Disruptive Effects of Natural Disasters: The 1906 San Francisco Fire*

Hanna Schwank†

Department of Economics, Boston University (BU)

This version: November 30, 2021
Link to newest version

JOB MARKET PAPER

Abstract

Natural disasters are growing in frequency globally. Understanding how vulnerable populations respond to these disasters is essential for effective policy response. This paper explores the short- and long-run consequences of the 1906 San Francisco Fire, one of the largest urban fires in American history. Using linked Census records, I follow residents of San Francisco and their children from 1900 to 1940. Historical records suggest that exogenous factors such as wind and the availability of water determined where the fire stopped. I implement a spatial regression discontinuity design across the boundary of the razed area to identify the effect of the fire on those who lost their home to it. I find that in the short run, the fire displaced affected residents, forced them into lower paying occupations and out of entrepreneurship. Experiencing the disaster disrupted children's school attendance and led to an average loss of six months of education. While most effects attenuated over time, the negative effect on business ownership persists even in 1940, 34 years after the fire. Therefore, my findings reject the hope for a "reversal of fortune" for the victims, in contrast to what is found for more recent natural disasters such as hurricane Katrina.

Keywords: Natural Disasters; Internal Migration; Economic History, Regional and Urban Economics.

JEL Classification Numbers: N91, N31, Q54, O15, J61, J62 .

*I am grateful to Ran Abramitzky, Martha Bailey, Samuel Bazzi, James Feigenbaum, Martin Fiszbein, Edward Glaeser, Claudia Goldin, Robert Margo, Petra Moser, Emi Nakamura, Daniele Paserman, Paul Rhode, Johannes Schmieder, and participants of the Harvard Economic History Workshop, the FAU Nuremberg Micro Workshop, the 2021 European Meeting of the Urban Economics Association, the 2021 NBER SI Development of the American Economy, the 2021 Annual Meeting of the EHA, and quantitative workshop at Florida State University for helpful comments. All errors are my own.

†Contact: Hanna Schwank, Boston University Department of Economics, 270 Bay State Road, Boston MA. Email: hschwank@bu.edu. Web: https://sites.google.com/bu.edu/hannaschwank/.

1 Introduction

Natural disasters are growing in frequency globally. Today, about one hundred million people are affected by natural disasters every year (EMDAT, 2020). Due to more intense weather events and a fast growing world population, this trend is likely to accelerate even further in the future. Understanding how victims cope with such an adverse event is crucial for designing and targeting disaster relief policies. How are individual's economic decisions affected by such disasters? Are the effects transitory or are they persistent, altering entire life trajectories? The consequences of today's natural disasters will unfold in the future. Studying a historical event affords us a unique opportunity to study the long-run effects on affected populations. I therefore turn to one of the worst natural disasters of the past century in the United States, the 1906 San Francisco Fire (Blanchard, 2008).

In 1906, a 7.8 magnitude earthquake led to a massive fire that destroyed nearly half of the residential buildings in San Francisco. In the razed area, the fire burned almost the entire building stock to the ground. About 250,000 individuals evacuated the city immediately after the disaster, seeking refuge with family and friends in nearby Bay Area counties or in quickly erected tent cities. While 95 percent of the buildings held fire insurance, the majority of the population consisted of renters who did not benefit from these payouts. The city was rebuilt in less than ten years, but not all evacuees returned, either by choice or because they could not afford to do so.

I analyze the short- and long-run consequences of the 1906 San Francisco Fire by following San Francisco residents and their sons from 1900 to 1940[1]. Implementing a spatial regression discontinuity design across the boundary of the razed district to identify the effect of the fire, I find that those who lost their home in the fire were 6.3 percentage points more likely to reside outside of San Francisco in 1910. While men were forced into lower paying occupations and out of business ownership, their wives entered the labor market as additional breadwinners. Using a novel measure of kin presence, I find suggestive evidence that informal risk sharing among members of extended family in different parts of the city mitigated the adverse effects of the fire. Previous research on "modern" natural disasters finds a reversal of fortune for individuals who were displaced away from areas they were badly matched to. However, there was no reversal of fortune for those hit by the 1906 San Francisco Fire. The negative effects on entrepreneurship are highly persistent, and while affected residents recovered from occupational downgrading,

[1]To follow individuals over time, I use linked US Census records. Standard linked methods rely on characteristics that do not change over time, such as surnames. Since women tend to change their surname upon marriage, I focus on fathers and sons in this paper.

there is no indication that they outpaced their unaffected neighbors in any metric. Persistence fades in the children's generation: while sons of affected fathers obtain six fewer months of schooling than their unaffected counterparts, I do not find significant differences in migration patterns, wages, or self-employment.

Treatment status is determined by 1900 household location. To obtain this information, I geocode the address string given in the Census using GIS software. Additionally, I georeference and digitize nineteen photographs of election precinct maps showing the 1900 enumeration districts of San Francisco. I use this information to verify the coordinates and to proxy the location of residence where the address is missing or unreliable. The precise household location data allows me to create a novel measure of presence of extended family in San Francisco, which I base on the geographic proximity of families that share surname and race. Using this measure I explore the extent of risk sharing among extended family.

Historical records suggest that exogenous factors such as wind and the availability of water determined where the fire stopped. However, where it started was not a coincidence: Candles that illuminated storefronts in the business district and open fires in laundries sparked the flames when the earthquake knocked them over. Fueled by gas from damaged pipes, the fire was able to spread in the early morning. Exploiting the exogeneity of the fire border, I implement a spatial regression discontinuity (spatial RD) design across the boundary of the razed district to control for the fact that individuals living in the burned and unburned areas had average differences in demographic characteristics.

I begin the analysis by analyzing the short-run effects of the fire. I find that the disaster displaced families, so that they were 6.3 percentage points more likely to live outside of San Francisco in 1910. The location patterns are complex: My findings suggest out-of-state migration as one coping mechanism. At the same time, affected individuals are 5.2 percentage points more likely to live in the Bay Area. This implies that four years after the fire, many families still stayed in the place they evacuated to, unable to return to San Francisco or migrate away. Losing not only a home, but an entire neighborhood to a fire, having to evacuate and not being able to return for months or even years can disrupt a number of aspects of life, and impact a wide range of economic decisions. I find that affected men were 6.3 percentage points more likely to switch to a lower paying occupation, resulting in 7.4 log points lower occupational income in 1910. The fire did not only destroy residential buildings, but also businesses including inventory, machinery and documents. Four years after the fire, affected individuals were 10 percentage points less likely to be self-employed and 5.3 percentage points less likely to have employees compared to unaffected individuals.

My results are qualitatively robust to a wide range of alternative specifications and sample restrictions, including controlling for a rich set of pre-treatment characteristics, different bandwidth choices, alternative running variables, restricting the sample to whites and younger individuals, and implementing an alternative linking approach.

After establishing my baseline results, I investigate how families reacted to the disaster, whether renters and home owners were affected differentially, and the extent to which risk sharing among extended family in San Francisco mitigated the effects of the fire. I find that both wives and children of affected men were pulled into the labor force. The fire delayed marriages, but had an insignificant effect on fertility. There is some evidence that relative to home owners, renters were more likely to be displaced and experienced stronger economic losses. My findings are suggestive that the presence of extended family in the unaffected parts of the city was able to mitigate the effects of the disaster. One standard deviation more relatives in the unaffected parts of San Francisco is associated with a 4.6 percentage points higher likelihood of living in San Francisco in 1910. Affected individuals with more extended family tend to have higher occupational income and to be entrepreneurs in 1910.

Finally, I investigate the persistence of the effects. While individual level shocks can have long-run effects (e.g. Kahn, 2010), research on more recent natural disasters implies that they can have a silver lining: in a reversal of fortune, younger migrants benefited from destruction-induced migration after a volcano eruption in Iceland (Nakamura et al., 2020), and the income of those pushed out of New Orleans by hurricane Katrina outpaced the control group's income a few years after the disaster (Deryugina et al., 2018; Groen et al., 2016). However, the data does not support the notion of reversal of fortune for the victims of the 1906 San Francisco Fire. In the long run, the fire does not have an impact on migration patterns: between 1910 and 1920, evacuees are more likely to return to San Francisco than their non-burned counterparts, so that by 1920, affected individuals are neither more nor less likely to live outside of San Francisco or have moved out of state. Similarly, the differences in occupational income attenuate significantly by 1920. The negative effects on self-employment and having employees are remarkably persistent, both in sign and in magnitude. This is because entrepreneurship itself is exceptionally persistent over the life cycle, with only approximately 20 percent of individuals in my sample switching in or out of self-employment in any given decade. While children of affected San Francisco residents had on average six fewer months of schooling than children of unaffected residents, there is no clear evidence for difference in self-employment. I suggest that this is due to counteracting forces: while entrepreneurship is strongly transmitted across generations (Lindquist et al., 2015), research shows that adverse economic events during childhood can alter children's believes towards an increase in the likelihood of entrepreneurship (Makridis and McGuire, 2019).

This paper makes several contributions to our understanding of the economic effects of natural disasters. The literature that studies the effects of natural disasters on their victims focuses on recent disasters such as hurricane Katrina, which caused the entire city of New Orleans to evacuate in 2005. Outcomes of interest are migration (Sastry and Gregory, 2014; Deryugina et al., 2018), earnings and labor supply (Vigdor, 2007; Groen et al., 2016; Deryugina et al., 2018), educational outcomes (Sacerdote, 2012), and homeownership (Bleemer and Van Der Klaauw, 2017). A related paper studies the evacuation-induced migration effects of a volcano eruption in Iceland (Nakamura et al., 2020). I make three distinct contributions to this literature. First, I use a spatial regression discontinuity design across the border of the razed district to identify the effect of the disaster. Due to data constraints and the nature of destruction, the effects in the hurricane Katrina literature are identified through difference-in-differences type of strategies, using propensity score matching or weighting to create a control group of either non-flooded New Orleans residents or individuals from nearby cities and counties. The spatial RD design allows me to identify the effect on the victims of the fire relative to a control group, which is highly comparable in both observed and unobserved characteristics, and that was subject to the same institutional, cultural and economic conditions before the disaster. Second, in studying the 1906 San Francisco Fire, I add a new context to the literature that studies the effects of natural disasters on their victims. The fire destroyed parts of a fast growing and prosperous city, the "Paris of the West", at a time with well-functioning insurance markets but before institutionalized disaster relief. New Orleans, in contrast, was characterized by a weak labor market and the hurricane might have enabled individuals to leave disadvantaged neighborhoods (Deryugina et al., 2018). Third, my paper is distinct in tracing the individual-level economic effects of a natural disaster for several decades, which is conceptually impossible at this time for a disaster that happened in the 2000s. Nakamura et al. (2020) observe victims of a volcano eruption that destroyed parts of fishing town in Iceland for many years. However, the setting in a small homogeneous society and with full reimbursement of the replacement value of destroyed houses and land allows them to use the destruction as an instrument for migration, assuming that direct effects of the disaster on lifetime earnings and investments into education are negligible.

In addition, my paper builds a bridge to the literature on the long-run effects of urban and natural disasters which, so far, has studied aggregate outcomes of affected geographical areas. A central question is the effect on migration flows and population growth (Boustan et al., 2012; Boustan et al., 2020), for example of the earthquake that caused the fire in San Francisco in 1906

(Ager et al., 2020), the Great Mississippi Flood of 1927 (Hornbeck and Naidu, 2014), the dust bowl in the Southern Plains region of the United States in the 1930s (Long and Siu, 2018; Hornbeck, 2020) or one of Europe's deadliest earthquakes, the Messina-Reggio Calabria Earthquake in 1908 (Spitzer et al., 2021). Previous research on urban fires investigate changes at the city-level, such as the effects of the 1906 San Francisco Fire on residential density (Siodla, 2015a), land-use (Siodla, 2017), business agglomeration (Siodla, 2021), or the effect of the Great Boston Fire of 1872 on land values and urban growth (Hornbeck and Keniston, 2017). To my knowledge, my paper is the first that follows the victims of a historical natural disaster, investigating the long-run and intergenerational effects on migration, occupational choice, entrepreneurship and investment in education. Focusing on the affected population and following both migrants and non-migrants is essential for a complete understanding the impact of a disaster.

The remainder of this paper is organized as follows: In Section 2, I give background information on the earthquake and fire. Section 3 presents my data sources and Section 4 outlines the empirical strategy. In Section 5, I show and discuss the results estimating the short-run effects of the shock, and in 6 I present the long-run effects on adults and children. Section 7 concludes.

2 Background

In this section I outline how the 1906 earthquake led to a massive fire that destroyed about half of San Francisco's residential buildings. I explain how the border of the burnt area was determined by exogenous factors. I conclude by describing the aftermath and reconstruction of the city.

2.1 The Earthquake and Fire

An earthquake of magnitude 7.8 struck the coast of Northern California at 5.12am on Wednesday, April 18 in 1906. Initially, there was little damage from the earthquake itself in San Francisco, but due to damaged gas pipes, overturned ovens and fallen chimneys, more than fifty fires broke out shortly after the earthquake (Kennedy, 1908). The fires would eventually burn for three and a half days.

The most dangerous fires started in the mercantile district around Market Street (Kennedy, 1908), where gas lamps were illuminating display windows, fires were kept burning in the furnaces of laundry shops. Most buildings were empty in the early morning at the time when the earthquake hit and the first fires started. Figure A1 shows the course of the fire. On the first day, the areas north and south of Market Street were destroyed. Unusual easterly winds sent the flames towards the west and southwest. Some fires were also caused by families starting to cook breakfast without knowing that their chimneys had been damaged by the earthquake[2], for example in the well-known "Ham and Egg Fire" on Hayes Street (Kennedy, 1908, position of the fire origin marked in Figure A1).

Fire Chief Engineer Dennis T. Sullivan had been fatally wounded during the earthquake and could, therefore, not coordinate the fire fighting efforts. The earthquake had damaged all major water pipes, so that there was little to no pressure in the water mains anywhere in the city for several days. Fire fighters were able to put out a few fires when they found water, for example the "Lippman Fire" in Lippman's dry good stores at the corner of 22nd and Mission Street (Kennedy, 1908, marked with "PUT OUT 10 A.M. APR. 18" in the bottom left of Figure A1), or another one at the corner of Golden Gate Avenue and Buchanan Street (Kennedy, 1908, marked with "PUT OUT AFTER A 5 HOUR FIGHT" in the upper center of Figure A1).

On the second day, the fire spread further south and destroyed parts of the mission district as well as traveling further west north of Market Street. On that day, the mansions of some of San Franciscos wealthiest citizens were destroyed at Nob Hill (Hansen and Hansen, 2013). The flames also destroyed the Stanford House at Powell Street which was set back from the street and should have been easy to save under normal circumstances (Kennedy, 1908). Eventually, the flames were stopped at Van Ness Avenue, halted by an increasing westerly wind, and by the width of the street. However, between Clay and Sutter Street, the fire had been able to cross this wide avenue along a five block stretch (Fradkin, 2005).

In the absence of water, firefighters tried to demolish buildings in order to create fire breaks. By doing so, they often started even more fires because they were untrained in the use of dynamite as an article in the *Mining and Scientific Press* describes: "The use of high-grade explosives by people ignorant of their strength and proper application, was instrumental in destroying a vast amount of property without the result desired, and in many cases it actually spread the conflagration" (Rickard, 1906: 109).

One example is the last major fire, which was ignited on the third day. The flames had been more or less halted around Russian Hill, when "the unexpected happened", as Russian Hill resident Jerome Landfield describes it. "Suddenly I heard a dull explosion, followed by a cloud of dust and debris just below me. Then came a fresh burst of flame. I went down at once to investigate. It appeared that the authorities had turned again to dynamiting and were using black powder. With this they had blown up the plant of the Viavi Company, a patent medicine

[2] About 90 percent of chimneys had been destroyed during the earthquake (Hansen and Hansen, 2013).

concern, and in doing this had ignited thousands of gallons of alcohol that were stored there." (Landfield, 1906, ch. 28). Unfortunately, the usual west wind, that had been lacking in the previous two days, picked up and blew the flames to the last remaining residential buildings east of Van Ness Avenue. The resulting fire eventually destroyed fifty housing blocks in the north east of San Francisco (Fradkin, 2005.)

The boundary of the razed district was determined by several exogenous factors such as the weather, the availability of extinguishing material, and geographic features. The direction of winds was a crucial determinant in the course of the fire. An unusual east wind on the first two days spread the flames towards the west while the prevailing west wind was absent at first, and, therefore, did not stop the fire's westward expansion. On the last day it came back, fueling the fire north of Russian Hill. While temperatures were unusually high in the first days, rain set in on Saturday April 21, extinguishing the last flames (Fradkin, 2005). Besides the often harmful use of explosives, the lack of water was the main reason that the fire destroyed such a large area of the city. Due to the damage to major water pipes, conventional water supply was unavailable. Those fighting the flames had to use water from alternative sources such as water flowing from burst mains, found in old underground cisterns or wells, and saltwater from the sea. In the rare case that water was available, firemen and volunteers often successfully fought the flames. Where it was not available, other materials such as sand from nearby construction sites or even barrels of vinegar and wine were used (Fradkin, 2005). In addition, some geographic factors such as hills, vacant lots and wide streets played a role in determining the spread and halt of the fire (Kennedy, 1908).

2.2 The Aftermath

The fire destroyed about 500 city blocks with more than 28,000 buildings (Tobriner, 2006). Figure 1 shows the burned area in the northeast of the city. Estimates suggest that up to 3,000 people might have died in connection with the earthquake and the fire[3]. Most damage to buildings, about 90 percent, was caused by the fire and not by the earthquake directly (Fradkin, 2005). While buildings with brick and stone walls were most susceptible to earthquake damage, no type of building material was fireproof (Fradkin, 2005). Since the fires were small initially and spread slowly, many building owners were able to remove inventory before the fire arrived. There was little looting due to heavy military presence and the strict order to shoot anyone caught in the act (Hansen et al., 2014).[4] In total, about 250,000 out of the 420,000 inhab-

[3]Estimates vary widely. U.S. Army relief operations (Greely, 1906) report about 500 casualties in San Francisco while Hansen and Condon (1989) estimate that more than 3,000 people died during or as a result of the disaster.

[4]The army withdrew its troops after 73 days of military presence on June 30 (Fradkin, 2005).

itants evacuated the city after the disaster, most of whom went to or passed through Oakland. Southern Pacific's railroad evacuation transported the vast majority of refugees with 129 trains leaving San Francisco and 739 trains out from Oakland (Hansen and Hansen (2013)). Others were evacuated by the Navy or fled by ferry or by foot. While abundant donations in the form of clothing and food arrived in the Bay Area, the refugees themselves received a mixed reception as communities were wary of destitute displaced residents of the "sinful city" (Fradkin, 2005.)

Rebuilding was quick. A reconstruction map from 1908 (Figure A2) shows that just two years after the fire, new construction covers almost the entire burnt area. By April 1907, just a year after the earthquake, about 435,000 people were living in San Francisco. In 1910, the rateable value of the city's properties was almost at 1905 level. In 1911, San Francisco won the competition to host the 1915 Panama Pacific International Exposition World's Fair. Reconstruction was completed by the time the fair opened. Siodla (2015b) shows that net residential density increased substantially in the burnt and reconstructed areas in the short run, and that a density differential persists until today. Rebuilt areas saw a shift out of residential use into non-residential use (Siodla, 2017) which was possible by constructing fewer single-family houses.

