Two Centuries of U.S. Innovation and the Capital Channel: Evidence from Climatic Shocks*

Mathias S. Kruttli, Noah Stoffman, Sumudu W. Watugala†

March 2025

Abstract

Using advanced machine learning methods, we construct a comprehensive, nearly-error-free database of the universe of around 12 million U.S. patents from 1836 to 2024. We analyze the resilience of innovation to disaster shocks using hurricane landfall data spanning two centuries. Major hurricanes destroy local innovative capacity for up to a decade and lead to permanent losses relative to the counterfactual. Relative to individual inventors, firm-based innovation increases the resilience of innovative output to disasters. Innovative capacity is not perfectly substituted across regions, leading to aggregate losses. Our findings reveal that capital destruction and capital constraints from natural disasters induce large, long-lasting innovation losses and can explain regional variation in economic outcomes.

JEL classification: G30, N1, N5, O30, R11

Keywords: Innovation, disasters, extreme weather, capital constraints, patents, LLM, big data

*We thank Brigitte Roth Tran, Dan Wilson, and Mike Woeppel for helpful comments.

†Kruttli: Kelley School of Business, Indiana University. Email: mkruttli@iu.edu. Stoffman: Kelley School of Business, Indiana University. Email: nstoffma@iu.edu. Watugala: Kelley School of Business, Indiana University. Email: sumudu@iu.edu.

1 Introduction

> *[The Congress shall have Power . . .] To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.*
>
> —U.S. Constitution, art. I, §8, cl. 8.

Innovation has been constitutionally protected in the United States since 1789. The Founders were visionary in recognizing that innovation is fundamental to the advancement of science, health, prosperity, public welfare, and national defense. Economists have similarly highlighted the importance of innovation, as it serves as a key driver of economic growth in numerous seminal theoretical models (e.g., Romer, 1990, 1994). Given the central role of innovation, it is crucial to understand how it has evolved and diffused across the U.S., as well as the mechanisms that may either hinder or promote innovation in different regions.

Studying these dynamics presents two key challenges for empirical researchers. First, a comprehensive analysis of the history of U.S. innovation requires a reliable and complete database of U.S. patents. Second, it is necessary to identify exogenous shocks that are recurrent throughout the entire time series of U.S. innovation, and that affect different regions over time.

In this paper, we address these challenges. First, we construct a new database of patent activity by harnessing recent advances in optical character recognition (OCR) and large language model (LLM) technologies that give, for the first time, a comprehensive view into inventor and assignee names, locations, and types for the universe of around 12 million U.S. patents from 1836. Second, we exploit the spatial and temporal variation of historical natural disaster shocks to analyze how exogenous local shocks experienced by inventors affect the amount and type of innovative output and its geographic distribution. Specifically, we combine the patent data with data spanning nearly two centuries on hurricane landfalls and storm paths, which we construct based on information obtained from the National Oceanic and Atmospheric Administration (NOAA). Adapting the research design of Kruttli, Roth Tran, and Watugala (2024), we exploit the quasi-exogenous variation in regional

climatic shocks for identification by using these hurricane landfalls as exogenous events that destroy local human and physical capital. We then analyze their long-term impact on local and national innovation.

The relationship between extreme weather events and innovation is also crucial for understanding how climate change affects economic growth and its implications for the social cost of carbon (e.g., Nordhaus (1977)). Climate scientists predict that extreme weather events will become more frequent and intense. If extreme weather events destroy innovative output in a region, and if there is imperfect substitution of innovation across regions in the U.S., then this opens another channel through which climate change can reduce aggregate innovation and contribute to economic costs.

We begin by analyzing the effects of hurricanes on county-level innovation. We use patents issued in each county and year to estimate a panel version of the local projections method developed by Jordà (2005). The coefficient estimates imply large and long-lasting negative effects on patents for a county that is in the landfall region of a hurricane. The effect is not immediate, but when comparing the number of patents in the third year after landfall to the year before landfall, the growth rate for hit counties is 4.3 percentage points lower than for control counties—counties outside of the landfall region. The effect is even larger for the fourth to the ninth year after landfall. When comparing the number of patents by county in the fifth year after landfall to the year before landfall, the hit counties' patent growth rate is 9.5 percentage points lower. The effect becomes insignificant only in the tenth year after landfall. The reduction in the growth rate of patents is most pronounced for counties located closest to the eye of the hurricane at landfall. Importantly, this reduction in the growth rate of patents for hit counties is a shortfall that does not seem to be compensated for by a higher growth rate once the hit county recovers.

Next, we analyze how the role of firms in the resilience of a location's innovative capacity. The majority of patents in the history of U.S. innovation are developed by individuals without an institutional affiliation. We call these "independent" patents. Most of the remaining

patents have an inventor who is backed by a company—that is, the patent is assigned to the company when it is granted. We call these "firm" patents. We hypothesize that independent innovation faces greater financial constraints compared to firm-backed innovation following a disaster and consequently, is less resilient to hurricane hits. Firms are more likely to have greater access to capital that is also regionally diversified. Our findings are consistent with this hypothesis. Counties for which the share of independent patents is large experience a disproportionately larger decrease in their innovation output. For the counties that mostly rely on firm patents, the effect is reversed, and they experience only a muted impact of the hurricane on their innovation output. This finding is in line with firms being better able to withstand destructive shocks to human and physical capital.

Finally, we examine whether the U.S. innovation network can perfectly substitute the patents shortfall in treated regions. On the one hand, innovation might decrease in hit regions, but inventors might relocate to unaffected counties, or those already in unaffected counties might increase their innovation output. In this case, aggregate innovation in the U.S. would not be affected. On the other hand, a region might be highly specialized in a certain type of innovation, and there may be imperfect substitution across regions for this type of innovation. Then, hurricanes would not only affect local but also aggregate innovation in the U.S. We find evidence for the latter. When patents in a class—we use the United States Patent Classification (USPC) system to identify patent classes—originate from a region that is hit by a hurricane, aggregate patent growth in the years after landfall in this class is up to 31 percentage points lower than for control classes.

We make two significant contributions to the literature. First, by developing a new methodology and constructing a new, highly accurate, comprehensive database of patent information (including inventors and their locations, technological characteristics, and assignee information), we enable research that answers important questions regarding two centuries of U.S. innovation. Second, by analyzing hurricane landfalls from 1851 to 2023, we show that hurricanes have a large negative effect on innovation output lasting several years for

counties in the landfall region. The effect is larger for counties in which the share of independent patents is large. Further, there exists imperfect substitution in innovation output across regions in the U.S. When a region that specializes in an innovation class is hit by a hurricane, aggregate patenting in this innovation class decreases in the years after hurricane landfall.

Our paper contributes to several strands of economic research. First, by constructing a new database for the universe of U.S. patents, we complement other papers that analyze innovation for years not covered by the NBER patent database (Hall, Jaffe, and Trajtenberg, 2001), which starts in 1976. For example, Kogan, Papanikolaou, Seru, and Stoffman (2017) process patent information for the subset of patents issued since 1926 for which the assignee is a publicly traded firm. Kelly, Papanikolaou, Seru, and Taddy (2021) measure the novelty of historical patents using textual analysis methods but do not extract information on the inventor and assignee names and location, and rely on the text provided by Google Patents. Petralia, Balland, and Rigby (2016) are more closely related to our paper in that their goal is to develop a database of historical patents that contains inventor and assignee names and location. However, as we discuss in Section 3, there are substantial discrepancies when comparing the Petralia, Balland, and Rigby (2016) data to our database. These discrepancies are likely driven by recent improvements in OCR and our ability to use LLM technology to extract information with fewer errors.

Second, we add to the literature on the economic risks associated with a changing climate. The risks are broadly categorized into two types: physical and transition risks (Carney, 2015). Physical risks encompass the destruction from more intense and frequent extreme weather events like hurricanes and heat waves (e.g., Kruttli, Roth Tran, and Watugala, 2024) or sea-level rise due to climate change. Transition risks refer to the risks associated with the transition to a low-carbon economy due to policies imposing a cost on firms and households (e.g., a cap-and-trade program). There exists an inherent trade-off between the two risks: imposing a price or a tax on greenhouse gas emissions increases transition risks but

reduces greenhouse gas emissions and consequently future physical risks (e.g., Ivanov, Kruttli, and Watugala, 2024). Therefore, a precise estimate of the cost of physical risks is key to determining how high the price or tax on greenhouse gas emissions should be set. Beginning with the seminal work of Nordhaus (1977), there is an extensive literature on the economic cost—in the form of reduced economic growth—of greenhouse gas emissions. The empirical studies on physical climate risks and economic growth have focused on temperature shocks (e.g., Dell, Jones, and Olken, 2012 and Burke, Hsieng, and Miguel, 2015) and other natural disasters (e.g., Deryugina, 2017; Boustan, Kahn, Rhode, and Yanguas, 2020; Roth Tran and Wilson, 2024). However, what is missing from this literature is an in-depth empirical assessment of the impact that physical risks have on innovation. Examining this innovation channel is particularly important given the integral role for innovation and technological change in economic growth. Noy and Strobl (2023) analyze the impact of hurricanes on innovation but rely on the patent data constructed by Petralia, Balland, and Rigby (2016), which has large discrepancies to our database. Moreover, our research questions differ due to our focus on independent versus firm patents and the substitution effects in innovative output across U.S. regions.

