Explorations in Economic History 77 (2020) 101342

Contents lists available at ScienceDirect

Explorations in Economic History

journal homepage: www.elsevier.com/locate/eeh

Shorter Article

How the 1906 San Francisco earthquake shaped economic activity in the American West[☆]

Philipp Ager[a,*], Katherine Eriksson[b], Casper Worm Hansen[c], Lars Lønstrup[d]

[a] *University of Southern Denmark and CEPR, Denmark*
[b] *UC Davis, Stellenbosch University, CEPR and NBER, United States*
[c] *Casper Worm Hansen, University of Copenhagen, Denmark*
[d] *University of Southern Denmark, Denmark*

ARTICLE INFO

JEL classification:
N91
O15
O40
R11
R12

Keywords:
Economic geography
Location of economic activity
Migration
Natural disasters

ABSTRACT

This paper examines the long-run effects of the 1906 San Francisco Earthquake on the spatial distribution of economic activity in the American West. Using variation in the potential damage intensity of the earthquake, we show that more severely affected cities experienced lower population increases relative to less affected cities until the late 20th century. The earthquake left a long-lasting mark mainly because it interrupted existing migrant networks. Less affected areas became more attractive migrant destinations in the immediate aftermath of the earthquake, which permanently changed relative city sizes in the American West.

1. Introduction

The location of economic activity is highly skewed. In 2020, the ten largest U.S. metropolitan areas accounted for roughly one-third of America's GDP. As economic activity is clustered within a few core metropolitan areas, a central question in economics is what determines the distribution of city sizes. Economists typically think of locational fundamentals and agglomeration economies as key determinants for the locations and relative sizes of cities. While theories based on locational fundamentals would predict that even large temporary shocks would leave the distribution of city sizes unchanged (e.g., Davis and Weinstein, 2002, agglomeration economies can give rise to multiple equilibria implying path dependence in city sizes (e.g., Bleakley and Lin, 2012; Bleakley and Lin, 2015; Krugman, 1991).

One possible mechanism of path dependence is the existence of migrant networks. Historically, migrants' decision to settle in a specific location depended primarily on established networks of relatives and friends (i.e., chain migration). Such social networks

---

[☆] We would like to thank the editor (Marianne H. Wanamaker) and two anonymous referees for helpful and constructive comments. We also thank Alberto Alesina, Sascha Becker, Leah Boustan, Carl-Johan Dalgaard, Francesco Drago, James Fenske, Walker Hanlon, Marc Klemp, Christopher Meissner, Stelios Michalopoulos, Nathan Nunn, Paul Sharp, Daniel Sturm, Holger Strulik, David Weil, Nico Voigtlaender, and seminar participants at the University of Copenhagen, the Fresh Meeting at the University of Southern Denmark, the Spanish Economic Association Meeting in Girona 2015, the 1st CREA workshop in Luxembourg, and at the University of Goettingen. The research of Philipp Ager and Casper Worm Hansen is generously funded bythe Danish Research Council reference No. DFF – 4182-00043.

[*] Corresponding author.

E-mail addresses: phag@sam.sdu.dk (P. Ager), kaeriksson@ucdavis.edu (K. Eriksson), casper.worm.hansen@econ.ku.dk (C.W. Hansen), loe@sam.sdu.dk (L. Lønstrup).

https://doi.org/10.1016/j.eeh.2020.101342
Received 18 October 2019; Received in revised form 11 June 2020; Accepted 11 June 2020

reduced the costs of moving for newcomers by providing valuable information about job opportunities, housing, laws and customs in the destination country (e.g., Ferrie and Hatton, 2015; Moretti, 1999; Wegge, 1998). The value of locating near other people with the same cultural background can explain persistence in relative city sizes, even if the factors that had determined the location choice of pioneer settlers became obsolete.

In this paper, we investigate if a major natural disaster is able to break up the path dependence of chain migration by interrupting existing migrant networks. In particular, we examine how the 1906 San Francisco Earthquake, a large but temporary shock, affected the long-run distribution of economic activity in the American West. The earthquake happened during a period of mass migration in the United States and at a time when the American West was still at the beginning of its development process. Large-scale settlement of the American West had begun in the second half of the 19th century, when job opportunities in mining and improvements in transportation transformed the region into a popular destination for foreigners and migrants from other parts of the United States (e.g., Clay and Wright, 2012; Hedges, 1926; White, 2011). One might expect that the destructive intensity of the 1906 San Francisco Earthquake was large enough to disrupt the social networks leading to chain migration.

Our approach to answer this question starts by evaluating whether the 1906 San Francisco earthquake affected the distribution of city sizes in the American West by using newly-collected data on the population size of cities and towns in California, Oregon, and Nevada for every decade from 1870 to 1970. Our estimation strategy compares city sizes before and after the earthquake between more and less affected cities. Estimates from this differences-in-differences strategy reveal that more affected cities experienced a substantial decline in population size relative to less affected cities after the earthquake occurred. This conclusion is robust to controlling for possible time-varying effects of a wide range of local geographical characteristics, such as proximity to San Francisco and Los Angeles. Importantly, the relative negative long-run effect of the earthquake also remains in place when allowing for a different adjustment process of affected cities located inside the Bay Area, or when considering cities with different access to markets and credit. We also show that the persistent effect of the earthquake is independent of city size. Our baseline estimate implies that a one-standard deviation increase in earthquake intensity is associated with a decrease in relative city size of around 36 percent in 1970, which corresponds to a relative decline of 0.6 percentage points in annualized population growth.

Next, we study the settlement patterns of migrants arriving in the American West around the time of the earthquake to evaluate the “chain-migration” hypothesis. This analysis is based on a repeated cross-section of newly-digitized complete-count U.S. Census samples. We show that the earthquake had significant larger effects on net-migration rates in areas with larger preexisting migrant networks, suggesting that that earthquake disrupted the path-dependence of chain migration, which helps to explain the persistence of the shock. Finally, we create a linked sample of adult males from 1900 to 1910, which allows us to construct in- and out-migration rates for affected and non-affected areas. The analysis based on the linked sample reveals that our results are driven by negative effects on in-migration rates.

This finding relates to empirical studies on the economic consequences of natural disasters in the United States during the first half of the 20th century, documenting that people moved away from the affected areas as a response to the shock (Boustan et al., 2012; Ferrie and Naidu, 2014; Hornbeck, 2012; Long and Siu, 2018).[1] Recent work by Boustan et al. (2017) evaluates how natural disasters affected migration rates and other economic outcomes at the county level between 1920 and 2010. The authors show that counties hit by severe disasters experienced falling land prices, greater out-migration rates, and higher poverty rates. Our migration analysis, on the other hand, focuses on whether in-migration rates responded to a natural disaster that happened in an relatively attractive area at that time. Our findings suggest that the 1906 San Francisco Earthquake left a long-lasting mark, mainly because it interrupted the path-dependence of chain migration, which diverted migrants to less affected areas of the American West. We also contribute to recent studies that investigate the effect of fires on city development. Ferrie and Keniston (2017) focus on the Great Boston Fire of 1872 and (Siodla, 1906; 2017) on the 1906 San Francisco Fire. Both studies find beneficial long-run effects on city development from these shocks. While these papers investigate the development of a single city after reconstruction, the present paper evaluates how a shock to the placement of people affected the relative size of cities and towns in a whole region over a relatively long time period. In this respect, our finding speaks also to a broader literature that examines the effect of natural disasters on economic growth which finds, in general, mixed results (e.g., Cavallo et al., 2013; Imaizumi et al., 2016; Loayza et al., 2012; Strobl, 2011).