The fire and earthquake caused damages of \$315 to \$350 million (Fradkin, 2005), of which about \$235 million were recovered from insurance companies (Grossi and Muir-Wood, 2006). About 95 percent of properties carried insurance (Whitney, 1972) and more than 90 percent of the face value of fire insurance policies were collected (James, 2011). Earthquake damages were not covered by any insurance, but many companies paid for them nonetheless in order to preserve their reputation. The very poor were disadvantaged in dealing with insurance companies which took advantage of families' liquidity constraints, trying to settle at low rates (Fradkin, 2005).

3 Data

I use individual level data from the US Federal Census of the years 1900 to 1940 (IPUMS, Ruggles et al., 2021). I link Census to Census records to follow San Francisco residents and their children over time. Treatment status is determined by household location in pre-disaster San Francisco. However, household location coordinates are not available in the Census microdata. To create this novel variable, I geolocate the address strings that are recorded in the Census using GIS software. In addition, I digitize photographs of a 1900 enumeration district map into machine readable format. This allows me to validate the coordinates, and to impute locations for missing addresses with the enumeration district centroid coordinates.

3.1 Linked Census Samples

I use individual data from the complete count samples of the US Census to follow San Francisco residents from 1900, right before the fire, up to 1940[5], 34 years after the disaster. The Census is taken every ten years, so that I observe an individual up to five times in the data. In practice, I only observe about a small fraction of individuals in all five Census years. ABE crosswalks provide links to at least one Census record in 1910, 1920, 1930 or 1940 for 36 percent of individuals who were between ten and sixty-five years old in 1900. Among those, I observe 56 percent of individuals at least three times.

Record Linkage. In order to follow individuals over time, it is necessary to link their records across Census years. Unfortunately, unique identifiers (such as social security numbers or a specific person identifier) are not available in the historic Census samples. Consequently, it is common practice to rely on *stable characteristics* to link individuals across Censuses. The assumption is that some individual characteristics are stable over time and can therefore be helpful in uniquely identifying an individual across Census samples. Most linking methods use first and last name, year of birth, sex, and birth place. In practice, these characteristics are not perfectly stable due to typos, transcription errors and name changes. An additional challenge is that for common names such as John Smith the characteristics mentioned are not sufficient to uniquely identify an individual.

Therefore, record linkage amounts to identifying the most likely corresponding record while discarding all cases where there are several likely candidates. There are several approaches to Census linking. I work in my baseline with ABE (Abramitzky, Boustan and Eriksson, 2012, 2014, 2019) standard links with phonetic cleaning of names. The crosswalks are provided by the Census Linking Project (Abramitzky et al., 2020) and available online. The ABE approach first converts names into phonetic codes using the NYSIIS algorithm.[6] Starting from all men recorded in the 1900 Census, it then restricts the sample to people who are unique by first and last name phonetic code, birth year, and place of birth.[7] The next step is to look for potential matches in each of the Census waves from 1910, 1920, 1930, and 1940. A record pair is kept if

[5] Access to the complete count samples of the Census (including names) is restricted for 72 years, so, at the time of writing, 1940 is the most recent full count sample available.

[6] The New York State Identification and Intelligence System (NYSIIS) algorithm converts names into phonetic codes, so that similar soundings names receive the same code, for example John and Jan are both assigned JAN.

[7] Due to surname changes upon marriage, women cannot be linked using standard methods, and are therefore excluded from my analysis.

there is an exact match and discarded if there are multiple exact matches for one 1900 record. If there is no unique match, the procedure is repeated allowing for one year of difference in birth year, and then for two years of difference in birth years. Using ABE links has two distinct advantages. First, the algorithm is fully automated and therefore perfectly replicable, so that I analyze the same sample of individuals as other research using ABE links. Second, the algorithm does not rely on any location information, which could otherwise under-represent internal migrants.

Adult Sample. The adult sample starts with individuals who reside in San Francisco in 1900. Out of the 340,874 individuals residing in San Francisco in that year, 334,975 lived outside of institutions such as prisons, hospitals and military bases. Of these, 271,124 were between 10 and 65 years old, and out of those, 146,349 were men. More than a third of those - 51,835 individuals - can be linked at least once to another record in 1910, 1920, 1930 or 1940[8].

Child Sample. The child sample consists of males born between 1891 and 1920 to fathers who lived in San Francisco in 1900. I combine two approaches. First, I link children who lived with their fathers in 1900 directly to their 1940 records. In addition, I am interested in those born after 1900, in particular after 1906, to explore the presence of indirect treatment effects. Therefore, I turn to the 1900-1910, 1900-1920, 1900-1930, and 1900-1940 adult samples and identify all fathers who had own children in their households. I then look for those children in the corresponding full count samples.

Finally, I link all children to their 1940 records. Thus, I have their fathers' information in 1900 (including household location in San Francisco) and their own outcomes in 1940.[9] I am mostly interested in the 1940 outcomes for two reasons. First, years of schooling information is not available in any of the earlier Census samples, but a disruption during childhood likely has a direct impact on educational attainment. Second, since I include individuals born until 1920 in my sample, focusing on the 1940 Census allows me to observe all children as adults.

Census Variables. The historical Census samples record a rich set of individual and household level variables which I either use directly or as inputs to construct additional measures. Table A1 (in the Appendix) shows an overview of all variables used in this paper. In the following, I briefly describe the most important ones.

[8]The 1900-1910 linked sample has 34,495 observations, the 1900-1920 sample has 27,307 observations, the 1900-1930 has 21,590 observations and the 1900-1940 sample has 15,469 observations.

[9]Since linking relies on the use of surnames, this restricts my sample to sons.

Post-Disaster Migration and Location. The first outcome that I am interested in is migration and location choice. My main measure of migration indicates whether an individual resided outside San Francisco in 1910 or later. An alternative measure of migration indicates whether the individual has gone as far as to leave California. Due to errors in the linking process I expect both types of migration rate to be upwardly biased: a false link is more likely to be located outside of San Francisco than by chance in San Francisco. However, since I am interested in the effect of the fire on the propensity to migrate, the linking-induced measurement error in migration would only be problematic if it was systematically affected by the fire. Linking is based on name, birth year and birth place, the reporting of which should not be affected by exposure to the fire.

Work and Income. The Census provides several individual and household level characteristics. While the occupation and labor force status is recorded, the 1940 Census is the first to ask about wages[10]. IPUMS (Ruggles et al., 2021) provides occupational standing measures based on the reported occupation, which quantify economic differences between occupations. I focus on OCCSCORE, which assigns the year 1950 median income in each occupation for all individuals with positive income. The occupational income score takes values from 1 to 80 and allows the econometrician to quantify occupational income differences. It can only vary on the occupational level so that no within-occupation variation can be captured. The informative value of the occupational income score decreases as the year for its construction becomes earlier, because the relative incomes of occupations changed over time.[11]

Additional variables show people's labor force status, and for those who work, whether they are salary workers, employers, or work on their own account. Months of unemployment was only recorded in 1900.

Household and Family. The Census enumerated the whole household, so that we know the number of each person's children and siblings living with them, whether a person was the household head, and the marital status of all individuals age twelve or older.[12] The Census also records whether the family owned or rented their home, and whether home owners paid a mortgage or some other loan for the dwelling.

Education. Information on literacy is available from 1900 to 1930, after which it was not recorded anymore. Instead, educational attainment was recorded in 1940 for the first time and asked for the highest year of school or degree completed. A school attendance measure is

[10]This excludes any non-wage income such as income from self-employment.

[11]IPUMS (Ruggles et al., 2021) does not classify all occupation strings in the 1900-1930 Census samples and therefore OCCSCORE is missing for about 15 to 25 percent of the observations, depending on the year and the exact sample. I exclude the unclassified individuals from the analysis of occupational income.

[12]Age 14+ in 1940.

available in all relevant years and indicates whether a person attended some type of school in a given reference period.[13]

3.2 Treatment Status and Geographic Information

Treatment is determined by household location in 1906: Households living in the area that burned down during the fire were "treated", those who lived outside that area were "untreated". Treatment is measured with some error since the Census was taken in 1900, and some families might have moved within San Francisco or left the city between 1900 and 1906. I locate households using their address information given in the Census records.

Border of the Razed Area. The boundary of the area destroyed by the fire is depicted in several historic maps. I use a shapefile (GIS) retrieved from DataSF.org, San Francisco's public data portal. The border is based on the paper map "San Francisco, California, showing the areas destroyed by fire, April 18-21, 1906" (R.J. Waters & Co., 1906) which is available at the Harvard Map Collection.

Geolocation of 1900 San Francisco Households. To locate the residences of all households in my data, I rely on ArcMap's geolocation tool. Geolocation means that ArcMap takes the street name and street number of a household that I provide as an input, and returns the corresponding coordinates of that street address using a current map. I then verify the accuracy of the coordinates by checking whether they are contained in the correct enumeration district.[14] The GIS software might return the wrong coordinates if the street name has been misspelled or the street has been renamed in the past 120 years. Additionally, the street number might be incorrect or the street numbering may have changed since 1900.

I am able to directly obtain coordinates by geolocation for 62.8 percent of all households living in San Francisco in 1900, and I can impute coordinates for an additional 13.3 percent using neighbors' information. The remaining 23.9 percent of households are assigned the coordinates of the enumeration districts they are living in. Appendix Section C explains the process of geolocation and imputation in detail. The extraordinarily high rate of geolocated addresses is

[13] The reference period varies from one census to the next: in 1900 it is the preceding 12 months, while in 1940, it is the preceding month only.

[14] A Census enumeration district is the smallest geographical level provided in the data, besides the actual street name. In order to make use of this information, I digitized photographs of historical maps into machine-readable form. Appendix Section B gives more information on this process.

due to two favorable conditions: I find that despite the fire, the street layout and street names have remained almost unchanged since 1900[15]. Additionally, the address information (street name and street number) provided in the Census data are of exceptional quality for historical data.

4 Estimation Strategy

4.1 Identifying the Effect of the Fire

To estimate the effect of the fire on its victims, I could simply run a simple OLS regression of the form

$$y_{it} = \alpha + \beta \, \text{inside}_i + \epsilon_i \tag{1}$$

where β is the coefficient of interest indicating the effect of residence in the razed district on outcomes indicated by y_{it}.

The fire was caused by an earthquake, which was an unpredictable event, and from historical records we know that exogenous factors such as wind and the availability of water determined where it stopped. However, due to the fact that the fires started in and near the main business district, the destroyed and the unaffected areas might be different on average. This is exacerbated by the fact that households do not randomly locate within a city. Instead, they tend to locate close to households with similar characteristics, thereby forming neighborhoods. Therefore, if we divide a city randomly in two halves, we cannot expect the resulting two groups of residents to have the same average demographic characteristics.

Indeed, observable characteristics of the treated and the non-treated group are *imbalanced*. That is, average differences between the two groups are statistically significant. Table 1 shows the means for all available pre-treatment characteristics, measured in 1900, for both groups. Individuals who lived in the part of the city that got destroyed by the fire were on average a bit older, less likely to be born in the US and born in California, had fewer children, lived in smaller households, were less likely to be home owners, and more likely to be in the labor force. The differences are statistically significant and economically meaningful.

If any observed or unobserved pre-treatment characteristics have an impact on the outcome of interest, for example on the propensity to migrate, the OLS estimate $\hat{\beta}$ would be biased.

[15] It is unclear whether this is also true for street numbers.

4.2 Spatial Regression Discontinuity Design

To account for observable and unobservable differences between groups, I leverage the rich geographic household location data that I created, and implement a spatial regression discontinuity design across the boundary of the razed area.

Suppose that we can express ϵ_i as

$$\epsilon_i = f(lat_i, lon_i) + u_i$$

where $E[u_i|lat_i, lon_i] = 0$ and $f(.)$ a continuous function in space, that is, of latitude and longitude.

I collapse the two-dimensional space into a one-dimensional running variable which is the linear distance of a household from the border of the razed area, distance$_i$, which I allow to be different on each side of the border.[16] In addition, I control for 30 border segment fixed effects Φ_b, in order to account for the complex geography of the city. I obtain the border segments by splitting the fire border into equal length sections.

Thus, I implement the spatial regression discontinuity design as follows:

$$y_{i,t} = \alpha + \beta \, \text{inside}_i + \lambda_1 \text{distance}_i + \lambda_2 \text{inside}_i \times \text{distance}_i + \Phi_b + u_i \tag{2}$$

To further account for spatial clustering of households within the city, I report standard errors corrected for arbitrary spatial correlation which allow for spatial dependence of an unknown form (Colella et al., 2019).

I implement two sets of regression weights. First, *Inverse Propensity Weights* (IPW) account for the linked sample not being representative of the general population. For example, being linked from 1900 to 1910 requires being alive ten years later, which is more likely for younger individuals, making the linked sample younger than the underlying population. Bailey et al. (2020) propose constructing IPW by predicting linking success with initial covariates such as birth year, name length and middle initial. Appendix D.1 describes the approach in detail.

Second, *Optimal Balancing Weights* (OBW) account for the specific geography of the razed area. The area is located in the Northeast of San Francisco, which means that there are border segments with many observations on one side, but very few on the other side.[17] These parts of the border are not very informative. Instead of excluding these observations, I reweight them in a way that gives more weight to those parts of the border with many observations on each side,

[16] For robustness, instead of using distance to the border as a running variable, I follow Dell (2010) and control for a linear function of longitude and latitude of household location.

[17] This is particularly the case in the Northwestern part of the border which runs close to the ocean.

following Heuermann et al. (2019). Appendix D.2 describes the construction of the weights.

While classic regression discontinuity designs leverage observations very close to the cutoff, I implement a variation called *donut RD*. This type of empirical approach excludes the observations that are closest to the cutoff (the donut hole). Often used as a robustness check, I implement the donut strategy because both the fire border as well as the household location coordinates are not precise to the meter.[18] Therefore, I exclude observations that are located within 100 meters from the fire border.

At the baseline, all individuals who are located up to one mile (1,600 meters) away from the border are part of the regression sample. For robustness, I implement different bandwidths.

4.3 Validity of the Approach

To test the validity of this estimation strategy, I carry out placebo regressions with pre-treatment characteristics as the dependent variable, meaning that I regress all available 1900 characteristics on the spatial RD specification from Equation 2.[19] Table 2 shows the coefficient on the treatment indicator. The sample comprises all individuals who were between 10 and 65 years old in 1900, and who got linked to their record in 1910.[20] The difference between treated and untreated individuals is statistically and economically insignificant for many relevant socioeconomic characteristics such as occupational income, illiteracy, and mortgage payments among home owners. In addition, the differences in age, race, nativity, measures name complexity and related outcomes are small and insignificant.[21]

Some imbalance remains, however. Affected individuals were more likely to be married, and therefore had more children, and lived with fewer of their own siblings. In addition, their households were smaller, which made them more likely to be household heads. Does this discontinuity imply that unmarried individuals were better able to fight the flames? This is unlikely. If the border was truly random, in expectation there are no discontinuities in predetermined characteristics. However, the realized border is only one random draw, which can lead to the existence of some discontinuities by chance. To account for these differences, which are all interrelated, I control for marital status and household size in 1900 in the baseline regressions.

[18] See Section 3.2 for more details.

[19] Unfortunately, almost the entire microdata of the 1890 Census was destroyed in a fire in 1921. Therefore it is impossible to construct first differences between 1890 and 1900, and I only test the balance in levels.

[20] I repeat the placebo analysis for a sample of all residents who get linked to at least one of their records in 1910, 1920, 1930 or 1940, see Appendix Table A2.

[21] Individuals who live in the razed district are 1.6 percentage points less likely to have a mother who was born in California, and the difference is significant at the 90 percent level.

In addition, home ownership rates were significantly lower in the part of San Francisco that was hit by the fire, in particular around the diagonal part of Market Street. Table 2 shows that the homeownership rate was about 15 percentage points lower in the razed area relative to the area outside the razed area. Since it makes a difference whether a family owned or rented the dwelling that burnt down, I control for home ownership status in all regressions and explore treatment effect differences between owners and renters. Finally, to capture the differences in treatment of the white and non-white population, I also control for race. Especially the Chinese community might have been affected differentially. The final estimation equation has the following form:

$$y_{i,t} = \alpha + \beta \, \text{inside}_i + \lambda_1 \text{distance}_i + \lambda_2 \text{inside}_i \times \text{distance}_i + \Phi_b + \mathbf{X}'_{i,1900}\gamma + e_i \tag{3}$$

Here, $\mathbf{X}_{i,1900}$ includes the pre-treatment characteristics marital status, household size, home ownership, and race. I show that my results are robust to the inclusion of a richer set of pre-treatment characteristics and the omission of all pre-treatment characteristics in Section 5.3.

5 The Short Run

In this section I discuss the short-run effects of the fire as measured in 1910, four years after the disaster. First, I present the effects on migration, occupational income and entrepreneurship. After discussing various robustness checks, I investigate coping strategies and mechanisms.

The fire was a shock to the affected residents along many dimensions. The loss of the home was both an emotional as well as a wealth shock, and due to insurance a liquidity shock for home owners. The complete destruction of the area meant that affected families had to evacuate the city immediately, and most went to stay with friends and family in the Bay Area (Fradkin, 2005). Evacuation without no prospects to return anytime soon is a major disruption for the victims of a catastrophe, because not only are families uprooted from their home, but so are many people around them, resulting in the fragmentation and destruction of social networks. Individuals remaining in or returning to San Francisco in the months or the years after the fire faced a short contraction of the economy.[22] This was followed by an increase of labor demand in particular in the construction and related industries that were involved in the rebuilding of the destroyed area. Due to the temporary shortage in housing along with a fast growing pop-

[22] The gold inflow necessary to pay for the insurance claims triggered the banking panic of 1907 which was followed by a short recession in the US (ODELL and WEIDENMIER, 2004). More locally, the destruction of most of San Francisco's business district decreased labor demand temporarily.

ulation, rents increased considerably after the fire.[23] In contrast, the nearby Bay Area counties faced a sharp increase in labor supply and housing demand. For example, Oakland's population more than doubled over night from the inflow of refugees displaced by the fire.[24]

5.1 Migration

While the destroyed area in San Francisco was soon rebuilt (Section 2.2), many evacuees are said to have stayed away permanently. They may either have settled in the originally temporary evacuation location, Oakland or another nearby county for most refugees, or have moved on to more distant locations. On one hand, the fire reduced the cost of migration considerably by destroying homes and businesses, and pushing people out of San Francisco, potentially inducing optimal migration. On the other hand, the wealth shock decreased the capacities for migration, in particular for costly, long-distance migration. To understand the location choices of affected families, I first estimate the effect of the fire on migration and location choice in 1910.

Panel (a) in Figure 3 shows graphical evidence of the fire's effect on across-county migration (i.e. living outside of San Francisco in 1910). The x-axis shows the distance of an individual's pre-fire home address to the border of the razed area in meters. The border is denoted by 0, values to the right show households away from the border in the razed area, values on the left show households away from the border in the unaffected area. The figure shows results from two specifications. First, the dots are coefficients of a non-parametric regression of migration rates on distance-to-border bins and the covariates in equation 3. The hollow dots are individuals in the donut hole which I exclude from the spatial RD regression. Second, the solid lines show the fitted values given by the coefficients on the distance-to-border measure from the spatial RD specification in equation 3, where the discontinuity in the middle shows that the fire increased the propensity to migrate by 6.3 percentage points. This estimate is somewhat lower than the 15.1 percentage points increase in migration induced by a volcano eruption in Iceland in 1973 (Nakamura et al., 2020). However, the former measures the effect on the 1900-1910 migration rate where the disaster happened in 1906, while the latter measures the effect on migration between the time of the volcano eruption and two years later. Hence, the impact on migration is fairly comparable between the two disasters.

