar burned areas and socioeconomic indicators, resulting in new regions that are smaller than a state. Geographical units are clustered using the max-p-region algorithm proposed by Duque et al. (2012) and described in detail in section 3.1. The validity of this aggregation is evaluated using customary statistical tests.[2]

Aiming to obtain sound estimates as in Canavire-Bacarreza et al. (2016), we must ensure that variables aggregated at the municipality level are reliable. We follow the

[2]The tests are a Welch test for the difference in means for continuous variables and a χ^2 test.

guidelines from the INE when a potentially unrepresentative subsample is available. Their suggestion is to check the coefficient of variation, which is an indicator of how reliable the resulting aggregated variables (or estimates) are. Our model takes as input averages of the observed variables. These averages correspond to rates for binary variables (such as occupation and poverty) and a measure of central tendency for the continuous and multinomial ones. For every variable W, we compute the coefficient of variation for every municipality i and every year t, which is the standard error of the variable's average divided by its sample average multiplied by 100 percent

$$CV_{it}\left(\overline{W}\right) = \frac{\mathrm{se}\left(\overline{W}_{it}\right)}{\overline{W}_{it}} \times 100\%.$$

The results are reported in Table A1. The INE states that the precision is good below 10 percent, acceptable between 10 and 20 percent, and unreliable over 20 percent. Accordingly, we drop the variables with a coefficient of variation above 20 percent: social security, marital status, and medical insurance. Enrollment in education is also dropped, because its coefficient of variation is 19 percent.

3 Methodology

3.1 Regional clusterization

Performing estimations at the municipality level could yield potentially sensitive and inaccurate results, given that our data are not representative at such a granular level. A potential solution is to aggregate neighboring municipalities, aiming to obtain representative data. However, this would only yield an improvement in the presence of spatial dependence—specifically, if adjacent municipalities experienced similar forest loss in the study period and share a similar macroeconomic environment. If adjacent municipalities are dissimilar, the aggregation process may not enhance representativeness. This is because in such cases, both the population and the sample size increase, but the samples would be nearly independent across municipalities, so there is no information gain.

Consider a scenario where we want to study the income in the region A ($\boldsymbol{Y}_A$), which is composed of four municipalities (1, 2, 3, and 4). We observe household income ($\boldsymbol{Y}_{A,1}, \boldsymbol{Y}_{A,2}, \boldsymbol{Y}_{A,3}$, and $\boldsymbol{Y}_{A,4}$) at the household level for municipalities 1, 2, 3, and 4, respectively, where $\boldsymbol{Y}_{A,i}$ is a vector of length n_i representing the incomes of n_i households in region i. If household income within region A is homogeneous, a smaller sample is required. However, in the presence of heterogeneity a larger sample is needed. Accordingly, if the income distribution is similar across municipalities 1 to 4, the number of households we need to observe in region A is smaller than if it varies substantially. Consequently, if the income process is similar across municipalities, we can hope that the sample $n_A = \sum_{i=1}^{4} n_i$ is large enough to learn about $\boldsymbol{Y}_A$ and, in the best-case scenario, is representative of the population in A.

To aggregate neighboring municipalities, we use the max-p-region algorithm (Duque et al., 2012). This algorithm clusters adjacent municipalities for a given threshold, minimizing within-attribute heterogeneity while maximizing heterogeneity between the new regions. To illustrate how the max-p-region algorithm works, we take the example from Duque et al. (2012), where the objective is to group areas based on housing prices using the number of houses in the resulting regions as a threshold. In the authors' example, there are nine areas for which both the average house price and the number of houses are observed. The optimal clusterization solution for a threshold of 120 houses[3] is depicted in Figure 1. The algorithm finds two regions such that the average housing price is homogeneous within the regions and heterogeneous between both regions. Notice that in Figure 1, the northeast region clusters areas where the price is low, and the southwest region clusters areas with high prices.

We conduct regional clusterization to group municipalities that experienced similar forest loss in the study period and share a similar macroeconomic environment. This allows us to obtain a sample for the resulting regions that is large enough to be representative. In our framework, the max-p-region algorithm takes as inputs the 330 municipalities in Bolivia, their spatially extensive attributes, and a specified threshold. Section 4.2.1 gives further details.

3.2 Counterfactual estimators

Consider that we observe $\{\{\boldsymbol{Y}_{it}, \boldsymbol{X}_{it}, D_{it}\}_{i=1}^{N}\}_{t=1}^{T}$ for region i at time t, where $\boldsymbol{Y}_{it}$ represents the outcomes of interest (poverty, income, agricultural income, and occupation), $\boldsymbol{X}_{it}$ denotes aggregate observed characteristics serving as controls, and D_{it} is the treatment indicator defined from the proportion of burned areas, as outlined in section 3.3. We aim

[3]That is, the optimization problem includes the constraint that the resulting regions must have at least 120 houses.

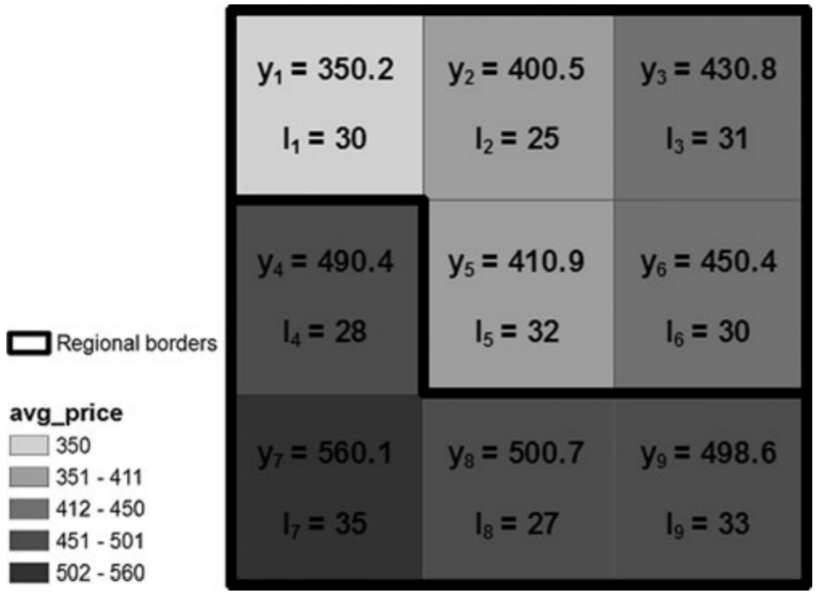


Figure 1: Optimal solution of the max-p-region algorithm for a threshold of 120 houses (Taken from Figure 2 in Duque et al. (2012)).

to estimate the effect of severe wildfires on socioeconomic outcomes in Bolivia, so that our object of interest is

$$ATT_s = \mathbb{E}\left[\delta_{it} | D_{i,t-s} = 0, D_{i,t-s+1} = ... = D_{i,t} = 1, C_i = 1\right], \quad s > 0, \tag{1}$$

where $\delta_{it} := Y_{it}(1) - Y_{it}(0)$ and $C_i = 1$ if there exist t and t' such that $D_{i,t} = 0$ and $D_{i,t'} = 1$ and 0 otherwise. The estimand in equation (1) is the ATT at the s-th period after the treatment's onset (ATT_s). This object captures the effect of wildfires on the outcome Y_{it} after s periods.