Our paper also advances the literature that highlights the importance of historical events for understanding the present-day locations of economic activity. Bleakley and Lin (2012) and Henderson et al. (2017) show that fundamentals that were historically important at the time when cities were founded leave a permanent mark on the spatial distribution of economic activity due to scale economies. Kline and Moretti (2014), Jedwab and Moradi (2016), Hanlon (2017), and Michaels and Rauch (2018) all obtain similar conclusions by finding persistent effects of historical events on the geographic distribution of economic activity. Schumann (2014) even shows that a different resettlement policy for German World War II expellees in the American and French zone can have long-lasting consequences for population levels. In contrast, Davis and Weinstein (2002), Brakman et al. (2004) and Miguel and Roland (2011) find no long-run effects of war-related events on relative city sizes.

2. Historical background

The California Gold Rush in 1848 unleashed a massive wave of migration to the American West. Over the next decades, an expanding mining sector, the construction of a transcontinental railroad, coupled with the distribution of public land to settlers via

---

[1] Ferrie and Naidu (2014) find for a sample of rural counties in the U.S. South that the Great Mississippi Flood of 1927 did not effect total population, possibly due to the in-migration of white labor, yet the flood triggered black out-migration and agricultural modernization.

the Homestead Act of 1862, attracted more migrants from the rest of the United States and abroad. The expansion of the railroad network towards the end of the 19th century further promoted the westward expansion by substantially reducing transportation costs of goods and people (e.g., Donaldson and Hornbeck, 2016; North, 1966; Walker, 2001; White, 2011). In 1900, for example, every second person in California, Nevada and Oregon was born out of state, and more than 20 percent were born abroad. During the 20th century, the American West experienced a rapid structural transformation (Malone and Etulain, 1989; Olmstead and Rhode, 2017). Overall, the population of the three states in our sample grew from around 2 million in 1900 to 22 million in 1970. During the same period, the share of the labor force working in agriculture decreased from around 28 percent to less than four percent and GDP per capita increased by around two percent per year (Turner et al., 2013).

Earthquakes occurred in the region before 1906, but were accepted as a nuisance in daily life. According to the U.S. Geological Survey, it was the 1906 earthquake that marked the onset of a scientific revolution in earthquake research, meaning that the timing and the location of the earthquake were unanticipated (Zoback, 2006). This fact is expressed by Andrew Lawson, at that time a professor in Geology at the University of California, Berkeley, who wrote in the university newspaper in 1904 *"History and records show that earthquakes in this locality have never been of a violent nature, as so far as I can judge from the nature of recent disturbances and from accounts of past occurrences there is not occasion for alarm at present"* (cited in Zoback, 2006, p. 4).

The earthquake struck on April 18, 1906 at 5:12 a.m. local time without warning. The total length of the rupture was 477 km (296 miles), and the main epicenter was located about 3 km off the shore of San Francisco. While the estimated moment magnitude places the 1906 San Francisco Earthquake just in the top 20 of the largest earthquakes in the history of the United States, the death toll, economic cost, and the associated fires classify it as one of the worst natural disasters on American soil. A contemporary U.S. Army relief report recorded 498 deaths in San Francisco, 64 deaths in Santa Rosa, and 102 deaths in and around San Jose (Greely, 1906). Revised numbers estimate an overall death toll of circa 3,000 individuals (Hansen and Condon, 1989). Besides the casualties, more than 225,000 people became homeless and 28,000 buildings were destroyed with an estimated economic damage of 10.5 billion in current U.S. dollars (Algermissen, 1972).

In comparison, there were less than 200 earthquake casualties in California from 1812 to 1901 and the property damages of the next significant earthquake after 1906, the Long Beach Earthquake of 1933, was around 750 million in current U.S. dollars (Coffman et al., 1982; Toppozada and Branum, 2004). From the 1930s until today, around 200 people have been killed by earthquakes in California (Toppozada and Branum, 2004). Together, these facts paint a picture of the 1906 earthquake as an unparalleled natural disaster in American history.[2]

Overall, the 1906 earthquake is naturally associated with the city of San Francisco, since it was almost entirely destroyed and was the worst affected area. However, as also pointed out by (Ellsworth, 1990), the intensity of the earthquake was comparable to that felt in the city of San Francisco in many other places close to the San Andreas Fault line. ((Coffman et al., 1982), p.150) wrote that the region of destructive intensity of the 1906 earthquake extended to 400 miles around the epicenter and was felt in most of California and parts of Oregon and Nevada; the three states included in our sample.

3. Data and estimation strategy

3.1. Data

Our measure of earthquake intensity is based on Boatwright and Bundock (2005) ShakeMap. The ShakeMap is a smooth measure of the so-called Modified Mercalli Intensity (MMI) of the 1906 earthquake and is deduced from damage reports compiled by Lawson and Reid (1908) and augmented with intensities inferred from additional historical sources. This collection amounts to more than 600 sites with information on the *potential* damage intensity of the earthquake. Boatwright and Bundock deploy the ShakeMap methodology to produce a "heat map" from these sites, where the potential damage intensity ranges from none to very heavy (Wald et al., 1999).[3]

Every city is assigned an MMI value by overlaying the geographic coordinates (longitude and latitude) for every sampled city in California, Oregon, and Nevada with the ShakeMap using the software QGIS. Cities in the sample that are not depicted on the ShakeMap are assigned the value zero and, therefore, regarded as non-affected control cities. Fig. 1 displays the location of the cities in the sample together with the potential damage intensity of the earthquake. In the empirical analysis, we standardize the potential damage intensity to be mean zero and standard deviation of one.

For the main outcome of interest, population size, we compile a newly digitized data set containing the population histories of cities and towns in the American West. Our source is (Moffat, 1996), who collected data on sizes of cities and towns for every decade between 1850 and 1990. The U.S. Census population figures by minor civil divisions are the primary source for this collection. These records contain population figures of all incorporated cities and towns. Moffat supplements these records in some instances with information provided by state and territorial censuses.

For unincorporated communities, Moffat draws on the population figures for unincorporated census designated places with over 1,000 inhabitants (published by the U.S. Census Bureau since 1950) and the Rand McNally Commercial Atlas and Marketing Guide, which was published annually since 1876.[4] It should be noted that these population figures are estimates. In the results section, we

---

[2] The U.S. Geological Survey provides a full list of the earthquakes in the region for the period 1800–1970; see, (Coffman et al., 1982).

[3] Using the *potential* and not the actual earthquake damage should mitigate concerns that variation in the earthquake damage variable is endogenous to the size of cities.

[4] Rand McNally obtained population estimates for unincorporated places by surveying municipal officials, postmasters, chamber of commerce and planning officials.

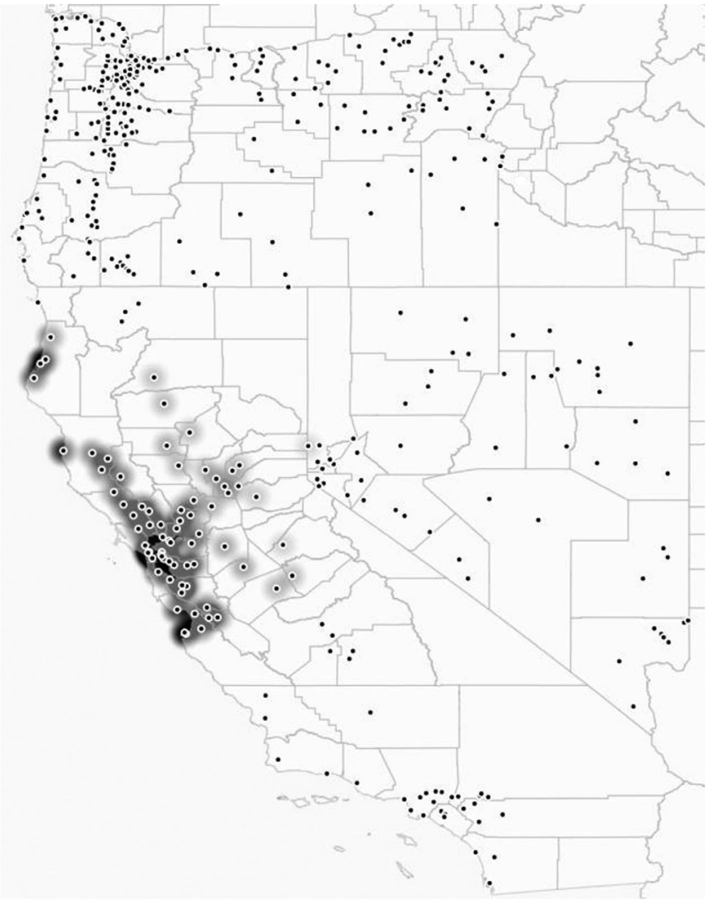


Fig. 1. Earthquake Intensity and Sampled Cities. *Notes:* This map shows the cities in the American West included in the empirical analysis. Darker shaded areas indicate a higher earthquake intensity.