While the relationship of distance and migration is flat for those unaffected by the fire,[25]

[23] Similarly, after the 1872 Fire in Boston, land values increased and the affected area was rebuild with higher quality buildings.

[24] While Oakland had less than doubled in the twenty years from 1880 to 1900, it almost tripled in the decade including the fire, from 67,000 inhabitants in 1900 to a population of over 150,000 in 1910, according to the Census.

[25] The coefficient on distance_i is slightly negative but statistically not significant for those unaffected by the fire.

migration rates are increasing as we move into the razed area. This is likely because the extent of the disruption is even stronger the further away from the unaffected area an individual had lived. While everyone in the affected area lost their homes to the flames, the closer someone lived to the border, the higher the expected share of their social network that lived in the unaffected part of town. Families living deep inside the razed district were much less likely to know someone across the border who could help them by providing shelter and other assistance.

The increase in migration could imply a better match of individuals to their new location, but it could also reflect individuals still living in the place that they evacuated to, due to inertia or to the inability to afford returning to San Francisco. I find evidence of the latter. Panel (b) in Figure 3 shows a small and statistically insignificant effect of the fire on leaving California, but it increases the likelihood of affected individuals to live in the Bay Area by 4.6 percentage points (Panel (c) in Figure 3). Taken together, these results imply that the fire displaced individuals away from San Francisco, but did not necessarily induce better location matches.

5.2 Economic Effects

Losing a home, evacuating the city and relocating to a new location is a major shock. How does this disruption impact the labor market trajectories and economic decisions of affected individuals? Given the economic turmoil following the fire, most of the evacuees had likely lost their job or business, and, with diminished social networks, were less likely to find a new one. I find that individuals were 6.3 percentage points more likely to have occupational downgrading, that is, they switched to a lower paying job between 1900 and 1910 (Table 3). Panel (a) of Figure 4 shows the effect on occupational income, which, in the absence of wage or income information, is a measure to quantify economic differences of occupations.[26] Affected individuals have 7.4 log points lower occupational income.[27] One typical example of such an occupational switch would be a downgrading from being a tailor in 1900 to a bartender in 1910.

The fire did not only destroy residential buildings but also businesses, including their inventory, tools or machines, and important documents. Insurance was partially able to refund these losses, but usually only covered about two thirds of the insured values. This implies that affected business owners might have been able to rebuild their business either in San Francisco or elsewhere, but some might have been unable to do so, due to lack of sufficient financial

[26]The occupational income score is constructed by IPUMS. It assigns all occupations the median income of that occupation in 1950 (when income for all individuals is available in the Census for the first time), in 1950-dollars, divided by 100. The occupational income score ranges from 1 to 80. Since it represents income, I perform a log transformation.

[27]The slope on distance is not significantly different from zero on either side of the border.

means. In addition, financial losses might have deterred aspiring entrepreneurs from starting a business. The Census asks about a worker's class since 1910, where the three possible values are "employer", "working on own account", and "works for wages or salary". By combining the first two categories, I create a variable measuring self-employment. Panel (b) of Figure 4 shows that affected individuals were 10 percentage points less likely to be self-employed in 1910. This difference is both statistically significant and economically meaningful. Unfortunately, information on self-employment before the fire is not available in Census data. However, it is unlikely that the difference in 1910 is driven by differences in 1900: since the business district was located in the part of San Francisco that was destroyed in the fire. If anything, I expect affected individuals to have higher rates of self-employment than unaffected individuals.[28] To further test this assumption, I proxy self-employment in 1900 by interacting occupation dummies with home ownership status. Controlling for this does not qualitatively change my results (Panel (d) in Table 4).

What are the economic implications of a lower rate of self-employment? If it implies more stable employment, it could be a welfare improvement for affected individuals. By constructing a variable indicating whether an individual was an employer in 1910, I show that the fire reduced the likelihood of an individuals to have employees in 1910 by 5.3 percentage points (Panel (c) in Figure 4). If having employees indicates the ownership of a larger business, this finding suggests a decrease in economic standing and is less likely to reflect a welfare-enhancing effect.

5.3 Robustness

To test the robustness of the spatial regression discontinuity design, I address a series of potential concerns, including robustness to pre-treatment controls, to different bandwidth choices, and to different sample restrictions. In addition, I use a sample that was linked based on a different method, and I implement a matching approach as an alternative identification strategy. Tables 4 and 5 show the baseline coefficients of the treatment effect on migration, occupational income, self-employment, and having employees. The panels below implement different robustness checks.

Pre-Treatment Controls. Omitting all pre-treatment controls (Panel (b) of Table 4) increases the magnitude of the treatment effects on migration, occupational income, self-employment,

[28] Pooley and Turnbull (1999) show that the majority of individuals in Great Britain at the beginning of the 20th Century lived very close to where they worked.

and having employees marginally. In Panel (c) of Table 4, I control for occupational income score, the number of children, labor force participation, a dummy for US born, and age in addition to the baseline controls race, home ownership, marital status and household size in 1900, with little effect on the coefficients. Similarly, controlling for the 1900 value of the dependent variable has little influence on the treatment effect estimates (Panel (d) of Table 4). I control for migration status in 1900 with an indicator whether a person was born outside of California. I proxy self-employment and having employees by including the full interaction of occupation dummies with home ownership status. This proxy exploits the fact that certain occupations had higher rates of self-employment than others.[29]

Weights. The baseline uses both inverse propensity weights (IPW) to account for the non-representativeness of the linked sample (Bailey et al., 2020), as well as optimal balancing weights (OBW) to focus on the most informative parts of the fire border (Heuermann et al. (2019)). Panel (e) of Table 4 omits the IPW with little influence on the coefficients. In Panel (f) of Table 4, instead of using OBW I manually drop the eastern parts of the fire border which have few or no observations outside the razed area, along the waterfront and the Mission Creek Channel. Figure A4 shows the included household locations. The treatment effect estimates on this sample are smaller in magnitude, but qualitatively unchanged.

RD Bandwidth. In Figure 5 I plot the treatment coefficient of the spatial RD design and 90 percent confidence intervals for 27 different bandwidth choices in steps of 50 meters. That is, the coefficient on the very right of each panel shows the treatment effect of the baseline specification with a bandwidth of 1600 meters (one mile) whereas the coefficient on the very left is based on a bandwidth of 300 meters. All specifications are donut RDs, excluding individuals who live less than 100 meters away from the boundary of the razed area.

The coefficients are stable to shrinking the bandwidth, but decrease for smaller bandwidths. At this point the confidence intervals increase, so that the baseline estimate is included in virtually all 90 percent confidence intervals. For very small bandwidths, the regression is likely overfitting the somewhat imprecise distance measure.

Function of Location. As explained in Section 4.2, I collapse the two-dimensions of space into a one-dimensional variable that measures the distance to the boundary of the razed area. Alternatively, I can implement a spatial RD with two running variables, latitude and longitude,

[29]The in-sample prediction of self-employment status based on this interaction is correct in more than 92% of individuals in a random Census sample of men who live in US cities in 1910.

following Dell (2010). Panel (b) of Table 5 shows that the treatment effects estimates are qualitatively stable, with some changes of magnitudes in both directions.

No Donut. Panel (c) of Table 5 shows results of a full RD regression, where I also include individuals living less than 100 meters away from the border of the razed district. Due to imprecision in the household coordinates, I expect serious measurement error of the treatment indicator for these individuals. Due to attenuation bias, the coefficients in Panel (c) have a smaller magnitude compared to the baseline.

Race. More than 90 percent of San Francisco's residents in 1900 were white, and the biggest minority group were Asians living in Chinatown. Due to racial animus, the Chinese population might have been affected differently than the white majority. Instead of controlling for race, I restrict the sample to whites in Panel (d) of Table 5. The coefficients remain virtually unchanged, with the exception of the effect on occupational earnings which drops from -7.4 log points to -6.6 log points.

Age. The baseline sample includes every linked man who was between 20 and 75 years old in 1910. Since labor force participation rates start dropping for men older than 50 years old, including older individuals might attenuate the occupation and employment effects. Indeed, restricting the sample to individuals aged 20 to 50 increases the magnitude of the effects on occupational income, self-employment, and having employees (Panel (e) of Table 5).

Alternative Linking Approach. To link a person's Census records across years, I use ABE (Abramitzky, Boustan and Eriksson, 2012, 2014, 2019) standard links with phonetic cleaning of names. The advantage of this fully automated linking approach is the reproducibility and the fact that it entirely relies on stable characteristics. However, due to the sparseness of the linking characteristics, many potential links are not uniquely identified and get therefore discarded. In addition, full automation is rather inflexible and does not allow incorporating more detailed knowledge on how information of one individuals might be recorded differently across Census waves. An example of this would be census takers recording a person's first name in one year and their nickname in another year. Machine learning based approaches, such as the one proposed by Feigenbaum (2016), allow for greater flexibility. As a robustness check, I rerun the analysis on a 1900-1910 linked sample with crosswalks based on Price et al. (2021) available from the Census Linking Project (Abramitzky et al., 2020).

Price et al. (2021) use a dataset of record links created by genealogists from FamilySearch to train a supervised machine learning algorithm. The "ground truth" from the hand-linked records inform the approach on how to pre-process the data and which features to include to identify new links. The algorithm then shows which features matter most. The advantage of the approach are high match rates with low to moderate false positive rates. However, some of the included features are non-stable characteristics, such as the geographic distance between the residences of the records in each Census wave or indicators whether the relation to the household head and marital status are the same between censuses. This is problematic because the resulting sample will both be under-representative of individuals who have experienced such changes, as well as linking errors will be biased towards greater stability in these characteristics. This is particularly problematic when studying populations that have been affected by a disruptive shock such as the 1906 San Francisco Fire.

The results from the analysis using Price et al. (2021) linking crosswalks are qualitatively similar to the baseline (Panel (f) of Table 5). As expected, the effect on migration is somewhat attenuated. While the magnitude of the effects on occupational income and self-employment are remarkably similar, the estimate of effect on having employees is somewhat smaller compared to the baseline.

5.4 Mechanisms

To gain a further understanding of the ways that the fire affected its victims, I first analyze how families reacted to the disaster, then I study heterogeneity in treatment effects between pre-disaster renters and owners, and test whether the presence of unaffected extended family in San Francisco mitigated the effects of the fire.

Families' Economic Decisions. Taken together, my main results imply a negative short run effect on occupation and workplace choice. Destruction of businesses and relocation due to evacuation likely limited employment options for affected residents in the short run. In the following, I present evidence for how families reacted to these economic hardships (Table 3).

First of all, I find evidence for an added worker effect: wives of affected individuals were 2.9 percentage points more likely to be part of the labor force than wives of unaffected individuals, where the labor force participation rate of the latter group was just five percent. In 1900, wives who lived in the affected area had somewhat higher labor force participation rates, although the coefficient is not significant. Controlling for wife's labor force status in 1900, the effect on wife's labor force participation in 1910 reduces to 2.7 percentage points.[30] In addition, I find that in affected households, children were 2 percentage points more likely to be part of the labor force in 1910 and 1.2 percentage points less likely to go to school. However, these estimates are not statistically significant. Besides women or children entering the labor force to support household income, individuals might have reacted by delaying or reducing marriages and childbearing. Indeed, affected individuals were 1.9 percentage points less likely to get married between 1906 and 1910, which is a sizable effect given that one in ten unaffected men got married over this time span. Besides a delay of marriages because of financial considerations, this reduction is likely due to a deterioration of the marriage market related to the disruption to communities and social networks. While negative, the effect on fertility is small and statistically insignificant.

Renters and Home Owners. The financial implications of a destroyed house are very different for renters than for owners. While both lost furniture and other mobile possessions, insured home owners received payments from their fire insurance of, on average, two thirds of the insured value.

To detect potential treatment effect heterogeneity, I interact an indicator for home ownership status in 1900 with all regressors in equation 3. Table 6 shows that while among renters, the fire increases the propensity to migrate by 7.9 percentage points, the effect is smaller among renters. However, the interaction of treatment status and home ownership is statistically insignificant, indicating that I cannot rule out equality of the effects. There is no difference in treatment effect on occupational income (Column 2) or having employees (Column 4). Column 3 suggests that renters see a stronger negative effect on self-employment, but the standard error on the interaction term is too large to rule out equality of the effect.

Informal Risk Sharing: Kin Presence in San Francisco. Social networks such as the extended family can be important determinants for how families react to a shock such a natural disaster. This was even more true at the beginning of the 20th Century in the absence of federal disaster relief. Extended family living in the unaffected parts of San Francisco might have provided food and shelter to those who lost their homes. In the months after the disaster, relatives could have been able to provide both financial and immaterial resources such as job referrals or employment in their own businesses.

The Census only records kinship relations for individuals who are cohabiting.[31] Therefore,

[30] The regression includes men who were unmarried in 1900 by controlling for a factor variable that takes values for the wife's labor force status or indicates the absence of a wife.

[31] More precisely, we know in which way individuals are related to the household head, and who belongs to

I construct a novel measure of kin presence which leverages information on race, surnames, and the exact geographic location of other households in San Francisco. The measure is the name-commonness adjusted sum of same-surname individuals of the same race, weighted by distance to the household.[32] Unlike existing approaches selecting an arbitrary cutoff beyond which households sharing a last name and race are not considered related, I take into account all isonymic households[33] of the same race living somewhere in the city. I assume that the further apart two households are located, the lower the probability that their members are related to one another. To account for common surnames, I adjust with a penalty term for surname commonness.

To formalize this, I define $P(R_{ij}|s_i = s_j, r_i = r_j, d_{i,j}) \equiv P(R_{ij}|d_{i,j})$ as the probability that household i and j are related, given that they have same surname and race, and live distance $d_{i,j}$ away from each other. $P(R_{ij}|d_{i,j})$ should have the following properties:

1. $P'(R_{ij}|d_{i,j}) < 0$, $P''(R_{ij}|d_{i,j}) > 0$
2. $P(R_{ij}|0) = 1$
3. $\lim\limits_{d_{i,j} \to \infty} P(R_{ij}|d_{i,j}) = 0$

Under Property (1), $P(R_{ij}|d_{i,j})$ decreases with the distance between two households, and it decreases faster in close proximity. With Property (2) I assume that cohabiting individuals who share the same surname and race are always related.[34] Finally, Property (3) is an intuitive

sub-families within the household.

[32]Since R_{irs} is based on same surnames, it can only measure *patrilineal* kinship ties: women usually take their husband's surname upon marriage, so that their surnames do not permit tracing their side of the family. In addition, not all individuals who share a surname will be directly related to each other, in particular those with common last names. At the same time, if Census enumerators recorded last names slightly differently, for example *Abbott* and *Abbot*, these individuals will not be considered related to each other.

[33]Isonymic households are households with the same name. Previous research on kin propinquity focuses on households that live in very close geographic proximity (Nelson, 2020; Hacker and Roberts (2017); Smith (1989)), exploiting sequential enumeration under the assumption that proximity on the enumeration form correlates with geographic proximity. The limitation of these existing measures is that they are hyper local, because they focus on households who live within the range of just a few buildings, and in the same enumeration district. For example, Smith (1989) considers next door households or the five closest nearby households, while Hacker and Roberts (2017) identify mothers in law based on name, age, marital status in the ten nearby households. Given that the average San Francisco enumeration district has 200 to 300 households in 1900, this likely underestimates the number of relatives, in particular for those households living close to the borders of enumeration districts. Crucially, these types of approaches disregard most potential relatives living across the fire border.

[34]Among all non-group quarter households in California in 1900, 0.8 percent have a one or more members that share race and surname with the household head, but are not related with them. This is mostly driven by Chinese households where this share is 10.5 percent while only 0.5 percent of white households have a non-related isonymic householder.

regularity condition stating that $P(R_{ij}|d_{i,j})$ approaches zero as the distance between households i and j approaches infinity.

A functional form meeting all the aforementioned conditions is $\frac{1}{(x+1)^k}$, where the parameter k illustrates my assumption about kinship propinquity: the higher k is, the faster $P(R_{ij}|d_{i,j})$ declines with distance between households i and j. I set the parameter $k = 1.5$, results are qualitatively robust towards other parametrizations. In addition, I normalize the measure to have a mean of zero and a standard deviation of one for ease of interpretation.

Simply calculating the distance-weighted sum of isonymic households would overestimate the number of relatives for families with very common surnames like Smith, for example: In 1900, 2511 individuals with that surname lived in San Francisco, and it is unlikely that they were all related to each other. To account for common names, I divide the probability weighted sum of isonymic relatives by the expected number of individuals with surname s and race r in San Francisco. I derive this quantity by multiplying the number of individuals in San Francisco, A, by the share of persons with with surname s and race r who live in California, excluding San Francisco, η_{rs}. The measure of kin presence R_{irs} can be expressed as

$$R_{irs} = \frac{1}{A\eta_{rs}} \sum_{j \neq i \in S_{rs}} \frac{F_{rs}}{(dist_{i,j}+1)^k} \tag{4}$$

where F_{rs} is the number of individuals in a given household j that is located $dist_{i,j}$ meters away from household i. By multiplying with and dividing by $|S_{rs}|$, the measure of the set of individuals with surname s and race r in San Francisco, R_{irs} can be rewritten as

$$R_{irs} = \underbrace{\frac{|S_{rs}|}{A\eta_{rs}}}_{\substack{\text{excess population} \\ \text{of name-race in SF}}} \times \underbrace{\sum_{j \neq i \in S_{rs}} \frac{1}{|S_{rs}|} \frac{F}{(dist_{i,j}+1)^k}}_{\text{mean kin probability}} \tag{5}$$

which illustrates the interpretation of the kin measure as the product of the excess population of surname s and race r in San Francisco with the mean kin probability derived by the settlement patterns of these individuals within the city.

I construct R_{irs} separately for kin in the affected part of the city, R^{in}_{irs}, and kin in the unaffected part, R^{out}_{irs}. After converting them into z scores for ease of interpretation, I include both measures in the analysis to uncover differential effects.[35] The analysis is restricted to individuals with positive kin measures, therefore uncovering intensive margin effects. Table 7 reports the results. Column 1 shows that, for unaffected individuals, a one standard deviation higher measure of unaffected kin is associated with a 2.1 percentage points lower propensity to mi-

[35]Notice that the correlation between R^{in}_{irs} and R^{out}_{irs} is only 0.05 in my regression sample.

grate. The coefficient on the measure of affected kin has the same magnitude, but is more noisily estimated. This is consistent with previous literature which finds that strong family ties are associated with lower geographic mobility (Alesina and Giuliano, 2014). The interaction terms $R^{out}_{irs}\times$ Inside and $R^{in}_{irs}\times$ Inside represent the additional effect of unaffected and affected kin on those individuals who themselves were affected by the fire. While statistically insignificant, the magnitude of the interaction term implies that more unaffected kin lowered the propensity to migrate almost twice as much for affected than for unaffected kin. In contrast, the stronger presence of affected kin has no differential effect on migration rates of affected individuals. The coefficient in the last row, "Inside Razed Area", shows the treatment effect of the fire for an individual with average values of both kin measures. Notice that the regression discontinuity design gives a causal interpretation to the treatment coefficient, while the number and location of extended family in San Francisco could be endogenous to the outcomes of interest.

