St. John's University

The Peter J. Tobin College of Business

School of Risk Management, Insurance, and Actuarial Science

Race Discrimination in the Adjudication of Claims: Evidence from Earthquake Insurance

Mark J. Browne
Annette Hofmann
Xiao (Joyce) Lin

SRM Working Paper Series
No. 4

May 2020

School of Risk Management, Insurance, and Actuarial Science
101 Astor Place | New York, NY 10003

Race Discrimination in the Adjudication of Claims: Evidence from Earthquake Insurance*

Mark J. Browne † Annette Hofmann ‡ Xiao (Joyce) Lin §

May 29, 2020

Abstract

Catastrophic events affect many people simultaneously. We exploit this claim characteristic to test for evidence of racial discrimination in the adjudication of insurance claims. There were eight earthquakes in Oklahoma between 2010 and 2016, mostly linked to oil and gas drilling activities (fracking). Using data from the Oklahoma department of insurance, the U.S. Geological Survey, and the U.S. census, this article tests whether claim resolutions differ among zip-code areas with different racial compositions, all else equal. Controlling for policy-level and zip-code level characteristics, we find evidence that claims from areas with higher percentages of Black population were less likely to result in payment, and when those claims did get paid, payments were lower in those areas. We further investigate the mechanisms through which such discrimination may exist. Although we cannot rule out taste-based discrimination, or existence of supply-side (insurer) prejudice, we find evidence that the tendency to file questionable claims is higher in zip-codes with higher Black population, suggesting statistical discrimination by insurers as a response to this filing behavior.

JEL Classification: C13; G22; H12.

Keywords: Discrimination; catastrophic risk; insurance claims adjudication; earthquake insurance.

*__Acknowledgments:__ We would like to thank Olga Kanj, Ty Leverty, and participants of the St. John's Brown Bag seminar, the 2019 ARIA Annual Meeting in San Francisco, and the 2019 EGRIE Annual Meeting in Rome.

†School of Risk Management, The Peter J. Tobin College of Business, St. John's University, 101 Astor Place, New York, New York 10003, USA. *brownem1@stjohns.edu*

‡School of Risk Management, The Peter J. Tobin College of Business, St. John's University, 101 Astor Place, New York, New York 10003, USA. *hofmanna@stjohns.edu*

§Corresponding author. School of Risk Management, The Peter J. Tobin College of Business, St. John's University, 101 Astor Place, New York, New York 10003, USA. *linx@stjohns.edu*

1 Introduction

Racial discrimination is a topic of interest to economists not only because of the social justice implications but because it suggests the possibility of market failure. Since taste-based discrimination, which stems from an animus toward a racial group, is costly for businesses, competition should preclude it (Becker (1957)).[1] Statistical discrimination, in contrast, could exist in competitive markets. Statistical discrimination stems from decision-makers using observable characteristics (for example, race or gender) to proxy for economically relevant but unobservable information (Phelps (1972); Arrow (1973)). Statistical discrimination is based on an expected difference in group means or an expected difference in the variance or the reliability of a proxy signal (Aigner and Cain (1977)). Accurate statistical discrimination can be economically efficient. Although statistical discrimination is "fair" at the margin; it is inequitable on average and it poses an equity-efficiency tradeoff. Both taste-based and statistical discrimination can result in an anticipation of future discrimination. This can lead to behavioral changes in the group whose members perceive themselves to be discriminated against, raising additional equity concerns (Loury (1992); Coate and Loury (1993)). The self-fulfilling pattern of discrimination also applies to second-order statistical discrimination (Klumpp and Su (2013)). The debates among economists on the theory, empirical evidence, and policy implications of discrimination have been at times contentious (Arrow (1998); Heckman (1998)).

Prior studies on discrimination have mainly focused on the labor market (Ashenfelter and Rees (1973); Cain (1986)) and the credit market (Munnell et al. (1996); Harrison (1998); Blanchflower et al. (2003); Pope and Sydnor (2011)). A widely used approach for testing for discrimination is the audit approach, which identifies discrimination in a controlled setting or in a field experiment (Bertrand and Mullainathan (2004); Ayres and Siegelman (1995); Castillo et al. (2013)). Evidence of discrimination is often interpreted with caution, as it is hard to disentangle taste-based from statistical discrimination,[2] except in a few limited settings.[3].

Despite the large body of prior studies, research on racial and other forms of demographic discrimination in the insurance industry has been somewhat sparse. Testing for discrimination in insurance markets is

[1] In the case of labor markets, compared to employer discrimination, co-worker- or consumer-based discrimination may be harder to be competed away, Becker (1993) asserted in his Nobel Lecture "of greater significance empirically is the long run discrimination by employees and customers".

[2] Heckman (1998) argues that audit studies can both identify taste-based discrimination when it does not exist and fail to recognize it when it does.

[3] See, for instance, Castillo et al. (2013) as well as Stephens-Davidowitz (2014)

challenging, because risk underwriting is partly built on statistical discrimination.[4] Tests for discrimination in the pricing of insurance do not report evidence of racial discrimination (Harrington and Niehaus (1998); Klein and Grace (2001)). These findings are consistent with the hypothesis that competitive personal lines insurance markets preclude taste-based discrimination (Klein (1989); Cummins and Tennyson (1992)).

The current paper tests for evidence of racial discrimination in the adjudication of earthquake insurance claims in Oklahoma, following earthquakes arising from oil and gas drilling activities known as hydraulic fracturing. We find that controlling for the earthquake event, distance to epicenter, policy-level differences, and other zip-code level differences, insurers are less likely to pay claims from zip-codes with higher percentages of Black population. Conditional on paid claims, we also find that payment is lower for those filed from zip codes with higher percentages of Black population. We do not find evidence that this discrimination is self-fulfilling. Quite the opposite, among those outside the shake areas of an earthquake, zip-codes with higher percentages of Black population are found to file earthquake claims at a higher rate. These findings suggest that claims adjusters could be engaging in statistical discrimination, though taste-based discrimination cannot be ruled out. Even if statistical discrimination may be efficient at the margin, discriminatory claims adjudication practices are not fair on an individual basis for a claimant who suffered an insured loss.

We contribute to the economic literature on discrimination by offering two unique perspectives. First, we consider the settlement of insurance claims, rather than the pricing and sale of insurance. Unlike at the point of sale, the pressures of a competitive marketplace - manifest through multiple insurers competing on price - do not come to bear on the immediate outcome; the claim settlement process is conducted between a single insurer and an insured. The outcome is influenced by the legal environment in which the claim is resolved. The studies closest to our paper - Doerpinghaus et al. (2003) and Doerpinghaus et al. (2008) - consider auto liability claims settlement practices, in contrast to the prior referenced studies on insurance pricing and availability. In both studies, the authors used the 1997 Insurance Research Council Closed Claim Survey to test for gender and age discrimination. They report evidence of differences in assignment of fault in bodily injury automobile claims between genders and across age groups. They find, while controlling for the degree of actual fault, that claims adjusters assign greater degrees of fault in automobile accidents to female, young and elderly drivers, and that female drivers receive lower payment than their male counterparts. They attribute their findings to claims adjusters' perceived differences in

[4] Gender, for instance, is a risk factor that may be used as an underwriting criteria for automobile insurance in some jurisdictions but not in others.

risk aversion, negotiation costs, and possible discrimination. Different from Doerpinghaus et al. (2003) and Doerpinghaus et al. (2008), we test for discrimination using zip-code level proxies.