Estimating the $ATTs$ requires a methodology that accounts for the nature of a wildfire treatment. The vast majority of causal inference methods focus on applications with staggered adoption (Callaway et al., 2024), where a unit, once treated, remains treated for subsequent periods. However, this is not a valid assumption in the case of wildfires. A region exposed to a wildfire event in one period can never receive the treatment again or become treated again after an undetermined number of periods. Therefore, we need a methodology allowing for the switching on and off of the treatment indicator, as is the case with the counterfactual estimation methodology of Liu et al. (2022).

Given our framework and aim to utilize the maximum amount of available information, we prefer the counterfactual estimation methodology of Liu et al. (2022) over alternative novel proposals. On the one hand, Sun and Abraham (2021) and Callaway et al. (2024) allow for heterogeneous treatment effects (and continuous treatment in the case of Callaway et al. (2024)), but staggered adoption is required. On the other, De Chaisemartin and d'Haultfoeuille (2020) allow for non-staggered adoption, but they only use observations one period before or after the treatment's onset or exit, leading to the exclusion of many observations in the estimation process. Moreover, the estimator of Liu et al. (2022) provides more reliable causal estimates than conventional linear two-way fixed effects (TWFE) estimators (Imai and Kim, 2021) when treatment effects are heterogeneous and unobserved time-varying confounders are present.

The estimator of Liu et al. (2022) allows us to estimate the $ATTs$ by directly imputing counterfactual outcomes for treated observations. Treatment variables within this framework are binary and are allowed to switch back and forth. The estimator under this framework takes observations under the treatment condition as missing, uses data under the control condition to build models, and imputes counterfactuals of treated observations based on the estimated models. Three estimators are allowed within this framework, i.e., Fixed Effects (FEct), Matrix Completion (MCct), and Interactive Fixed Effects (IFEct). We chose the IFEct estimator base on Cross-Validation. However, our findings are robust to the use of any of these alternatives, as well as to the variance estimator of the ATT.

To estimate equation (1) using the IFEct, we need to estimate $Y_{it}(0)$ for the treated units. As proposed by Gobillon and Magnac (2016), we impute $Y_{it}(0)$ by estimating the following equation:

$$Y_{it}(0) = \boldsymbol{X}_{it}'\boldsymbol{\beta} + \alpha_i + \xi_t + \boldsymbol{\lambda}_i f_t + \epsilon_{it}, \quad i = 1, ..., N, \quad t = 1, ..., T, \tag{2}$$

where $\boldsymbol{X}_{it}$ represents observed characteristics; $\boldsymbol{\beta}$ denotes location parameters; α_i and ξ_t are unit and period fixed effects, respectively; f_t represents common factors with a heterogeneous impact on each unit captured by $\boldsymbol{\lambda}_i$; and ϵ_{it} is the idiosyncratic error term. For notation purposes, let $\boldsymbol{U}_{it} \coloneqq \alpha_i + \xi_t + \boldsymbol{\lambda}_i f_t$.

The three underlying assumptions in equation (2) are the following:

$$Y_{it}(0) = f(\boldsymbol{X}_{it}) + h(\boldsymbol{U}_{it}) + \epsilon_{it}, \tag{3}$$

$$f(\boldsymbol{X}_{it}) = X_{it}'\boldsymbol{\beta}, \tag{4}$$

$$h(\boldsymbol{U}_{it}) = \alpha_i + \xi_t + \boldsymbol{\lambda}_i f_t, \tag{5}$$

$$\epsilon_{it} \perp \{D_{js}, \boldsymbol{X}_{js}, \boldsymbol{U}_{js}\} \, \forall i, j \in \{1, ..., N\}, s, t \in \{1, ..., T\}, \tag{6}$$

where $f(\cdot)$ and $h(\cdot)$ are known parametric functions. Equation (3) implies additive separability, while equation (4) assumes linearity of $f(\cdot)$. Equation (5) requires a low-dimensional decomposition of $h(\boldsymbol{U}_{it}) : h(\boldsymbol{U}_{it}) = \boldsymbol{\Lambda F}$, with rank$(\boldsymbol{\Lambda F}) << \min\{N, T\}$, while equation (6) represents the strict exogeneity assumption, translating into quasi-randomization conditional on $\boldsymbol{X}$ and $\boldsymbol{U}$.

The proposal of Liu et al. (2022) consists of the following four steps: (i) estimate equation (2) on the subset of untreated observations with the linear factor method from Bai (2009) and obtain $(\hat{\beta}, \hat{\alpha}_i, \hat{\xi}_t, \hat{\lambda}_i, \hat{f}_t)$; (ii) compute $\hat{Y}_{it}(0)$ for each treated unit in equation (2) using the previous estimates; (iii) estimate the individual treatment effect $\hat{\delta}_i t = Y_{it}(1) - \hat{Y}_{it}(0)$; and (iv) estimate the ATT_s in equation (1) as

$$A\hat{T}T_s = \frac{1}{|\mathcal{S}|} \sum_{i,t \in \mathcal{S}} \hat{\delta}_{it}, \tag{7}$$

where $\mathcal{S} = \{(i,t) | D_{i,t-s} = 0, D_{i,t-s+1} = ... = D_{i,t} = 1\}$, and $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$.

3.3 Treatment definition

The counterfactual estimation methodology proposed by Liu et al. (2022) requires a binary treatment variable. Defining a proper treatment indicator proves challenging in the context of our study. Directly using the proportion of burned areas relative to the total region size is unsuitable, due to the nonrandom occurrence of wildfires. Wildfires tend to be geographically concentrated in the northern and northeastern regions of Bolivia, with minimal incidence in the south. Moreover, the continuous presence of wildfires can lead to stronger adaptation and mitigation strategies in affected areas, potentially masking the true effects of wildfires on socioeconomic outcomes if these strategies are not accounted for.