Affected individuals with more unaffected kin have higher occupational incomes and are both more likely to be self-employed and to have employees (Columns 2, 3 and 4). While none of the coefficients is significant at conventional levels, the results lend suggestive support to the hypothesis of informal risk sharing as a mitigation mechanism after the disaster. While kin presence in the destroyed part of San Francisco is positively associated with occupational income and self-employment for unaffected individuals, likely reflecting valuable connections to the business district, this is not the case for individuals who lost their jobs, businesses, and social networks due to the fire.

6 The Long Run

6.1 Adults

What are the long-run effects of the fire? Previous research shows that economic shocks such as entering the labor market during a recession can have long-run effects on wages and other labor market outcomes (Kahn, 2010). But studies on the 2005 hurricane Katrina suggest that after a few years, affected residents actually do better than the control group. A suggested mechanism for this reversal of fortune is migration out of a weak labor market and disadvantaged neighborhoods (Groen et al., 2016; Deryugina et al., 2018). But San Francisco was fast-growing and prosperous in the early twentieth century, and while the city took a hit from the fire and earthquake devastation, population levels rebounded quickly and reconstruction was swift. In addition, the conflagration happened before the introduction of federal disaster relief, making the context less favorable for positive long-run effects.

In the following, I estimate the effect of the fire not only in the year 1910, but also for 1920, 1930, and 1940, to see how the effect evolves over time. I focus on a sample of individuals who were between 10 and 35 in 1900 to rule out cohort effects. Notice that while this sample follows a consistent *cohort*, the composition of the sample is not entirely consistent over time, as only a small set of individuals can be linked to all Census years. First, I trace how the fire affects location choice over time, and then I investigate the persistence of the effect on occupations and entrepreneurship.

Panel (a) of Figure 6 shows the effect of the fire on living outside of San Francisco. While affected residents were clearly more likely to have lived outside the city in 1910, the effect decreases considerably by 1920 and is statistically indistinguishable from then onwards. This could reflect the fact that unaffected individuals caught up in migration rates between 1910 and 1920 or that some evacuees were able to return to San Francisco eventually. I find evidence for the latter: among those who lived outside of the city in 1910, affected individuals were more likely to have returned to San Francisco by 1920 (Panel (b) of Figure 6).

While individuals recover from occupational downgrading over time, the data does not support the notion of reversal of fortune for the victims of the San Francisco Fire. The magnitude of the negative effect on occupational income is highest in 1910, and it decreases somewhat in 1920 and 1930 (Panel (a) of Figure 7). The magnitude increases again for 1940, but is statistically indistinguishable from the other coefficients. It is possible that selective attrition from the labor force widens the occupational income gap in 1940. This hypothesis is supported by the fact that the fire increases the likelihood for occupational downgrading in 1910 and 1940, but not in 1920 and 1930 (Panel (b) of Figure 7). Due to financial constraints, individuals who were pushed to lower paying occupations might have been more likely to stay in the labor force and delay retirement as they got older.

Entrepreneurship is an important factor for economic development (James A. Schmitz, 1989), and it can increase income mobility for some (Halvarsson et al., 2018; Åstebro et al., 2011; Hamilton, 2000). However, there is only little research about the long-run effects of historical shocks on the entry to or exit from entrepreneurship.[36] Section 5 finds that the fire decreases a person's likelihood of being self-employed and having employees four years after the disaster. Panels (c) and (d) of Figure 8 show that these effects are highly persistent with magnitudes changing little over time. This might be explained by a lasting effect of the negative wealth shock which makes it more difficult both for former business owners as well as for aspiring entrepreneurs to start a new enterprise. In addition, it is possible that the experience of the

[36] Makridis and McGuire (2019) is a notable exception.

fire affected preferences away from higher risk higher return entrepreneurship to more stable forms of employment. Entrepreneurship itself is remarkably persistent: in any given decade, only about 20 percent of individuals in my sample switch into or out of self-employment, making it possible for a shock to have a very long-run effect.

6.2 Children

How did the disaster affect those who were children at the time or not even born yet? Nakamura et al. (2020) find that the displacement effects of a volcano eruption are unequally distributed across generations: While the parental generation is slightly worse off, those who were younger than 25 at time of displacement enjoyed higher lifetime earnings and invested more in education. The authors argue that the beneficiaries of displacement are those who were a bad fit in the old location, and that being young at time of migration offered the opportunity to re-optimize. Is the same true for children who were directly or indirectly affected by the 1906 Fire?

To answer this question, I analyze the effect of the disaster on children born between 1891 and 1920. Treatment status is assigned via father's household location in 1900. Children born before 1906 experienced the fire themselves, and were therefore *directly affected*. Children of 1900 San Francisco residents who were born after 1906 were *indirectly affected* if their father had lived in the razed area. For comparability, I will focus on 1940 outcomes when all individuals in the children sample had reached adulthood. To uncover potential differences in being directly or indirectly affected, I will interact the treatment variable and all control variables with an indicator for being born before the year 1906.

While I've shown in Section 5 that the disaster forced adults into lower paying jobs, decreased entrepreneurship and delayed marriages, the fire might have interrupted school attendance not only in 1906, but the added worker effect might have kept children out of school in following years, too. If children were unable to make up for this, the fire could have a persistent negative effect on educational attainment. On the other hand, research implies that forced migrants experience shifts in preferences towards higher investment in human capital (Becker et al., 2020). The loss of material possession in the fire could have had a similar effect. In addition, Sacerdote (2012) finds a silver lining for children displaced by hurricane Katrina: after a short decline in school performance, their grades improve over their pre-hurricane performance.

Table 8 reports the effect on total years of schooling, high school graduation and college attendance. I implement the spatial RD specification, additionally controlling for age of the child to account for cohort effects in educational attainment (Goldin and Katz, 1999). Column 1 shows that indirectly affected children obtained 0.52 fewer years of education than unaffected children. The very small magnitude of the interaction term suggests that children who personally experienced the fire were not differentially affected, but the large standard errors do not allow to reject a heterogeneous treatment effect. Indirectly (directly) affected children were 3.9 percentage points (6.8 percentage points) less likely to graduate from high school and 12 percentage points less likely to attend college. Notably, the negative effect on college attendance is only 5.1 percentage points for directly affected children, with the difference between direct and indirect treatment effects statistically significant. The effect sizes are similar to the in utero effect of the 1918 flu pandemic, which decreased educational attainment of children of infected mothers by about five months and lowered high school graduation rates by twelve to 15 percent (Almond, 2006).

Entrepreneurship is strongly transmitted across generations, with children of entrepreneurs being 60 percent more likely to be entrepreneurs themselves (Lindquist et al., 2015). At the same time, research shows that adverse economic conditions during childhood, such as the Great Depression, can alter children's believes and increase the likelihood of entrepreneurship (Makridis and McGuire, 2019). Given the persistent negative effect on entrepreneurship among the parent generation, the effect of the fire on the following generation's entrepreneurship is ambiguous.

I find no significant effect on self-employment or having employees, but given the standard errors, I cannot rule out economically meaningful effects in either direction (Table 8). However, the lower bound of the 90 percent confidence interval of the effect on entrepreneurship is -0.044 for indirectly and -0.057 for directly affected children, which is a smaller magnitude than the point estimate of the effect on adults in 1940. The same is true for the effect on having employees. This suggests support for Makridis and McGuire (2019)'s finding that negative economic shocks can foster entrepreneurship in the younger generation, hence counteracting the negative effect operating through lowered entrepreneurship of the fathers' generation.

7 Conclusion

In this paper I study the short- and long-run effects of the 1906 San Francisco Fire on residents who lost their home in the natural disaster. Caused by an earthquake in April 1906, the fire destroyed nearly half of San Francisco's residential buildings. About 250,000 out of the 420,000 inhabitants evacuated the city after the disaster. I link US Census records from 1900 to 1910, 1920, 1930, and 1940 to follow 1900 San Francisco residents and their sons over time. To account for non-random household location within the city, I use a spatial regression discontinuity design across the border of the razed area as my identification strategy.

I find that the fire displaced individuals in the short run, but there is no persistent effect on location choice in the long-run. Affected residents were forced into lower paying occupations and out of entrepreneurship after the disaster. Due to this negative impact on household income, wives of affected residents are more likely to be part of the labor force. Overall, my findings suggest that the victims of the fire were able to recover from the short-run economic effects over time, but the reduction in entrepreneurship is highly persistent. Affected children have lower educational attainment in 1940 which amounts to a half a year less schooling. My findings suggest that the effects are not more pronounced for sons who experienced the fire themselves compared to those who were born to affected fathers after 1906.

In light of climate change, urbanization and continuing population growth, the number of people threatened by natural disasters is expected to grow. My findings emphasize the need for well-targeted and appropriately sized disaster relief: while modern relief programs have proven successful in mitigating the economic impacts of hurricane Katrina (Bleemer and Van Der Klaauw, 2017; Vigdor, 2007), the absence of such programs likely played a role in the slow recovery of victims after the 1906 San Francisco Fire. In addition, my paper suggests high importance ex-ante mitigation spending that is particularly targeted to places that create high levels of individual welfare, such as counties identified to have positive causal effects on intergenerational mobility (Chetty and Hendren, 2018). In that regard, ex-post relief programs that take the quality of the affected place into account will either facilitate out-migration, which has proven to be beneficial after hurricane Katrina (Deryugina et al., 2018) or a volcano eruption in Iceland (Nakamura et al., 2020), or support individuals in returning to their old home as soon as possible which might have mitigated the impact of the fire in San Francisco. Ultimately, whether a natural disaster induces "moving to opportunity" depends on whether opportunity awaits in the origin location or somewhere else.

References

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration. *American Economic Review*, (5):1832–1856, 2012. doi: 10.1257/aer.102.5.1832. URL http://dx.doi.org/10.1257/aer.102.5.1832http://dx.doi.org/10.1257/aer.102.5.1832.

Ran Abramitzky, Leah Platt Boustan, and Katherine Eriksson. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration. *https://doi.org/10.1086/675805*, 122(3):467–506, jul 2014. doi: 10.1086/675805. URL https://www.journals.uchicago.edu/doi/abs/10.1086/675805.

Ran Abramitzky, Leah Boustan, and Katherine Eriksson. TO THE NEW WORLD AND BACK AGAIN: RETURN MIGRANTS IN THE AGE OF MASS MIGRATION. *ILR Review*, 2019. doi: 10.1177/0019793917726981.

Ran Abramitzky, Leah Boustan, and Myera Rashid. Census Linking Project: Version 1.0 [dataset], 2020. URL https://censuslinkingproject.org.

Philipp Ager, Katherine Eriksson, Casper Worm Hansen, and Lars Lønstrup. How the 1906 San Francisco earthquake shaped economic activity in the American West. *Explorations in Economic History*, 77:101342, jul 2020. ISSN 0014-4983. doi: 10.1016/J.EEH.2020.101342.

Alberto Alesina and Paola Giuliano. Family Ties. *Handbook of Economic Growth*, 2:177–215, jan 2014. ISSN 1574-0684. doi: 10.1016/B978-0-444-53538-2.00004-6.

Douglas Almond. Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *https://doi.org/10.1086/507154*, 114(4):672–712, jul 2006. doi: 10.1086/507154. URL https://www.journals.uchicago.edu/doi/abs/10.1086/507154.

Thomas Åstebro, Jing Chen, and Peter Thompson. Stars and Misfits: Self-Employment and Labor Market Frictions. *http://dx.doi.org/10.1287/mnsc.1110.1400*, 57(11):1999–2017, sep 2011. doi: 10.1287/MNSC.1110.1400. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1110.1400.

Martha Bailey, Connor Cole, Morgan Henderson, and Catherine Massey. How Well Do Automated Linking Methods Perform? Lessons from U.S. Historical Data. Technical report, 2019.

Martha J Bailey, Connor Cole, Morgan Henderson, Catherine Massey, Eytan Adar, George Alter, Jeremy Atack, Hoyt Bleakley, Leah Boustan, John Bound, Charlie Brown, Matias Cattaneo, William Collins, Dora Costa, Shari Eli, Katherine Eriksson, James Feigenbaum, Joseph Ferrie, Katie Genadek, Tim Guinanne, Mary Hansen, Kris Inwood, Maggie Levenstein, Bhash Mazumder, Jorgen Modalsli, Adriana Lleras-Muney, Jared Murray, Joseph Price, Paul Rhode, Evan Roberts, Steve Ruggles, Mel Stephens, Sarah Anderson, Garrett Anstreicher, Ali Doxey, Meizi Li, Shariq Mohammed, Paul Mohnen, Mike Ricks, and Hanna Zlotnick. How Well Do Automated Linking Methods Perform? Lessons from US Historical Data †. *Journal of Economic Literature*, (4):997–1044, 2020. doi: 10.1257/jel.20191526. URL https://doi.org/10.1257/jel.20191526.

Sascha O. Becker, Irena Grosfeld, Pauline Grosjean, Nico Voigtländer, and Ekaterina Zhuravskaya. Forced migration and human capital: Evidence from post-WWII population transfers†. *American Economic Review*, 110(5):1430–1463, may 2020. ISSN 19447981. doi: 10.1257/aer.20181518.

Wayne Blanchard. Deadliest U.S. Disasters - Top Fifty. Technical report, FEMA Emergency Management, 2008.

Zachary Bleemer and Wilbert Van Der Klaauw. Disaster (Over-)Insurance: The Long-Term Financial and Socioeconomic Consequences of Hurricane Katrina. Technical report, 2017.

Leah Platt Boustan, Matthew E. Kahn, and Paul W. Rhode. Moving to higher ground: Migration response to natural disasters in the early twentieth century. *American Economic Review*, 102 (3):238–244, 2012. ISSN 00028282. doi: 10.1257/aer.102.3.238.

Leah Platt Boustan, Matthew E Kahn, Paul W Rhode, and Maria Lucia Yanguas. The effect of natural disasters on economic activity in US counties: A century of data. *Journal of Urban Economics*, 118:103257, 2020. doi: 10.1016/j.jue.2020.103257. URL https://doi.org/10.1016/j.jue.2020.103257.

Raj Chetty and Nathaniel Hendren. THE IMPACTS OF NEIGHBORHOODS ON INTERGENERATIONAL MOBILITY II: COUNTY-LEVEL ESTIMATES *. *The Quarterly Journal of Economics*, pages 1163–1228, 2018. doi: 10.1093/qje/qjy006. URL https://academic.oup.com/qje/article/133/3/1163/4850659.

Fabrizio Colella, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. Inference with Arbitrary Clustering. Technical report, 2019. URL https://acregstata.weebly.com.

Melissa Dell. The Persistent Effects of Peru's Mining Mita. *Econometrica*, 78(6):1863–1903, 2010. ISSN 0012-9682. doi: 10.3982/ecta8121.

Tatyana Deryugina, Laura Kawano, and Steven Levitt. The economic impact of hurricane Katrina on its victims: Evidence from individual tax returns. *American Economic Journal: Applied Economics*, 10(2):202–233, 2018. ISSN 19457790. doi: 10.1257/app.20160307.

EMDAT. OFDA/CRED International Disaster Database, 2020.

James J Feigenbaum. A Machine Learning Approach to Census Record Linking. *Working Paper*, pages 1–34, 2016. doi: 10.1109/3DUI.2016.7460053.

PL Fradkin. *The great earthquake and firestorms of 1906: how San Francisco nearly destroyed itself*. 2005.

Claudia Goldin and Lawrence F Katz. The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940. *JEP ja03 Journal of Economic Perspectives*, 13:37–62, 1999.

Greely. Special Report of Maj. Gen. Adolphus W. Greely. Technical report, 1906. URL https://history.army.mil/documents/sfearthquake/GreelyReport.htm.

Jeffrey Groen, Mark J Kutzbach, and Anne Elise Polivka. Storms and Jobs: The Effect of Hurricanes on Individualss Employment and Earnings Over the Long Term. *SSRN Electronic Journal*, 2016. doi: 10.2139/ssrn.2782038.

Patricia Grossi and Robert Muir-Wood. The 1906 San Francisco Earthquake and Fire: Perspectives on a Modern Super Cat. page 28, 2006.

J. David Hacker and Evan Roberts. The impact of kin availability, parental religiosity, and nativity on fertility differentials in the late 19th-century United States. *Demographic research*, 37(34):1049, oct 2017. doi: 10.4054/DEMRES.2017.37.34. URL /pmc/articles/PMC5926193//pmc/articles/PMC5926193/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5926193/.

Daniel Halvarsson, Martin Korpi, and Karl Wennberg. Entrepreneurship and income inequality. *Journal of Economic Behavior and Organization*, 145:275–293, jan 2018. doi: 10.1016/J.JEBO.2017.11.003. URL https://doi.org/10.1016/j.jebo.2017.11.003.

Barton H Hamilton. Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment. 2000.

G Hansen and E Condon. *Denial of Disaster: the Untold Story and Unpublished Photographs of the San Francisco Earthquake and Fire of 1906*. 1989.

Gladys Hansen, Richard Hansen, and William Blaisdell. *Earthquake, Fire & Epidemic: Personal Accounts of the 1906 Disaster*. 2014.

Richard Hansen and Gladys Hansen. *Images of America: 1906 San Francisco Earthquake*. Arcadia Publishing, 2013.

Daniel F Heuermann, Johannes F Schmieder, and Philipp vom Berge. Do Active Labor Market Policies Help Workers ? Evidence from a Spatial Regression Discontinuity Design. 2019.

J. J. Hoag. Guide map of the city & county of San Francisco, compiled from latest official surveys, 1907.

Richard Hornbeck. Dust Bowl Migrants: Identifying an Archetype. 2020.

Richard Hornbeck and Daniel Keniston. Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872). *American Economic Review*, 107(6):1365–1398, 2017. ISSN 0002-8282. doi: 10.1257/aer.20141707. URL http://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.20141707.

Richard Hornbeck and Suresh Naidu. When the levee breaks: Black migration and economic development in the American South. *American Economic Review*, 104(3):963–990, 2014. ISSN 00028282. doi: 10.1257/aer.104.3.963.

Robert James. Six Bits or Bust: Insurance Litigation Over the 1906 San Francisco Earthquake and Fire. *Western Legal History*, 24(2), 2011.

Jr. James A. Schmitz. Imitation, Entrepreneurship, and Long-Run Growth. *https://doi.org/10.1086/261624*, 97(3):721–739, oct 1989. doi: 10.1086/261624. URL https://www.journals.uchicago.edu/doi/abs/10.1086/261624.

Lisa B. Kahn. The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316, apr 2010. ISSN 0927-5371. doi: 10.1016/J.LABECO.2009.09.002.

Lawrence J. Kennedy. The progress of the fire in San Francisco April 18th-21st, 1906 : as shown by an analysis of original documents. Technical report, 1908.

Jerome Landfield. *Operation kaleidoscope: melange of personal recollections: "A City in Ruins"*. San Francisco, 1906.