also report outcomes by city size (above and below 2,500 inhabitants in 1900) and use aggregated county-level instead of city-level population data to show that our results are not only driven by small cities where population estimates are less reliable.

Our database contains the population size of cities covering the states of California, Oregon, and Nevada for every decade from 1870 to 1970. We include a city in the sample if it is at least observed in 1900 and 1910 (i.e., the immediate decades before and after the earthquake occurred) implying that the maximum number of cites included in the analysis is 413 (similar results are obtained for a balanced sample of cities). We further use county-level population and manufacturing data from (Haines, 2010) to measure the effects on labor productivity. We geocoded the cities based on their names and obtained various geographical controls from (Fishback et al., 2011). Distance to San Francisco, Los Angeles, or to the epicenter (N37.75, W122.55) are calculated using the "geodist" command in STATA.

The migration analysis is based on newly digitized complete-count U.S. Census samples for the decades from 1900 to 1940. The data are retrieved from IPUMS (Ruggles et al., 2017) and consist of a repeated cross-section of individuals that reported their birthplace (state or country of birth) and place of residence (available at the county level). We use this information to calculate net-migration rates for people born outside of California, Nevada, and Oregon and to measure the size of migration networks by birthplace and county.

The complete-count U.S. Census data further contain information about individuals' age, and their first and last names. With this information at hand, we can construct a linked sample of adult males (aged 18–55), applying the fully automated matching approach pioneered by Ferrie (1996) and further developed by Abramitzky et al. (2012, 2014). We use a conservative matching strategy of this linking technique requiring individuals to be unique by name and state of birth within a five-year age band. This conservative approach reduces the likelihood that individuals are matched to the wrong person with similar attributes by around 50 percent (Abramitzky et al., 2019), which is crucial when studying migration decisions.

We match all men aged 18–55 in 1900 to the Census of 1910, and then restrict to individuals living in one of our three states in 1900 or in 1910. The initial matched sample results in a match rate of about 15 percent, and a sample of over 137,000 men after imposing the state of residence restriction. We then aggregate the data at the county level to construct in- and out-migration

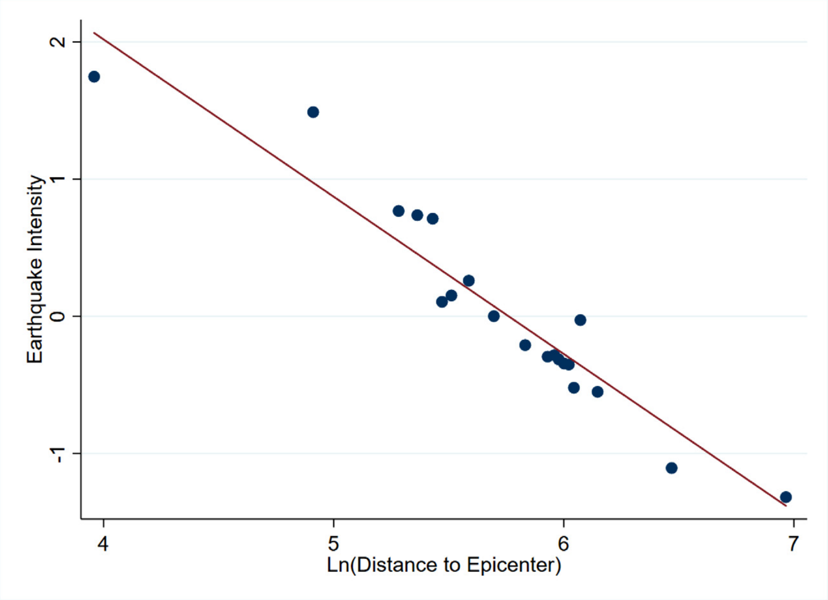


Fig. 2. First-stage Relationship between Earthquake Intensity and Distance to Epicenter. *Notes:* This figure (binscatter plot) displays the first-stage relationship between earthquake intensity and the natural logarithm of the distance to the epicenter controlling for longitude and latitude in logarithmic units.

rates between 1900 and 1910, the decade in which the earthquake occurred. Summary statistics of the main variables are shown in Appendix Table 1.[5]

3.2. Empirical strategy

Our econometric model follows a differences-in-differences approach with a continuous measure of treatment using city-level data for the decades 1870–1970. We use event-study regressions to evaluate how relative city sizes vary by standardized earthquake intensity across locations in California, Oregon, and Nevada before and after the disaster occurred. The baseline estimation equation takes the following form:

$$y_{ct} = \sum_{j=1870}^{1970} \beta_j Quake_c \times I_t^j + \mathbf{X}'_{ct}\Gamma + \lambda_c + \lambda_t + \varepsilon_{ct}, \tag{1}$$

where y_{ct} is *ln* population size in city *c* in period *t; Quake*$_c$ is the standardized potential damage intensity in city *c* interacted with a full set of time-period fixed effects, $\sum I_t^j$, where 1900 is the omitted year of comparison; $\mathbf{X}'_{ct}$ includes controls for latitude and longitude measured in logarithmic units interacted by time-period fixed effects, λ_c denotes city fixed-effects, λ_t time-period fixed effect, and ε_{ct} is the error term. The standard errors are Huber robust and clustered at the city level.

The parameters of interest are the $\beta'_j s$, which can be interpreted as the effect of the standardized earthquake intensity on the change in *ln* population size from 1900 to period *j*. For example, negative values of $\hat{\beta}_j$ for $j > 1900$ indicate that more affected cities experienced smaller population increases relative to less affected cities after the earthquake occurred. The main identifying assumption is that cities with different potential damage intensities would have evolved similar to non-affected cities had the earthquake not occurred. In the next section, we provide evidence that changes in relative city sizes in the decades before the earthquake are not systematically correlated with damage intensity, suggesting that pre-earthquake trends in observables are indeed parallel, indirectly supporting the identifying assumption.[6]

One concern is that the potential damage intensity is likely to be measured with error, which would bias the parameters of interest, $\beta'_j s$, towards zero. We address this issue by instrumenting the potential damage intensity with *ln* distance to the epicenter. It is plausible to assume that measurement error in the potential damage intensity is unrelated to measurement error in the distance to the epicenter. Fig. 2 presents a binscatter plot of the relationship between standardized potential damage intensity and *ln* distance to the epicenter controlling for latitude and longitude. The relationship is negative and highly statistically significant and shows no indication of instrumental weakness (the t-statistic is −28.86).

---

[5] The replication file for this study can be found at https://doi.org/10.3886/E119841V1 (Ager et al., 2020).