Second, we evaluate claims resulting from a cause of loss, an earthquake, that affects many people at the same time. Unlike with audit studies, we do not need to control for a variety of individual differences that may be confounded with the variable of interest in a discrimination test. We do, however, need to control for the intensity of a natural process - earthquake. In a study related to this paper, Baker and McElrath (1996) use quantitative and qualitative data collected after Hurricane Andrew, to show that a statistically significant bias in the insurance claims process exists. They find ethnic differences in the timing of insurance payments. In this paper, we control for different earthquake events and the distance to epicenter for each zip-code to identify whether and how the race composition of a zip-code is related to claims adjudication.

Lastly, by focusing on earthquake events and investigating the claims resolution process, we shed light on the level of equality in the process of compensating for harm after disastrous events. Institutional inequality has a negative impact on the recovery ability of a society (Anbarci et al. (2005)). The implications from discrimination in insurance claims adjudication after earthquakes, whether statistical or taste-based, could be far-reaching.

The paper is organized as follows. The next section discusses the earthquake peril in Oklahoma. Section 3 presents our theory. Section 4 details the datasets used in this paper. Section 5 discusses our empirical approach, and Section 6 presents our empirical results. Section 7 concludes the paper with a summary of our findings.

2 Background

Oklahoma was not a state prone to earthquakes prior to the shale gas boom of the last two decades.[5] As hydraulic fracturing has increased, researchers found that it changes the natural stress between faults and triggers shallow earthquakes, which can be large enough to cause damage (Rubinstein and Mahani (2015); Liu et al. (2017)). The number of magnitude 3.0 and higher earthquakes in Oklahoma has grown exponentially since the first one in 2007. There were 20 in 2009, 579 in 2014, and 903 in 2015

[5] The United States has experienced an extraordinary shale gas boom in the past decade. The Energy Information Administration reports that the annual U.S. natural gas withdrawals from shale gas has grown from 1.99 trillion cubic feet in 2007 to 16.58 trillion cubic feet in 2016, representing more than half of the total U.S. natural gas production in 2016, up from about 8% in 2007.

(Oklahoma Geological Survey). Though very few of these earthquakes were damaging, the sheer frequency of earthquakes is notable.

Oklahoman's exposure to earthquake risk has led to their purchase of earthquake insurance. Because a homeowners insurance policy does not cover losses from an earthquake, residents who want this coverage have to purchase an additional earthquake endorsement or a separate policy. The take-up rate on earthquake coverage among Oklahoman residents has increased from 2% in 2011 to 15% in 2015.[6] This is quite significant, considering that only about 10% of Californian homeowners have earthquake insurance (Lin (2016)). Earthquake insurance deductibles are a percentage of the insured value of a home and not the usual $200 to $2000 deductibles found in homeowners insurance. The lowest deductible plans are set at 2 percent, with premiums from about 30 cents to 71 cents per $1,000 in coverage. A high deductible of 25 percent has premiums that range from 22 cents to 63 cents per $1,000 in coverage.

Earthquake insurance coverage in Oklahoma has led to disputes between insurers and insureds. Consumers have complained that insurers denied earthquake claims based on a "man-made earthquake exclusion". In March 2015, the Oklahoma Insurance Commissioner issued a Bulletin cautioning earthquake insurers that his department would "take appropriate action to enforce the law" if insurers were found "denying claims based on the unsupported belief that these earthquakes were the result of fracking or injection well activity". He cited an "extraordinary denial rate of the earthquake claims" based on a preliminary 2014 data, indicating only 8 out of about 100 claims were paid in that year.[7]

As scientific evidence became overwhelming, state regulators finally agreed that these earthquake phenomena are probably linked to unconventional oil and gas drilling activity, or hydraulic fracturing. The Oklahoma Insurance Commissioner issued another Bulletin in October 2015, asking insurers to "clarify the issue of coverage by furnishing written notice to insureds and producers", acknowledging that "the majority of the numerous quakes experienced by Oklahomans are, more than likely, the result of waste water injection into disposal wells". The majority of insurers ended up clarifying that they do cover natural *and* man-made earthquakes. According to the Oklahoma Insurance Department, more than 90 percent of the insurance market covers man-made and natural earthquakes. Even with the "man-made earthquake exclusion" out of the way, a media report in March 2017 still claims "3 in 20 claims approved in Oklahoma since 2010", and that the denial rate is even higher for claims filed following the 2016 Pawnee

[6] Property Insurance Law Observer. Retrieved from `http://propertyinsurancelawobserver.com/2015/04/02/oklahoma-insurance-commissioner-dont-deny-earthquake-claims-as-man-made-by-linking-them-to-fracking/`

[7] Oklahoma Insurance Department: Earthquake Insurance Bulletin No. PC 2015-02.

earthquake than for those filed following the 2011 Prague earthquake.[8] To date, public dissatisfaction with insurers' claims handling persists.

3 Theory

Since an insurance claim can ultimately result in a civil action for resolution, expectations of insureds and insurers on how they will fare in court will affect their perceptions of the value of a claim. Racial bias in the legal system could, therefore, influence the value that both the insurer and the claimant assign to a claim.[9] We consider each in turn below.

(1) The Insured

An insured, having suffered what is believed to be an insured loss, determines whether or not to initiate a claim. The perceived value of the claim and the expected cost of pursuing it, both in terms of time and money, influence this decision. In some cases, a claim is relatively quickly and easily evaluated and processed by an insurer with an outcome with which the insured is satisfied. In other situations, just the opposite is the case; disagreement over the value of a claim results in protracted legal action. Since both parties recognize the possibility that a claim will result in legal action; negotiation over settlement occurs in the shadow of the potentially lengthy and costly process of litigation. The potential claimant's assessment of the value of the claim, therefore, depends on the perceived likelihood of prevailing in a legal action, should that become necessary, and the accompanying costs.

Browne and Puelz (1999) state that a plaintiff's decision to file a legal action is determined by its benefits and costs. The plaintiff will file a legal claim if the cost of litigation (C^p) is less than the subjectively perceived benefits. The benefits are the sum of the (perceived) probability of achieving an out-of-court settlement (π_{so}) times the value of the out-of-court settlement (S) and – in the case of a trial – the probability of prevailing at trial (π_t) times the value of the court award (A) in that case. Since each of these is unknown a priori, the plaintiff's individual perception of values weighs heavily in the decision to

8 See `http://www.tulsaworld.com/earthquakes/earthquake-insurance-in-claims-approved-in-oklahoma-since/articlede588725-1475-592c-9025-bdcfbf9b8bcd.html`.