Matthew J Lindquist, Joeri Sol, and Mirjam Van Praag. Why Do Entrepreneurial Parents Have Entrepreneurial Children? *Source: Journal of Labor Economics*, 33(2):269–296, 2015. doi: 10.1086/678493.

Jason Long and Henry Siu. Refugees from dust and shrinking land: Tracking the dust bowl migrants. *Journal of Economic History*, 78(4):1001–1033, dec 2018. ISSN 14716372. doi: 10.1017/S0022050718000591.

Christos Makridis and Erin McGuire. Refined by Fire: The Great Depression and Entrepreneurship. *SSRN Electronic Journal*, jul 2019. doi: 10.2139/SSRN.3371991. URL https://papers.ssrn.com/abstract=3371991.

Emi Nakamura, Jósef Sigurdsson, and Jón Steinsson. The Gift of Moving: Intergenerational Consequences of a Mobility Shock. Technical report, 2020.

Matt A. Nelson. The decline of patrilineal kin propinquity in the United States, 1790–1940. *Demographic Research*, 2020. doi: 10.2307/26967815. URL https://www.jstor.org/stable/10.2307/26967815?seq=1{&}cid=pdf-.

KERRY A. ODELL and MARC D. WEIDENMIER. Real Shock, Monetary Aftershock: The 1906 San Francisco Earthquake and the Panic of 1907. *The Journal of Economic History*, 64(4): 1002–1027, 2004. ISSN 1471-6372. doi: 10.1017/S0022050704043062. URL https://www.cambridge.org/core/journals/journal-of-economic-history/article/real-shock-monetary-aftershock-the-1906-san-francisco-earthquake-and-the-pa AB2C757125EAE9DECB624073CA300097.

Colin G Pooley and Jean Turnbull. The journey to work: a century of change. *Area*, 31:281–292, 1999.

Joseph Price, Kasey Buckles, Jacob Van Leeuwen, and Isaac Riley. Combining family history and machine learning to link historical records: The Census Tree data set. *Explorations in Economic History*, 80:101391, apr 2021. ISSN 0014-4983. doi: 10.1016/J.EEH.2021.101391.

Punnet Brothers. Map of part of San Francisco, California, April 18, 1908 : showing buildings constructed and buildings under construction during two years after fire of April 18, 1906., 1908.

T.A. Rickard, editor. *After Earthquake and Fire*. Mining and Scientific Press, San Francisco, 1906.

R.J. Waters & Co. San Francisco, California, showing the areas destroyed by fire, April 18-21, 1906, 1906.

Steven Ruggles, Sarah Flood, Sophia Foster, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek. IPUMS USA: Version 11.0 [dataset], 2021. URL https://doi.org/10.18128/D010.V11.0.

Bruce Sacerdote. When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita. 2012. doi: 10.1257/app.4.1.109. URL http://dx.doi.org/10.1257/app.4.1.109.

Narayan Sastry and Jesse Gregory. The Location of Displaced New Orleans Residents in the Year After Hurricane Katrina. *Demography*, 51(3):753–775, 2014. ISSN 15337790. doi: 10.1007/s13524-014-0284-y.

James Siodla. A Non-Random Walk in the City: The 1906 San Francisco Disaster and Industry Localization. pages 1–30, 2015a.

James Siodla. Razing San Francisco: The 1906 disaster as a natural experiment in urban redevelopment. *Journal of Urban Economics*, 89:48–61, 2015b. ISSN 00941190. doi: 10.1016/j.jue.2015.07.001. URL http://dx.doi.org/10.1016/j.jue.2015.07.001.

James Siodla. Clean slate: Land-use changes in San Francisco after the 1906 disaster. *Explorations in Economic History*, 65(April):1–16, 2017. ISSN 10902457. doi: 10.1016/j.eeh.2017.04.001. URL http://dx.doi.org/10.1016/j.eeh.2017.04.001.

James Siodla. Firms, fires, and firebreaks: The impact of the 1906 San Francisco disaster on business agglomeration. *Regional Science and Urban Economics*, 88(February):103659, 2021. ISSN 18792308. doi: 10.1016/j.regsciurbeco.2021.103659. URL http://dx.doi.org/10.1016/j.regsciurbeco.2021.103659.

Daniel Scott Smith. "All in Some Degree Related to Each Other": A Demographic and Comparative Resolution of the Anomaly of New. *Source: The American Historical Review*, 94(1):44–79, 1989. URL https://www.jstor.org/stable/1862077.

Yannay Spitzer, Gaspare Tortorici, and Ariell Zimran. International Migration Responses to Natural Disasters: Evidence from Modern Europe's Deadliest. 2021. URL http://www.nber.org/data-appendix/w27506.

Stephen Tobriner. An EERI reconnaissance report: Damage to San Francisco in the 1906 earthquake - A centennial perspective. *Earthquake Spectra*, 22(SPEC. ISS. 2):10–41, 2006. ISSN 87552930. doi: 10.1193/1.2186693.

Jacob L Vigdor. The Katrina Effect: Was There a Bright Side to the Evacuation of Greater New Orleans? 2007. URL http://www.nber.org/papers/w13022.

Albert Whitney. ON INSURANCE SETTLEMENTS INCIDENT TO THE 1906 SAN FRANCISCO FIRE, 1972.

Figures

Figure 1: The Area Burnt in the 1906 Fire in San Francisco, Hoag (1907)

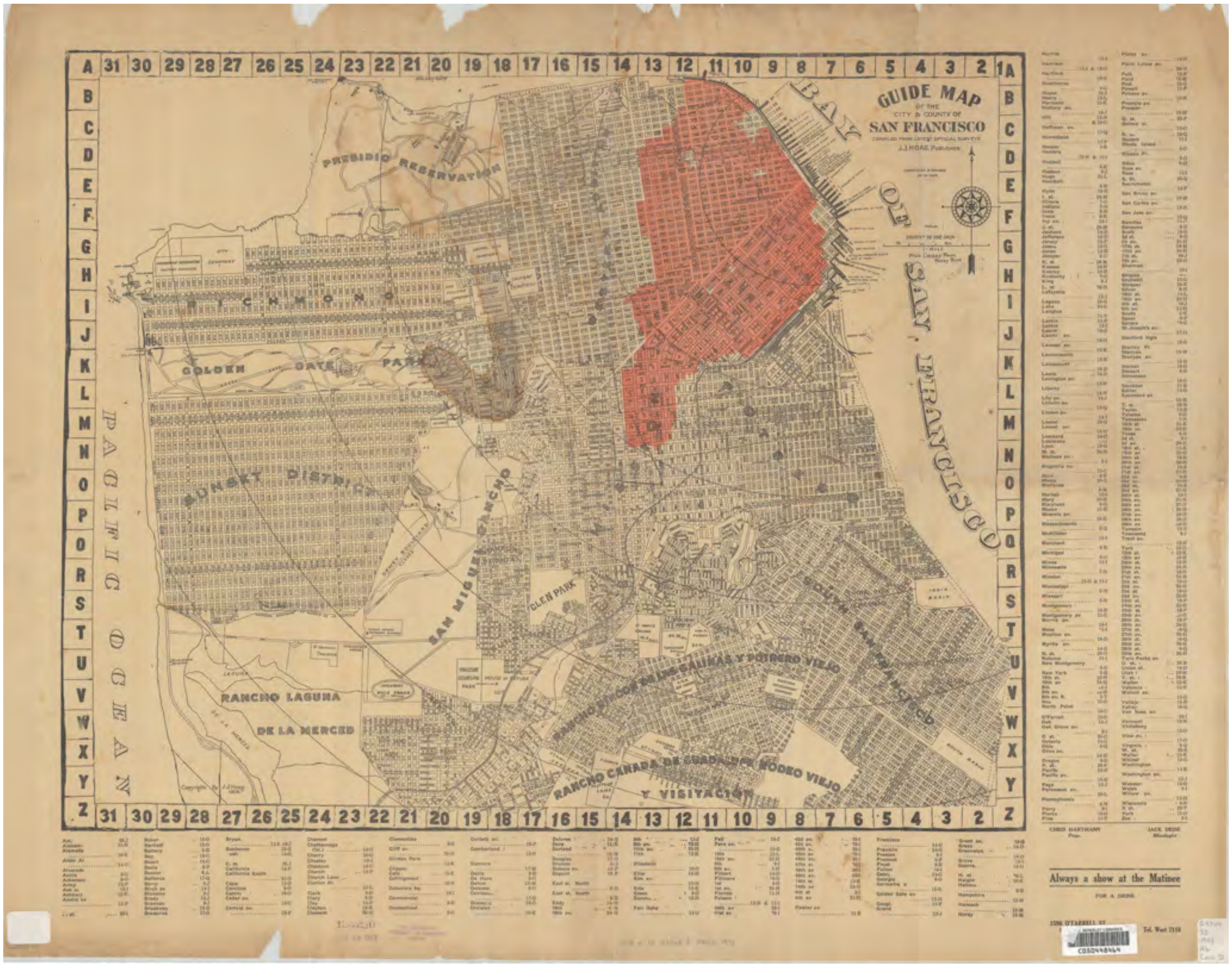
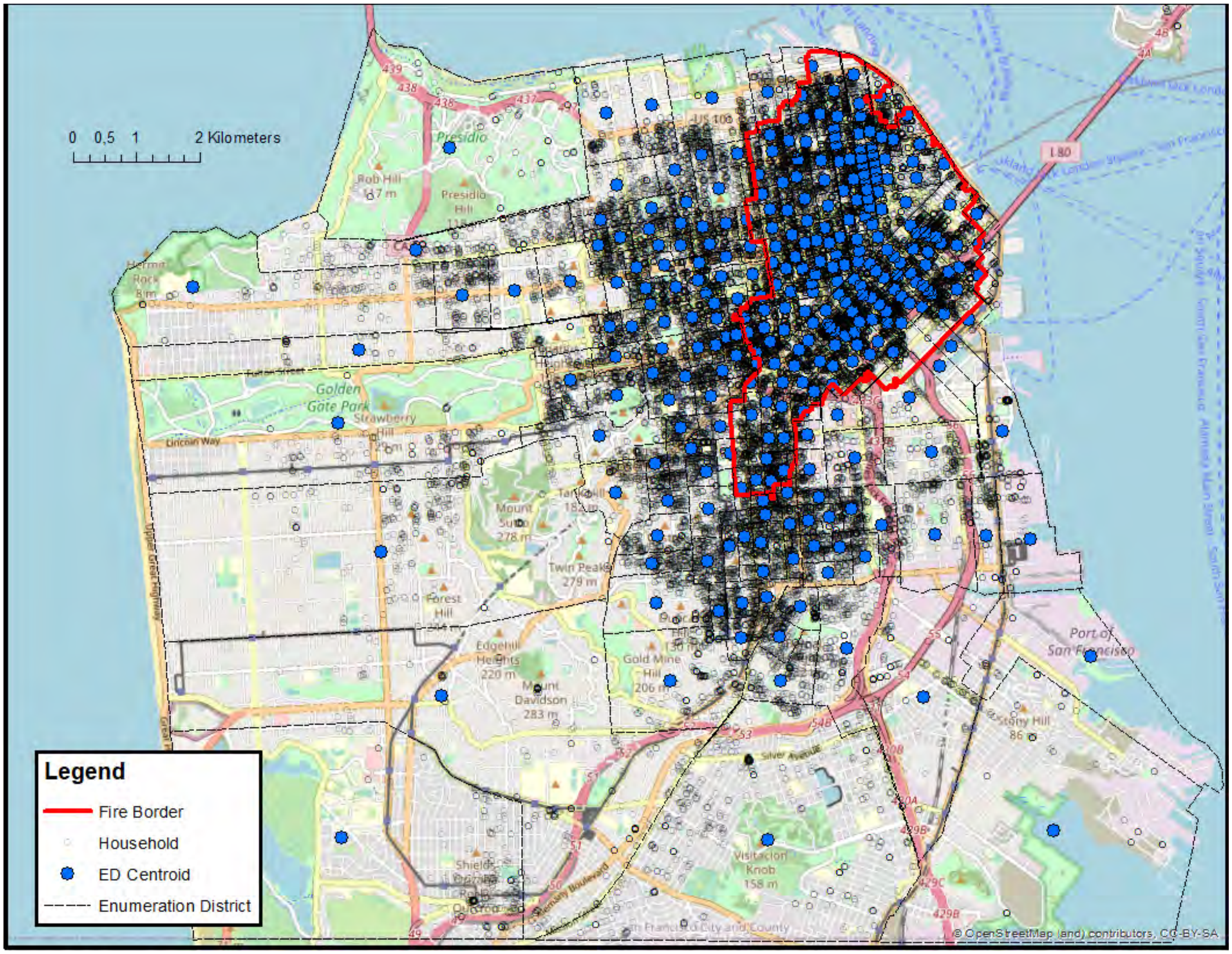
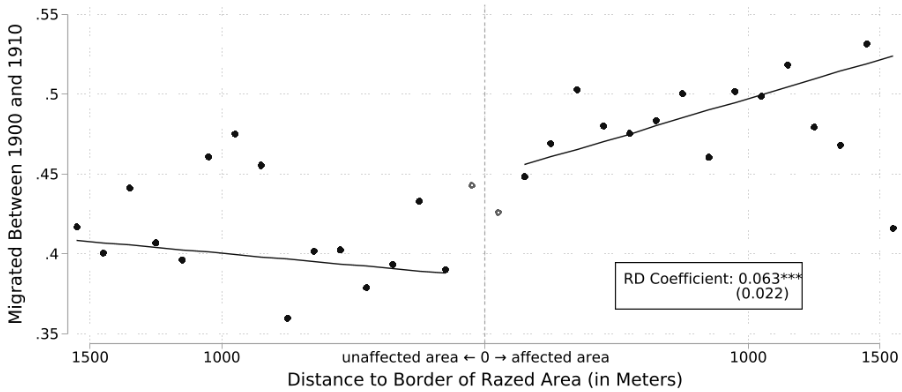


Figure 2: Location of Households in 1900, on Modern Map

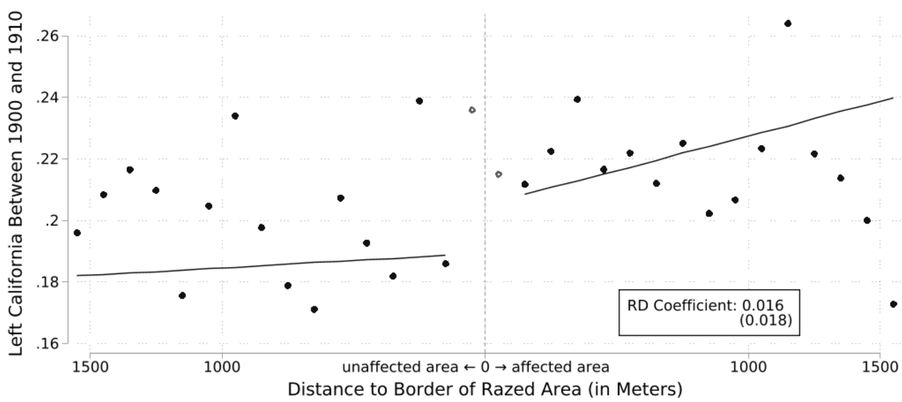


Note: Source: 1900 Census, Author's calculations

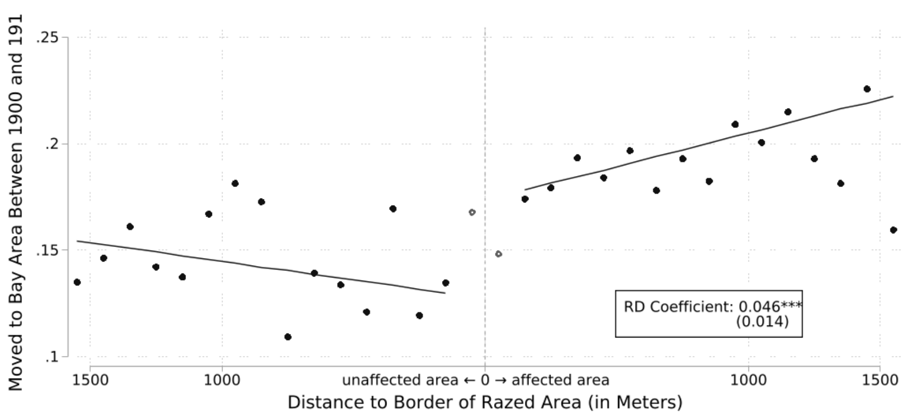
Figure 3: The Effect on Migration in 1910



(a) Dependent Variable: Living Outside of San Francisco



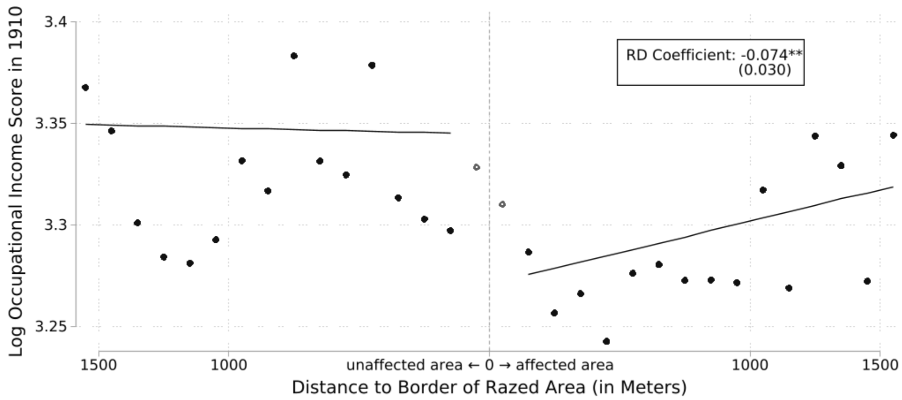
(b) Dependent Variable: Living Outside of California



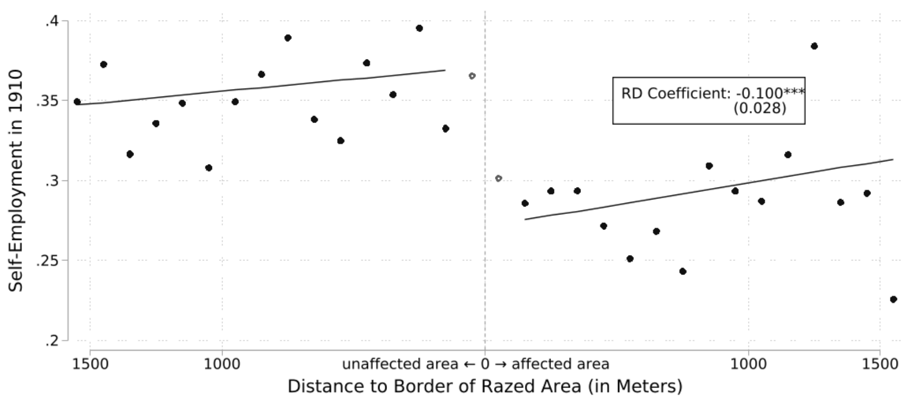
(c) Dependent Variable: Living in Bay Area

Note: Dots plot coefficients of a non-parametric regression of outcome on distance-to-border bins and the covariates in equation 3. The hollow dots are individuals in the donut hole which I exclude from the spatial RD regression. The solid lines show the fitted values given by the coefficients on the distance-to-border measure from the spatial RD specification in equation 3, controlling for race, marital status, home ownership and household size in 1900. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019).