[6] It is important to acknowledge that the empirical setup only allows us to study if the earthquake influenced *relative* city sizes in the three states, since the Stable Unit Treatment Value Assumption (SUTVA) is obviously not satisfied. A similar approach is taken in Hanlon (2017).

4. Results

4.1. Effects on city population and manufacturing

Fig. 3 presents the event-study estimates for *ln* city population as the outcome variable. All specifications (Panels A-F) include fixed effects for city and time and control for *ln* latitude and *ln* longitude interacted by time-period fixed effects. Panel A presents the baseline results of estimating Eq. (1) using least squares as the estimation method. This specification shows that more affected cities experienced smaller population increases relative to less affected cities after the 1906 earthquake. While it appears that the gap in population size widens until the 1940s, the earthquake-effect on annual population growth is not increasing over time: For example, the estimate for 1910 implies that a unit increase in standardized damage intensity is associated with a decrease in annual population growth of $0.085/10 \approx 0.0085$ from 1900 to 1910, while the corresponding 1920 event-estimate suggests a decrease of $0.135/20 \approx 0.0065$ from 1900 to 1920. Overall, the estimates imply that the effect of the earthquake on annual population growth was decreasing over the sample period. The event-study displayed in Panel A also reveals that there was no statistically significant relationship between earthquake intensity and changes in relative city sizes before the disaster took place.

Panel B displays our preferred specification. The potential damage intensity is now instrumented by *ln* distance to the epicenter. While the two-stage least squares coefficients have the same sign, they turn out to be larger (in absolute terms) for any post-earthquake decade, which possibly reflects classical measurement error in the earthquake intensity variable. The estimated magnitude of β_{1940} implies that a one standard-deviation increase in earthquake intensity is associated with a decrease in relative city size by around 28 percent relative to 1900, translating into a relative decrease in annual population growth of 0.8 percentage points. Calculated over the entire period from 1900 to 1970, the annualized effect on population growth is 0.6 percentage points. It is important to note that this quantification is relative to less affected cities in the sample, meaning that we can only conclude that the earthquake changed the distribution of city sizes in the American West and not the aggregate number of people living there.

The pre-earthquake event estimates are small in magnitude and not individually nor jointly statistically significant different from zero, indirectly supporting the main identifying assumption. It is worth pointing out that our results are not sensitive to using Conley standard errors to account for spatial correlation instead of clustering standard errors at the city level (Appendix Fig. 1 Panels A–F present the results for different spatial cutoffs).

Panels C-F present modified specifications of the two-stage least squares approach used in Panel B. The event-study reported in Panel C drops any city located within 50–260 miles from San Francisco and, in Panel D, we drop cities with low and intermediate damage intensities. Reassuringly, dropping these cities do not affect our results. While including latitude and longitude in the baseline specification (interacted with time-period fixed effects) is one attempt to ensure that cities are similar in terms of natural characteristics and to address the issue of future correlated shocks, we now take additional steps. In particular, the specification in Panel E takes into account that cities in the proximity of San Francisco and Los Angeles might have evolved differently over time irrespective of the potential earthquake damage (e.g., the port of Los Angeles opened its gates in December 1907 and could have attracted population independently of the earthquake). Even when controlling for the *ln* distance to San Francisco and Los Angeles (both interacted by time-period fixed effects), the relative decline in population size of more affected cities after the earthquake occurred remains in place, albeit the precision of the estimated coefficients is somewhat lower. In light of the close proximity of the epicenter to San Francisco (the raw correlation is −0.82), it is unsurprising that these coefficients are less precisely estimated.[7] The estimated coefficients reveal a similar pattern as shown in Panel B when taking potential mean reverting effects into account. In particular, Panel F displays this result when *ln* city population 1870, 1880, 1890, and 1900 (all interacted by time-period fixed effects) are added to estimating Eq. (1).[8]

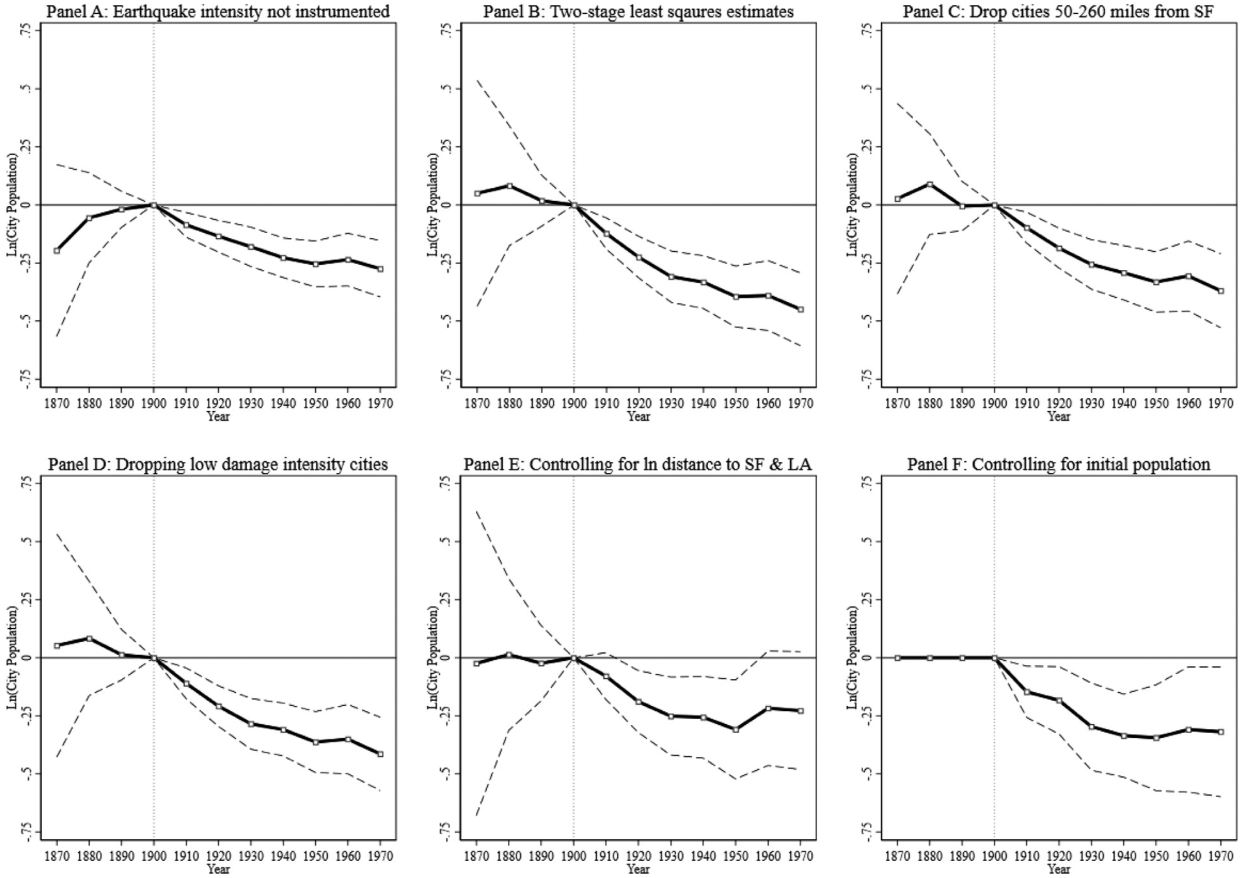
Appendix Fig. 2 presents further robustness checks. We include additional geographic controls (access to water, climatic conditions, altitude, and cotton suitability) each interacted by time-period fixed effects; we control for the diffusion of new knowledge about higher earthquake risk (i.e., a dummy if a city is located in a county on the San Andreas Fault line interacted by time-period fixed effects); or we account for changes in the trading costs between any city in our sample and San Francisco. Importantly, none of these additional control variables affect our main conclusion. The final specification, reported in Panel F, includes all the control variables mentioned so far. The general conclusion from the robustness analysis is that the estimated effects remain stable when we control for additional factors that possibly influenced city sizes during the sample period. Finally, we use county-level data from Haines (2010) to cross-check our city level data. Reassuringly, column (1) of Appendix Table 3 shows similar effects on county population size, indicating that the relative change in the distribution of people occurred in more affected areas as a whole.