Third, we empirically test predictions of theoretical models on innovation. Innovation is a key driver of economic growth in many seminal theoretical models (e.g., Romer, 1990, 1994; Aghion and Howitt, 1992), and capital in the form of infrastructure and labor is needed to generate innovation in these models. However, the impact of changes in available capital on innovation (i.e., the capital channel) has been challenging to identify empirically because long time series measures of both innovation and exogenous shocks to capital have thus far not been collected and analyzed. In an ideal experiment, one would exploit exogenous variation in the capital stock of a (treated) region and analyze how innovation in this region changes compared to the contemporaneous innovation in other (control) regions. Our research design approaches this setting by using hurricane landfall data that dates back to 1851 as exogenous shocks to capital combined with comprehensive data on patents—arguably the most

commonly used measure of innovation and technological change. Hurricanes are destructive shocks to the capital stock of a region. For example, hurricanes can both damage a region's infrastructure and lead to an out-migration of the labor force (Deryugina, Kawano, and Levitt, 2018). Importantly, whether a hurricane makes landfall in a specific location is not caused by the prevailing economic conditions of that location.

The remainder of the paper is organized as follows. In the next section, we describe the empirical design. Section 3 discusses the data, including the methodology for constructing the novel patent database covering two centuries. Section 4 presents the results. Section 5 concludes.

2 Empirical strategy

Our identification strategy relies on hurricanes as exogenous shocks to regions and is adapted from Kruttli, Roth Tran, and Watugala (2024). Hurricanes make landfall on the Atlantic and Gulf Coasts of the U.S. and the landfall region typically spans several counties in one or more states. Hurricanes have made landfall over major population and economic centers in a range of states in this region. Figure 4 plots the landfall regions of four hurricanes in our sample, which we construct by processing text files on hurricane paths that are available from 1851 onwards from NOAA.

Figure 5A presents a stylized example illustrating inventor location within counties that are exposed and unexposed to a particular hurricane landfall. We consider counties located within a hurricane landfall region as treated and counties outside of it as controls. This dimension of spatial variation in the empirical design gives us cross-sectional variation. Because counties are hit by hurricanes infrequently, the time series variation allows us to analyze within county effects.

Using the hurricane landfall regions and the location of the inventors for a given patent, we can estimate how a hurricane affects innovation activity in a county. For our baseline

estimates, we adapt the local projection estimator of Jordà (2005) for our panel of county-year observations.[1] The regression specification is:

$$\log\left(\frac{NPatents_{c,t+h}}{NPatents_{c,t-1}}\right) = \beta_1 Hit_{c,t} + \sum_{r=-5,r\neq 0}^{h} \rho_r Hit_{c,t+r} + \sum_{r=1}^{5} \kappa_r \log NPatents_{c,t-r} + \mu_c + \theta_t + \epsilon_{c,t}. \quad (1)$$

The dependent variable is the change in the log of the number of patents, $NPatents$, of a county c from the year before the hurricane hit, $t-1$, to h years after the hurricane hit, $t+h$. A patent is assigned to county c if all the inventors of the patent are located in that county. A patent is assigned to the year in which the patent is issued. The exceptions are patents issued between January and May. These patents count for the previous year to align patents to the correct event because Atlantic Coast hurricanes only make landfall during the hurricane season from June to November.[2] The independent variable of interest is the variable $Hit_{c,t}$, which is an indicator variable that takes the value one when a county c is in the landfall region of a hurricane in year t and zero otherwise. In addition to county and year fixed effects, we include several control variables. First, we control for a county being hit by hurricanes that make landfall during the preceding 5-year period from $t-5$ to $t-1$. We also control for a county being hit by hurricanes that make landfall during the years from $t+1$ to $t+h$ These controls are important to account for the possibility of staggered and multiple treatment of counties (De Chaisemartin and d'Haultfoeuille, 2020; Athey and Imbens, 2022; Baker, Larcker, and Wang, 2022). Second, we capture potential pre-trends in a county's innovation activity by including the log of patents lagged up to five years.[3] The standard errors are double clustered at the county and year levels. Of all the states in the

[1] A similar regression specification is used by Roth Tran and Wilson (2024).

[2] Omitting this adjustment leads to qualitatively similar results.

[3] The estimates are qualitatively similar when including additional lags, and including additional lags does not materially improve the fit of the regression model.

US, 32 have at one point been hit by a hurricane from 1851 to 2023 based on our landfall region estimations. We only include counties from these 32 states in our sample.

In addition to the local projection regression in equation 1, we estimate the effect of hurricanes on innovative output through an event study specification. The event study regression takes the form:

$$\log\left(\frac{\overline{NPatents}_{c,t+h}}{\text{NPatents}_{c,t-1}}\right) = \beta_1 Hit_{c,t} + \sum_{r=1}^{5} \kappa_r \log \text{NPatents}_{c,t-r} + \mu_c + \theta_t + \epsilon_{c,t}. \tag{2}$$

For the event study specification, the dependent variable is the difference between the logs of the 5-year average of the number of patents from county c after h years following a hurricane hit, $\overline{NPatents}_{c,t+h}$, and the number of patents from county c the year before the hurricane hit, $NPatents_{c,t-1}$. This regression is jointly estimated for all the hurricanes in our sample. County and year fixed effects in addition to lagged patents are included as in equation 1. To ensure that the control counties are not hit by hurricanes that make landfall outside of year t, we exclude hit counties from the control group ten years before and after the hit. The standard errors are again double clustered by year and county. This event study specification is more parsimonious than the local projection specification and allows us to interact the treatment variable Hit with other variables to investigate cross-sectional heterogeneity in how hurricanes affect innovation in a county.

3 Data

3.1 Novel patents database

To facilitate studying the longest possible time series of hurricanes, we construct a new patent database. The data include information on grant date and inventor location (city and state) spanning 1836 through 2023. For data beginning in 1976, we obtain data from PatentsView, which is maintained by the Office of the Chief Economist at the USPTO. For

data from 1836 through 1975, we extract data directly from the PDFs of patent documents on the USPTO website by taking advantage of modern OCR algorithms and ChatGPT, as we explain below. Additional detail is provided in the Appendix.

Modern patent documents include the title and abstract of the invention; application and grant dates; the name(s) and locations of residence for all inventors and any assignees; a detailed description of the invention; drawings of the invention; and the claims, which define the scope of legal protection provided by the patent. An inventor must be a person, but patents can be assigned to a corporation or another person. Generally, when an invention is developed in a company, the inventor will assign the patent to their employer. Beginning in February 1947, patent documents began consistently including citations to existing patents as references to "prior art" (Nicholas, 2010). Older patents don't typically contain all of this information, although the grant date and inventor location have always been included.

While the layout of patent documents has changed over time, the required information has mostly remained constant. Since 1976, the USPTO has digitally recorded newly-granted patents and made machine-readable files publicly available. For the period prior to that, information must be extracted from scanned documents, an example of which is shown in Figure 1.

The launch of Google Patents in 2006 made it easier to search the text of the pre-1976 historical patent documents. These data served as the source for several influential papers, including Kogan, Papanikolaou, Seru, and Stoffman (2017), who created a widely used database, and Kelly, Papanikolaou, Seru, and Taddy (2021). However, the quality of the text extracted from scanned documents using Optical Character Recognition (OCR) available in Google Patents is quite variable, especially for patents granted before 1950. The importance of the errors in these data depends greatly on how the text is being used. For our paper, correctly identifying the inventor's location is of paramount importance, so correctly capturing this information is critical.

As an example of the OCR errors that can be found in the Google data, Figure 2 shows how the text for the patent in Figure 1 is rendered by Google (left column). Note that this text does not include anything that resembles the name of the inventor, his city of residence, or the title of the invention. The entire first paragraph of the text is missing, as is the beginning of the next paragraph. Although not all of the OCR in Google Patents is of this poor quality, many hundreds of thousands of documents have significant degradation in the extracted text. Even the use of exceptionally flexible textual analysis techniques cannot overcome the "garbage-in-garbage-out" problems associated with such poor OCR quality.

OCR algorithms have vastly improved in recent years (Correia and Luck, 2023) but the text available from Google Patents has not been updated. Advances in computer vision have given OCR algorithms the ability to identify parts of documents, patterns in text layout, and words in ways that far surpass what is available in Google Patents. Moreover, ChatGPT is a game changer when it comes to extracting information from text. Even in the presence of OCR errors, spelling mistakes, typos, or missing information, it can draw on its vast training data to answer carefully-crafted questions about the (intended) meaning of text.

As a point of comparison with our novel data, we compared the locations of inventors in our data with those in the "HistPat" data created by Petralia, Balland, and Rigby (2016). Of the 3.9 million unique patents between 1836 and 1975, HistPat is missing over 600,000 patents, which appears to be mainly due to the exclusion of foreign inventors. Of the remaining 3.3 million patents, just over 400,000 have a discrepancy between inventor locations; this appears to be because the assignee's location is often misreported as the inventor's location in the HistPat data. Despite the requirement that all patents have an inventor, another 115,000 patents in HistPat have an assignee but no inventor. We also observe significant time series variation in differences between the two datasets. For example, the rate of missing patents in HistPat relative to our dataset is below 10% through 1880 before jumping to the teens for most of the period between 1880 and 1965. This missing rate increases to 19%, 24%, and 50% in 1965, 1970, and 1975, respectively.

Using our new inventor location data, aggregated to the county level, Figure 3 illustrates the expected westward shift in innovation over the last two hundred years.

3.2 Two centuries of hurricane landfalls

We obtain the hurricane path data for all Atlantic and Gulf Coast hurricanes making landfall from 1851 to 2023 from NOAA's Atlantic HURDAT2 database, which covers all known tropical cyclones and subtropical cyclones and is a part of the Re-analysis Project (Landsea and Franklin, 2013). These data document the latitude and longitude of the eye of a tropical cyclone at least every 6 hours before it dissipates. This allows us to calculate hurricane paths and landfall dates and times. We use these data to calculate the list of counties that are within 50, 100, and 200 miles of the landfall location and path of the eye of a hurricane as in Kruttli, Roth Tran, and Watugala (2024). Figure 4 shows the landfall regions for four hurricanes in our sample. The counties that are within 50 miles of a hurricane eye would likely be the hardest "hit," while those that fall only within 200 miles of a hurricane eye would experience winds at lower speeds and likely suffer less destructive impacts. We use this idea in our empirical analysis and show results separately with assignment to treatment measured at different landfall radii.