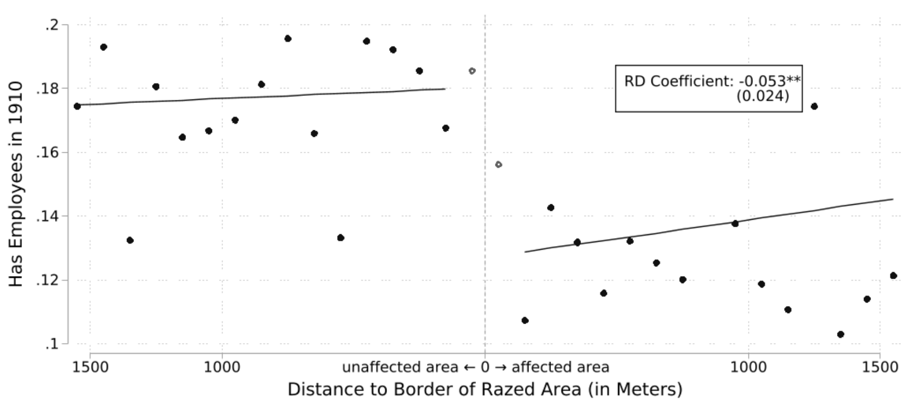
Figure 4: The Effect on Occupation and Entrepreneurship in 1910



(a) Dependent Variable: Log Occupational Income



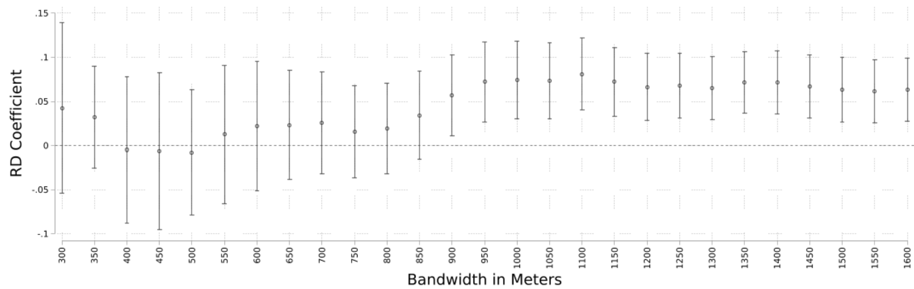
(b) Dependent Variable: Self-Employment



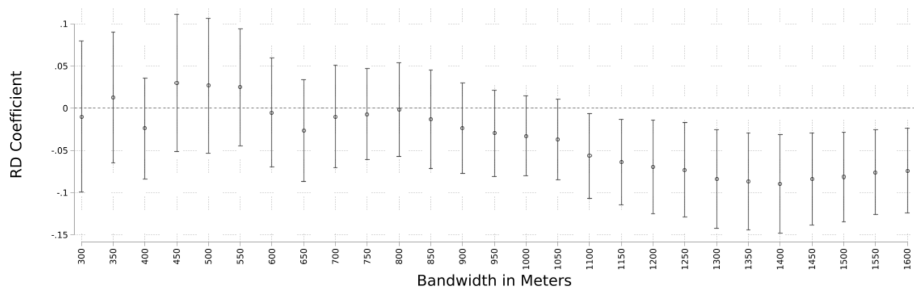
(c) Dependent Variable: Having Employees

Note: Dots plot coefficients of a non-parametric regression of outcome on distance-to-border bins and the covariates in equation 3. The hollow dots are individuals in the donut hole which I exclude from the spatial RD regression. The solid lines show the fitted values given by the coefficients on the distance-to-border measure from the spatial RD specification in equation 3, controlling for race, marital status, home ownership and household size in 1900. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019).

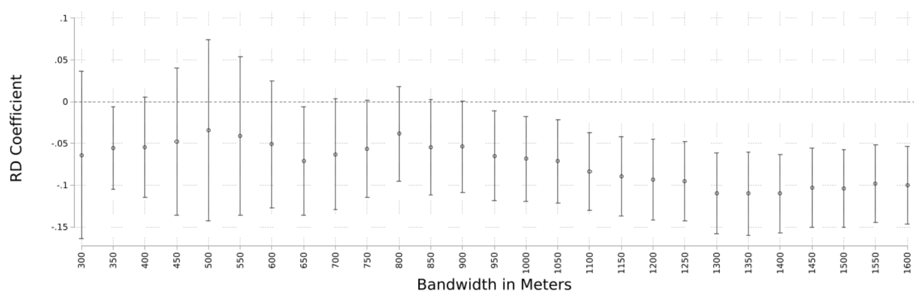
Figure 5: Bandwidth Choice and Short-Run Treatment Effects



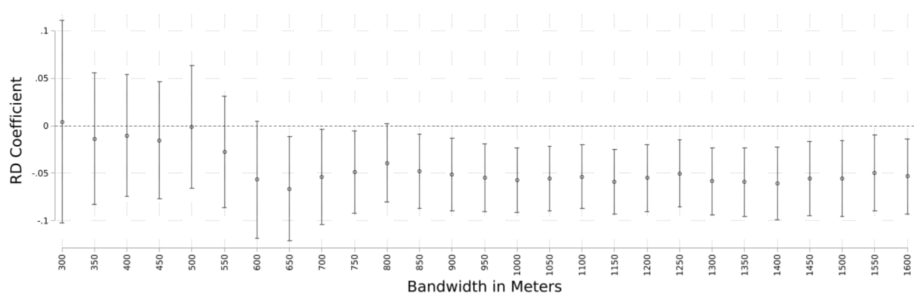
(a) Dependent Variable: Migration



(b) Dependent Variable: Log Occupational Income



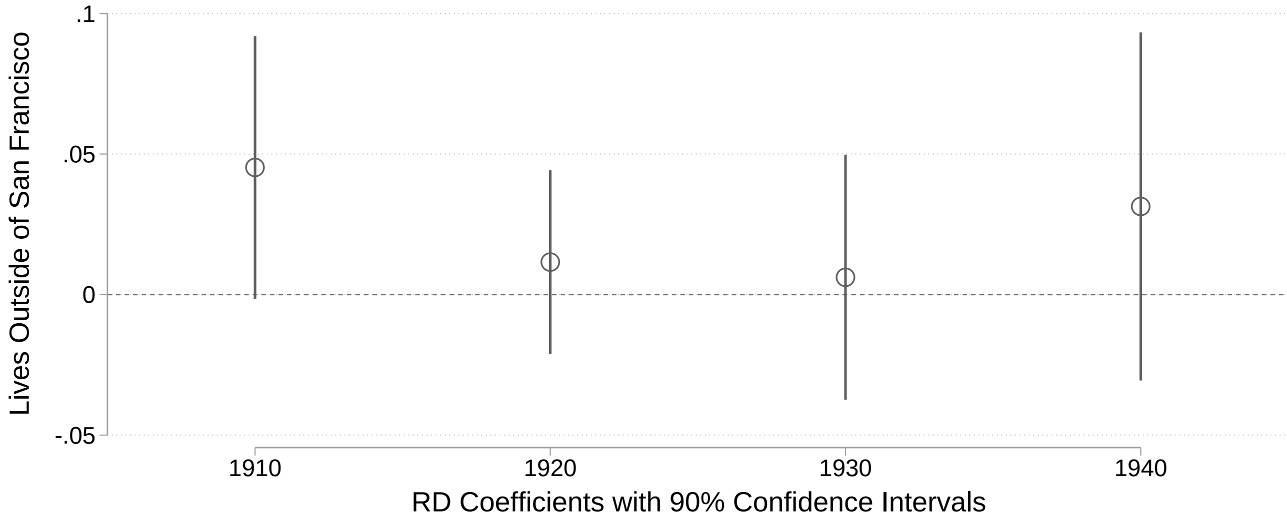
(c) Dependent Variable: Self-Employment



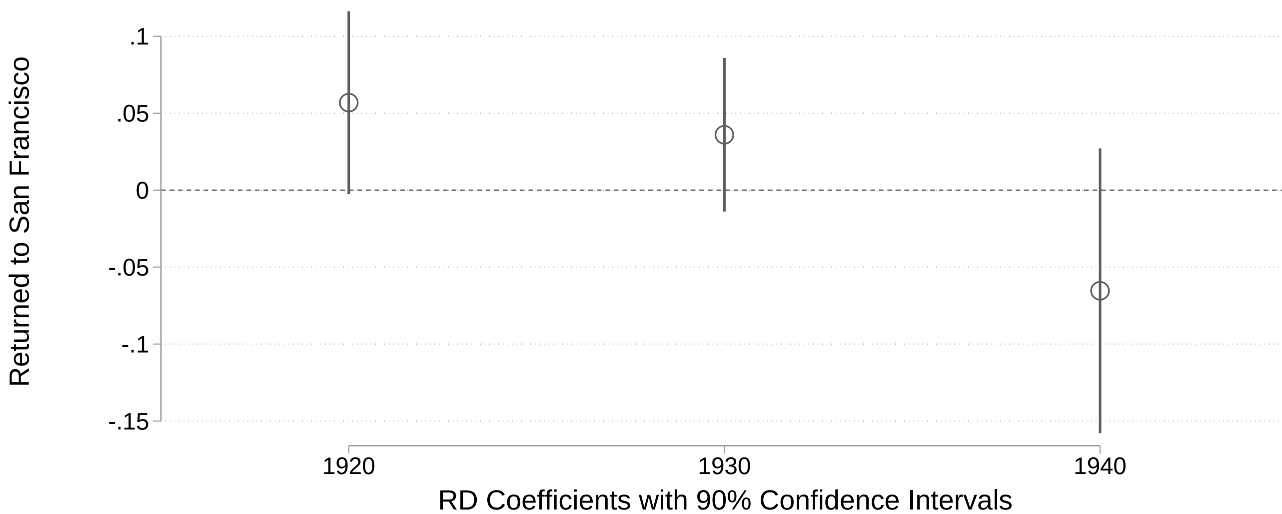
(d) Dependent Variable: Having Employees

Note: Dots plot coefficients of a non-parametric regression of outcome on distance-to-border bins and the covariates in equation 3. The hollow dots are individuals in the donut hole which I exclude from the spatial RD regression. The solid lines show the fitted values given by the coefficients on the distance-to-border measure from the spatial RD specification in equation 3, controlling for race, marital status, home ownership and household size in 1900. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019).

Figure 6: Long-Run Effect on Migration



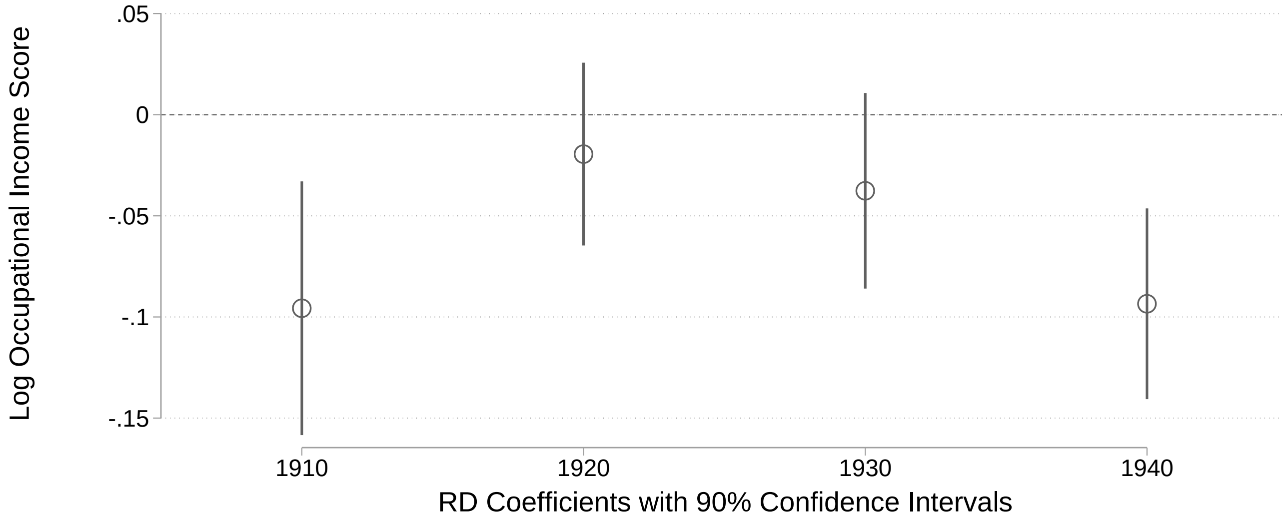
(a) Outcome: Living Outside of San Francisco



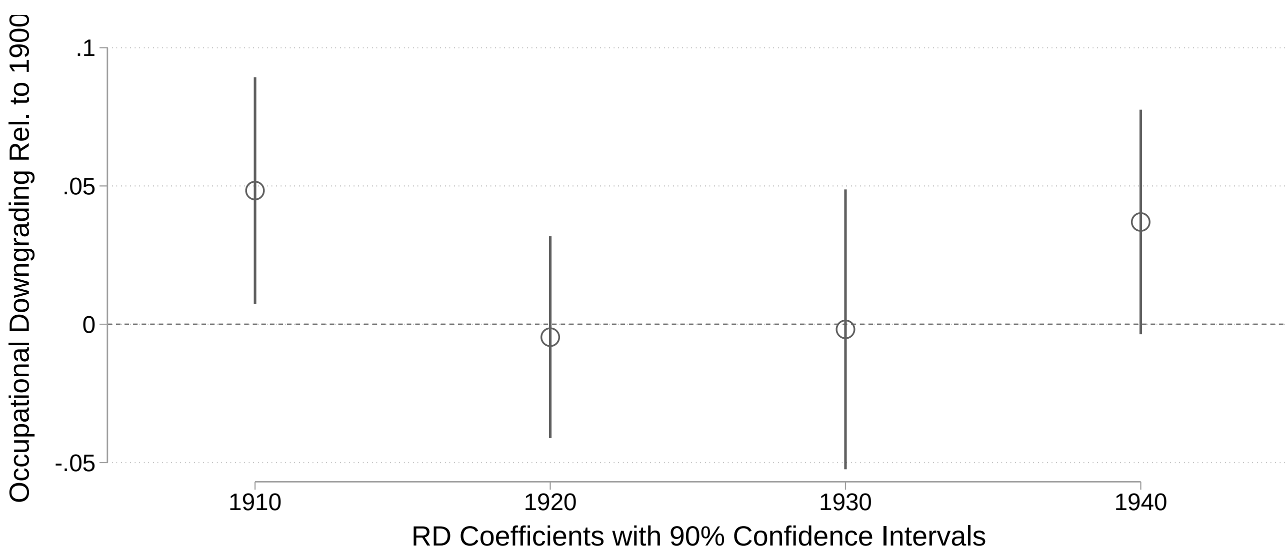
(b) Outcome: Migrant Returned to San Francisco

Note: Figure shows the regression coefficients on a treatment indicator for the spatial regression discontinuity design and the 95% confidence interval. Standard errors allow for arbitrary spatial correlation of unknown form with a distance cutoff of 300m. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border, and who were between 10 to 35 years old in 1900. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019). Dependent variable in Panel (a) indicates whether individual lives outside of San Francisco, dependent variable in Panel (b) indicates residence in San Francisco for those who lived outside of San Francisco in 1910.

Figure 7: Long-Run Effect on Occupational Income



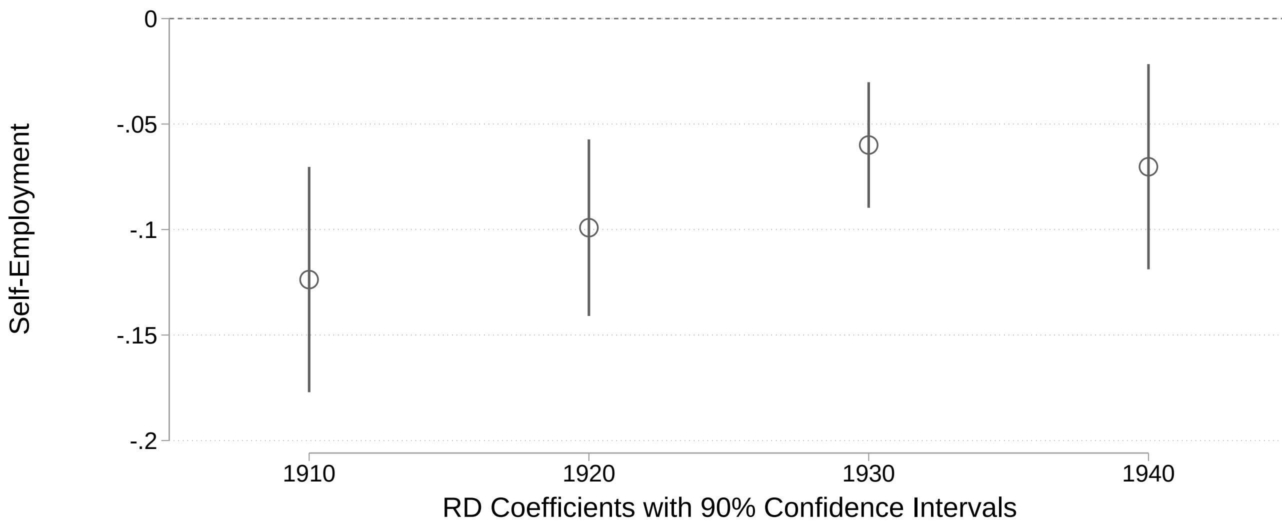
(a) Outcome: Log Occupational Income



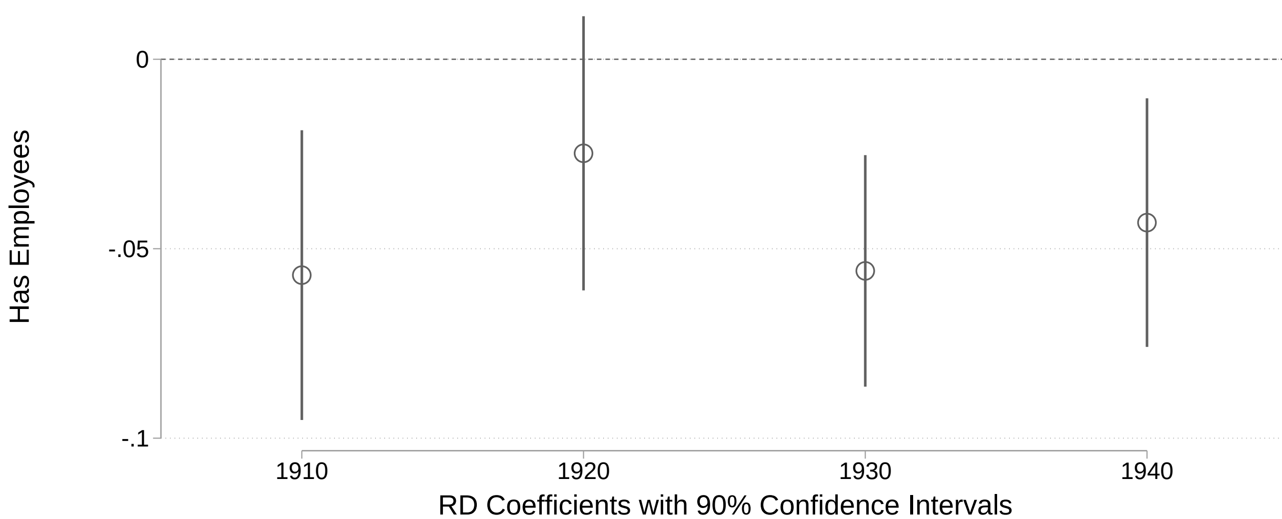
(b) Outcome: Occupational Downgrading

Note: Figure shows the regression coefficients on a treatment indicator for the spatial regression discontinuity design and the 95% confidence interval. Standard errors allow for arbitrary spatial correlation of unknown form with a distance cutoff of 300m. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border, and who were between 10 to 35 years old in 1900. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019). Dependent variable in Panel (a) is log occupational income score, dependent variable in Panel (b) indicates whether an individual has a lower occupational income relative to 1900.