Next, we investigate how the 1906 earthquake influenced the manufacturing sector. The estimation equation is similar to Eq. (1), but the unit of analysis is the county level. The sample spans all counties in the states of California, Oregon and Nevada for the decades 1880 to 1970 (no manufacturing data are available in 1910). All specifications include fixed effects for county and time and control for *ln* latitude and *ln* longitude interacted by time-period fixed effects. The estimates reported in Panel A of Fig. 4 show that the earthquake was associated with a decrease in relative productivity as measured by *ln* wages per manufacturing worker. The 1920 event estimate suggests that a standard deviation increase in damage intensity is associated with a relative decrease in wages per worker of around 6.7 percent, which is equal to an annualized growth rate effect of 0.4 percentage points. The corresponding estimates for later years are of similar magnitude and the effect on annual productivity growth is thus decreasing over the sample

---

[7] Results are very similar to the estimates reported in Panel B when only controlling for the *ln* distance to Los Angeles instead (not reported).

[8] Appendix Table 2 reports the corresponding estimates with standard errors for all event-study graphs displayed in Fig. 3.


Panel A: Earthquake intensity not instrumented

Panel B: Two-stage least sqaures estimates

Panel C: Drop cities 50-260 miles from SF

Panel D: Dropping low damage intensity cities

Panel E: Controlling for ln distance to SF & LA

Panel F: Controlling for initial population

Fig. 3. Event-study of the effect of earthquake intensity on city population. *Notes:* This figure shows the dynamic effects of earthquake intensity on city population. The solid line depicts the effect on city population relative to the base year (1900). The different panels show modifications of the baseline estimates (Panel A). The method of estimation is least squares in Panel A and two-stage least squares in Panels B-F. Dashed lines indicate 95% confidence intervals. Standard errors are clustered at the city level.

period. Panel B of Fig. 4 shows that manufacturing activity, as proxied by thy number of manufacturing establishment per capita, also decreased more in affected areas relative to less (or non) affected areas. Both set of results are indicative of small production scale effects in the manufacturing sector.[9]

4.2. Heterogeneous effects by city characteristics

Fig. 5 reports treatment heterogeneity effects by different city characteristics.[10] First, we study if cities located in the Bay Area were affected differently by the disaster, since the Bay Area is one of the most densely populated areas in the US today.[11] The estimates, reported in Panel A, show that we still find a negative and long-lasting impact of the earthquake on affected cities in the Bay Area, albeit the magnitude of the effect is quantitatively much smaller than for other affected places in the American West.

Next, we investigate if the effect of the shock varies by initial city size. While our sample contains many small cities, there is a sufficient number of cities with more than 2,500 inhabitants in 1900 to estimate the effects separately for cities above this initial population threshold (the number of cities with more (less) than 2,500 inhabitants is 53 (360)). Panel B shows the impact of the earthquake is independent of initial city size, also indicating that the measurement of population size in smaller cities is not an issue in our setting.[12]

---

[9] Appendix Fig. 3 displays the same results but using Conley standard errors instead.

[10] Fig. 5 only reports estimates from 1890 onward. The 1870 and 1880 coefficients are estimated with great imprecision and including them distorts the readability of the figure. The estimates are instead reported in Appendix Table 4.

[11] Cities located in the counties of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma are coded as “Bay Area cities”.

[12] Related research shows that a temporary positive population shock (the resettlement of German refugees after WWII) led to a persistent population increase in small and medium sized German villages (Schumann, 2014).

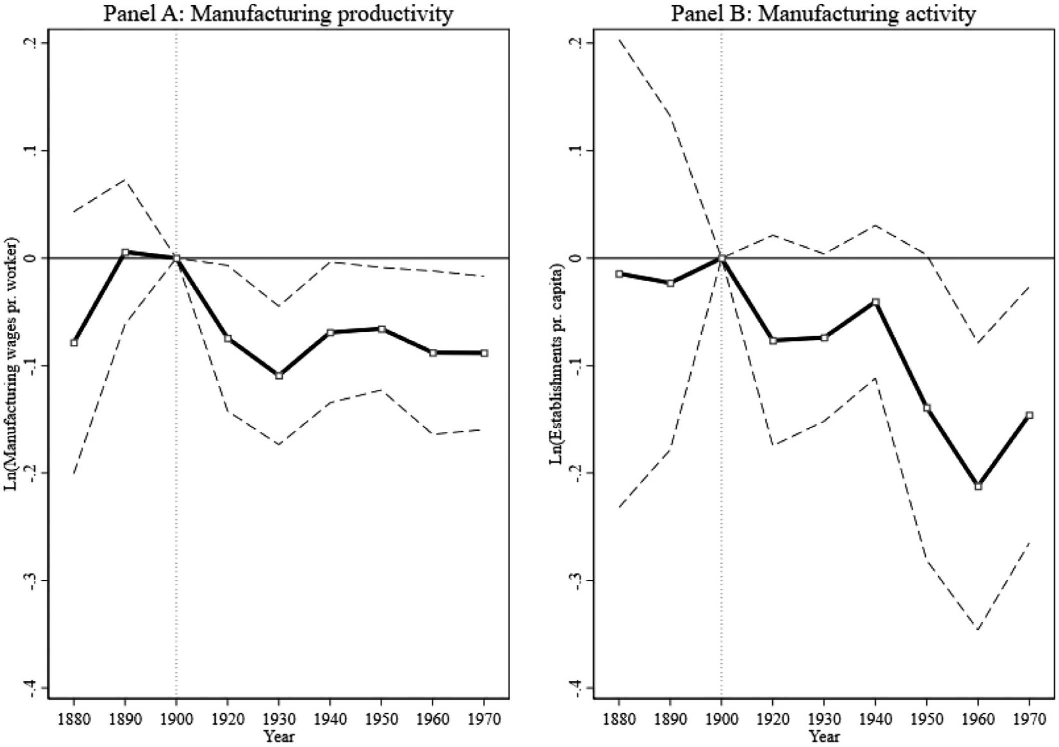


Fig. 4. Event-study of the effect of earthquake intensity on the manufacturing sector. *Notes:* This figure shows the dynamic effects of earthquake intensity on manufacturing outcomes. The method of estimation is two-stage least squares using 1900 as the reference year (omitted). Both specifications include fixed effects for county and time and control for *ln* latitude and *ln* longitude interacted by time-period fixed effects. The solid line depicts the effect on city population relative to the base year (1900). The outcome in Panel A is *ln* wages per manufacturing worker and in Panel B *ln* manufacturing establishments per capita. Dashed lines indicate 95% confidence intervals. Standard errors are clustered at the county level.

Panels C and D report heterogeneity results by access to credit and markets. Cities with banks might have recovered faster from the earthquake as they could obtain financial resources needed for reconstruction locally. We digitized data from the Rand McNally Banking Directory for the year 1900 to evaluate whether this was the case. The directory contains detailed information about locations of banks in the three states included in our sample. In Panel C, we display results by cities with and without a bank in 1900. While cities with banks were relatively less affected by the earthquake compared to cities with no banks, they still did not completely recover by 1970. Panel D reports results by cities above and below median market access in 1890.[13] Although the estimates of the earthquake a numerically larger for cities located in counties with below median market access, these differences are not statistically significant.