For our main analysis, we focus on the set of deadliest tropical storms with more than 25 fatalities during our sample period. We construct this set starting with NOAA's report on "The deadliest, costliest, and most intense United States tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts)" that covers the 1851-2010 period (Blake, Landsea, and Gibney, 2011). We manually augment these data for more recent years to cover the period from 1851 to 2023. The list of 64 deadliest hurricanes in our sample is shown in Table B.2. We focus on this set of storms for two key reasons. First, these are likely the hurricanes with the most reliable data for the early period of our sample. Second, these are likely the hurricanes that significantly impacted human and physical capital in the landfall regions. The number of fatalities is one reliable measure of the destructive impact

of a hurricane, which does not rely on as many assumptions as, potentially, the monetary value of damages.

The summary statistics for the variables described in Section 2 and Section 4 are presented in Table 1.

4 Results

This section presents the results of our analyses. First, we discuss the baseline effects. Second, we analyze how reliance on household versus company innovation differentially affects the impact of a hurricane on a county's innovative activity. Third, we test whether regions not hit by a hurricane can compensate for the shortfall of patents in the hit region. In other words, how resilient is U.S. innovation to local shocks?

4.1 Baseline effects

We estimate the regression in equation (1) to test if a hurricane hit adversely affects innovation in that county, how large the effects are, and whether the effects are transitory or permanent. When identifying hit and control counties, that is, for which counties the variable Hit is one, we measure landfall regions based on three radii around the eye of the hurricane. We expect that the effects are largest for counties that lie within 50 miles of the eye of the hurricane. The effects are thought to weaken if the county lies farther away, that is, within 100 or 200 miles of the eye of the hurricane. We estimate the regression separately for each radius. These radii correspond to the radii used in Kruttli, Roth Tran, and Watugala (2024) and line up with NOAA's measurements on the average distance from the eye within which wind speeds are at hurricane strength. For all estimations, including those using smaller radii, we exclude from the control group any counties located within the 200-mile landfall region.

Figure 6 plots the coefficient estimate of the variable Hit for different horizons h. The negative effects on patents do not appear immediately after landfall. The drop in the number of patents from a hit county, relative to the year before landfall, starts three years after landfall. However, the effect intensifies as time passes and the decrease in innovation is most pronounced between five and eight years after the hurricane landfall. Ten years after landfall, the coefficient estimates become insignificant.

The effects are larger for counties closer to the eye of the hurricane at landfall. The magnitude of the coefficients for the 50-mile radius reaches as low as -0.094 in five years after landfall. This magnitude implies that the number of patents in a hit county is 9.4% lower in the fifth year after the hurricane hit than in the year before the hurricane hit. For the 200-mile radius, the coefficient magnitude decreases but still goes as low as -0.044. The lower coefficient magnitude is consistent with the idea that hurricanes are more destructive closer to the eye of the storm.

The negative and significant coefficient estimates for several years post-landfall show that a county in a hurricane landfall region experiences a substantial shortfall in patents. This shortfall can affect aggregate U.S. innovation output if innovation is not perfectly substitutable across regions. Section 4.3 investigates this hypothesis.

Table B.1 in the appendix reports the estimates of all the coefficients in equation (1). The first row of the table shows the coefficient on the independent variable of interest, $Hit_{c,t}$, which is plotted in Figure 6. The coefficient estimates on the lagged patent variables are highly significant. The coefficient on the first lag is negative with a magnitude between -0.67 and -0.78. This estimate implies that the time series of county-level patents is no longer explosive after differencing the dependent variable. However, the positive and significant coefficient estimates on the additional patent lags suggest that the undifferenced time series is a unit root process. We confirm the unit root in county-level patents with an augmented Dickey-Fuller test (unreported) for all U.S. patents from 1936 to 2023. Therefore, differencing

the dependent variable is essential when estimating regression models with county-level time series of patents.

We estimate the event study regression in equation (2) and find that the estimates of the magnitudes and horizons for the decline in patent output are qualitatively similar to the estimates of the local projection model. The estimates of the event study regression are reported in Table 2. As with the local projection regression, the coefficient estimates decrease for landfall regions based on a larger radius around the eye of the hurricane. Also, the coefficients on the lagged patents are of the same sign and close in magnitude to the coefficients in Table B.1.

4.2 Firm versus independent inventors and the resilience of local innovation

U.S. patents come from three main sources. The majority of innovations that receive a patent are developed by individuals who are unaffiliated with a company. The second largest category is patents that are developed by employees of a company and assigned to that company. The third category is substantially smaller than the other two and stems from employees who work for government or educational institutions.

Because individuals without any company backing likely have considerably less capital, a hurricane could have a more detrimental effect on counties that rely on "independent" versus "company" patents. A company can absorb a shock to capital and insulate innovative activity, which is more challenging for individual inventors. We test this hypothesis by adapting the regression specification in equation (2):

$$\log\left(\frac{\overline{NPatents}_{c,t+h}}{NPatents_{c,t-1}}\right) = \beta_1 Hit_{c,t} + \beta_2 Hit_{c,t} \times FirmShare_{c,t-1:t-5}$$
$$+\beta_3 FirmShare_{c,t-1:t-5} + \sum_{r=1}^{5} \kappa_r \log NPatents_{c,t-r} + \mu_c + \theta_t + \epsilon_{c,t}. \quad (3)$$

This specification adds the term $FirmShare_{c,t-1}$, which is interacted with the treatment variable $Hit_{c,t}$. The variable $FirmShare_{c,t-1}$ measures the share of patents from that county for which the assignee was a firm over the five-year window before the hurricane hit. The variable is standardized to facilitate the interpretation of the coefficients.

Table 3 reports the coefficient estimates. In counties where innovative activity is driven by firms, the impact of a hurricane on patents is muted. In fact, a one standard deviation increase in $FirmShare_{c,t-1}$ cancels out the negative coefficient on the $Hit_{c,t}$ variable. The coefficient estimate on the interaction term is positive for all the specifications, but the statistical significance is stronger for the landfall region based on the 100 and 200 miles around the eye of the hurricane. This increase in statistical significance is likely due to more counties falling within the larger landfall regions, making the estimates more precise.

Table 4 reports the coefficient estimates for the regression in equation (3) but replacing the variable $FirmShare_{c,t-1}$ with the variable $IndShare_{c,t-1}$. This variable is constructed in the same way as $FirmShare_{c,t-1}$ but instead captures the share of independent innovations, that is, patents for which the inventor is not affiliated with any firm. The estimates yield the same finding as those in Table 3: in counties that rely on independent innovative activity, a hurricane hit is more detrimental to future innovation output.

4.3 Substitution effects and the resilience of aggregate innovation

In Section 4.1, we estimate large shortfalls in patents in the years after a hurricane makes landfall. A natural question that follows is whether this shortfall in patents from one region is compensated for by an increase in patents in another region, or whether the local shortfall in innovation after a hurricane hit affects aggregate U.S. innovation. This question is particularly important when a region hit by a hurricane is specialized in a certain type of innovation.

To answer this question, we estimate a regression at the patent classification-year level. For the patent classification, we use the USPC codes from the USPTO. The regression

equation is given by:

$$\log\left(\frac{\overline{NPatents}_{k,t+h}}{NPatents_{k,\ t-1}}\right) = \beta_1 Hit_{k,t} + \sum_{r=1}^{5} \kappa_r \log NPatents_{k,\ t-r} + \mu_k + \theta_t + \epsilon_{k,t} \tag{4}$$

$Hit_{k,t}$ is an indicator variable that is equal to one if more than 25% of the geographic footprint of innovation within a particular patent classification k is "hit" by a hurricane in year t and is zero otherwise. Specifically, $Hit_{k,t} = 1$ if $ClassShareHit_{k,t} \geq 0.25$, where $ClassShareHit_{k,t}$ is a continuous variable between zero and one defined as:

$$ClassShareHit_{k,t} = \sum_{c=1}^{N} (Hit_{c,t} \times CountyClassShare_{c,k,t-5:t-1}). \tag{5}$$

The variable $ClassShareHit_{k,t}$ measures for each USPC patent class k the share of all patents issued in the preceding 5 years that come from a region hit by a hurricane at time t. In other words, the variable measures the extent to which the hit region specialized in that type of innovation.[4] The regression in equation (4) includes patent class and year fixed effects. The standard errors are double clustered at the class and year levels.

We first note that individual patent classes heavily rely on innovation output from a few counties, and the composition of these counties changes relatively little over time. Of the top ten contributing counties to a patent class, around 50% of the counties stay the same over 5-year increments.

The estimates of the regression in equation (4) are reported in Table 5. The coefficient estimates are negative and significant for the landfall regions based on the 50- and 100-mile radius around the hurricane eye. The coefficient estimates are also economically large. For the 50-mile radius landfall regions, the coefficient estimate is up to 0.31. This magnitude implies that the number of patents in a hit classification is 31% lower between five and nine years after hurricane landfall compared to the year before hurricane landfall. For the

[4]We only include a class in the regression once the class reached 50 patents in a year. This filter is imposed to avoid including nascent classes in the regression. Such nascent classes are more likely to be identified as hit due to the small number of counties that innovate in these classes.