Figure 8: Long-Run Effect on Entrepreneurship



(a) Outcome: Self-Employment



(b) Outcome: Has Employees

Note: Figure shows the regression coefficients on a treatment indicator for the spatial regression discontinuity design and the 95% confidence interval. Standard errors allow for arbitrary spatial correlation of unknown form with a distance cutoff of 300m. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border, and who were between 10 to 35 years old in 1900. Observations weighted by inverse propensity weights (Bailey et al., 2020 and optimal balancing weights (Heuermann et al., 2019). Dependent variable in Panel (a) indicates whether individual is self-employed, dependent variable in Panel (b) indicates whether an individual has employees.

Tables

Table 1: Differences in Pre-Treatment Characteristics (Measured in 1900)

	Observations	**Inside Razed District**	**Standard Error**
Age	25492	0.422	(0.601)
White	25492	-0.126	(0.099)
Born in the US	25492	-0.110*	(0.057)
Born in California	25492	-0.093***	(0.033)
Mother Born in the US	25492	-0.060	(0.039)
Father Born in the US	25492	-0.042	(0.037)
Mother Born in California	25492	-0.014**	(0.006)
Father Born in California	25492	-0.003	(0.004)
Married	25492	-0.078***	(0.025)
Number of Own Children	25492	-0.439***	(0.091)
Number of Own Children Below Age 5	25492	-0.065***	(0.024)
Number of Siblings in Household	25492	-0.383***	(0.091)
Household Size	25492	0.778	(0.709)
Household Head	25492	-0.125***	(0.030)
Children in Labor Force	2491	0.042***	(0.014)
Wife in Labor Force	8704	0.019	(0.019)
Kin Presence	25492	0.017	(0.034)
Middle Initial	25492	-0.075***	(0.027)
Length First Name	25492	-0.344*	(0.204)
Length Last Name	25492	-0.368	(0.283)
Home Ownership	22800	-0.242***	(0.031)
Mortgage	5224	-0.096***	(0.028)
Log Occupational Income Score	15971	-0.085***	(0.023)
Months Unemployed in 1899	20196	0.264***	(0.079)
In Labor Force	21839	0.037**	(0.017)
School Attendance	25492	-0.077***	(0.021)
Illiterate	25492	0.023*	(0.013)

Note: $^{*}\ p < 0.1$ $^{**}\ p < 0.05$, $^{***}\ p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Each row represents a separate regression of pre-treatment outcomes (measured in 1900) on treatment indicator, without additional controls. Implements inverse propensity weights (Bailey et al., 2020).

Table 2: Placebo RD with Pre-Treatment Characteristics (Sample: Linked 1900-1910)

	Observations	**Inside Razed District**	**Standard Error**
Age	24760	0.583	(0.514)
White	24760	0.011	(0.012)
Born in the US	24760	-0.032	(0.030)
Born in California	24760	-0.036	(0.029)
Mother Born in the US	24760	-0.031	(0.037)
Father Born in the US	24760	-0.003	(0.035)
Mother Born in California	24760	-0.016*	(0.009)
Father Born in California	24760	0.001	(0.006)
Married	24760	0.071***	(0.014)
Number of Own Children	24760	0.099**	(0.041)
Number of Own Children Below Age 5	24760	0.025	(0.017)
Number of Siblings in Household	24760	-0.263***	(0.096)
Household Size	24760	-2.201***	(0.725)
Household Head	24760	0.104***	(0.013)
Children in Labor Force	2419	0.021	(0.024)
Wife in Labor Force	8442	0.038	(0.025)
Kin Presence	24760	-0.071	(0.062)
Middle Initial	24760	0.022	(0.053)
Length First Name	24760	-0.089	(0.077)
Length Last Name	24760	-0.171	(0.116)
Home Ownership	22155	-0.153***	(0.046)
Mortgage	5142	0.007	(0.052)
Log Occupational Income Score	15488	-0.002	(0.033)
Months Unemployed in 1899	19578	0.071	(0.122)
In Labor Force	21203	0.028	(0.025)
School Attendance	24760	-0.019	(0.023)
Illiterate	24760	-0.004	(0.006)

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Each row represents a separate placebo spatial regression discontinuity regression. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border. Dependent variables measured in 1900. No individual level controls.

Table 3: The Short-Run Effect of the 1906 San Francisco Fire, Spatial Regression Discontinuity Design

	Observations	**Inside Razed District**	**Standard Error**
Lives Outside of SF in 1910	24760	0.063***	(0.022)
Lives Outside of CA in 1910	24760	0.016	(0.018)
Migrant Lives in Bay Area in 1910	24760	0.046***	(0.014)
Log Occupational Income in 1910	18413	-0.074**	(0.030)
Difference Log Occupational Income 1900-1910	11755	-0.044	(0.039)
Occupational Downgrading 1900-1910	16746	0.063**	(0.027)
Self-Employment in 1910	17933	-0.100***	(0.028)
Has Employees in 1910	17933	-0.053**	(0.024)
Wife's LFP in 1910	12951	0.029**	(0.012)
Child's LFP (Age 10-15) in 1910	3610	0.020	(0.019)
Child's School Attendance (Age 6-15) in 1910	7274	-0.012	(0.021)
Got Married 1906-1910	24760	-0.019**	(0.009)
Children Born 1906-1910	24760	-0.025	(0.018)

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Each row represents a separate spatial regression discontinuity regression, controlling for race, marital status, home ownership and household size in 1900. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border.

Table 4: Robustness of Main Short-Run Results (Part I)

	Lives Outside San Francisco	**Log Occupational Income**	**Self-Employment**	**Has Employees**
	(1)	(2)	(3)	(4)
	(a) Baseline Specification			
Inside Razed Area	0.063***	-0.074**	-0.100***	-0.053**
	(0.022)	(0.030)	(0.028)	(0.024)
Observations	24,760	18,413	17,933	17,933
	(b) No Pre-Treatment Controls			
Inside Razed Area	0.070***	-0.073**	-0.104***	-0.058**
	(0.021)	(0.031)	(0.028)	(0.025)
Observations	24,760	18,413	17,933	17,933
	(c) Add Controls: Occscore, Number of Children, LFP, Born US, Age			
Inside Razed Area	0.064***	-0.070**	-0.092***	-0.050**
	(0.022)	(0.029)	(0.025)	(0.024)
Observations	24,760	18,413	17,933	17,933
	(d) Add Controls: Lagged Dependent Variable			
Inside Razed Area	0.063***	-0.074**	-0.085***	-0.049**
	(0.022)	(0.030)	(0.021)	(0.021)
Observations	24,760	18,413	17,933	17,933
	(e) No Inverse Probability Weights			
Inside Razed Area	0.065***	-0.072**	-0.101***	-0.056**
	(0.022)	(0.030)	(0.027)	(0.023)
Observations	24,760	18,413	17,933	17,933
	(f) No Optimal Balancing Weights			
Inside Razed Area	0.037*	-0.059**	-0.045*	-0.034*
	(0.019)	(0.025)	(0.023)	(0.018)
Observations	18,028	13,427	13,047	13,047

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Baseline specification in Panel (a) is spatial regression discontinuity regression, controlling for race, marital status, home ownership and household size in 1900. Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border. Panel (b) omits pre-treatment controls from the regression, Panel (c) additionally includes 1900 values of occscore, number of children, LFP, born US, and age, Panel (d) controls for the lagged value of the dependent variable. Panel (e) does not use inverse propensity weights, Panel (f) does not use optimal balancing weights and excludes individuals along the Eastern part of the fire border.

Table 5: Robustness of Main Short-Run Results (Part II)

	Lives Outside San Francisco	Log Occupational Income	Self-Employment	Has Employees
	(1)	(2)	(3)	(4)
(a) Baseline Specification				
Inside Razed Area	0.063***	-0.074**	-0.100***	-0.053**
	(0.022)	(0.030)	(0.028)	(0.024)
Observations	24,760	18,413	17,933	17,933
(b) Latitude and Longitude				
Inside Razed Area	0.069***	-0.044**	-0.070***	-0.045**
	(0.018)	(0.022)	(0.021)	(0.018)
Observations	24,760	18,413	17,933	17,933
(c) Include the Donut Hole Observations				
Inside Razed Area	0.046**	-0.061***	-0.086***	-0.046**
	(0.018)	(0.022)	(0.021)	(0.019)
Observations	27,486	20,429	19,898	19,898
(d) Restrict to Whites				
Inside Razed Area	0.063***	-0.066**	-0.098***	-0.053**
	(0.021)	(0.029)	(0.027)	(0.025)
Observations	23,059	17,238	16,764	16,764
(e) Age 20 - 50 in 1910				
Inside Razed Area	0.054**	-0.078**	-0.114***	-0.062***
	(0.026)	(0.034)	(0.026)	(0.024)
Observations	19,209	14,545	14,444	14,444
(f) Alternative Linking Method				
Inside Razed Area	0.045**	-0.075***	-0.095***	-0.035*
	(0.018)	(0.026)	(0.030)	(0.021)
Observations	34,512	25,610	24,874	24,874

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Baseline specification in Panel (a) is spatial regression discontinuity regression, controlling for race, marital status, home ownership and household size in 1900. Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border. Panel (b) uses longitude and latitude as the running variables. Panel (c) includes individuals who live less than 100m away from the border. Panel (d) restricts the sample to whites, and Panel (e) restricts the sample to individuals who are between 20 and 50 years old in 1910. Panel (f) implements baseline specification on a sample linked with Price et al. (2021) crosswalks.

Table 6: Effect of Fire in 1910 by 1900 Home Ownership

	Lives Outside San Francisco	**Log Occupational Income**	**Self-Employment**	**Has Employees**
	(1)	(2)	(3)	(4)
Inside Razed Area	0.079***	-0.075***	-0.098***	-0.054**
	(0.026)	(0.027)	(0.033)	(0.027)
Home Ownership × Inside	-0.051	-0.023	0.035	-0.020
	(0.058)	(0.042)	(0.030)	(0.032)
Observations	22,155	16,472	16,056	16,056

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Each column represents a separate spatial regression discontinuity regression, controlling for race, marital status, home ownership and household size in 1900. Treatment indicator, running variable, and controls are interacted with 1900 home ownership indicator. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border.

Table 7: Effect of Fire and Kin Presence

	Lives Outside San Francisco	**Log Occupational Income**	**Self-Employment**	**Has Employees**
	(1)	(2)	(3)	(4)
Unaffected Kin	-0.021***	0.014**	0.022***	0.030***
	(0.004)	(0.006)	(0.004)	(0.004)
Affected Kin	-0.020	0.138***	0.103**	-0.001
	(0.021)	(0.041)	(0.049)	(0.025)
Unaffected Kin × Inside	-0.014	0.006	0.030	0.026
	(0.032)	(0.018)	(0.023)	(0.017)
Affected Kin × Inside	0.004	-0.138***	-0.108**	0.004
	(0.023)	(0.039)	(0.050)	(0.026)
Inside Razed Area	0.075***	-0.096***	-0.101***	-0.059**
	(0.024)	(0.033)	(0.031)	(0.024)
Observations	15,529	11,471	11,183	11,183

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to 1910, with 1900 household location within 100m to 1600m distance to the fire border. Each column represents a separate spatial regression discontinuity regression, controlling for race, marital status, home ownership and household size in 1900. Treatment indicator, running variable, and controls are interacted with R_{irs}^{out} and R_{irs}^{in} which measure presence of unaffected and affected kin, respectively. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border.

Table 8: Effect of Fire on Children's Generation in 1940

	Years of Schooling (1)	**High School Degree** (2)	**College Attendance** (3)	**Self-Employment** (4)	**Has Employees** (5)
Inside Razed Area	-0.521**	-0.039	-0.121***	-0.005	-0.007
	(0.240)	(0.034)	(0.042)	(0.024)	(0.007)
Inside × Before	0.008	-0.029	0.069*	-0.018	0.005
	(0.150)	(0.033)	(0.039)	(0.025)	(0.016)
Observations	13,119	13,119	13,119	11,684	11,684

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to sons of 1900 San Francisco residents, with 1900 household location within 100m to 1600m distance to the fire border. Spatial regression discontinuity regression controls for father's for race, marital status, home ownership, household size in 1900, and child's age in 1940. All covariates interacted with indicator for child born before 1906. Outcomes measured in 1940. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border.

A Census Variables

Table A1: Definition of Census Variables

Variable	IPUMS Variable	Definition	Available Years
Number of children	NCHILD	Number of own children in household.	1900 - 1940
Number of children under age 5	NCHLT5	Number of own children under age 5 in household.	1900 - 1940
Number of siblings	NSIBS	Number of own siblings in household.	1900 - 1940
Age	AGE	Age in years.	1900 - 1940
Marital status	MARST	Married (spouse present), married (spouse absent), separated, divorced, widowed, never married.	1900 - 1940
Duration of marital status	DURMARR	Duration of current marital status in years.	1900, 1910
Married	Based on MARST	Equals 1 if person is married (spouse present or absent), and 0 otherwise.	1900 - 1940
Household size		Number of individuals in househld.	1900 - 1940
Race	RACE	White, black, Chinese, Japanese.	1900 - 1940
White	Based on RACE	Equals 1 if person is white, and 0 otherwise.	1900 - 1940
Born in US	Based on BPL	Equals 1 if person was born in US, and 0 otherwise.	1900 - 1940
Born in California	Based on BPL	Equals 1 if person was born in California, and 0 otherwise.	1900 - 1940
Mother born in US	Based on MBPL	Equals 1 if person's mother was born in US, and 0 otherwise.	1900 - 1940
Mother born in California	Based on MBPL	Equals 1 if person's mother was born in California, and 0 otherwise.	1900 - 1940
Father born in US	Based on FBPL	Equals 1 if person's father was born in US, and 0 otherwise.	1900 - 1940
Father born in California	Based on FBPL	Equals 1 if person's father was born in California, and 0 otherwise.	1900 - 1940

Continued on next page

Table A1 – continued from previous page

Variable	IPUMS Variable	Definition	Available Years
School attendance	SCHOOL	Respondent attended school during a specified period.	1900 - 1940
Literacy	Based on LIT	Equals 1 if person know to read and write, and zero otherwise.	1900 - 1930
Years of education	Based on variables EDUC and HIGHGRADE	Years of completed education.	1940
Wife's labor force participation	Based on LABFORCE	Indicates whether a person's coresiding wife participated in the labor force.	1900 - 1940
Self employed	Based on CLASSWKR	Equals 1 if person was recorded as an "employer" or "working on own account", and zero otherwise.	1910 - 1940
Employer	Based on CLASSWKR	Equals 1 if person was recorded as an "employer", and zero otherwise.	1910 - 1940
Months unemployed	MOUNEMP	Number of months unemployed, out of the last 12 months.	1900
Occupational income score	OCCSCORE	See section 3.1.	1900 - 1940
Home ownership	OWNERSHP	Indicates whether home was rented or owned.	1900 - 1940
Mortgage	MORTGAGE	Indicates whether an owner-occupied housing unit was owned free and clear or was encumbered by a mortgage, loan, or other type of debt.	1900 - 1920
Middle initial		Indicates whether individual has a middle name or middle initial.	1900 - 1940
Length first name		Length of first name.	1900 - 1940
Length middle name		Length of middle name.	1900 - 1940
Length surname		Length of surname.	1900 - 1940

B Enumeration Districts

A Census enumeration district is a geographical unit that was intended to be covered by one census taker during the census period. Since these districts covered approximately the same number of households, they varied in size from just a few city blocks in urban areas up to a whole county in rural areas. Aside from the street name and street number, the enumeration district is the smallest geographical level that is reported in the Census.

There are 319 enumeration districts in San Francisco for the 1900 Census. Of these, 11 consist entirely of hospitals, prisons, asylums or military bases, which I exclude from the analysis. The remaining 308 districts have about 1,100 inhabitants on average.

I leverage the information on enumeration districts to localize each 1900 San Francisco resident within the city and with respect to the razed area. While I use the street name and street number reported by each household to infer the exact location of the dwelling, I check the validity of this localization using the enumeration district. I impute residence location coordinates with enumeration district centroid coordinates for 24 percent of all 1900 households.

Census enumeration district maps for 1900 are not available in a GIS-usable format, so I employ images of the paper maps retrieved from the genealogy platform FamilySearch. The maps themselves show 1898 election districts; the 1900 Census enumeration district borders and numbers are later handwritten additions. Figure A3 shows a typical example. The maps are relatively well preserved although in some cases damage to the paper obscures parts of the map.

I first georeference the photographs using ArcMap's georeferencing tool. Then I create a polygon layer (i.e. a machine-readable district boundary map) by hand which includes all 308 enumeration districts of San Francisco. From this newly created boundary map I derive each district's size, centroid coordinates, location relative to the razed area (inside, outside, intersect) and the centroid's shortest distance to the fire area boundary.

C Geolocation of 1900 Household Addresses

There are 87,130 unique households who report living in San Francisco county in the 1900 Census. Three percent (2,606) of the households do not report a street name and the street number is missing for 522 additional households. The original street name strings have been cleaned by IPUMS and are therefore mostly complete and legible.

In a first step, I geolocate all available addresses (street name and street number if available) using the geocoding tool in ArcMap. The software returns coordinates that correspond to the correct enumeration district for 59.3 percent of all households. For the remaining households, I update the street names that have changed between 1900 and today. In many cases, only parts of the streets have been renamed. This first step will therefore identify the location of households residing in the non-renamed parts of renamed steets. In the second step, I repeat the geolocation on the updated street names. An additional 1,090 addresses can now be geolocated in the correct enumeration district. This brings the total share of household addresses can be geolocated directly to 62.8 percent.

Third, I impute coordinates of neighbors' addresses if they reside in the same enumeration district as the household with the missing coordinates, and I recover the neighbors' coordinates such that they correspond to the correct enumeration district in steps 1 or 2. I identify neighbors by the household's serial number (Census variable SERIAL). Serial numbers are assigned by the order in which the Census takers enumerated households (Nelson, 2020). Therefore, an adjacent serial number usually corresponds to a household residing in the same building or the building next door. In some cases Census takers might have "jumped" within the enumeration district depending on street layout. I assume that Census takers tried to minimized their walking distance and that jumps rare. Inspection of the addresses in the order of serial numbers confirms that Census takers tended to walk from house to house without skipping buildings.

For the imputation I proceed via a simple algorithm:

1. Assign the coordinates of one the immediately neighbors (starting with the neighbor identified by serial number + 1, then serial number - 1) if there is an exact match on the street name. *Imputed addresses: 4,455*.

2. Impute coordinates for all missing households with an exact match for enumeration district and address (street name *n*umber) using coordinates obtained from step 1. *Imputed addresses: 1,223*.