9 Wright (2018) discusses how racial bias in jury composition can be seen in the ability of prosecutors and defense attorneys to block potential jurors without needing to give the court any reason for their exclusion. This is reflected in prosecutors use of options to remove nonwhite jurors who are statistically more likely to acquit from a jury. Wright (2018) refers to recently published research on juror removal in North Carolina conducted with colleagues at the Wake Forest University School of Law. This research shows with statewide evidence that such peremptory challenges "are indeed a vehicle for veiled racial bias that results in juries less sympathetic to defendants of color".

file a claim. As a consequence, a claim will be filed by the plaintiff whenever,

$$C^p < \pi^p_{so} \times S^p + \pi^p_t \times A^p, \tag{1}$$

where the superscript p indicates the plaintiff and $S^p \geq 0$ and $A^p \geq 0$ include the cases of zero settlement or zero court award, whenever the other one is nonzero.

(2) Insurer

When presented with a claim, an insurer has a duty to evaluate and potentially pay the claim in a timely fashion. In determining whether or not a claim should be paid, an insurer needs to investigate the circumstances of the claim. Important considerations include the cause of loss, time of loss, and place of loss as well as the terms of the insurance policy under which the claim was made. The amount of the insured loss may be clear to both the insured and the insurer. An example of this would be the destruction of a work of art insured with a stated value or the death of an insured covered by a life insurance policy. Alternatively, the amount of the loss may be subject to dispute. Examples include many liability losses as well as complex property losses with a business interruption component. Similarly, whether or not a loss actually occurred or is covered within the scope of the insuring agreement may be subject to dispute.

When payment of a loss or the amount to be paid is subject to dispute, an insurer's determination of its claim settlement offer will presumably take into consideration not only the loss characteristics and insurance policy terms, but also the perceived costs of claim management associated with offers to settle at different amounts. The perceived costs of claim management in turn depend on the perceived likelihood that a claimant will file a legal action and the perceived costs should a legal action be filed. The insurer's cost to settle the claim can be written:

$$C^i = f(\text{policy characteristics, loss characteristics, claim management costs}). \tag{2}$$

If the insurer does not pay the loss, the policyholder may hire an attorney and try to obtain an of out-of-court settlement or go to trial. Each of these options is associated with additional costs for the insurer. The insurer will settle the policyholder's claim if the cost of doing so is below his perceived cost of denying settlement, which means anticipating the likelihood with which the policyholder will use legal assistance that might lead to a legally enforced settlement either inside or outside of court. As a consequence, the insurer will pay the claim whenever the cost of paying the claim is below the perceived cost of policyholder action:

$$C^i < \pi^i_{so} \times S^i + \pi^i_t \times A^i, \tag{3}$$

where $S^i \geq 0$ and $A^i \geq 0$ include the cases of zero settlement or zero court award, whenever the other one is nonzero.

As can be seen from the above, in the case that the insured's and insurer's perceptions about chances and outcomes of out-of-court and court settlements are equal, the insured will file and the insurer will settle as long as their perceived costs are below the the legal cost associated with a dispute. However, in the case that the perceived probabilities and financial outcomes differ between policyholder and insurer, a valid claim may be denied or over-/underpaid depending on individual perceptions.

In summary, both the insured and the insurer, when considering actions they might take on a claim, evaluate the costs and benefits of their decisions through the filter of their own perceptions, including their subjective assessment of the legal system's likelihood of reaching a decision upholding their position on the matter.

We model claims resolution with the graphs below. In each, the x-axis represents the expected costs if a claim is disputed and the y-axis represents expected settlement.

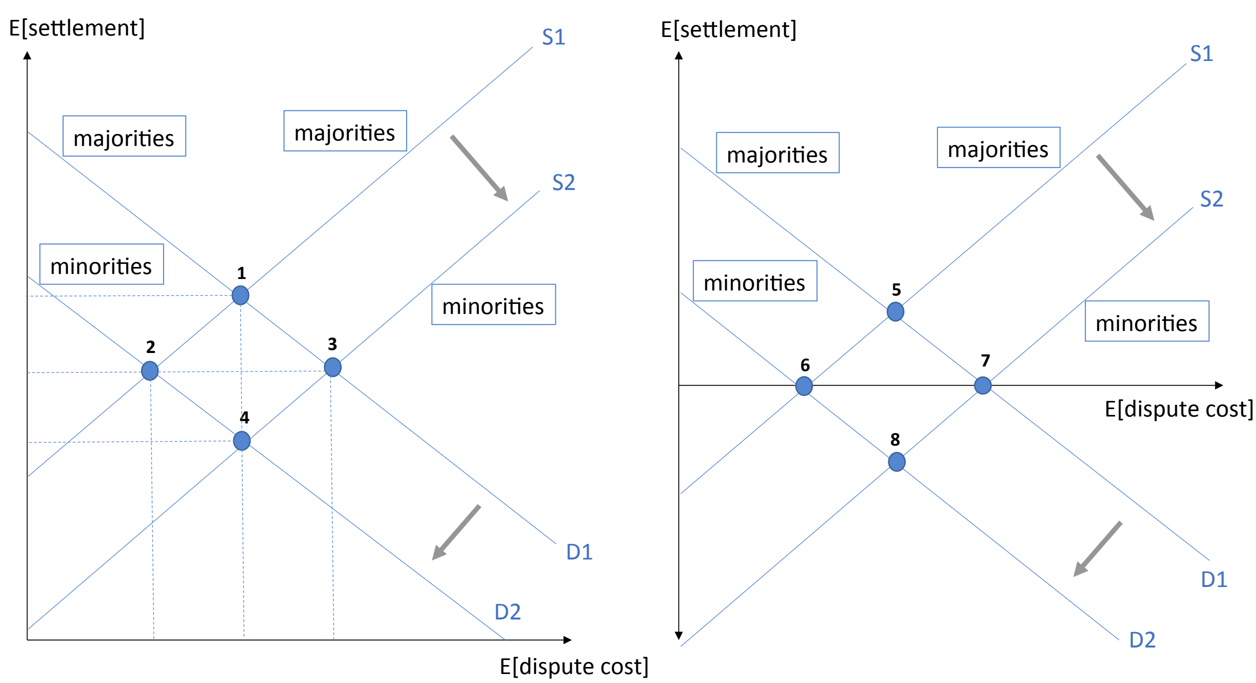

Figure 1 Schematic illustration of impact of minorities on dispute cost and settlement equilibria.