100-mile radius landfall regions, the magnitude of the coefficient estimates is lower, which is consistent with counties experiencing less destructive force further away from the eye of the hurricane. For landfall regions based on the 200-mile radius, the estimates are statistically insignificant.

Overall, these results show that there is imperfect substitution of innovation between different regions in the US. When a center for a class of innovation is hit by a hurricane, aggregate innovation output decreases for this class in the years following hurricane landfall.

5 Conclusion

Innovation supports continued economic growth and is an essential driver of an economy maintaining a competitive edge. In this paper, we focus on the capital channel and analyze the impacts on innovation from quasi-exogenous local disasters that cause human and physical capital destruction. To do so, we first construct a comprehensive, nearly-error-free database of all U.S. patents from 1836 to 2024 using advanced OCR and LLM technologies. We then analyze the resilience of regional innovation to disaster shocks using hurricane landfalls spanning two centuries.

We find that major hurricanes destroy local innovative capacity for up to a decade following landfall and lead to permanent counterfactual losses. A higher presence of innovation backed by firms increases the resilience of local innovative output to disasters, suggesting that firms are important in easing constraints on the capital required to support innovation. We analyze the variation in the evolution of aggregate innovation levels between patent classifications that are significantly exposed to a hurricane hit compared to those that are not. We find that innovative capacity is not perfectly substituted across regions, leading to aggregate losses in innovative output following hurricane hits. Our findings reveal that the capital destruction and capital constraints induced by natural disasters generate large, long-lasting innovation losses and can explain regional variation in economic outcomes.

Our findings are highly relevant for policymakers and regulators focused on protecting and promoting innovative activity. Our results on how the capital destruction and constraints induced by local natural disaster shocks have long-lasting effects on innovation are of particular importance to current debates regarding how to ensure the competitiveness of the U.S. economy while keeping it resilient to unexpected disasters and shocks.

Harnessing almost two centuries of data, we reveal a mechanism that can generate divergence in regional economic prosperity. This channel is important for local and national policymakers to factor in when choosing policies to promote economic growth and prosperity.

Further, if extreme weather events like hurricanes become more damaging or frequent in the future given the trends in demographics, construction, or the environment, academics and policymakers alike will need to better understand the impact of extreme weather events on innovation to comprehensively assess the economic costs of disaster shocks. Our findings will be crucial for policymakers who seek to balance the costs and benefits of adaptation and resilience to natural disasters.

References

Aghion, Philippe, and Peter Howitt, 1992, A model of growth through creative destruction, *Econometrica* 60, 323–351.

Athey, Susan, and Guido W Imbens, 2022, Design-based analysis in difference-in-differences settings with staggered adoption, *Journal of Econometrics* 226, 62–79.

Baker, Andrew C, David F Larcker, and Charles CY Wang, 2022, How much should we trust staggered difference-in-differences estimates?, *Journal of Financial Economics* 144, 370–395.

Blake, Eric S, Chris Landsea, and Ethan J Gibney, 2011, The deadliest, costliest, and most intense united states tropical cyclones from 1851 to 2010 (and other frequently requested hurricane facts), NOAA technical memorandum NWS NHC-6 National Weather Service, National Hurricane Center.

Boustan, Leah Platt, Matthew E. Kahn, Paul W. Rhode, and Maria Lucia Yanguas, 2020, The effect of natural disasters on economic activity in us counties: A century of data, *Journal of Urban Economics* 118, 103257.

Burke, Marshall, Solomon Hsieng, and Edward Miguel, 2015, Global non-linear effect of temperature on economic production, *Nature* 527, 235–239.

Carney, Mark, 2015, Breaking the tragedy of the horizon - climate change and financial stability, *Speech given at Lloyd's of London (29 September)*.

Correia, Sergio, and Stephan Luck, 2023, Digitizing historical balance sheet data: A practitioner's guide, *Explorations in Economic History* 87, 101475.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille, 2020, Two-way fixed effects estimators with heterogeneous treatment effects, *American Economic Review* 110, 2964–2996.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, 2012, Temperature shocks and economic growth: Evidence from the last half century, *American Economic Journal: Macroeconomics* 4, 66–95.

Deryugina, Tatyana, 2017, The fiscal cost of hurricanes: Disaster aid versus social insurance, *American Economic Journal: Economic Policy* 9, 168–198.

——— , Laura Kawano, and Steven Levitt, 2018, The economic impact of hurricane katrina on its victims: Evidence from individual tax returns, *American Economic Journal: Applied Economics* 10, 202–233.

Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg, 2001, The NBER patent citation data file: Lessons, insights and methodological tools, NBER working paper.

Ivanov, Ivan T., Mathias S. Kruttli, and Sumudu W. Watugala, 2024, Banking on carbon: Corporate lending and cap-and-trade policy, *Review of Financial Studies* 37, 1640–1684.

Jordà, Òscar, 2005, Estimation and inference of impulse responses by local projections, *American economic review* 95, 161–182.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, 2021, Measuring technological innovation over the long run, *American Economic Review: Insights* 3, 303–320.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, 2017, Technological innovation, resource allocation, and growth, *Quarterly Journal of Economics* 132, 665–712.

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa, 2024, Large language models are zero-shot reasoners, in *Proceedings of the 36th International Conference on Neural Information Processing Systems* NIPS '22 Red Hook, NY, USA. Curran Associates Inc.

Kruttli, Mathias S., Brigitte Roth Tran, and Sumudu W. Watugala, 2024, Pricing poseidon: Extreme weather uncertainty and firm return dynamics, *Journal of Finance, forthcoming.*

Landsea, Christopher W, and James L Franklin, 2013, Atlantic hurricane database uncertainty and presentation of a new database format, *Monthly Weather Review* 141, 3576–3592.

Nicholas, Tom, 2010, The role of independent invention in U.S. technological development, 1880-1930, *Journal of Economic History* 70, 57–82.

Nordhaus, William D., 1977, Economic growth and climate: The carbon dioxide problem, *American Economic Review* 67, 341–346.

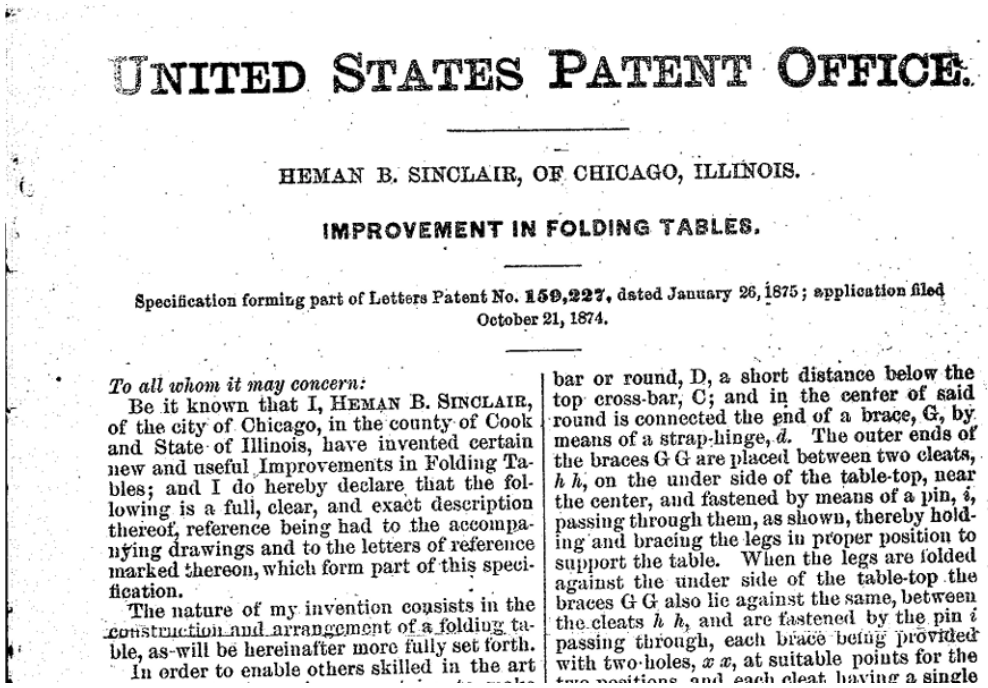
Noy, Ilan, and Eric Strobl, 2023, Creatively destructive hurricanes: Do disasters spark innovation, *Environmental and Resource Economics* 84, 1–17.

Petralia, Sergio, Pierre-Alexandre Balland, and David L. Rigby, 2016, Unveiling the geography of historical patents in the united states from 1836 to 1975, *Scientific Data* 3, 160074.

Romer, Paul M., 1990, Endogenous technological change, *Journal of Political Economy* 98, 71–102.

——— , 1994, The origins of endogenous growth, *Journal of Economic Perspectives* 8, 3–22.

Roth Tran, Brigitte, and Daniel J Wilson, 2024, The local economic impact of natural disasters, .



UNITED STATES PATENT OFFICE.

HEMAN B. SINCLAIR, OF CHICAGO, ILLINOIS.

IMPROVEMENT IN FOLDING TABLES.

Specification forming part of Letters Patent No. **159,227**, dated January 26, 1875; application filed October 21, 1874.

To all whom it may concern:

Be it known that I, HEMAN B. SINCLAIR, of the city of Chicago, in the county of Cook and State of Illinois, have invented certain new and useful Improvements in Folding Tables; and I do hereby declare that the following is a full, clear, and exact description thereof, reference being had to the accompanying drawings and to the letters of reference marked thereon, which form part of this specification.

The nature of my invention consists in the construction and arrangement of a folding table, as will be hereinafter more fully set forth.