To overcome this issue, we propose identifying severe wildfire events by considering the historical evolution of wildfires in each geographical unit in the spirit of Fingado and Poelhekke (2023).[4] Specifically, we construct a Wildfire Index (WI) that relates current deviations (with respect to the historical mean) in burned areas in year t in location i $(burned_{i,t} - mean_i)$ to the maximum (max_i) and minimum values (min_i) ever observed

[4] These authors use a similar approach to define severe droughts in Africa based on the normalized difference vegetation index (NDVI).

at that location:[5]

$$\mathrm{WI}_{i,t} = \frac{burned_{i,t} - mean_i}{max_i - min_i}. \tag{8}$$

A wildfire event (WE_{it}) in location i at time t is classified as severe if its deviation from the historical mean is sufficiently high, i.e., if its corresponding Wildfire Index ($\mathrm{WI}_{i,t}$) exceeds a certain threshold, δ. Accordingly, a unit i experiencing a severe wildfire at time t is considered treated. Given that the treatment switches on and off for certain units, a potential concern is the violation of the no-carryover effects assumption, which posits that the potential outcome could be influenced by its treatment status in earlier periods. To address this issue, and considering the dynamic nature of the effects of wildfires on socioeconomic outcomes, we assume that after a severe wildfire, the unit remains treated for the next k periods. For example, if a severe wildfire event is identified in 2015 and $k = 2$, we assume that the treatment variable for the unit takes the value of 1 in 2015, 2016, and 2017 and becomes 0 afterward. In our baseline specification, we set $\delta = 0.8$ and $k = 2$. We report the robustness of our results concerning these parameters in section 4.3.

3.4 Identification

The identification of the causal effects of wildfires on the socioeconomic outcomes of interest relies on the unexpected nature of severe wildfires. In our definition of severe wildfires, we disregard predictable fluctuations in burned areas within a region and focus on specific events that significantly deviate from the historical norm. Therefore, we assume these events are unexpected and could not have been predicted in advance, allowing us to consider the treatment variable as exogenous. Anticipation of such events could prompt regions to invest more in adaptation or mitigation strategies to fight fires, potentially mitigating the impact on socioeconomic outcomes. In this scenario, the estimated effects would represent a conservative lower bound of the actual effect.

A potential challenge to identification arises from the primary source of ignition in Bolivia. As discussed in Bustillo et al. (2021) and Devisscher et al. (2019), wildfires in Bolivia often originate from the slash-and-burn agricultural practice, which can escalate and lead to uncontrolled fires. Additionally, other human activities, such as pasture management, waste burning, and hunting, contribute to the overall wildfire risk. This scenario suggests that areas with fewer institutional resources to combat fires, less skilled

[5] Eq. (8) is a standard normalization technique, where instead of subtracting the minimum from each observation, we subtract the mean, as we are interested in substantial deviations from the normal wildfire conditions, considering the shock as unexpected.

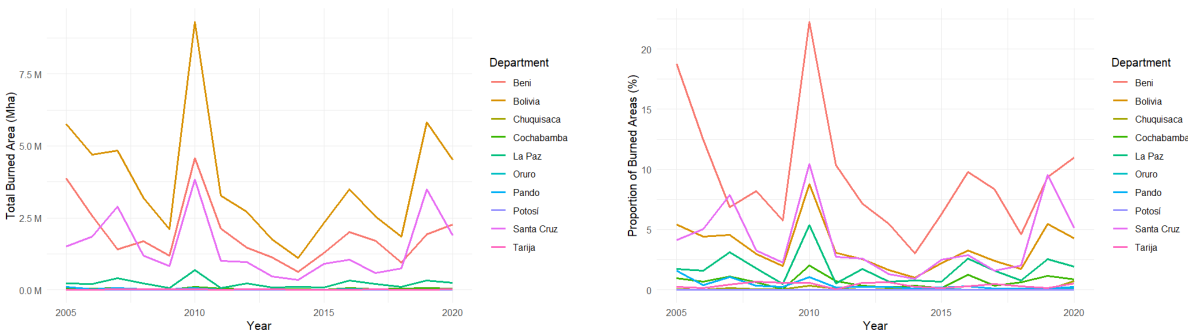
farmers to contain them, or more challenging geographical and climatic conditions may experience worse outcomes and are more likely to be treated. By defining the treatment as deviations from the historical norm we partially address this source of bias, in the sense that idiosyncratic factors that influence typical absolute outcomes in specific locations are removed. Our focus lies on particularly severe events that are unexpected given the wildfire history of the location and the current economic conditions. Moreover, to enhance the robustness of our analysis, we include in the vector of control variables $\boldsymbol{X}$ controls as years of education, literacy rates, age, and spoken language, ensuring an additional degree of conditional exogeneity in our assumptions.

Our empirical analysis addresses another potential sources of bias and endogeneity. To account for unobserved heterogeneity and control for time-invariant characteristics, the counterfactual estimators include region fixed effects in the regression equations. These fixed effects capture any time-invariant differences across geographical locations that may influence both the occurrence of wildfires and the outcomes we are examining. Time-varying heterogeneity is controlled by the inclusion of time fixed effects. Measurement error is another potential concern when using remote-sensing data to measure the extent of burned areas. While there may be some errors in the measurement of burned areas generated by mistakes in the algorithm classifying areas as either burned or not burned, we do not expect these errors to be systematically correlated with the economic conditions of the municipalities. Therefore, any measurement errors are likely to be non-differential and should not bias our estimated effects.

4 Results

4.1 Wildfires over time and space

On average, between 2005 and 2020, the period of analysis, approximately 3.71 Mha burned annually in Bolivia, equivalent to 3.5 percent of the country's territory. Total burned areas for the entire country and each department vary considerably across time and space. The year 2010 was particularly severe, with a record 9.32 Mha burned, 8.7 percent of the territory. Other years with severe wildfires include 2005 and 2019, with total burned areas exceeding 5 Mha. As depicted in Figure 2, Beni and Santa Cruz are the departments most severely affected by wildfires. These departments are located in northern and northeastern Bolivia, respectively, and contain three of the country's main ecoregions: the Bolivian Amazon, the Beni Savanna, and the Chiquitano seasonally dry tropical forest. All of these ecoregions lie at the southwestern edge of the Amazon basin. Notably, Beni experienced wildfires that affected approximately 22.3 percent of its territory in 2010 and 18.8 percent in 2005. In Santa Cruz, the proportion of burned areas has historically been smaller than in Beni, but it can still reach high values, such as 10.45 percent in 2010 and 9.53 percent in 2019. According Bustillo et al. (2021), the primary cause of fire ignition in the area is human-made, primarily attributed to the common practice of slash-and-burn. This practice involves cutting trees, low vegetation, and agricultural residuals, followed by burning the biomass to clear land for various purposes such as agriculture, livestock, or logging. The inherent risk lies in the potential for this practice to get out of control and give rise to the large fires.



(a) Annual burned areas (Mha)

(b) Proportion of burned areas

Figure 2: Evolution of wildfires in Bolivia 2005–2020

Figure 3 depicts the proportion of burned areas at the municipality level for every year in the study period. Consistent with our earlier observations, the maps reveal a clear pattern: municipalities with more than 10 percent of their land area burned are predominantly located in the northern and northeastern regions of the country, while municipalities in the south show very limited instances of wildfires. These visual representations highlight the regional concentration of wildfires in Bolivia, with certain municipalities experiencing significant burning each year and others remaining largely unaffected. Figure 4 allows us to contrast the maps of burned areas by municipalities with Unsatisfied Basic Needs vulnerability maps from 2012. We observe that municipalities with limited or no instances of wildfires tend to have medium to high vulnerability, while municipalities with more-frequent wildfires exhibit medium to low vulnerability levels, with some exceptions.