Consider a claim with a set of characteristics known to both the claimant and the insurer; for instance, a damaged home following an earthquake that has been inspected by both. The claimant's settlement demand is given by line D1, which shows the settlement demand is greater when the perceived dispute costs are lower. The insurer's settlement offer is given by line S1, which shows the settlement offer amount increasing as the perceived costs of dispute increase. Perceptions of the settlement amount and expected dispute costs of the claimant and insurer depend on their beliefs of the cost of prevailing in a possible legal action to resolve the claim. When a claimant feels that the legal system will be more favorable to his or her argument, the perceived costs of prevailing will be less and a larger settlement amount will be demanded. When the insurer perceives the courts to be more favorably disposed to its argument, then its perceived costs of prevailing will be less and its settlement offer will be less.

Claimants who demand less than the insurer's offer will accept the offer. Those with a higher demand than the insurer's offer will dispute the claim. All claims will eventually be resolved either through the parties reaching an agreement on the value of the claim or by a court deciding. The dispute costs incurred by each side will impact the final value of the claim, regardless of how it is resolved. The resolution of claims will inform claimants' and insurers' expectations when new claims arise.

Next, consider a minority population whose assessment of the judicial system leads it to perceive higher dispute costs than the majority population for similar claims. The expected settlement costs and expected dispute costs for this population are given by D2, which lies to the southwest of D1, indicating higher dispute costs per settlement dollar. If the insurer believes that in a dispute with a minority claimant it incurs less in costs per settlement dollar than in a comparable dispute with a majority claimant, the insurer's settlement offer line moves southeast to S2. We call the movement of D1 to D2 in Figure 1's left-hand-side panel "demand-side discrimination" or "self-fulfilling discrimination" and the movement of S1 to S2 "supply-side discrimination." With only demand-side discrimination, the point of equilibrium moves from 1 to 2. If there is only supply-side discrimination, the point moves from 1 to 3. With both demand-side and supply-side discrimination, the point moves from 1 to 4. We hypothesize that discrimination, either supply-side or demand-side, leads to lower settlement amounts on paid claims for a population that is discriminated against.

In Figure 1's right-hand-side panel the initial point of equilibrium is 5. Discrimination leads to points 6, 7, and 8, depending on the type of discrimination. Since settlements can not be negative, all three points represent settlements of $0. Therefore, we hypothesize that discrimination, either supply-side or demand-side, leads to a lower likelihood that a claim made by a member of a population that is discriminated against is paid.

4 Data and Summary Statistics

We obtained a recent dataset of individual earthquake insurance claims from the Oklahoma Insurance Department (OID). The OID has been collecting earthquake claims data on an ongoing basis since 2010 by surveying major insurers in the state. Our dataset includes policy-level claim information from 2010 through the end of 2017. We identify eight noticeable earthquake events between 2010 and 2017 that occured in Oklahoma, and we link earthquake insurance claims to those events based on each claim's date of loss.[10] We exclude "earthquake claims" that report a dollar deductible rather than a percentage deductible. This is because homeowners may file earthquake claims, not knowing that loss from earthquakes is excluded from their homeowners insurance, which carries a dollar deductible; but, is covered by a separate earthquake insurance policy or earthquake endorsement, which always carries a percentage deductible.

Table 1 Earthquake events and corresponding claims. Average claims payment is calculated conditional on a claim being paid.

EVENT	Number of Claims filed	Number of Claims Paid (%)	Average Claims Paid ($)	Largest Claim Paid ($)
10/13/2010	30	20 (66.67%)	14,696	113,370
11/5/2011	236	111 (47.03%)	13,318	204,543
4/16/2013	20	4 (20.00%)	7,854	11,860
12/7/2013	25	5 (24.00%)	6,600	23,582
7/27/2015	15	3 (16.67%)	13,869	21,435
12/29/2015	36	7 (19.44%)	10,716	31,206
9/3/2016	389	67 (17.48%)	10,733	146,298
11/6/2016	95	33 (35.79%)	19,231	165,765
Total	846	253 (29.91%)	13,215	204,543

In total we are able to match 846 earthquake claims to the eight earthquake events. [11] Table 1 summarizes the number of claims filed after each earthquake, along with the number of claims paid and the average claim payment if a claim was paid. Among all claims submitted, the majority are not paid. The one exception to this is the first earthquake event in 2010. Among all of the paid claims, the average claim payment is $13,215, and the single largest payment is $204,543. Two earthquakes - the 2011 Prague

[10] In most cases, the date of loss is the date of the earthquake. For some earthquakes, aftershocks occurred in the same location for a few more days; therefore, we include claims with a loss date within a few days of the main quake. If a main quake occurred close to midnight, then we include earthquake insurance claims filed within two days – before and after midnight.

[11] There are still about two hundred earthquake claims that we are unable to match to one of the eight identified earthquake events based on the date of loss information. We suspect these claims were filled after one of the hundreds of small earthquakes that occurred in Oklahoma during the study period. We drop these observations because these claims do not fit our definition of a "positively correlated" event and they are similar to homeowners or auto insurance claims.

earthquake and the 2016 Pawnee earthquake - account for the majority of earthquake claims filed and paid.

Figure 2 Earthquake insurance claims filed and paid. Upper left: number of earthquake claims filed following the 2011 Prague earthquake; lower left: number of earthquake claims paid following the 2011 Prague earthquake; upper right: number of earthquake claims filed following the 2016 Pawnee earthquake; lower right: number of earthquake claims paid following the 2016 Pawnee earthquake;