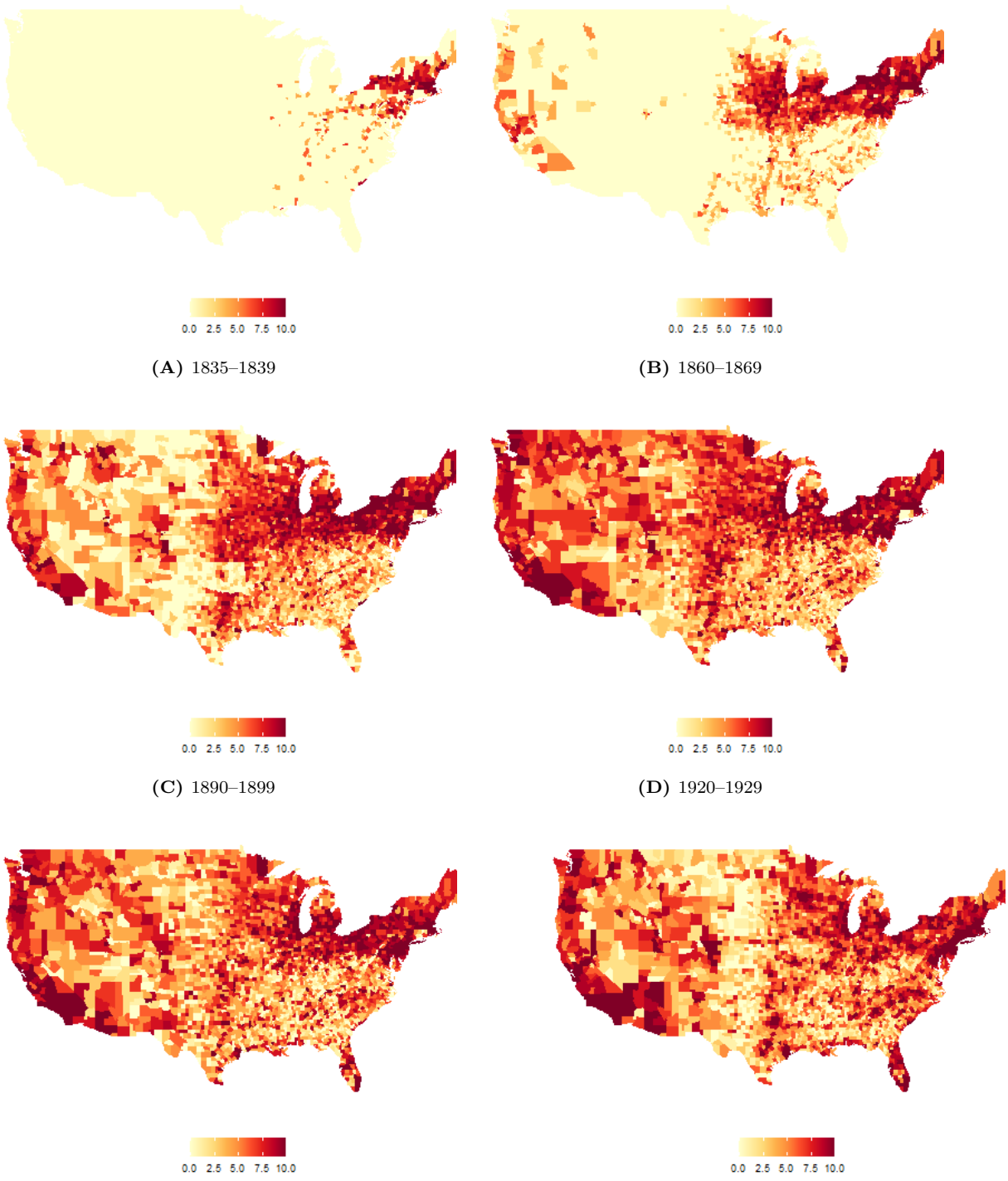
In order to enable others skilled in the art

bar or round, D, a short distance below the top cross-bar, C; and in the center of said round is connected the end of a brace, G, by means of a strap-hinge, *d*. The outer ends of the braces G G are placed between two cleats, *h h*, on the under side of the table-top, near the center, and fastened by means of a pin, *i*, passing through them, as shown, thereby holding and bracing the legs in proper position to support the table. When the legs are folded against the under side of the table-top the braces G G also lie against the same, between the cleats *h h*, and are fastened by the pin *i* passing through, each brace being provided with two holes, *x x*, at suitable points for the

Figure 1: Example USPTO scan of a patent (number 159,227)

Google Patents OCR	**Our New OCR**
iNITED,	UNITED STATES PATENT OFFICE.
’ Bnl-uns n. srNoLAin, on, ourense, immers.	HEMAN B. SINCLAIR, OF CHICAGO, ILLINOIS.
inn-Psovsmsnr ns Forense tastes.	IMPROVEMENT IN FOLDING TABLES.
Specification formingpart of Letters Patent	Specification forming part of Letters Patent
No. 159,227. dated January 26, li875;	No. 159,227, dated January 26, 1875
application-filed October 21, 1874. f f	application filed October 21, 1874.
runstnnctionjindrrangement ofraQldiUE. tn-	To all whom it may concern:
	Be it known that I, HEMAN B. SINCLAIR,
	of the city of Chicago, in the county of Cook
	and State of Illinois, have invented certain
	new and useful Improvements in Folding Ta-
	bles; and I do hereby declare that the fol-
	lowing is a full, clear, and exact description
	thereof, reference being had to the accompa-
	nying drawings and to the letters of reference
	marked thereon, which form part of this speci-
	fication.
	The nature of my invention consists in the
	construction.and arrangement of a folding ta-
ble, as-will be hereinafter more fully set forth.	ble, as-will be hereinafter more fully set forth.

Figure 2: Comparison of OCR rendering of patent number 159,227



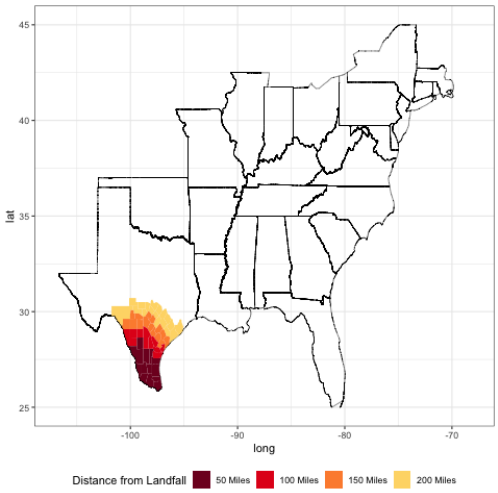
(A) 1835–1839 **(B)** 1860–1869

(C) 1890–1899 **(D)** 1920–1929

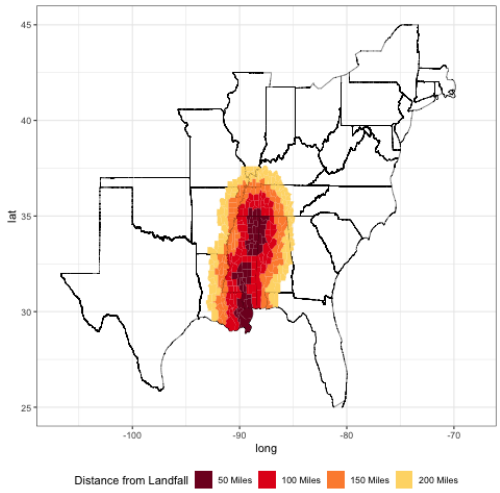
(E) 1950–1959 **(F)** 2010–2019

Figure 3: Locations of inventors

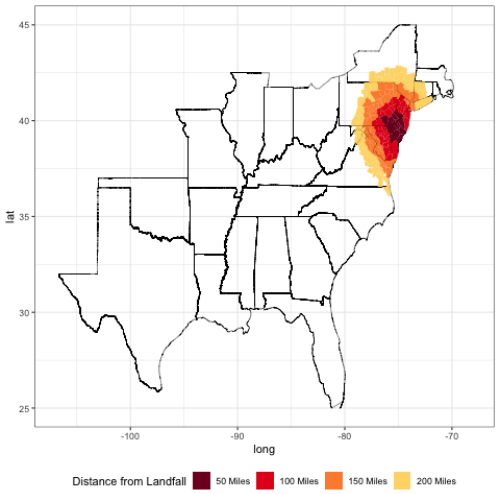
This figure shows the location of inventors who have patents granted during different periods over the last two centuries. Inventor locations are aggregated to counties that are then sorted into deciles in each period based on the number of inventors residing in each county. The darker the shade, the greater the number of active inventors in a county.