3. Assign coordinates of one of the immediately adjacent neighbors regardless of street name match. Assign coordinates of one of the two-doors-away neighbors (serial number +/- 2)

if there is a match for the stemmed street name[37]. *Imputed addresses: 5,512*.

The final step assigns the remaining 23.9 percent of households (20,103) the coordinates of the geometric centroid of their enumeration district. The average enumeration district is 0.63 square kilometers (about 156 acres), with enumeration districts further away from the center (and the razed area) being considerably larger. Among all enumeration districts within 1km of the fire border, the average size is only 0.14 square kilometers (34 acres). Imputing enumeration district centroid coordinates is therefore a decent approximation. I exclude households in enumeration districts that are intersected by the fire border from this imputation. This pertains to 5,001 households, so that the final sample of households includes coordinates for 94 percent of the observations (79,048). Figure 2 shows all households locations, enumeration district borders and geographic centroids, and the border of the razed district.

D Weights

D.1 Inverse Propensity Weights

Following Bailey et al. (2020), I construct inverse propensity weights (IPW) in three simple steps. First, I estimate a probit model for the dependent variable that indicates whether in individual was linked from 1900 to the year of interest, for example 1910. The covariates of this model are

- Indicator for middle initial, length of first, middle, last names,
- quadratic in age and month of birth,
- presence of siblings, number of siblings.

Second, I predict the conditional probability of being linked, p_i, from the results of the probit. Third, I construct the weight as

$$IPW_i = \frac{1 - p_i}{p_i} \tag{6}$$

[37] I stem street names by retaining the first word (e.g. "Octavia" in "Octavia Street" or "Octavia St".)

D.2 Optimal Balancing Weights

Following Heuermann et al. (2019), I reweight observations around each border segment according to

$$OBW_b = \frac{n_{b,inside} \times n_{b,outside}}{n_{b,inside} + n_{b,outside}} \tag{7}$$

where $n_{b,inside}$ and $n_{b,outside}$ are the number of observations in each border segment b, inside and outside the razed area, respectively.

E Historical Maps

Figure A1: The Course of the 1906 Fire in San Francisco

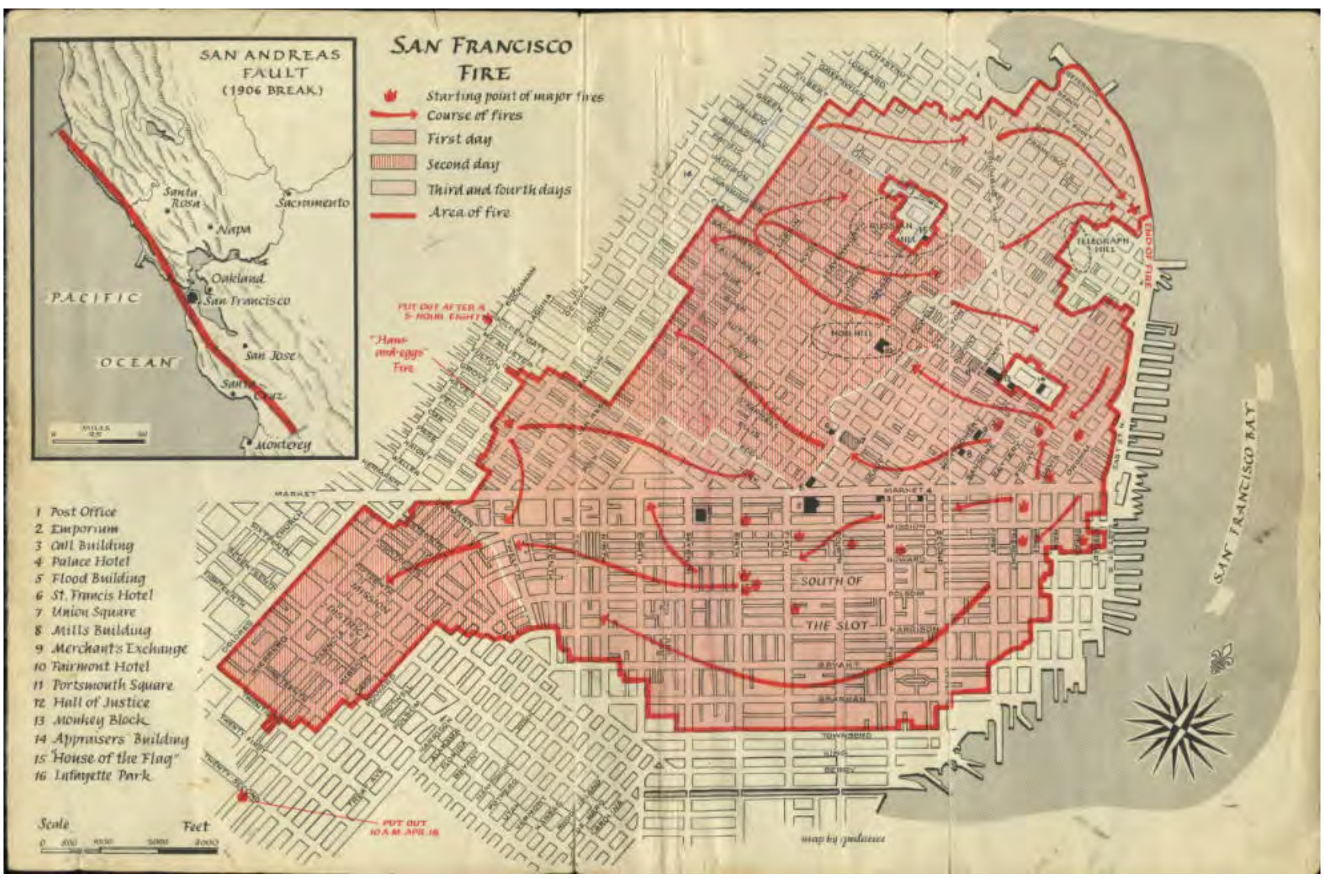


Figure A2: San Francisco Reconstruction Map, Punnet Brothers (1908)

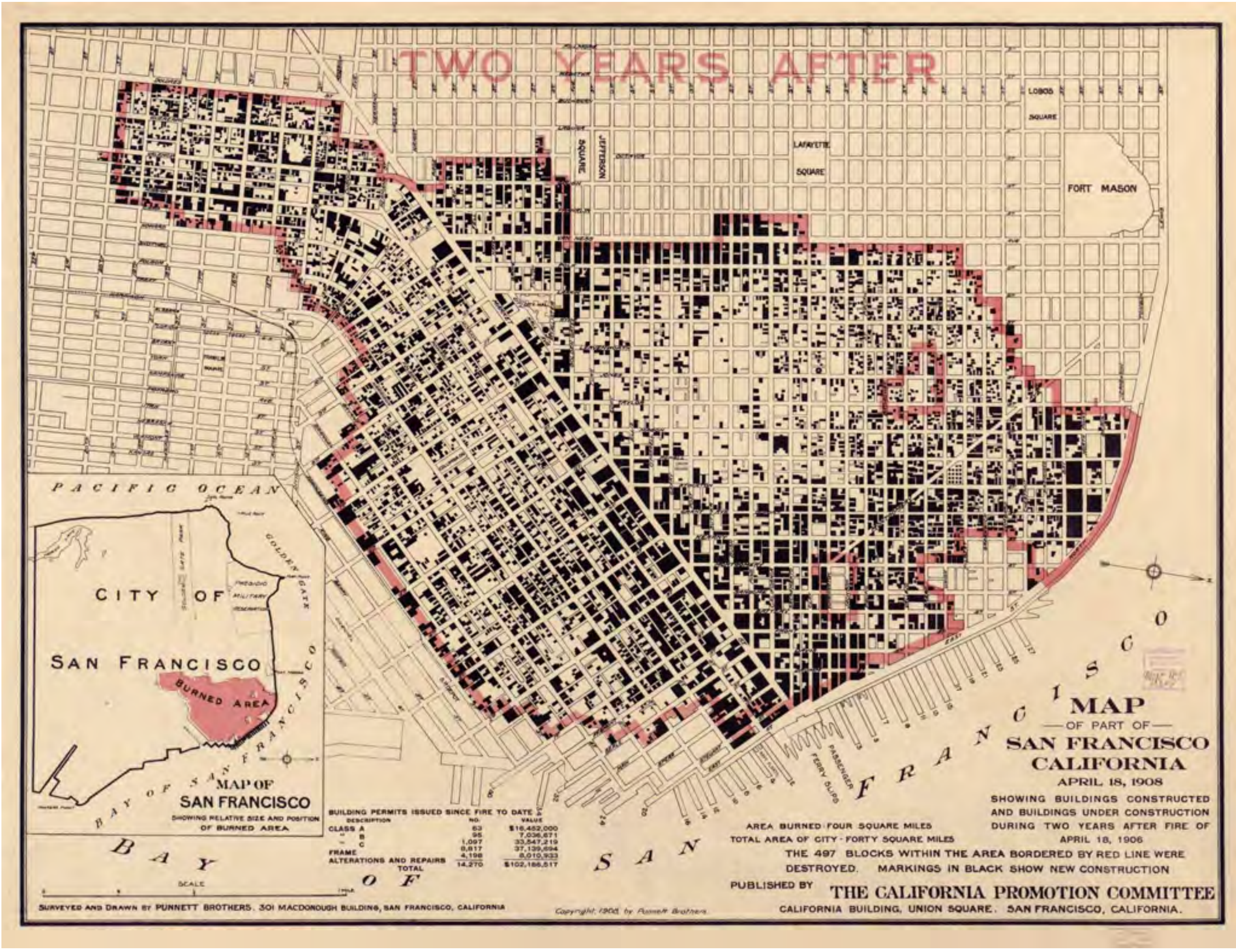
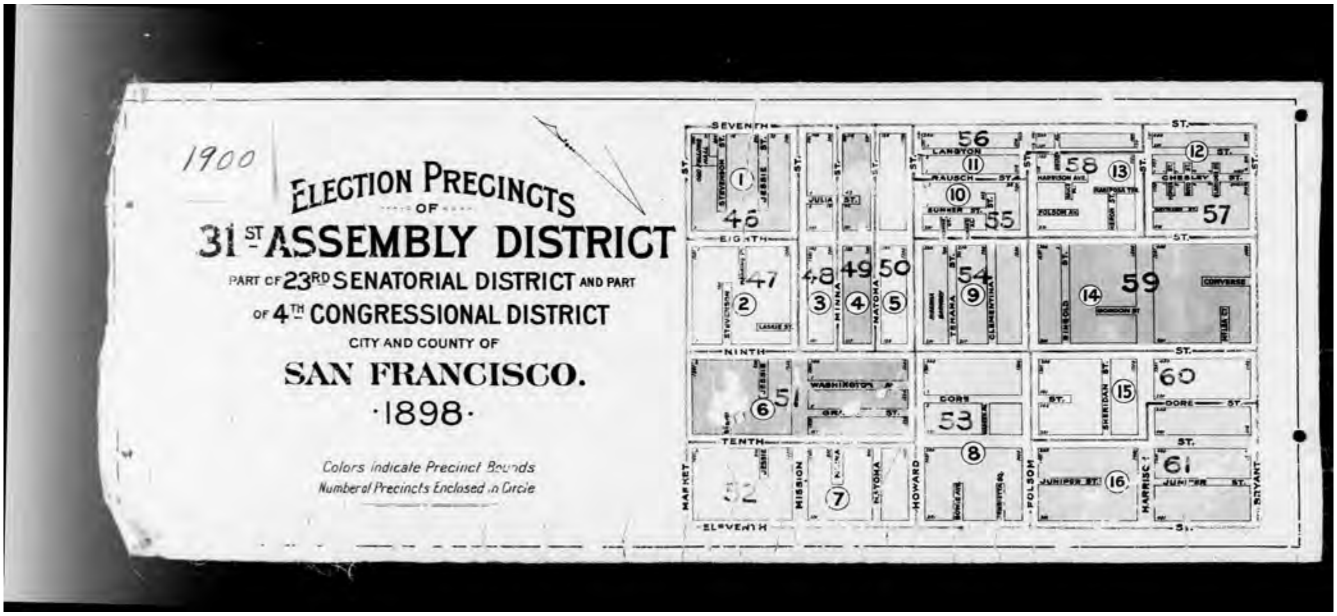


Figure A3: Photo of a Map Showing 1900 San Francisco Enumeration Districts 46 - 61



F Additional Results

Table A2: Placebo RD with Pre-Treatment Characteristics (Sample: Linked at Least Once)

	Observations	**Inside Razed District**	**Standard Error**
Age	37176	-0.005	(0.453)
White	37176	0.011	(0.010)
Born in the US	37176	-0.036	(0.031)
Born in California	37176	-0.027	(0.031)
Mother Born in the US	37176	-0.031	(0.035)
Father Born in the US	37176	0.006	(0.032)
Mother Born in California	37176	-0.012	(0.009)
Father Born in California	37176	0.002	(0.006)
Married	37176	0.043^{***}	(0.011)
Number of Own Children	37176	0.057	(0.045)
Number of Own Children Below Age 5	37176	0.022^{*}	(0.012)
Number of Siblings in Household	37176	-0.167^{*}	(0.087)
Household Size	37176	-1.264^{**}	(0.605)
Household Head	37176	0.076^{***}	(0.016)
Children in Labor Force	3224	0.033	(0.023)
Wife in Labor Force	11927	0.022	(0.024)
Kin Presence	37176	-0.090^{*}	(0.049)
Middle Initial	37176	0.025	(0.051)
Length First Name	37176	-0.079	(0.082)
Length Last Name	37176	-0.195^{*}	(0.106)
Home Ownership	32917	-0.144^{***}	(0.045)
Mortgage	7328	-0.014	(0.049)
Log Occupational Income Score	23396	-0.002	(0.025)
Months Unemployed in 1899	29524	0.088	(0.121)
In Labor Force	31867	0.018	(0.025)
School Attendance	37176	-0.008	(0.019)
Illiterate	37176	-0.001	(0.005)

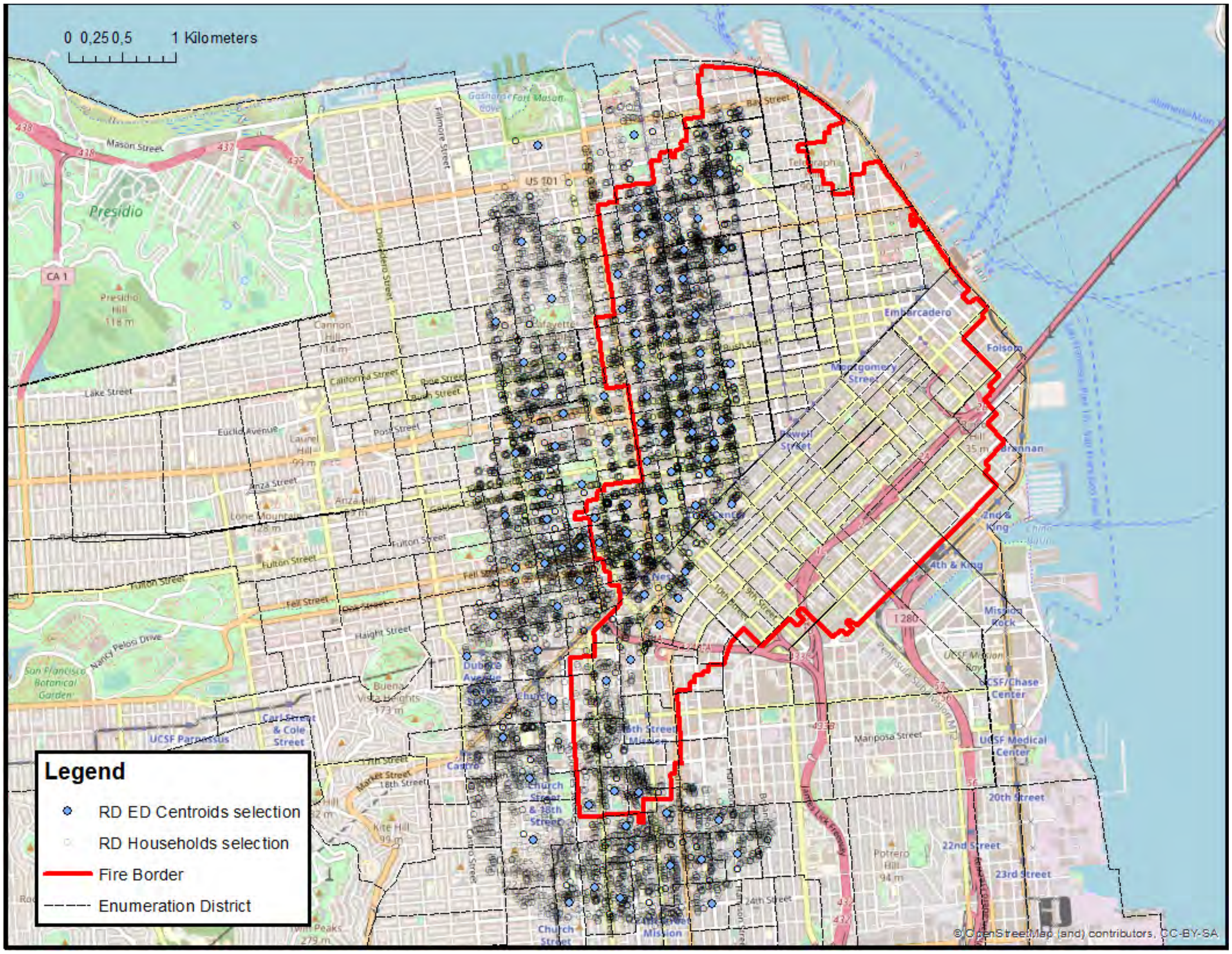
Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to individuals linked from 1900 to at least one of the years 1910, 1920, 1930, or 1940, with 1900 household location within 100m to 1600m distance to the fire border. Each row represents a separate placebo spatial regression discontinuity regression. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border. Dependent variables measured in 1900. No individual level controls.

Table A3: Effect of Fire on Children's Generation in 1940, Additional Results

	Migrated (SF)	**Migrated (CA)**	**Urban**	**Log Occscore**	**Log Wages**
	(1)	(2)	(3)	(4)	(5)
Inside Razed Area	0.021	-0.003	0.013	-0.042	-0.027
	(0.016)	(0.026)	(0.011)	(0.030)	(0.075)
Inside × Before	0.008	-0.004	-0.012	0.022	-0.010
	(0.008)	(0.040)	(0.017)	(0.035)	(0.071)
Observations	13,282	13,282	13,282	11,722	9,669

Note: * $p < 0.1$ ** $p < 0.05$, *** $p < 0.01$. Standard errors allow for arbitrary spatial clustering with a distance cutoff of 0.4 km (in parentheses). Sample restricted to sons of 1900 San Francisco residents, with 1900 household location within 100m to 1600m distance to the fire border. Spatial regression discontinuity regression controls for father's for race, marital status, home ownership, household size in 1900, and child's age in 1940. All covariates interacted with indicator for child born before 1906. Outcomes measured in 1940, wages only include non-selfemployed individuals. Observations weighted by inverse propensity weights (Bailey et al., 2020) and optimal balancing weights (Heuermann et al., 2019), running variable is distance to the border.

Figure A4: Location of Households in 1900 (Spatial RD Sample Manually Selected), on Modern Map



Note: Source: 1900 Census, Author's calculations