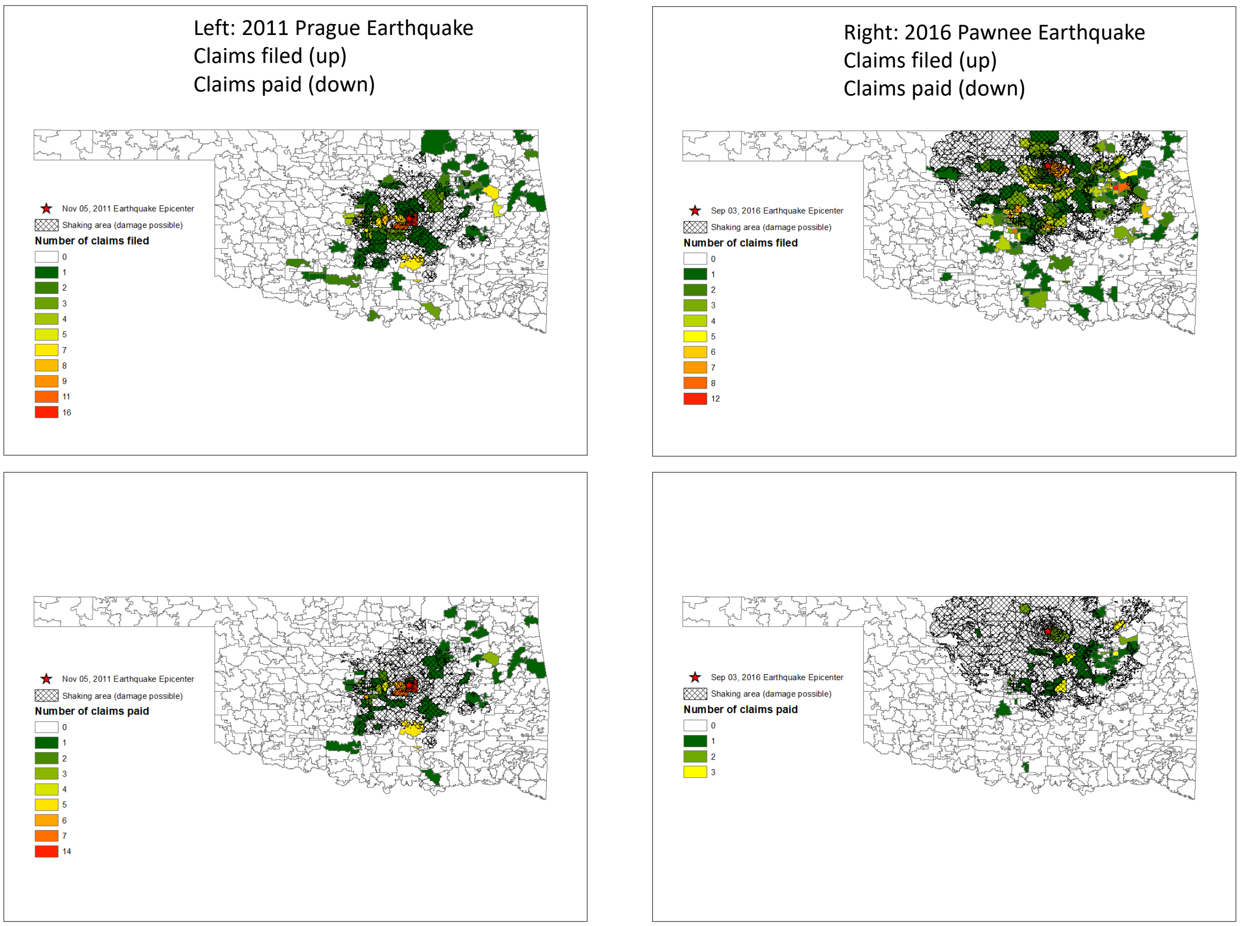

Figure 2 depicts the epicenters of the two most significant earthquakes and where claims were filed and paid. The locations of claims are identifiable at the zip-code level. The two maps on the left-hand-side indicate claims associated with the 2011 Prague earthquake, and the two maps on the right-hand-side indicate claims associated with the 2016 Pawnee earthquake. The Prague earthquake measured 5.7 on the Richter scale, and the Pawnee 5.8. The maps in the upper panel indicate where and how many claims were filed, and the maps in the lower panel indicate where and how many claims were paid. Both earthquakes had a very large geographic impact given their scores on the Richter scale. A scientific explanation is that earthquakes triggered by human gas and oil drilling activities tend to be much shallower than those triggered by natural force; therefore, they impact a greater area for a given magnitude, which measures the total energy released. The geographic locations of claims filed and paid seem quite random, suggesting factors other than the distance to epicenter affect the filing of claims and their settlement.

The shaded zone in Figure 2 indicates the area of "damage possible". The boundaries of these zones are of irregular shape and are defined by the U.S. Geological Survey (USGS). The USGS defines a ground movement acceleration measure, and categorizes earthquake-affected areas starting from the epicenter as having experienced: 1, very strong ground acceleration ($> 1.8m/S^2$); 2, strong ground acceleration ($0.92 - 1.8m/S^2$); or 3, moderate ground acceleration ($0.4 - 0.92m/S^2$). The USGS states that the potential damage in areas outside of the moderate ground acceleration zone is *none*; in other words, only areas that experience at least $0.4m/S^2$ ground acceleration can possibly experience property damage, according to the USGS. We use GIS software to intersect the USGS-defined "damage possible" zone with zip-codes and label a zip-code as "within damage possible zone" if any part of it is within or touches upon the boundaries of a damage possible zone. We do this for each of the eight earthquakes.

Each earthquake claim comes with the following policy-level information: zip-code where the claim is filed, policy limit that approximately represents the value of the insured home, policy deductible, date of loss (i.e., date of the earthquake that causes damage), date of notice of claim, date of initial claim payment (if any), amount of claims paid (if any), date of the close of claim, if a legal counsel is involved, if an engineering report was obtained from a licensed structural engineer (which is likely ordered by the insurer), and if a public adjuster was retained by the claimant in the claim settlement process, in addition to the claims adjuster assigned by the insurer.

Table 2 summarizes these variables. The 846 earthquake claims filed since 2010 are from policies that have an average policy limit of $191,650 and a median limit of $169,607, with the largest limit being a little over $1.3 million dollars. Recall that an earthquake insurance policy deductible is a percentage of the policy limit. More than half of the earthquake policies have a 2% deductible, the lowest possible deductible option. Translated into dollar terms, the average deductible amount is $6,251 and the median deductible is $3,760, which is much larger than the typical deductible in a standard homeowners insurance. In terms of delay in the notice of claim after an earthquake, half of the claims were submitted within nine days of the date of loss. About two-thirds of these claims include an engineer's report, and in just a little over 1% of the claims a public claims adjuster was retained by the claimant. Legal counsel was involved in less than 1% of the claims (or 5 out of 846 claims).

Table 2 Summary statistics of claims data obtained from the Oklahoma Insurance Department (OID).

VARIABLES	Obs	Mean	Median	Std. Dev.	Min	Max
Policy limit ($)	846	191,650	169,607	116,841	12,000	1,309,886
Deductible (% of policy limit)	846	3.28%	2.00%	2.88%	2.00%	25.00%
Deductible ($)	846	6,251	3,760	7,633	240	95,040
Delay in notice of claim (days)	846	52	9	115	0	1473
Engineer report (1=yes, 0=no)	846	66.78%	1	0.471		
Public adjuster (1=yes, 0=no)	846	1.18%	0	0.108		
Legal Counsel (1=yes, 0=no)	846	0.59%	0	0.0767		

Next, we link the policy-level claims dataset to the census dataset by zip-code. For each claim, we then have information on the neighborhood characteristics of the zip-code area. Available census information includes a zipcode area's population, population density, median household income, median home value, percentage of college graduates among those at least 25 years old, percentage of population that is Black, Hispanic, Asian, White, Native or American Indian. There are 649 zip-codes in Oklahoma. Earthquake claims in our dataset come from 192 different zip-code areas. Table 3 summarizes the characteristics of these 192 zip-codes.

Table 3 Summary statistics of census data for the zip-code areas in the claims dataset.