4.3. Why did the 1906 earthquake matter in the long run?

One possible explanation why we observe a long-run effect of the 1906 earthquake is related to the path-dependence of chain migration. If the earthquake disrupted established migrant networks, which otherwise catalyzed chain migration, one would expect such long-run effects of a temporary shock. This is a plausible mechanism in our setting, because the American West was an important migrant destination in the United States since the late 19th century.

Our approach to answer this question starts by studying whether the earthquake affected net-migration rates of people born outside the three sample states by breaking the persistence of existing migrant networks. The following model is based on full-count U.S. census data in 1900 and 1910:

$$\frac{\Delta migrant_{nc}}{Pop_{c,1900}} = \beta_0 Quake_c + \beta_1 Quake_c \times Natshare_{nc,1900} + \beta_2 Natshare_{nc,1900} + \mathbf{X}_c'\Gamma + \lambda_n + \varepsilon_{cn} \tag{2}$$

where $\Delta migrant_{nc}/Pop_{c,1900}$ is the change in the number of migrants from 1900 to 1910 of birthplace n living in county c scaled with the initial county population, which is equal to the net-migration rate of birthplace n (we refer to the Online Appendix for further details). The variable $Natshare_{nc,1900}$ measures the strength of the network of birthplace n in county c relative to other counties in the three sample states (i.e., the number of individuals from birthplace n in county c is divided by the total number of people from

---

[13] The data are retrieved from Donaldson and Hornbeck (2016), who provide information about general market access at the county level in 1890.

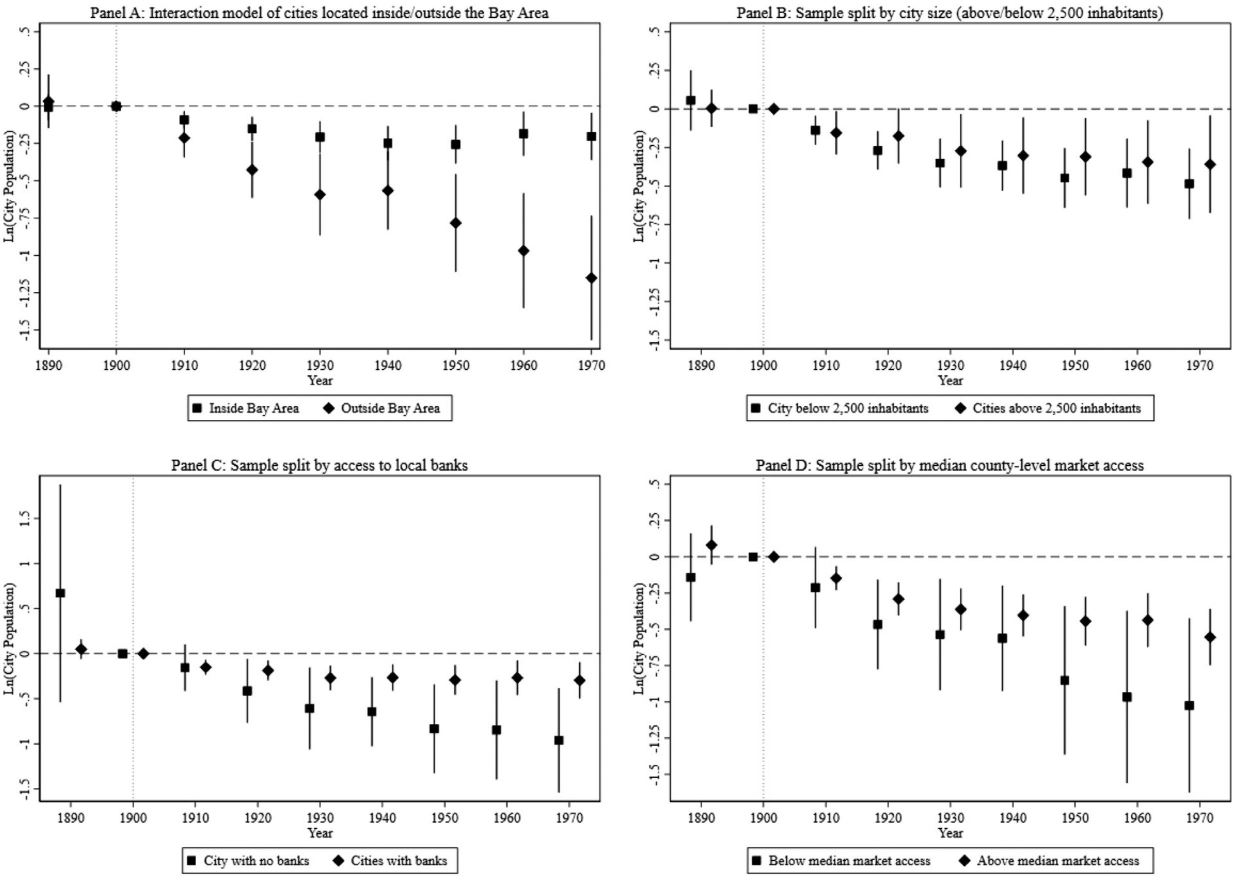


Fig. 5. Treatment Heterogeneity Effects. *Notes:* This figure reports treatment-heterogeneity effects by different city characteristics. The results are based on the specification used in Fig. 3 Panel B. The method of estimation is two-stage least squares using 1900 as the reference year (omitted). The visualization of the estimates in Panel A is different from Panels B-D, because a sample spilt was not possible due to the limited number of observations in the Bay Area. Instead, we report the effects of an interaction model, where quake intenstiy is interacted with a dummy whehter a city is located in the Bay Area. Panel A displays estimates for cities located inside vs. outside the Bay Area. The estimates referring to the Bay Area (rectangle) display the joint effect of quake intensity and quake intensity interacted with a dummy == 1 if a city is located in the Bay Area. Panels B-D report sample splits: Panel B displays cities by initial population size of cities (above vs. below 2,500 inhabitants in 1900). Panel C reports effects by access to local banks (with and without banks in 1900) and Panel D shows results for cities above vs. below median market access in 1890. Each estimate is plotted together with a spike for the 95% confidence interval. Standard errors are clustered at the city level.

birthplace *n* in the sample). The vector $\mathbf{X}_c'$ includes our baseline geographical controls latitude and longitude and λ_n denotes birthplace fixed effects. Our sample includes 100 different birthplaces (i.e., people born in other US states or from foreign countries). Standard errors are clustered by birthplace and county.[14]

Columns (1) and (2) of Table 1 report the results of estimating Eq. (2). The method of estimation is two-stage least squares. Excluding treatment heterogeneity (i.e., $\beta_1 = \beta_2 = 0$), we find that the main effect is statistically significant at the one percent level, implying that a one standard deviation increase in damage intensity is associated with 0.11 fewer migrants per 100 people (column 1). Column (2) reports the results from the full interaction model. While we see that the main effect reduces in numerical magnitude, the interaction term is negative and statistically significant, consistent with the hypothesis that the earthquake diminished the importance of chain migration in the aftermath of the earthquake. The estimated coefficient on the interaction term implies that for each standard deviation increase in the earthquake intensity, the power of chain migration is reduced by approximately one third (−0.025/0.077). We also see evidence of chain migration being a strong agglomeration force during this period, since the estimated coefficient on the county-specific nationality network is positive and statistically significant at the 1-percent level. Overall, the estimates in column (2) provide strong evidence that the earthquake interrupted chain migration in the affected areas.

---

[14] As we demonstrate in the Online Appendix, the coefficient β_0 quantifies the average effect of the earthquake on the net-migration rate of birthplace *n* if one excludes treatment heterogeneity. The aggregated migration rate, which is useful to assess the importance of migration in explaining the population results, is obtained by multiplying the coefficient β_0 with the number of birthplaces in the sample.