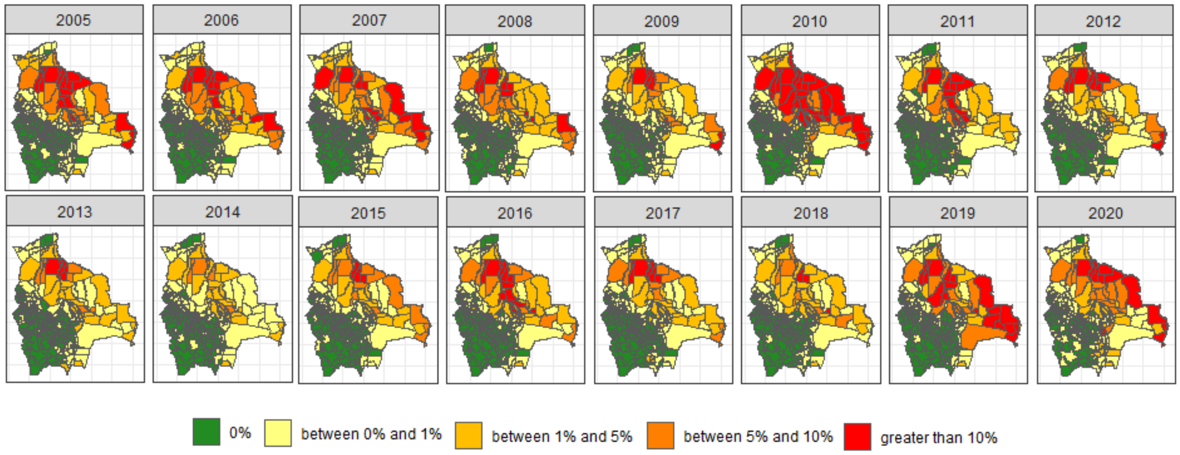


Figure 3: Annual burned areas as a proportion of municipality extension, 2005–2020

4.2 Wildfires and poverty: Baseline specification

4.2.1 Regional clusterization

We need to ensure that our data set is suitable for clustering, which happens if the attributes are similar between neighboring municipalities. To check this, two different statistical tests are performed: a Welch test for the difference in means for continuous variables, and a χ^2 test to compare the empirical cumulative distribution functions (cdf) for multinomial variables. The intuition behind both tests is that if we fail to reject that the averages are different between neighbors and the empirical cdfs are not different, we have evidence that neighboring municipalities share similar characteristics, making regional clustering reasonable. The results are reported in Tables A2 and A3.

For continuous variables, most neighboring municipalities have different average ages, years of education, and per capita incomes, with rejection rates at a 5 percent significance level being 0.40, 0.58, and 0.72, respectively. The rejection rate is calculated as the proportion of times we could not reject the null hypothesis of different means for a pair of neighbors. Table A3 indicates that for all neighboring municipalities, we could not reject the null hypothesis of differences between the classes in the samples, that is, we find that the distributions of the multinomial variables are similar across neighboring municipalities. Comparing the attributes of neighboring municipalities indicates how similar they are. However, two neighboring municipalities will not be clustered if they have different attributes. The previous exercise suggests that it is reasonable to make

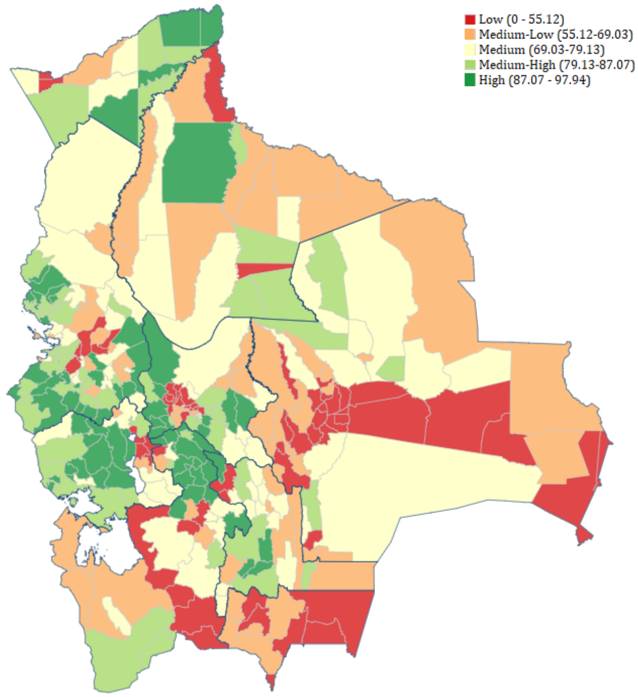


Figure 4: Vulnerability in Unsatisfied Basic Needs in Bolivia 2012

the clusterization. Nonetheless, what we require is that the clustered municipalities have similar attributes, which is why we perform the same exercise for municipalities in the same cluster once the clusterization is done. The results are reported in Tables A2 and A3, and find very similar results. Consequently, we perform the clusterization.

Municipalities are clustered based on both the dependent and independent variables considered in the empirical analysis. In our baseline specification, the clustering algorithm is employed by imposing a minimum of 6 municipalities per cluster, resulting in a total of 48 clusters corresponding to the panel units in our application. In Figure 5 we depict the resulting clusters from the max-p region algorithm. A panel of burned areas at the cluster level from 2005 to 2020 is constructed. Similarly, socioeconomic variables are aggregated using the available individual-level data for the municipalities encompassing each cluster. Because wildfires predominantly occur in rural areas, we remove the data associated with individuals living in the capital of each department before aggregation.