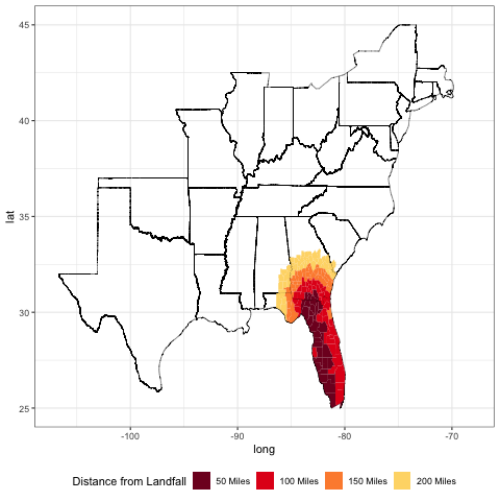
(A) 1999 Bret



(B) 2005 Katrina



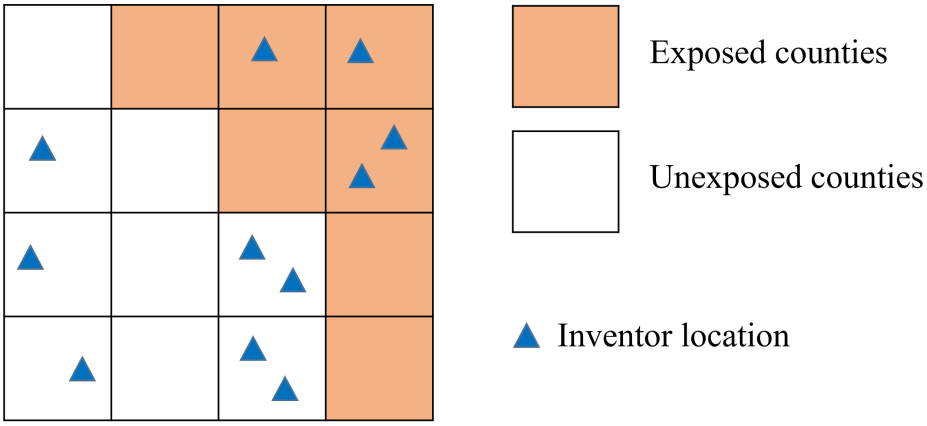
(C) 2012 Sandy



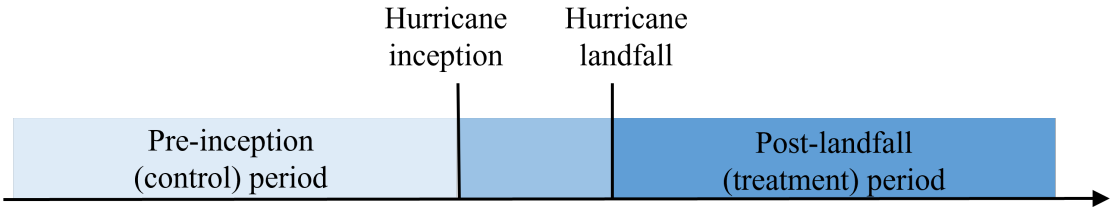
(D) 2017 Irma

Figure 4: Counties within a hurricane landfall region

This figure shows the counties within 50, 100, 150, and 200 miles of the hurricane eye for four hurricanes in our sample from 1851-2023.



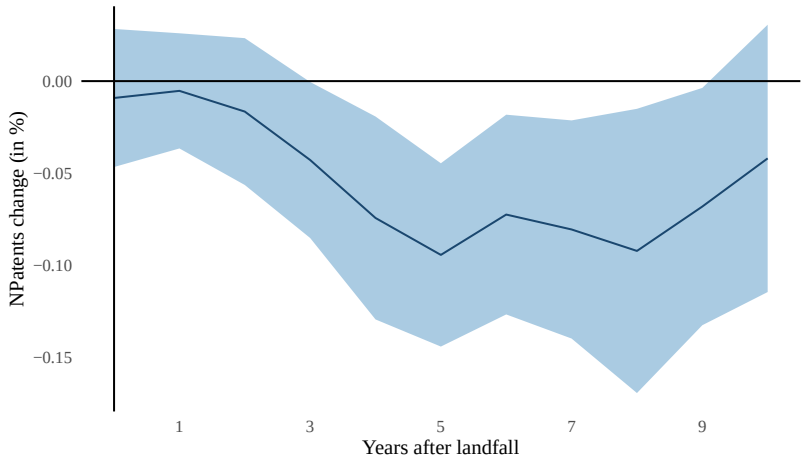
(A) Stylized example of spatial exposure (across variation)



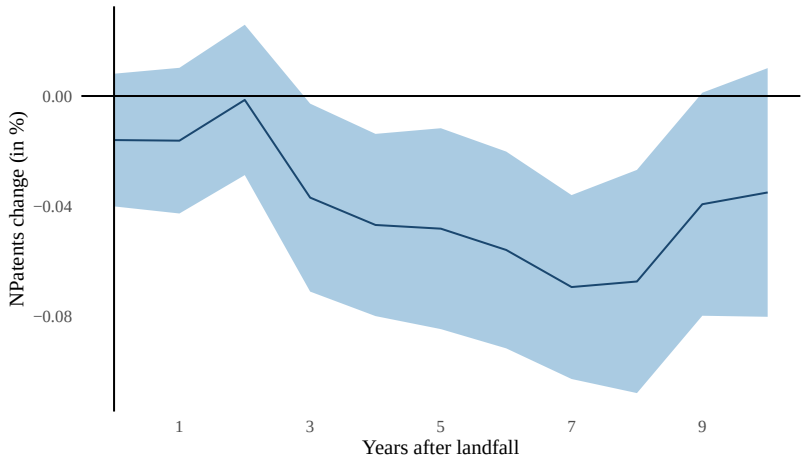
(B) The timeline of a hurricane (within variation)

Figure 5: Identification strategy

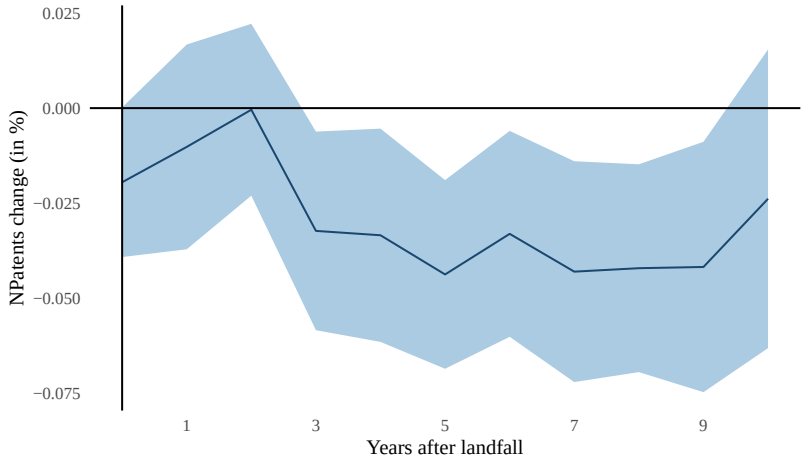
Panel A illustrates a stylized example of inventor locations and county exposure to a hurricane landfall region. Panel B illustrates the timeline of a hurricane.



(A) Counties within 50-mile radius



(B) Counties within 100-mile radius



(C) Counties within 200-mile radius

Figure 6: Coefficients from the baseline local projection estimation

This figure presents the β_1 estimate and its 95% confidence interval from estimating the regression given in equation (1) for horizons $h = 0, 1, 2, \ldots, 9, 10$. Panels A, B, C show the estimates with landfall regions determined at 50-, 100- and 200-miles, respectively, from the radius of the eye of a hurricane.

Table 1: Summary statistics

This table shows the summary statistics for the main variables used in the paper. For Panel A, the data are at the county-year level from 1851 to 2023. For Panel B, the data are at the patent class-year level from 1851 to 2014. R50 radius denotes that hurricane landfall regions are based on a 50-mile radius around the eye of the hurricane. All variables are described in Appendix Table B.3.

Panel A: County-year level

	Observations	Avg.	St. dev.	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
$NPatents_{c,t}$	164,383	23.615	100.239	1.000	1.000	4.000	11.000	40.000
$\Delta log(NPatents_{c,t})$	164,383	0.014	0.625	-0.693	-0.318	0.000	0.357	0.693
$Hit_{c,t}$ R50 miles	164,383	0.008	0.088	0.000	0.000	0.000	0.000	0.000
$Hit_{c,t}$ R100 miles	164,383	0.021	0.143	0.000	0.000	0.000	0.000	0.000
$Hit_{c,t}$ R200 miles	164,383	0.047	0.213	0.000	0.000	0.000	0.000	0.000
$IndependentShare_{c,t}$	162,633	0.652	0.331	0.172	0.338	0.724	1.000	1.000
$FirmShare_{c,t}$	162,633	0.335	0.327	0.000	0.000	0.250	0.635	0.815

Panel B: Patent class-year level

	Observations	Avg.	St. dev.	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
$NPatents_{k,t}$	41,630	138.196	202.113	21.000	46.000	87.000	163.000	284.000
$log(NPatents_{k,t})$	41,630	4.400	1.073	3.045	3.829	4.466	5.094	5.649
$\Delta log(NPatents_{k,t})$	41,630	0.006	0.330	-0.368	-0.175	0.000	0.190	0.380
$ClassShareHit_{k,t}$ R50 miles	41,630	0.003	0.015	0.000	0.000	0.000	0.000	0.000
$ClassShareHit_{k,t}$ R100 miles	41,630	0.009	0.042	0.000	0.000	0.000	0.000	0.005
$ClassShareHit_{k,t}$ R200 miles	41,630	0.016	0.061	0.000	0.000	0.000	0.000	0.026

Table 2: Baseline local effects - event study specification

This table presents results from estimating the regression specification given in equation (2). The dependent variable is $\log\left(\frac{\overline{NPatents}_{c,t+h}}{\text{NPatents}_{c,t-1}}\right)$. The main independent variable on interest is $Hit_{c,t}$. The data span from 1851 to 2023. Standard errors are clustered by year and county and shown in parentheses. The specifications include county and time fixed effects. The significance of each coefficient estimate is indicated by * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Landfall radius	50		100		200	
Years since hit	3	5	3	5	3	5
$Hit_{c,t}$	-0.084**	-0.086**	-0.069**	-0.067**	-0.056**	-0.061**
	(0.034)	(0.035)	(0.029)	(0.029)	(0.023)	(0.025)
$logNPatents_{c,t-1}$	-0.735***	-0.752***	-0.739***	-0.756***	-0.740***	-0.747***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)
$logNPatents_{c,t-2}$	0.182***	0.180***	0.183***	0.175***	0.175***	0.169***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
$logNPatents_{c,t-3}$	0.123***	0.128***	0.134***	0.132***	0.140***	0.137***
	(0.011)	(0.009)	(0.011)	(0.009)	(0.011)	(0.010)
$logNPatents_{c,t-4}$	0.132***	0.125***	0.115***	0.119***	0.133***	0.132***
	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
$logNPatents_{c,t-5}$	0.100***	0.103***	0.097***	0.101***	0.097***	0.098***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,287	16,911	18,297	17,999	17,816	17,574
R2	0.523	0.568	0.526	0.570	0.531	0.570