Table 1
Migration analysis (Two-Stage Least Squares Estimates)

	(1)	(2)	(3)	(4)
	Dependent Variable			
	Δ Migration Rate	Δ Migration Rate	In-migration Rate	Out-migration Rate
Quake	−0.001***	−0.001***	−0.113***	0.011
	(0.0002)	(0.0002)	(0.044)	(0.012)
$\textit{Nationality Share}_{1900} \times \textit{Quake}$		−0.025***		
		(0.005)		
$\textit{Nationality Share}_{1900}$		0.077***		
		(0.012)		
Observations	7,150	7,150	102	102
R-squared	0.036	0.048	0.155	0.070

Notes: This table shows the impact of the 1906 earthquake on migration for the period 1900–10. The dependent variable in columns 1–2 is the change in the number of migrants from 1900 to 1910 of birthplace *n* living in county *c* scaled with the initial county population; the out-of-state in-migration rate (column 3); and the out-migration rate (column 4). Columns 1–2 are based on complete-count Census data, while columns 3–4 are based on a linked sample of adult males age 18–55 in 1900. The variable of interests are *Quake*, the standardized potential earthquake damage intensity in county *c*, and the interaction of *Quake* with *Nationality Share*, the county-level network of birthplace *n* in 1900. The specifications in columns 1–2 also include birthplace fixed effects. All specifications control for *ln* longitude and *ln* latitude. The method of estimation is two-stage least squares. Standard errors in parentheses are clustered at the birthplace by county level in columns (1) and (2) and at the county level in columns (3)-(4): *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table 5 includes the estimates for the remaining full-count census decades (1910–1920, 1920–1930, 1930–1940). We pool the data, and consequently interact the baseline controls with a full set of time-period fixed effects and add birthplace-by-year fixed effects to estimation Eq. (2). This is equivalent to running separate regressions for each decade. We find that the main effect is negative and statistically significant at the 1-percent level up until 1920–1930 but insignificant in statistical and quantitative terms for the final period 1930–1940 (i.e., β_0 displayed in column 2). The interaction effect is stronger in numerical magnitude for the period 1900–1910 compared to any other period.

Next, we use a linked sample from 1900 to 1910, to assess if the effects for internal migrants are driven by in- and/or out-migration rates. The advantage of using a linked sample is that we can test whether the previous results are driven by a reduced inflow of newcomers or people moving out of the affected areas. It is important to note that this part of the analysis does not distinguish by birthplace, but estimates the aggregate effect of the earthquake on in- and out-migration rates at the county level.[15]

Columns (3)-(4) present the results for the linked sample using two-stage least squares as estimation method. The estimates reveal that our results are driven by out-of-state in-migration only. In particular, the estimated coefficient of interest reported in column (3) implies that a one standard deviation increase in damage intensity is associated with 11 fewer in-migrants per 100 people. The earthquake, however, did not trigger larger migratory responses of residents already living in the affected counties (column 4). Overall, these results confirm the disruptive effect that the earthquake had on existing migrant networks by diverting newcomers to other, less affected, areas in the American West.[16]

5. Concluding remarks

The 1906 San Francisco Earthquake fundamentally changed the distribution of economic activity in the American West over the course of the 20th century. More affected areas experienced a persistent decline in population size and manufacturing productivity relative to less and non-affected areas. We considered a specific mechanism of path dependence that can explain why such a temporary shock can have a long-lasting impact: chain-migration. While existing research highlighted the persistence of immigrant inflows to

---

[15] The out-migration rate is calculated as the share of individuals in the linked sample that moved out of their county of residence in 1900. In-migration rate is constructed as the number of out-of-state migrants in the linked sample that moved into a sampled county between 1900–1910 scaled with the corresponding county population in 1900.

[16] One challenge of linked data is the risk that match rates might differ across the treatment, in this case quake intensity. We do see that match rates are lower in places with higher earthquake intensity (corr = −0.147). Regressing the county-level match rate (% of those living in a county in 1900 who are found living anywhere in the U.S. in 1910) on our standardized quake intensity with the controls from the regression in Table 1 (columns 3–4), we find a statistically significant coefficient of −0.011 (s.e. = 0.005). Under certain assumptions this could bias the coefficient downwards in column (4). If those who migrate out of the county are less likely to be found, and there is more migration out of the more affected counties, we may under-state the out-migration rate in affected relative to non-affected counties. Unfortunately, this is not possible to test given the constraints of linked data.

local areas due to existing migrant networks (e.g., Card, 2001; Munshi, 2003), we show that a natural disaster, such as the 1906 earthquake, can break up such migration chains and, hence, path-dependence.

How important was the 1906 earthquake compared to other known determinants of the population distribution in the American West? Considering the long-run effects from 1900 to 1970 and using our most conservative specification (Panel F of Appendix Fig. 2 without initial population controls), we find that a one standard deviation increase in damage intensity is associated with a reduction in relative city size of 30 percent (see column 2 of Appendix Table 3). When comparing this magnitude to other determinants from this specification, we find, for example, that access to the Pacific coast is associated with an increase in relative city size of around 37 percent. Thus, the magnitude of the earthquake is almost comparable (in numerical size) to this well-known determinant of economic activity, suggesting that the damage intensity of the 1906 earthquake played an important role for the distribution of city sizes in the American West.[17]

Supplementary material

Supplementary material associated with Ager, Eriksson, Hansen, Lønstrup can be found, in the online version, at doi:10.1016/j.eeh.2020.101342.

References

Abramitzky, R., Boustan, L.P., Eriksson, K., 2012. Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration. Am. Econ. Rev. 102 (5), 1832–1856.

Abramitzky, R., Boustan, L.P., Eriksson, K., 2014. A nation of immigrants: assimilation and economic outcomes in the age of mass migration. J. Polit. Econ. 122 (3), 467–506.

Abramitzky, R., Boustan, L.P., Eriksson, K., Feigenbaum, J., Santiago, P., 2019. Automated Linking of Historical Data. NBER Working Paper 25825.

Ager, P., Eriksson, K. A., Hansen, C. W., Lønstrup, L., 2020. ICPSR how the 1906 san francisco earthquake shaped economic activity in the American West. https://doi.org/10.3886/E119841V1.

Algermissen, S.T., 1972. A Study of Earthquake Losses in the San Francisco Bay Area: Data and Analysis. US Department of Commerce, National Oceanic & Atmospheric Administration.

Bleakley, H., Lin, J., 2012. Portage and path dependence. Q. J. Econ. 127 (2), 587–644.

Bleakley, H., Lin, J., 2015. History and the sizes of cities. Am. Econ. Rev. 105 (5), 558–563.

Boatwright, J., Bundock, H., 2005. Modified mercalli intensity maps for the 1906 san francisco earthquake plotted in shakemap format.

Boustan, L.P., Kahn, M.E., Rhode, P.W., 2012. Moving to higher ground: migration response to natural disasters in the early twentieth century. Am. Econ. Rev. 102 (3), 238–244.

Boustan, L.P., Kahn, M.E., Rhode, P.W., Yanguas, M.L., 2017. The Effect of Natural Disasters on Economic Activity in Us Counties: A Century of Data. NBER Working Paper 23410.

Brakman, S., Garretsen, H., Schramm, M., 2004. The strategic bombing of german cities during world war II and its impact on city growth. J. Econ. Geogr. 4 (2), 201–218.

Card, D., 2001. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. J. Labor Econ. 19 (1), 22–64.

Cavallo, E., Galiani, S., Noy, I., Pantano, J., 2013. Catastrophic natural disasters and economic growth. Rev. Econ. Stat. 95 (5), 1549–1561.