VARIABLES	Obs	Mean	Median	Std. Dev.	Min	Max
Population	192	14,905	12,000	12,552	272	61,147
Population density (per square mile)	192	836	138	1,265	6.20	5,162
Median household income ($)	192	48,591	45,251	16,198	14,306	136,694
Median home value ($)	192	116,996	105,800	50,456	39,800	387,300
Education (% with college degree)	192	22.93	19.30	12.68	1.70	65.90
Percent Black (%)	192	7.98	3.56	14.12	0	81.31
Percent Native or American Indian (%)	192	7.41	5.49	5.99	0	46.93
Percent Asian (%)	192	1.42	0.63	1.96	0	8.52
Percent Hispanic (%)	192	7.85	4.95	9.66	0	62.00

Table A.1 in the Appendix summarizes the correlation among both policy-level characteristics and zip-code level characteristics. Not surprisingly, zip-code characteristics correlate with each other quite strongly: more densely populated zip-codes are associated with higher Black, Asian, and Hispanic populations but lower Native and American Indian populations; zip-code level minority population percentages are generally negatively correlated with zip-code median income, median home value, and education (an exception is the Asian population percentage). In terms of the correlation between policy-level and zip-code level characteristics, there is generally no strong correlation (e.g., the choice of deductible does not seem to differ by zip-code areas), except that the observed individual policy limit tends to be positively correlated with a zip-code's median income, median home value, and education level. This makes sense as a policy is only observed in our dataset when the policyholder files a claim, and the purchase of earthquake insurance is more common in more affluent areas.

5 Empirical Approach

In our baseline model, our identification strategy rests on the assumption that given the same earthquake event and controlling for the distance-to-epicenter from a zip-code centroid, neither whether an earthquake claim is paid nor how much is paid should differ by zip-code level racial composition, if there is no racial discrimination. Equation (4) is the baseline model.

$$y_{i,j,k} = R_j^{'}\beta + z_{i,j}^{'}\alpha_{1i} + \alpha_i + \varepsilon_{i,j,k} \tag{4}$$

Let i represent an earthquake event, j represent a zip-code area, and k represent a claimant. We define $y_{i,j,k}$ as the outcome of an earthquake policy claim k filed from zip-code j affected by event i. This outcome could be whether or not the claim filed by this individual is paid, or the amount of claim payment this individual receives, conditional on his/her claim being paid.

R_j is a vector of race component variables in zip-code j. The racial make-up of zip-codes in Oklahoma tends to be majority White and non-Hispanic. The percentage of any type of minority population is highly right-skewed. To illustrate, the left-hand-side of Figure A.1 shows the raw distribution of zip-code level Black percentage in our dataset, and the right-hand-side shows the distribution after taking the log.[12]

$z_{i,j}^{'}\alpha_{1i}$ is a term that controls for the distance from the centroid of zip-code j to the epicenter of earthquake i. We allow the coefficient to vary by earthquake because the amount of damage associated with the distance from an earthquake epicenter differs by earthquake. We also include earthquake event fixed-effects, α_i, in our equation.

Our baseline assumptions hold only if given the same earthquake event and controlling for the distance-to-epicenter from a zip-code centroid, the followings are true: (1) the characteristics of the underlying

[12] Our race variables are all in percentage terms, ranging from 0-100. To accommodate the zeros, we take the natural log of (variable +1).

insurance contract (e.g., deductible level) are uncorrelated with race; (2) the insured loss exposures of the property (e.g., building type) are uncorrelated with race; and (3) the quality and nature of a claim (e.g., no speculative claims) are uncorrelated with race.

The above may not hold. For one, race could be correlated with characteristics of an insurance contract. For example, higher-deductible policies may be more common in a majority Black area; therefore, any correlation between race and claim outcome may be driven not by racial discrimination but by policy characteristics. For another, it is possible, that the racial composition of an area is correlated with other neighborhood characteristics, which in turn correlate with whether or not a claim filed from that neighborhood should be paid. If that is the case, then even after controlling for earthquake event fixed-effects, distance-to-epicenter, and insurance policy characteristics; we may observe differences in claims handling that correlate with race, but which may be driven by other neighborhood characteristics such as income, home value, building type, home ownership rate, and perhaps education.

To address the above concerns, in equation (5) below, we include in addition a set of insurance policy-level variables ($X_{i,j,k}$) and a set of zip-code level variables (W_j). Specifically, $X_{i,j,k}$ is a vector that includes policy limit, policy deductible (in \$ term), whether an engineer's report was obtained, whether a public adjuster was retained by the claimant, and the claimant's delay in filing a claim following an earthquake. Delay in the filing of a claim may signal potential moral hazard or insurance fraud. W_j is a vector that includes zip-code median household income, median home value, population density, home ownership percentage, building type (percentage of homes built before 1949), and education level (measured by the percentage of college graduates among adults aged 25 or above).

$$y_{i,j,k} = R_j^{'}\beta + W_j^{'}\gamma_1 + X_{i,j,k}^{'}\gamma_2 + z_{i,j}^{'}\alpha_{1i} + \alpha_i + \varepsilon_{i,j,k} \tag{5}$$

After controlling for earthquake event fixed-effects, distance-to-epicenter by zip-code and earthquake, and an array of policy level characteristics and zip-code level characteristics, there may still be other omitted variables which correlate with both the zip-code level racial composition and the "indiscriminate" outcome of a claim from that zip-code, such as the distribution of types of insurers and access to lawyers. We are able to observe that only five out of 846 claims in our dataset involved a lawyer, which we include as a control variable in some of our regression models. While we are unable to observe which insurers operate in which zip-code areas, the assumption that market competition prevails is plausible.[13] Ideally, including zip-code fixed effects would take care of most zip-code level unobserved characteristics. Unfortunately, this is not possible since our race variables are at the zip-code level and do not vary over time. It is, however, possible to include zip-code random effects. A type of random-effects model – a multi-level model – is appropriate for our setting in which individual claims are clustered by zip-code areas and zip-code areas are clustered by earthquake events.[14] On the other hand, a multi-level model make certain assumptions about correlation structures among variables,[15] and so we provide estimation of this model side-by-side with our earthquake-event fixed-effects models.

Our empirical strategy of testing for possible racial discrimination does not require the assumption that earthquakes affect a random selection of zip-codes. We do not need to assume that the decision of where to live is random (e.g., by race) either. Rather, our empirical strategy rests on the random assignment of race *conditional on* having the same insured loss exposure to a particular earthquake. Our models focus on finding various ways to be as close to "having the same insured loss exposure to a particular earthquake" as possible.

Lastly, it is nonetheless possible that the quality and nature of a claim is correlated with race: the rejection rate may be higher for claims filed from a higher minority area because the insurer expects those claims are more likely to be "questionable". It is possible to test in our dataset whether earthquake claims, especially "questionable" claims, are more likely to come from certain areas, all else equal. We discuss this point in more detail in the following section.

---

[13] According to the Insurance Information Institute's *2020 Insurance Factbook*, Oklahoma's insurance industry is quite competitive with 31 property/casualty insurers competing for business with relatively stable market shares in 2018. Oklahoma is the 4th most expensive state in terms of home insurance rates. The annual average homeowners insurance premium is $1,875, which is significantly higher than the national average of $1,192 (2016).

[14] For fixed-effects models, standard errors can be clustered to correct for a possible nested structure in the dataset.