Table 3: Resilience of local innovation by firm share

This table presents results from estimating the regression specification given in equation (3). The dependent variable is $\log\left(\frac{\overline{NPatents}_{c,t+h}}{\text{NPatents}_{c,t-1}}\right)$. The interacted independent variable of interest is $Hit_{c,t} \times FirmShare_{c,t-1}$. $FirmShare$ is standardized. The data span from 1851 to 2023. Standard errors are clustered by year and county and shown in parentheses. The specifications include county and time fixed effects. The significance of each coefficient estimate is indicated by * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Landfall radius	50		100		200	
Years since hit	3	5	3	5	3	5
$Hit_{c,t}$	-0.070**	-0.078**	-0.066**	-0.068**	-0.055**	-0.043*
	(0.033)	(0.037)	(0.026)	(0.029)	(0.021)	(0.023)
$Hit_{c,t} \times FirmShare_{c,t-1}$	0.046	0.037	0.064***	0.043*	0.050***	0.037*
	(0.030)	(0.034)	(0.021)	(0.025)	(0.018)	(0.019)
$FirmShare_{c,t-1}$	0.046***	0.048**	0.042**	0.046**	0.053***	0.060***
	(0.017)	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)
$logNPatents_{c,t-1}$	-0.759***	-0.771***	-0.763***	-0.780***	-0.769***	-0.779***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)
$logNPatents_{c,t-2}$	0.166***	0.160***	0.165***	0.161***	0.162***	0.156***
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
$logNPatents_{c,t-3}$	0.115***	0.107***	0.116***	0.107***	0.117***	0.112***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
$logNPatents_{c,t-4}$	0.116***	0.113***	0.105***	0.112***	0.121***	0.124***
	(0.009)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)
$logNPatents_{c,t-5}$	0.090***	0.092***	0.091***	0.089***	0.093***	0.087***
	(0.008)	(0.007)	(0.009)	(0.008)	(0.009)	(0.008)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,061	13,852	15,739	14,435	15,438	14,192
R2	0.569	0.606	0.569	0.610	0.577	0.616

Table 4: Resilience of local innovation by independent inventor share

This table presents results from estimating the regression specification given in equation (3). The dependent variable is $\log\left(\frac{\overline{NPatents}_{c,t+h}}{\overline{\text{NPatents}}_{c,t-1}}\right)$. The interacted independent variable of interest is $Hit_{c,t} \times IndShare_{c,t-1}$. $IndShare$ is standardized. The data span from 1851 to 2023. Standard errors are clustered by year and county and shown in parentheses. The specifications include county and time fixed effects. The significance of each coefficient estimate is indicated by * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Landfall radius	50		100		200	
Years since hit	3	5	3	5	3	5
$Hit_{c,t}$	-0.070**	-0.079**	-0.067**	-0.068**	-0.054**	-0.042*
	(0.033)	(0.037)	(0.027)	(0.029)	(0.021)	(0.024)
$Hit_{c,t} \times IndShare_{c,t-1}$	-0.052*	-0.044	-0.068***	-0.047*	-0.051***	-0.038**
	(0.029)	(0.033)	(0.021)	(0.025)	(0.018)	(0.019)
$IndShare_{c,t-1}$	-0.057***	-0.060***	-0.051***	-0.056***	-0.062***	-0.069***
	(0.017)	(0.019)	(0.018)	(0.020)	(0.018)	(0.020)
$logNPatents_{c,t-1}$	-0.760***	-0.772***	-0.764***	-0.780***	-0.769***	-0.780***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)
$logNPatents_{c,t-2}$	0.165***	0.159***	0.164***	0.160***	0.161***	0.155***
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
$logNPatents_{c,t-3}$	0.114***	0.106***	0.115***	0.106***	0.117***	0.112***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
$logNPatents_{c,t-4}$	0.115***	0.112***	0.105***	0.112***	0.120***	0.123***
	(0.009)	(0.010)	(0.011)	(0.010)	(0.009)	(0.010)
$logNPatents_{c,t-5}$	0.090***	0.092***	0.091***	0.089***	0.093***	0.087***
	(0.008)	(0.008)	(0.009)	(0.008)	(0.009)	(0.008)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	15,061	13,852	15,739	14,435	15,438	14,192
R2	0.570	0.606	0.569	0.610	0.578	0.616

Table 5: Substitution effects and resilience of aggregate innovation

This table presents results from estimating the regression specification given in equation (4). The dependent variable is $\log\left(\frac{\overline{NPatents}_{k,t+h}}{NPatents_{k,\ t-1}}\right)$. The main independent variable on interest is $Hit_{k,t}$. The data span from 1851 to 2014. Standard errors are clustered by year and USPC classification category and shown in parentheses. The specifications include classification and time fixed effects. The significance of each coefficient estimate is indicated by * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Landfall radius	50		100		200	
Years since hit	3	5	3	5	3	5
$Hit_{k,t}$	-0.267***	-0.306*	-0.142***	-0.149***	0.014	-0.002
	(0.054)	(0.125)	(0.030)	(0.032)	(0.035)	(0.037)
$logNPatents_{k,t-1}$	-0.624***	-0.634***	-0.612***	-0.623***	-0.617***	-0.642***
	(0.110)	(0.105)	(0.102)	(0.097)	(0.047)	(0.049)
$logNPatents_{k,t-2}$	0.153***	0.146**	0.149***	0.142***	0.124**	0.138***
	(0.036)	(0.045)	(0.028)	(0.039)	(0.050)	(0.044)
$logNPatents_{k,t-3}$	0.109	0.150*	0.088	0.127*	0.110***	0.129***
	(0.057)	(0.068)	(0.056)	(0.064)	(0.032)	(0.032)
$logNPatents_{k,t-4}$	0.134*	0.138*	0.140*	0.144**	0.130***	0.145***
	(0.067)	(0.059)	(0.066)	(0.059)	(0.035)	(0.034)
$logNPatents_{k,t-5}$	0.095***	0.059	0.111**	0.077*	0.116*	0.082
	(0.025)	(0.032)	(0.032)	(0.039)	(0.055)	(0.053)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Classification FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,878	1,879	1,880	1,881	3,073	3,074
R2	0.527	0.572	0.550	0.591	0.554	0.593

Appendix A Constructing the patents database

We extract information from pre-1976 patents in three steps. First, we identify the title page of each patent document. Second, we perform OCR (optimal character recognition) on the title page of each patent to obtain the text of each title page. Third, we use ChatGPT to extract inventor and assignee names and locations. We manually checked the accuracy of the extracted data of a random sample of 500 patents by comparing them to the patent documents. Only one patent had a minor discrepancy, indicating an error rate of 0.2%.

Combining these two technologies, we construct a new database of patents.[5] Our process is as follows:

1. Identify the title-page in each patent document. Beginning in the early 1970s, the first page of a patent document is the title page and contains relevant biographical and other information. Prior to that, the initial pages of a patent were often one or more pages of drawings, so the first page of text appears later.
2. Apply state-of-the-art OCR on the relevant PDF pages. We extract the layout of the document, making it possible to, for example, read columns of text in the correct order.
3. Use ChatGPT to extract information from the OCR text. To date, we have extracted inventor and assignee names and locations. We do this using a fine-tuned version of the GPT model and use OpenAI's Python API to submit millions of batch requests.

Consider again the patent shown in Figure 1; the OCR output we obtain is shown in the right column of Table 2. The quality improvement over the Google Patents version of the text is readily apparent; this rendering perfectly matches the original patent document.

Next, we use ChatGPT to exactract the relevant information from the enhanced text. In particular, we use automated queries with questions about the text, as shown here along with ChatGPT's corresponding answers:

Input question	**ChatGPT response**
1) What is the title of this patent document?	IMPROVEMENT IN FOLDING TABLES
2) How many inventors are listed?	1
3) What is the name of each inventor?	Heman B. Sinclair
4) What city/state is each inventor from?	Chicago, IL
5) If the text mentions an "assignor" or "assignee," to what person or company was it assigned?	null
6) And if it was assigned, list any corresponding cities.	null

[5]To date, we have extracted information only for the title page of patent documents. We are in the process of securing funds to allow the creation of a complete and highly accurate historical database.

Given the improved quality of the OCR, one might wonder whether ChatGPT provides a significant benefit over simpler text analysis approaches. Determining the name and city/state of the inventor from this high-quality OCR text may appear to be quite straightforward using simple regular expression pattern matching. But ChatGPT can seamlessly handle far more complicated situations involving multiple inventors from different cities, multiple assignees, significant changes in formatting, and other edge cases that are hard to anticipate across millions of patents.

Our results were obtained using a recent ChatGPT model, `gpt-4o-mini`, which OpenAI describes as an "affordable and intelligent small model for fast, lightweight tasks." Out-of-the-box, this model performed quite well on our questions. We are able to elicit even higher-quality responses over a range of input types with two approaches. First, we adopt recent advances in prompt engineering. For example, we instruct the model to "take it step by step" before answering. Despite its apparent simplicity, this instruction has been shown to provide a significant improvement of an LLM's ability to "reason" through certain types of questions (Kojima, Gu, Reid, Matsuo, and Iwasawa, 2024). This approach is especially useful in patents with multiple inventors; in these patents we see an improvement in ChatGPT's answers once we require it *first* to count how many inventors there are, and *then* to name them and identify their locations.