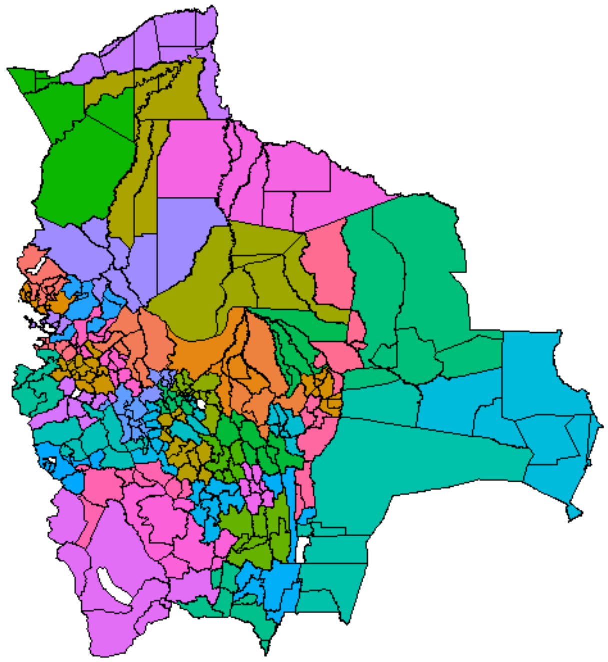


Figure 5: Regional clusterization from max-p region algorithm

4.2.2 Econometric estimates

Severe wildfire events are defined as those that exceed a value of 0.8 in the index outlined in equation (8). Approximately 3 percent of the recorded wildfires can be classified as severe under this threshold. It is important to remember that these wildfire events are not necessarily the events with the highest absolute magnitudes, but rather the ones with more pronounced deviations with respect to the historical evolution of wildfires in a given geographical cluster. For the treatment definition, we assume that the effects on socioeconomic outcomes persist for at least two years following a severe wildfire event. Therefore, the treatment variable at the cluster level is an indicator, taking the value of 1 for the year in which a severe wildfire occurs and for the subsequent two years and 0 otherwise. The control group is defined by imposing a clean-control condition. Specifically, in each period there might be clusters with positive values on the index, indicating burned areas exceeding their historical mean, that are not classified as treated because their index is below the defined threshold of 0.8. These observations 'close to being treated' are removed from the analysis, because their substantial burned areas make them unsuitable for comparison with the treated units.

The estimated effects of wildfires on the main socioeconomic outcomes in our baseline

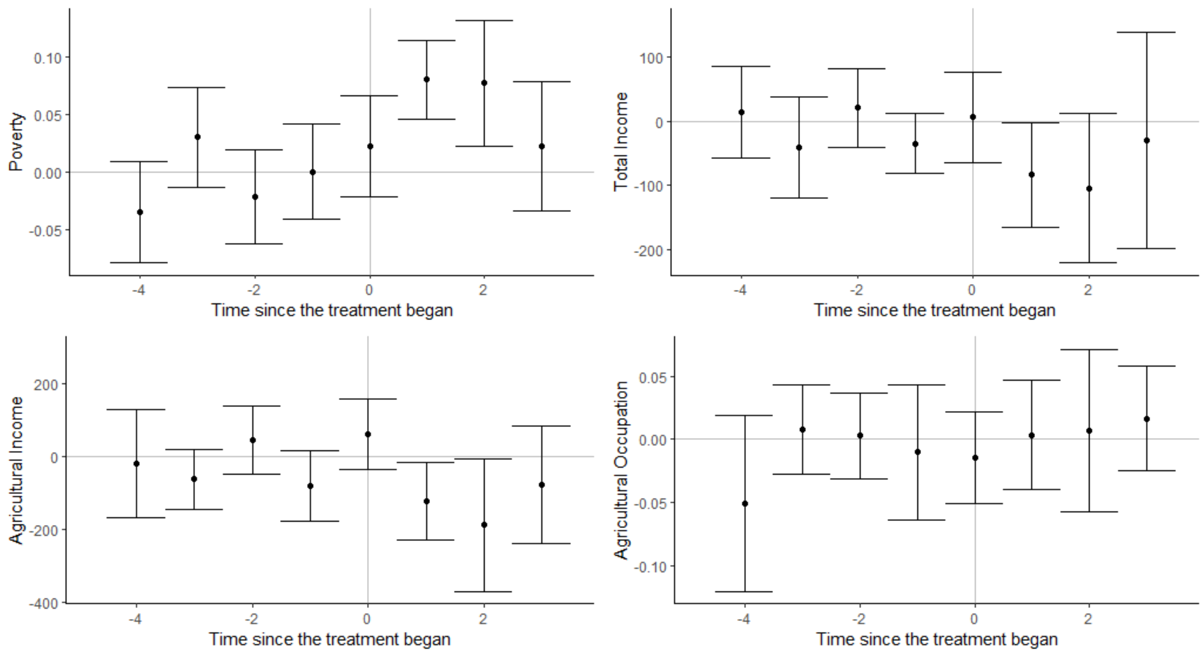


Figure 6: Effect of wildfire events on socioeconomic outcomes

specification are shown in Figure 6. The results are presented in the form of dynamic ATTs using an IFE estimator. We also report 95 percent confidence intervals, computed with 500 bootstrap replications. Our primary outcome of interest is aggregate poverty at the cluster level. Estimates indicate that severe wildfire events lead to a significant increase in poverty within the affected clusters (see top left panel of Figure 6). In the year following a wildfire, poverty increases by approximately 8 percent; two years later this effect is 7.7 percent. After three years, the effect becomes non-statistically different from 0, indicating that the impact of wildfires is short-lived. Importantly, at least 10 observations were used in the computation of these ATTs to ensure the reliability of the estimates and the confidence intervals.

To explore the mechanisms behind the positive effects of wildfires on poverty, we investigate the effects on other socioeconomic variables determining poverty. Starting with aggregate per-capita household income, as shown in the top right panel of Figure 6, we observe a significant reduction of approximately 84 Bolivianos (Bs) one year after a wildfire event. The point estimates in the top right panel of Figure 6 suggest a short-term effect on total income, as it decreases and the effect starts to vanish. We classify workers as employed in the agricultural and nonagricultural sectors and construct aggregate variables for per-capita agricultural and nonagricultural income, respectively. The effect of wildfires on agricultural income is illustrated in the left bottom panel of Figure 6. The estimates reveal that in the year following a wildfire event, agricultural income decreases by approximately 121 Bs. This effect escalates to around 186 Bs two years later.[6] The timing and direction of these effects allow us to infer that the observed increase in poverty is driven mainly by the reduction in agricultural income. This observation is rationalized by the damage to crops and agricultural infrastructure, resulting in a reduction of individuals' income following a severe wildfire. We also examine the effect on occupation status, formality, and forest loss. Our results indicate that wildfires do not lead to the creation or destruction of jobs in the affected areas.

4.2.3 Diagnosis

The plausibility of the identifying assumptions is analyzed. A first visual inspection of the estimated ATTs in Figure 6 suggests no pretrends: the estimates are statistically equal to zero for all dependent variables at horizons $s \leq 0$. This evidence is complemented using the F-test for the null of no pretrends proposed by Liu et al. (2022). According to the p-values reported in column (1) of Table 1, the null hypothesis of no pretrends is not rejected in our application. Additionally, we conduct a placebo test, assuming that the treatment starts two periods earlier than its actual onset for each unit in the treatment group. The estimated dynamic ATTs applying the same IFE counterfactual estimator as before are close to zero. The p-values of the corresponding F-test, reported in column (2) of Table 1, indicate that we cannot reject the null hypothesis that the placebo effect is zero. Finally, we test for the presence of carryover effects by hiding three periods right after the treatment ends and predicting the counterfactual outcome in those periods. The p-values, reported in column (3) of Table 1, suggest that there are no carryover effects, given that the average prediction errors in those periods are statistically zero. The results of the three tests in Table 1 support the validity of our identifying assumptions.