[15] For instance, multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent. This is hard to satisfy per Table A.1, especially since claims from higher-income zip-code areas also tend to have higher policy limits, which affect the outcome of how the claim is going to be handled. But Table A.1 is less explicit about the correlation between policy-level characteristics and zip-code level racial composition. We discuss this question further in the next section of the paper. On the other hand, dependency among variables of the same level (e.g., high correlations among zip-code characteristics according to Table A.1) is not a significant concern for multi-level models.

6 Empirical Results

6.1 Probability of a Claim Being Paid

This section focuses on whether or not a claim is paid. For each claim filed, we observe if it is paid. Table **??** reports the percentage of claims that result in a payment in the ten zip-codes with the highest percentage of Black population and the ten with the lowest percentage. We only include in the table those zip-code areas where at least two claims were filed. While the zip-codes listed in Table **??** are different in many other dimensions and each claim is different, the comparison is striking.

Table 4 Earthquake Claims Filed from Selected Zip-Codes

Zip-Codes with the Highest Black Population Percentage

Black Percentage	# of Claims Filed	# (%) of Claims Approved
81.3%	6	0 (0%)
74.9%	5	1 (20%)
69.5%	3	0 (0%)
68.7%	3	1 (33%)
67.8%	3	1 (33%)
67.2%	7	2 (29%)
54.0%	13	2 (15%)
31.7%	7	0 (0%)
28.7%	2	0 (0%)
27.7%	3	1 (33%)
Overall	52	8 (15%)

Zip-Codes with the Lowest Black Population Percentage

Black Percentage	# of Claims Filed	# (%) of Claims Approved
0.000%	3	2 (67%)
0.000%	5	4 (80%)
0.000%	3	0 (0%)
0.000%	2	0 (0%)
0.000%	3	3 (100%)
0.060%	13	8 (62%)
0.106%	4	2 (50%)
0.131%	2	1 (50%)
0.309%	4	2 (50%)
0.378%	4	1 (25%)
Overall	43	23 (53%)

Note: for a more effective comparison, we select zip-codes areas that have at least filed two claims.

Next, we run logit regressions by fitting a logit transformation of the binary outcome to obtain $y_{i,j,k}$. Table 5 and Table 6 present the regression results. Table 5 includes the full sample of 846 claims from 192 different zip-codes. Table 6 only includes a subset of claims filed from zip-codes within the USGS-defined "damage possible" zone. There are 600 claims and 138 zip-codes in this subset. We focus on this subset to minimize the influence of "questionable" claims. Overall, results in both tables are broadly similar.

Our baseline regressions are reported in column (1) of both tables. These models include only a set of race variables, earthquake event fixed-effects, and distance-to-epicenter controls (which vary by earthquake). The coefficients of the zip-code Black percentage variable are significantly negative, with a magnitude ranging from 0.22 to 0.24. The other race variables are for the most part not statistically significant.

Race variables may correlate with insurance policy-level characteristics, which ultimately affect the claim outcome. We, therefore, control in the models in column (2) of each table for the following policy-level characteristics: policy limit, deductible, delay in notice of claim, whether or not an engineer report is obtained, whether a legal counsel is obtained, and whether or not a public adjuster is retained by the claimant.[16] In column (3), we use a slightly different model specification: a random-effects multi-level model, which includes event level and zip-code level random effects.[17] After controlling for zip-code random effects, and accounting for potentially correlated residuals from claims filed from the same zip-code and zip-codes affected by the same earthquake, the coefficients of the Black percentage variable are still negative, albeit only marginally significant.

Models in columns (4)and (5) further control for zip-code level characteristics, including population density, median household income, median owner-occupied home value, percentage of college graduates among adults at least 25 years old, percentage of homes that are owner-occupied, and percentage of homes built prior to 1949. In column (5) we include all zip-code level and insurance policy level controls. The coefficient of the Black percentage variable is even more significant and doubles in magnitude despite the

[16] The estimates of these control variables are not shown in the table. As expected, lower deductibles and higher policy limits significantly increase the probability of a claim being paid. Both the presence in the claim file of an engineer report and the engagement of a public adjuster increases the probability of a claim being paid. It may not be intuitive why the insurer is more likely to pay claims after ordering an engineer report, but it may signal that the claim is worth investigating and at the same time more likely to result in payment.

[17] Multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent. We only include policy-level controls without additional zip-code level controls. This is because individual policy limits are found to be positively correlated with zip-code's median income, median home value, and education level (lower left panel in Table A.1). This indicates that the zip-codes we observe where claims are led are not really randomly drawn after an earthquake event based on the previously mentioned characteristics. Of course, questions remain on whether the observed zip-codes are randomly drawn after each earthquake conditional on the racial components of a zip-code (given the same earthquake and same distance-to-epicenter). We discuss this question further when we discuss our empirical findings. On the other hand, dependencies among variables of the same level (e.g., high correlations among zip-code characteristics according to Table A.1) can be accounted for by the multi-level model's covariance structure.

Table 5 Logit regression of Claim (1=paid, 0=denied) on zip-code race components, controlling for distance to epicenter, policy characteristics and other zip-code characteristics (FULL sample)

	Dependent Variable: If a Claim is Paid (YES=1)				
	(1)	(2)	(3)	(4)	(5)
Log of Black Percentage	-0.221**	-0.213**	-0.242*	-0.429***	-0.430***
	(0.108)	(0.096)	(0.130)	(0.111)	(0.094)
Log of Native and American Indian Percentage	-0.018	-0.035	-0.032	-0.074	-0.134
	(0.158)	(0.224)	(0.240)	(0.101)	(0.178)
Log of Asian Percentage	-0.263**	-0.284*	-0.312	-0.141	-0.196
	(0.133)	(0.166)	(0.202)	(0.123)	(0.173)
Log of Hispanic Percentage	-0.044	-0.052	-0.091	-0.376	-0.390
	(0.244)	(0.291)	(0.181)	(0.266)	(0.342)
Event Fixed/Random Effects	YES	YES	YES	YES	YES
Event FE*Distance to Epicenter	YES	YES	YES	YES	YES
Policy characteristics controls		YES	YES		YES
Other zip-code characteristics controls				YES	YES
Zip-code Random Effects			YES		
Observations	846	846	846	846	846
Number of different zip-codes	192	192	192	192	192

Note: All regressions are estimated using the full sample. All models include earthquake event fix-effects and distance-to-epicenter by earthquake controls. Columns (1)-(3) are estimated without other zip-code level controls; columns (4)-(5) are estimated with other zip-codes characteristics controls. Columns (2)(3)(5) are estimated with insurance contract characteristics covariates, including column (5) that includes all of the insurance contract characteristics and zip-code level controls. Column (3) is estimated assuming a multi-level structure of data and includes EQ event random effects and zip-code random effects in lieu of other zip-code level variables. Since multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent, and since our data suggests that policy limit has a strong correlation with zip-code median income, median home value, and college graduate percentage, we do not include zip-code level variables when estimating columns (3).