Clay, K., Wright, G., 2012. Gold rush legacy: American minerals and the knowledge economy. Daniel Cole and Elinor Ostrom, eds., The Evolution of Property Rights Related to Land and Natural Resources. Lincoln Land Institute Press, Cambridge. 67–95

Coffman, J.L., von Hake, C.A., Stover, C.W., 1982. Earthquake history of the United States. US Department of Commerce, National Oceanic and Atmospheric Administration.

Davis, D.R., Weinstein, D.E., 2002. Bones, bombs, and break points: the geography of economic activity. Am. Econ. Rev. 92 (5), 1269–1289.

Donaldson, D., Hornbeck, R., 2016. Railroads and American economic growth: a "market access" approach. Q. J. Econ. 131 (2), 799–858.

Ellsworth, W.L., 1990. The san andreas fault system, california. US Geol. Surv. Prof. Pap 1515, 153–185.

Ferrie, J.P., 1996. A new sample of males linked from the public use microdata sample of the 1850 US federal census of population to the 1860 US federal census manuscript schedules. Historical methods. J. Quant.Interdiscip. Hist. 29 (4), 141–156.

Ferrie, J.P., Hatton, T.J., 2015. Two centuries of international migration. In: Handbook of the Economics of International Migration, 1. Elsevier, pp. 53–88.

Ferrie, J.P., Keniston, D., 2017. Creative destruction: barriers to urban growth and the great boston fire of 1872. Am. Econ. Rev. 107 (6), 1365–1398.

Ferrie, J.P., Naidu, S., 2014. When the levee breaks: black migration and economic development in the american south. Am. Econ. Rev. 104 (3), 963–990.

Fishback, P.V., Troesken, W., Kollmann, T., Haines, M., Rhode, P.W., Thomasson, M., 2011. Information and the impact of climate and weather on mortality rates during the great depression. In: The Economics of Climate Change: Adaptations Past and Present. University of Chicago Press, pp. 131–168.

Greely, A.W., 1906. Earthquake in california. Annu. Rep. War Dep. 1, 189–205.

Haines, M. R., 2010. ICPSR historical, demographic, economic, and social data: The United States, 1790–2002. https://doi.org/10.3886/ICPSR02896.v3.

Hanlon, W.W., 2017. Temporary shocks and persistent effects in urban economies: evidence from british cities after the US civil war. Rev. Econ. Stat. 99 (1), 67–79.

Hansen, G. C., Condon, E., 1989. Denial of disaster Cameron and Company San Francisco.

Hedges, J.B., 1926. The colonization work of the northern pacific railroad. Mississippi Valley Hist. Rev. 13 (3), 311–342.

Henderson, J.V., Squires, T., Storeygard, A., Weil, D., 2017. The global distribution of economic activity: nature, history, and the role of trade. Q. J. Econ. 133 (1), 357–406.

Hornbeck, R., 2012. The enduring impact of the American dust bowl: short-and long-run adjustments to environmental catastrophe. Am. Econ. Rev. 102 (4), 1477–1507.

Imaizumi, A., Ito, K., Okazaki, T., 2016. Impact of natural disasters on industrial agglomeration: the case of the great kantö earthquake in 1923. Explor. Econ. Hist. 60, 52–68.

Jedwab, R., Moradi, A., 2016. The permanent effects of transportation revolutions in poor countries: evidence from africa. Rev. Econ. Stat. 98 (2), 268–284.

Kline, P., Moretti, E., 2014. Local economic development, agglomeration economies, and the big push: 100 years of evidence from the tennessee valley authority. Q. J. Econ. 129 (1), 275–331.

Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99 (3), 483–499.

Lawson, A.C., Reid, H.F., 1908. The California Earthquake of April 18,1906: report of the state earthquake investigation commission. Carnegie Institution of Washington.

Loayza, N.V., Olaberria, E., Rigolini, J., Christiaensen, L., 2012. Natural disasters and growth: going beyond the averages. World Dev. 40 (7), 1317–1336.

---

[17] Coefficients and standard errors from this long-difference specification are: *Quake* (0.36; SE=0.15), *ln latitude* (−0.81; SE=1.96), *ln longitude* (5.32; SE=5.14), *Distance rivers* (−0.01; SE=0.02), *Coast access* (0.47; SE=0.19), *Temperature* (0.07; SE=0.03), *Droughts* (−0.15; SE=0.03), *Cotton suitability* (0.45; SE=0.35), *Altitude* (−0.02, SE=0.01), *Fault line* (0.41, SE=0.27), *SF trade cost 1890* (−0.02; SE=0.01), *ln distance LA* (−0.16; SE=0.13), *ln distance SF* (0.18; SE=0.04).

Long, J., Siu, H., 2018. Refugees from dust and shrinking land: Tracking the dust bowl migrants. J. Econ. Hist. 78 (4), 1001–1033.
Malone, M.P., Etulain, R.W., 1989. The American West: A Twentieth-Century History. University of Nebraska Press.
Michaels, G., Rauch, F., 2018. Resetting the urban network: 117–2012. Econ. J. 128 (608), 378–412.
Miguel, E., Roland, G., 2011. The long-run impact of bombing vietnam. J. Dev. Econ. 96 (1), 1–15.
Moffat, R.M., 1996. Population History of Western U.S. Cities and Towns, 1850–1990. Scarecrow Press.
Moretti, E., 1999. Social networks and migrations: Italy 1876–1913. Int. Migr. Rev. 33 (3), 640–657.
Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the US labor market. Q. J. Econ. 118 (2), 549–599.
North, D.C., 1966. Growth and Welfare in the American Past: a New Economic History. Prentice-Hall.
Olmstead, A.L., Rhode, P.W., 2017. A history of california agriculture. Giannini Foundation of Agricultural Economics. University of California Agriculture and Natural Resources, Davis.
Ruggles, S., Genadek, K., Goeken, R., Grover, J., Sobek, M., 2017. Integrated public use microdata series: Version 7.0 [dataset].
Schumann, A., 2014. Persistence of population shocks: evidence from the occupation of west germany after world war II. Am. Econ. J. 6 (3), 189–205.
Siodla, J., 2015. Razing San Francisco: the 1906 disaster as a natural experiment in urban redevelopment. J. Urban Econ. 89, 48–61.
Siodla, J., 2017. Clean slate: land-use changes in san francisco after the 1906 disaster. Explor. Econ. Hist. 65, 1–16.
Strobl, E., 2011. The economic growth impact of hurricanes: evidence from US coastal counties. Rev. Econ. Stat. 93 (2), 575–589.
Toppozada, T., Branum, D., 2004. California earthquake history. Ann. Geophys. 47 (2–3).
Turner, C., Tamura, R., Mulholland, S.E., 2013. How important are human capital, physical capital and total factor productivity for determining state economic growth in the united states, 1840–2000? J. Econ. Growth 18 (4), 319–371.
Wald, D.J., Quitoriano, V., Heaton, T.H., Kanamori, H., Scrivner, C.W., Worden, B.C., 1999. Trinet “shakemaps”: rapid generation of peak ground motion and intensity maps for earthquakes in southern california. Earthq. Spectra 15 (3), 537–555.
Walker, R.A., 2001. California’s golden road to riches: natural resources and regional capitalism, 1848–1940. Ann. Assoc. Am. Geogr. 91 (1), 167–199.
Wegge, S.A., 1998. Chain migration and information networks: evidence from nineteenth-century Hesse-Cassel. J. Econ. Hist. 58 (4), 957–986.
White, R., 2011. Railroaded: the Transcontinentals and the Making of Modern America. WW Norton & Company.
Zoback, M.L., 2006. The 1906 earthquake and a century of progress in understanding earthquakes and their hazards. GSA Today 16 (4–5), 4.