4.2.4 Limitations

Our analysis has certain limitations that are worth mentioning. First, based on the available data, it is not possible for us to identify migration decisions due to wildfires.

[6] The dependent variable is expressed in levels of (total and agricultural) household per-capita income. Although applying a log-transformation led to similar conclusions, we chose to present the results for the outcomes in levels. This decision is motivated by the scale dependency of this class of models, as discussed in Athey and Imbens. (2006), which may invalidate identification assumptions after applying transformations.

Table 1: Diagnosis tests

Outcome	Pretrends	Placebo	Carry-over
Poverty	0.866	0.695	0.955
Income	0.750	0.893	0.721
Agricultural income	0.542	0.596	0.396
Occupation	0.742	0.636	0.271

Note: The table contains the p-values of the F-tests for the nulls of no-pretrends (column 1), no-placebo effects (column 2), and no-carry-over effects (column 3).

If migration is adopted as an adaptation decision, then wildfires occurring in a certain municipality could have effects in other unaffected places, and our construction of the control group is challenged. Regional clusterization helps us to partially alleviate this concern. If we assume that migration is costly and individuals who migrate tend to move to neighboring locations with similar economic conditions, our estimated causal effect at the cluster level would weight the local effects in both the origin and destination municipalities. Second, there is no total guarantee that the data are representative at the cluster level. However, if we want to exploit the regional variation in burned areas, grouping similar municipalities is a valid alternative and aggregates can be computed with a higher level of confidence. Thirdly, our identification relies on the unexpected nature of severe wildfires. If anticipation of such events occurs, regions are more prone to invest in adaptation or mitigation strategies to fight fires. As mentioned earlier, under this scenario the estimated effects would represent a conservative lower bound of the actual effect. Finally, one threat to the stable unit treatment values assumption (SUTVA) assumptions is the smoke effect on neighboring regions of a severe wildfire. If this effect is non-negligible, the potential outcome of the control units would be affected by the occurrence of a severe wildfire, potentially tainting our causal estimates.

4.3 Robustness

In this section, we study the robustness of our findings to the threshold defined for classifying severe wildfires (δ), the minimum number of municipalities in each cluster (m), the aggregation level (at the municipality or cluster-level), and the treatment persistence (k).

4.3.1 Severe wildfire threshold

In our baseline specification, a wildfire event is classified as 'severe' if the WI exceeds 0.8. To study the robustness of our findings, we consider a grid of alternative threshold values from 0.7 to 0.9 in 0.05 steps. For this robustness check, we allow the ATT to be estimated with less than 10 observations, in order to obtain estimates for the higher thresholds. Results are reported in Figure A1. We find that the direction of the effect is the same for all the socioeconomic outcomes. Moreover, the magnitudes of the ATTs are similar to the baseline of 0.8, except for $\delta = 0.9$, which overestimates the effect and presents much more uncertainty. This result is consistent with the fact that for such a high threshold, only a few observations are considered to be treated (high uncertainty) and are those that experience the most severe wildfires, which might have a bigger impact on socioeconomic outcomes (overestimation of the effect). We find that the magnitude of the effects is sensitive to the definition of the threshold for the Wildfire Index. Nonetheless, defining a smaller or bigger threshold does not translate into substantial deviations, because the point estimates are close to the baseline ones. Even though our qualitative results are robust to the threshold for the Wildfire Index, we find that the magnitudes depend on it. Accordingly, the policy implications derived from our results remain the same.

4.3.2 Minimum number of municipalities in a cluster

A critical step in our methodology is regional clusterization, particularly the parameter m, which determines the minimum number of municipalities per cluster imposed in the max-p algorithm. In our baseline specification, we set $m = 6$, resulting in 48 clusters of municipalities. Here, we report results for considering m from 2 to 8. The results in Figure A2 closely resemble those in our baseline specification, suggesting an increase in poverty and decreases in income and agricultural income following severe wildfire events. The magnitudes of the dynamic ATTs remain similar. Nonetheless, the significance of the effects changes substantially. On the one hand, m being small reflects the lack of representativeness of the data. On the other hand, for bigger values of m the spatial dependence vanishes as more municipalities are aggregated. Consequently, it is highly unlikely to find an effect, as we are clustering units with very heterogeneous wildfire events and socioeconomic outcomes. This exercise suggests that to effectively capture the effects of wildfires on socioeconomic outcomes, a higher level of granularity in defining regional units is required to prevent aggregation from obscuring the potential impact of those events. In summary, the choice of m significantly influences the observed effects of severe wildfires on socioeconomic outcomes.

4.3.3 Aggregation level

The motivation for the regional clusterization is the lack of representativeness at the municipality level, which potentially makes the results unreliable if data is aggregated at such level. If it were the case that our results are robust to aggregating data at the municipality level, we could perform the analysis without the max-p region algorithm clusterization. We compare the results of our baseline specification (Figure 6) with those resulting from aggregating at the municipality level in Figure A3. We find that none of the effects are significant anymore. Moreover, the point estimates substantially change, except for occupation which remains non-significant with similar and negligible point estimates. This result motivates the use of regional clusterization in the absence of representative data in our case study.

4.3.4 Treatment persistence

To capture the dynamics in the effects of severe wildfires on socioeconomic outcomes and mitigate concerns regarding carryover effects, we initially assumed that a unit remains treated for up to $k = 2$ years following a severe wildfire. We now define alternative values for the parameter k and assess the robustness of our baseline findings. In Figure A4, we present the dynamic ATTs on poverty, considering values of k from 0 to 8. Notice that larger values of k imply that the treatment adoption is closer to a staggered adoption setup. Whenever an ATT is not observed for a particular k, this is because there are less than 10 observations to estimate it. As observed, our estimates for every socioeconomic outcome remain virtually the same. Consequently, we find that our results are robust to the assumed persistence of the treatment.

5 Conclusion

Combining satellite data on burned areas with household surveys during the period 2005—2020 in Bolivia, we estimate the causal effects of severe wildfires on poverty and other aggregate socioeconomic outcomes. Our analysis reveals a significant short-term increase in poverty and an insignificant effect in the medium term. The mechanism that explains this effect is a decline in household per-capita income and, more specifically, a fall in the agricultural labor income in the years following a wildfire event. We find that our results are robust to the assumed persistence of the treatment. Moreover, our qualitative results remain the same when we define the minimum number of municipalities in each cluster (m) and the threshold that defines whether a wildfire is considered severe (δ). Even though the magnitudes of the ATTs change with m and δ, they do not deviate much from the baseline estimates. Accordingly, the policy implications from our results remain the same.

This is the first paper that documents the effects of wildfires on poverty in Bolivia, which allows the identification of the locational effects enhancing the government's ability to respond promptly to the dire situations faced by the most economically disadvantaged communities. Our findings carry significant policy implications, shedding light on the magnitude and duration of the wildfire effect. This information is crucial for understanding the welfare effects of severe wildfire events and gauging the potential damage to household incomes.