Second, we further improve the output by "fine-tuning" the model to our particular needs. This is done by providing the model with additional training examples of questions along with our desired output; the model then learns to adjust its output to match the target. Fine-tuning alters the learning environment from "zero-shot" to "few-shot" by showing the LLM what an appropriate response looks like. After providing 100 fine-tuning examples, we verify that our model performs exceptionally well. We manually checked the accuracy of the extracted inventor and assignee names and locations of a random sample of 500 patents by comparing them to the patent documents. Only one patent had a minor discrepancy, indicating an error rate of 0.2%, which can likely be reduced further with additional fine-tuning of the model.

Appendix B Additional tables

Table B.1: Baseline local effects - local projection estimation

This table presents results from estimating the regression specification given in equation (1) for horizons $h = 0, 1, 2, \ldots, 9, 10$. Panels A, B, C show the estimates with landfall regions determined at 50-, 100- and 200-miles, respectively, from the radius of the eye of a hurricane. The data span from 1851 to 2023. Standard errors are clustered by year and county and shown in parentheses. The specifications include county and time fixed effects. The significance of each coefficient estimate is indicated by * for $p < 0.10$, ** for $p < 0.05$, and *** for $p < 0.01$.

Panel A: Country treatment within 50-mile radius

Years since hit	0	1	2	3	4	5	6	7	8	9	10
$Hit_{c,t}$	-0.009	-0.005	-0.017	-0.043**	-0.074***	-0.094***	-0.072***	-0.081***	-0.092**	-0.068**	-0.042
	(0.019)	(0.016)	(0.020)	(0.022)	(0.028)	(0.025)	(0.028)	(0.030)	(0.039)	(0.033)	(0.037)
$NPatents_{c,t-1}$	-0.677***	-0.700***	-0.721***	-0.730***	-0.737***	-0.750***	-0.749***	-0.765***	-0.760***	-0.771***	-0.774***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
$NPatents_{c,t-2}$	0.206***	0.190***	0.190***	0.185***	0.174***	0.178***	0.169***	0.172***	0.160***	0.159***	0.155***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-3}$	0.138***	0.143***	0.141***	0.134***	0.136***	0.130***	0.131***	0.126***	0.124***	0.122***	0.114***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-4}$	0.112***	0.120***	0.113***	0.118***	0.111***	0.116***	0.106***	0.110***	0.105***	0.098***	0.103***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-5}$	0.109***	0.109***	0.118***	0.113***	0.115***	0.109***	0.106***	0.104***	0.102***	0.104***	0.096***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	80,813	79,263	77,797	76,288	74,993	73,589	72,388	71,197	69,936	68,531	67,264
R2	0.349	0.368	0.387	0.403	0.413	0.421	0.422	0.432	0.433	0.442	0.452

Panel B: Country treatment within 100-mile radius

Years since hit	0	1	2	3	4	5	6	7	8	9	10
$Hit_{c,t}$	-0.016	-0.016	-0.001	-0.037**	-0.047***	-0.048**	-0.056***	-0.069***	-0.067***	-0.039*	-0.035
	(0.012)	(0.013)	(0.014)	(0.017)	(0.017)	(0.019)	(0.018)	(0.017)	(0.021)	(0.021)	(0.023)
$NPatents_{c,t-1}$	-0.678***	-0.705***	-0.724***	-0.734***	-0.741***	-0.752***	-0.755***	-0.767***	-0.764***	-0.774***	-0.778***
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)
$NPatents_{c,t-2}$	0.203***	0.190***	0.187***	0.183***	0.173***	0.174***	0.169***	0.170***	0.160***	0.158***	0.153***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-3}$	0.138***	0.142***	0.143***	0.136***	0.134***	0.131***	0.131***	0.128***	0.124***	0.120***	0.115***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-4}$	0.112***	0.119***	0.114***	0.116***	0.113***	0.116***	0.107***	0.109***	0.105***	0.100***	0.103***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
$NPatents_{c,t-5}$	0.109***	0.111***	0.117***	0.116***	0.117***	0.109***	0.108***	0.105***	0.101***	0.102***	0.096***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	92,189	90,415	88,710	86,985	85,477	83,918	82,510	81,154	79,764	78,213	76,766
R2	0.348	0.369	0.387	0.403	0.413	0.419	0.422	0.430	0.432	0.440	0.450

Panel C: Country treatment within 200-mile radius

Years since hit	0	1	2	3	4	5	6	7	8	9	10
$Hit_{c,t}$	-0.019*	-0.010	-0.000	-0.032**	-0.034**	-0.044***	-0.033**	-0.043***	-0.042***	-0.042**	-0.024
	(0.010)	(0.014)	(0.011)	(0.013)	(0.014)	(0.013)	(0.014)	(0.015)	(0.014)	(0.017)	(0.020)
$NPatents_{c,t-1}$	-0.679***	-0.705***	-0.724***	-0.735***	-0.743***	-0.751***	-0.756***	-0.766***	-0.763***	-0.774***	-0.777***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
$NPatents_{c,t-2}$	0.202***	0.192***	0.187***	0.183***	0.175***	0.173***	0.169***	0.170***	0.160***	0.158***	0.153***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$NPatents_{c,t-3}$	0.139***	0.141***	0.142***	0.138***	0.134***	0.131***	0.133***	0.127***	0.123***	0.119***	0.116***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
$NPatents_{c,t-4}$	0.113***	0.119***	0.115***	0.115***	0.113***	0.117***	0.108***	0.110***	0.105***	0.101***	0.103***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.004)
$NPatents_{c,t-5}$	0.109***	0.112***	0.117***	0.116***	0.118***	0.108***	0.106***	0.103***	0.101***	0.101***	0.093***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	100,938	98,997	97,075	95,156	93,443	91,772	90,203	88,688	87,185	85,533	83,818
R2	0.348	0.369	0.386	0.403	0.414	0.418	0.421	0.428	0.430	0.439	0.448

Table B.2: List of deadliest hurricanes

This table presents the list of deadliest Atlantic and Gulf Coast hurricanes included in our sample based on data from NOAA. The first column indicates the main states of impact and/or, from 1954, the hurricane names.

Hurricane	Year	Fatalities
LA (Last Island)	1856	400
LA	1860	47
TX	1875	176
New England	1878	27
NC, VA	1879	46
GA/SC	1881	700
NC	1883	53
TX (Indianola)	1886	150
Texas	1886	27
Mid-Atlantic	1889	40
LA (Cheniere Caminanda)	1893	1100-1400
SC/GA (Sea Islands)	1893	1000-2000
SC, FL	1893	28
FL, GA, SC	1896	130
GA, SC, NC	1898	179
NC, SC	1899	50
TX (Galveston)	1900	8000
SE FL	1906	164
MS/AL/Pensacola	1906	134
LA (Grand Isle)	1909	350
TX (Velasco)	1909	41
SW FL	1910	30
LA (New Orleans)	1915	275
TX (Galveston)	1915	275
SW LA/Upper TX	1918	34
FL (Keys)/S TX	1919	287
FL (Miami)/MS/AL/Pensacola	1926	372
LA	1926	25
FL (SE/Lake Okeechobee)	1928	2500
TX (Freeport)	1932	40
S TX	1933	40
FL (Keys)	1935	408
New England	1938	256
GA/SC/NC	1940	50
Northeast U.S.	1944	64
SE FL/SE LA/MS	1947	51
Hazel (SC/NC)	1954	95
Carol (NE U.S.)	1954	60
Diane (NE U.S.)	1955	184
Connie (NC)	1955	25
Audrey (SW LA/N TX)	1957	416
Donna (FL/Eastern U.S.)	1960	50
Carla (N & Central TX)	1961	46
Hilda (LA)	1964	38
Betsy (SE FL/SE LA)	1965	75
Camille (MS/SE LA/VA)	1969	256
Agnes (FL/NE U.S.)	1972	122
Andrew (S FL, LA)	1992	26
Alberto (NW FL, GA, AL)	1994	30

Continued on the next page.

Table B.2: List of deadliest hurricanes (continued)

Hurricane	Year	Fatalities
Fran (NC)	1996	26
Floyd (Mid Atlantic & NE U.S.)	1999	56
Allison (SE TX)	2001	41
Ivan (NW FL, AL)	2004	25
Katrina (SE LA/MS)	2005	1200
Irene	2011	48
Sandy	2012	160
Matthew	2016	47
Harvey	2017	106
Irma	2017	96
Florence	2018	54
Michael	2018	59
Laura	2020	41
Ida	2021	92
Ian	2022	156

Table B.3: Variable definitions

This table presents definitions of the main variables. The first column gives the variable name. The second column includes a short description.

Variable Name	Description	Source
$Hit_{c,t}$	This variable captures if a county c is in the landfall region of a hurricane in year t.	NOAA, Census
$NPatents_{c,t}$	This variable counts the number of patents issued in year t for which all inventors reside in county c.	USPTO, Census
$FirmShare_{c,t}$	This variable captures the share of NPatents in year t and county c that have firms as assignees.	USPTO, Census
$IndependentShare_{c,t}$	This variable captures the share of NPatents in year t and county c that have no institution as assignees.	USPTO, Census
$CountyClassShare_{c,k,t}$	This variable measures how much of the patents of a patent class (based on the USPC system) come from county c five years before year t.	USPTO, Census
$ClassShareHit_{k,t}$	This variable sums $CountyClassShare_{c,k,t}$ across all counties that are hit by a hurricane in year t.	NOAA, USPTO, Census
$Hit_{k,t}$	This is an indicator variable that takes the value one if the variable $ClassShareHit_{k,t}$ is at least 0.25 and zero otherwise.	NOAA, USPTO, Census