Addressing the implications of wildfires on impoverished communities requires a comprehensive strategy that extends beyond immediate relief efforts. It involves a dual focus on short-term interventions and long-term resilience-building initiatives. Beyond addressing urgent needs, such as providing immediate relief and effective disaster response, attention must be given to bolstering education, healthcare, and economic opportunities. This holistic approach aims not only to alleviate the immediate impact of wildfires but also to create sustainable solutions for reducing vulnerability in these regions.

Moreover, our results emphasize the importance of implementing policies that stabilize household incomes and consumption, particularly in the aftermath of a fire event. Prioritizing short-term direct transfers to affected families, especially those engaged in small-scale agriculture – as our study identifies them as particularly vulnerable to such shocks – can serve as a practical tool to mitigate the negative economic impact. Additionally, by identifying the agricultural sector as the underlying mechanism exacerbating the vulnerability of communities to wildfires, our research facilitates the design of targeted public policies. This targeted approach allows the government to tailor interventions to the specific needs of the affected sector, potentially maximizing the efficiency and impact of mitigation efforts.

References

Athey, S. and G. W. Imbens. (2006). Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica 74*(2), 431–497.

Bai, J. (2009). Panel Data Models with Interactive Fixed Effects. *Econometrica 77*(4), 1229–1279.

Bustillo, M., M. Tonini, A. Mapelli, and P. Fiorucci (2021). Spatial Assessment of Wildfires Susceptibility in Santa Cruz (Bolivia) Using Random Forest. *Geosciences 11*, 224.

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2024). Difference-in-differences with a continuous treatment.

Canavire-Bacarreza, G., J. C. Duque, J. Urrego, et al. (2016). Moving Citizens and Deterring Criminals: Innovation in Public Transport Facilities. *CAF Development Bank Of Latinamerica*.

Carmenta, R., L. Parry, A. Blackburn, S. Vermeylen, and J. Barlow (2011). Understanding Human-Fire Interactions in Tropical Forest Regions: A Case for Interdisciplinary Research across the Natural and Social Sciences. *Ecology and Society 16*(1), 53.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996.

Devisscher, T., Y. Malhi, and E. Boyd (2019). Deliberation for Wildfire Risk Management: Addressing Conflicting Views in the Chiquitania, Bolivia. *Geographical Journal 185*(1), 38–54.

Duque, J. C., L. Anselin, and S. J. Rey (2012). The max-p-regions Problem. *Journal of Regional Science 52*(3), 397–419.

Fingado, N. and S. Poelhekke (2023). Droughts and Malnutrition in Africa. *CESifo Working Paper No. 10385*.

Giglio, L., L. Boschetti, D. Roy, M. Humber, and C. Justice (2018). The Collection 6 MODIS Burned Area Mapping Algorithm and Product. *Remote Sensing of Environment 217*, 72–85.

Gobillon, L. and T. Magnac (2016). Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls. *Review of Economics and Statistics 98*(3), 535–551.

Imai, K. and I. S. Kim (2021). On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data. *Political Analysis 29*(3), 405–415.

Jones, B. and S. McDermott (2021). The Local Labor Market Impacts of US Megafires. *Sustainability 13*, 9078.

Liu, L., Y. Wang, and Y. Xu (2022). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *American Journal of Political Science*.

Nielsen-Pincus, M., C. Moseley, and K. Gebert (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.

Singh, M., S. Sood, and M. Collins (2022). Fire Dynamics of the Bolivian Amazon. *Land 11*, 1436.

Sun, L. and S. Abraham (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics 225*(2), 175–199.

Appendix: Tables and Figures

Table A1: Summary of coefficient of variation

Variable	Mean	Median
Years of education	10.17	9.86
Literacy	5.56	4.96
Attendance in the education system	13.92	5.59
Age	9.23	9.72
Sex	12.82	12.68
Native tongue	16.81	8.17
Members	5.19	4.86
Occupation	9.13	8.40
Poverty by income	12.07	9.16
Economically active population	8.55	7.96
Medical insurance	22.03	18.60
Urban	4.15	0.42
Per-capita income	12.54	10.99
Marital status	21.77	19.91
Social security	53.39	48.48
Enrollment in education	19.13	17.38

Table A2: Summary results for $t-$ test at a 5% significance level

(a) Neighboring municipalities

Variable	Rejection rate
Age	0.40
Years of education	0.58
Per-capita income	0.72

(b) Municipalities in the same cluster

Variable	Rejection rate
Age	0.43
Years of education	0.65
Per-capita income	0.56

Table A3: Summary results for χ^2 test at a 5% significance level

(a) Neighboring municipalities

Variable	Rejection rate
Literacy	0.00
Education	0.00
Sex	0.00
Native tongue	0.00
Members	0.00
Occupation	0.00
Poverty by income	0.00
Economically active	0.00
Urban	0.00

(b) Municipalities in the same cluster

Variable	Rejection rate
Literacy	0.00
Education	0.00
Sex	0.00
Native tongue	0.00
Members	0.00
Occupation	0.00
Poverty by income	0.00
Economically active	0.00
Urban	0.00

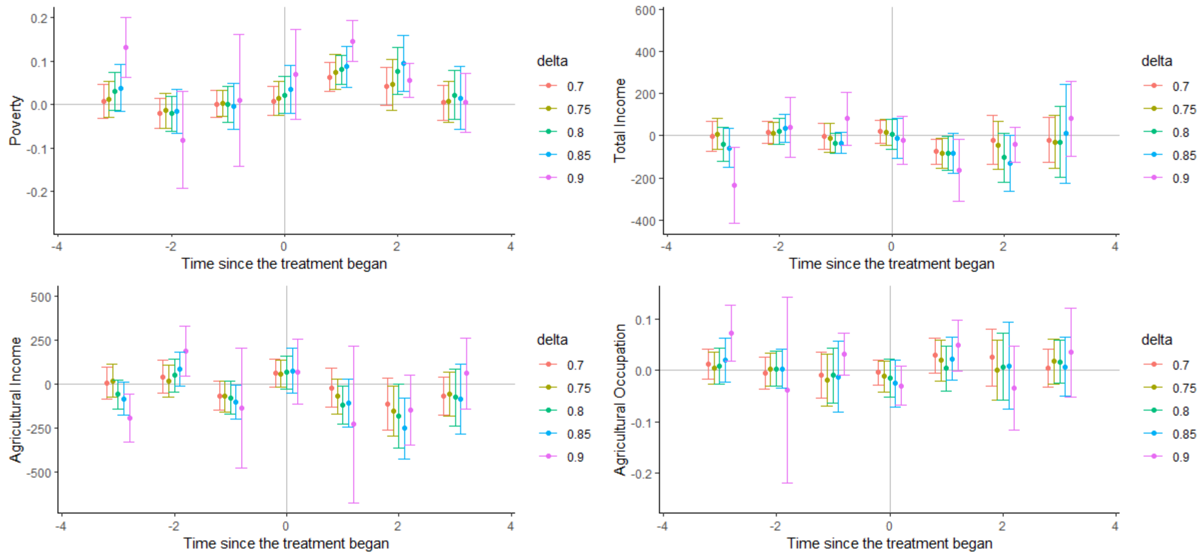


Figure A1: Effect of wildfire events on socioeconomic outcomes varying the Wildfire Index threshold

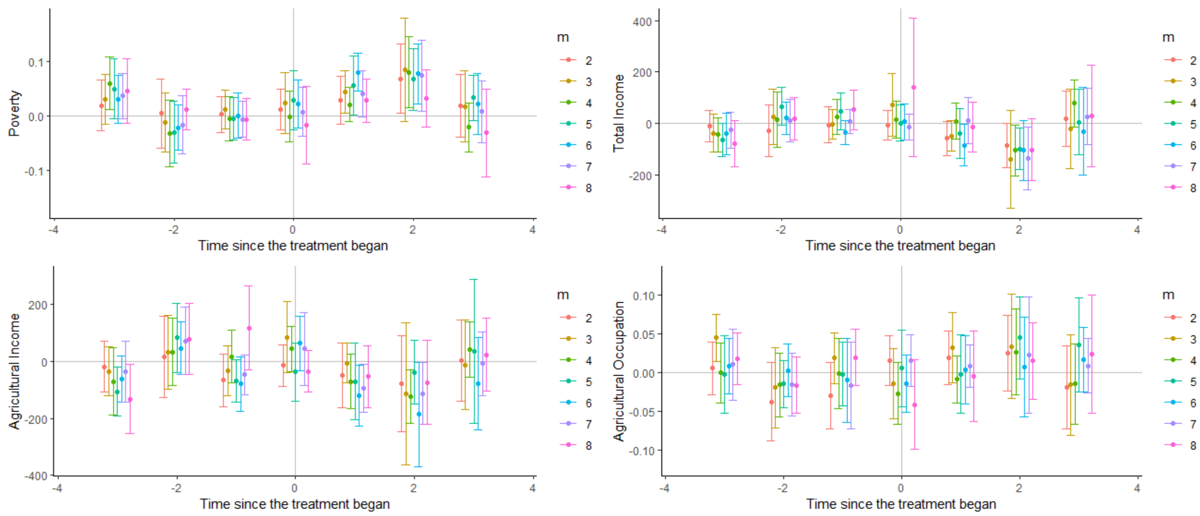


Figure A2: Effect of wildfire events on socioeconomic outcomes varying the minimum number of municipalities in a cluster

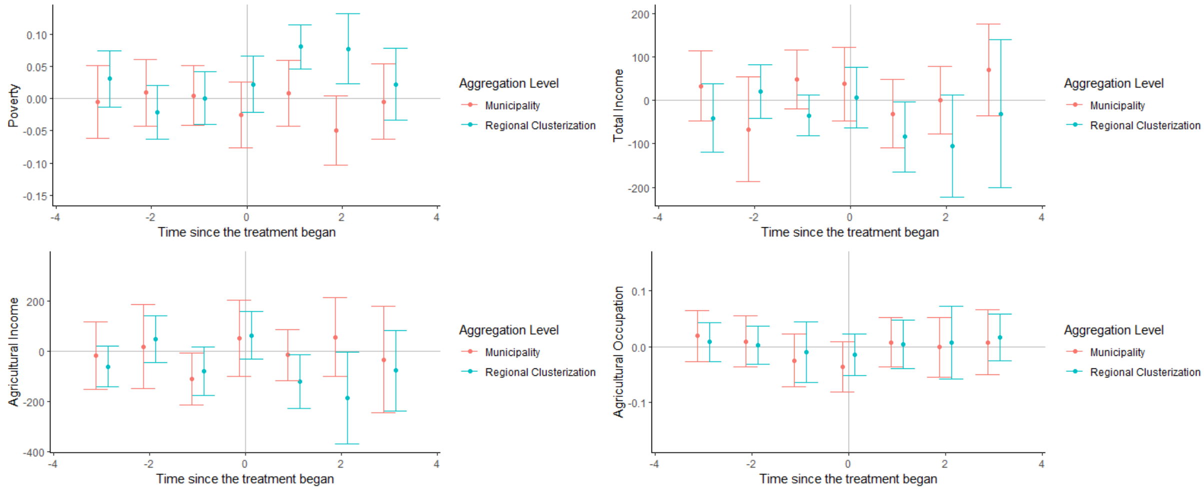


Figure A3: Effect of wildfire events on socioeconomic outcomes changing the aggregation level

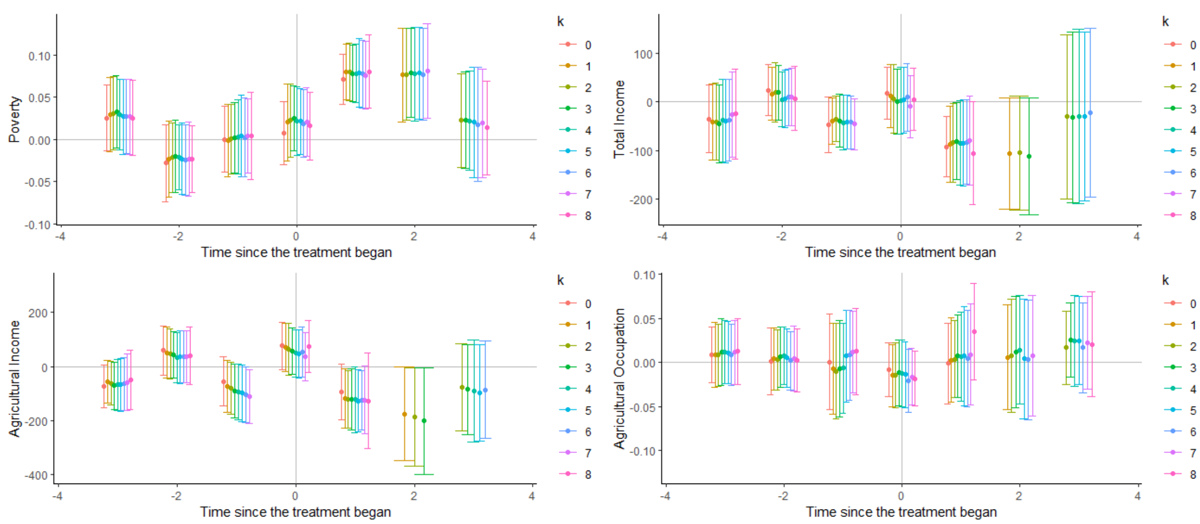


Figure A4: Effect of wildfire events on socioeconomic outcomes varying the persistence of the treatment