Robust standard errors clustered at the earthquake event level for fixed-effect models. Standard errors are naturally clustered for multi-level models.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 6 Logit regression of Claim (1=paid, 0=denied) on zip-code race components, controlling for distance to epicenter, policy characteristics and other zip-code characteristics (SUBSET of the full sample)

	Dependent Variable: If a Claim is Paid (YES=1)				
	(1)	(2)	(3)	(4)	(5)
Log of Black Percentage	-0.242***	-0.224***	-0.225	-0.426***	-0.424***
	(0.072)	(0.045)	(0.141)	(0.066)	(0.043)
Log of Native and American Indian Percentage	-0.079	-0.129	-0.068	-0.167**	-0.262*
	(0.090)	(0.181)	(0.267)	(0.067)	(0.146)
Log of Asian Percentage	-0.214*	-0.237*	-0.229	-0.099	-0.129
	(0.112)	(0.143)	(0.228)	(0.124)	(0.189)
Log of Hispanic Percentage	-0.060	-0.023	0.004	-0.573**	-0.568
	(0.241)	(0.305)	(0.198)	(0.256)	(0.368)
Event Fixed/Random Effects	YES	YES	YES	YES	YES
Event FE*Distance to Epicenter	YES	YES	YES	YES	YES
Policy characteristics controls		YES	YES		YES
Other zip-code characteristics controls				YES	YES
Zip-code Random Effects			YES		
Observations	600	600	600	600	600
Number of different zip-codes	138	138	138	138	138

Note: All regressions are estimated using a subset of claims filed from zip-codes within the USGS-defined “damage possible” zone. All models include earthquake event fix-effects and distance-to-epicenter by earthquake controls. Columns (1)-(3) are estimated without other zip-code level controls; columns (4)-(5) are estimated with other zip-codes characteristics controls. Columns (2)(3)(5) are estimated with insurance contract characteristics covariates, including column (5) that includes all of the insurance contract characteristics and zip-code level controls. Column (3) is estimated assuming a multi-level structure of data and include EQ event random effects and zip-code random effects in lieu of other zip-code level variables. Since multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent, and since our data suggests that policy limit has some strong correlation with zip-code median income, median home value, and college graduate percentage, we do not include these zip-code level variables while estimating columns (3).

Robust standard errors clustered at the earthquake event level for fixed-effect models. Standard errors are naturally clustered for multi-level models.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

addition of other zip-code level control variables. Therefore, the claim that the apparent lower rates of claim payment in zip-codes with higher Black population are driven by neighborhood characteristics other than race is not supported.

The zip-code level Black percentage variable is statistically significant at the 5% level or less in eight of the ten models in Tables 5 and 6, and at the 10% level in another. The sign of this coefficient indicates that all else equal, an earthquake claim filed from a zip-code with a higher percentage of Black population is on average less likely to be paid. Based on the baseline result in column (1) of Table 6, the average marginal effect of the Black percentage coefficient is -0.047 – meaning one percent increase in this x variable is associated with a 4.7 percentage point decrease in the probability of the claim being paid.[18]

6.2 Amount of Claim Payment

Conditional on a claim being paid, we now examine the size of the claim payment. The same regression models and control variables are used as in the previous section, except that the new dependent variable is the log of the dollar amount of claim payment. Results are presented in Table 7 and Table 8. It is noteworthy that the number of observations is smaller given that less than a third of the claims are paid. There are 253 paid claims from 103 different zip-code areas in our full sample, and 212 paid claims from 81 different zip-code areas in our sub-sample of zip-code areas within the USGS-defined "damage possible" zone.

Overall, the zip-code Black percentage variable is consistently negative and statistically significant in both samples. In six out of the ten model specifications in both tables, this coefficient is significantly negative at the 5% level; in three out of the ten specifications, this coefficient is significantly negative at the 10% level; in all but one specification, the significance level reaches 10%. As a result, for an increase in the proportion of Black by 1%, we expect the amount of claims paid to decrease by 0.2%-0.5% (model (1) in Table 7 versus model (5) in Table 8).

6.3 Tendency to File Claims

In this section we conduct tests on the possible correlation between zip-code characteristics and residents' tendency to file earthquake claims – specifically, whether policyholders from higher minority percentage

[18] Zip-code Black population percentage ranges from 0.41% to 81% in our dataset. A variation from 0.41% to 81% approximately represents a 4 unit increase in our transformed independent variable, which, multiplied by the marginal effect of -4.7, is equal to about a 19% decrease in probability.

Table 7 Regression of the $ amount of claims paid (conditional on a claim being paid) on zip-code race components, controlling for distance to epicenter, policy characteristics and other zip-code characteristics (FULL sample)

Dependent Variable: Log of Claims ($) paid conditional on a claim being paid

	(1)	(2)	(3)	(4)	(5)
Log of Black Percentage	-0.215**	-0.206*	-0.212*	-0.288*	-0.294**
	(0.084)	(0.092)	(0.118)	(0.149)	(0.122)
Log of Native and American Indian Percentage	-0.008	-0.040	0.076	-0.045	-0.089
	(0.166)	(0.118)	(0.216)	(0.196)	(0.155)
Log of Asian Percentage	-0.157	-0.269**	-0.095	-0.376*	-0.386**
	(0.124)	(0.096)	(0.200)	(0.164)	(0.158)
Log of Hispanic Percentage	0.053	0.125	0.078	0.097	0.104
	(0.200)	(0.180)	(0.162)	(0.102)	(0.141)
Event Fixed/Random Effects	YES	YES	YES	YES	YES
Event FE*Distance to Epicenter	YES	YES	YES	YES	YES
Policy characteristics		YES	YES		YES
Other zip-code demographic				YES	YES
Zip-code Random Effects			YES		
Observations	253	253	253	253	253
Number of different zip-codes	103	103	103	103	103

Note: All regressions are estimated using the full sample. All models include earthquake event fix-effects and distance-to-epicenter by earthquake controls. Columns (1)-(3) are estimated without other zip-code characteristics controls; columns (4)-(5) are estimated with other zip-codes characteristics controls. Columns (2)(3)(5) are estimated with insurance contract characteristics covariates, including column (5) that includes all of the insurance contract characteristics and zip-code characteristics controls. Column (3) is estimated assuming a multi-level structure of data and include EQ event random effects and zip-code random effects in lieu of other zip-code level variables. Since multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent, and since our data suggests that policy limit has some strong correlation with zip-code median income, median home value, and college graduate percentage, we do not include additional zip-code level variables while estimating columns (3).

Robust standard errors clustered at the earthquake event level for fixed-effect models. Standard errors are naturally clustered for multi-level models.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8 Regression of the $ amount of claims paid (conditional on a claim being paid) on zip-code race components, controlling for distance to epicenter, policy characteristics and other zip-code characteristics (SUBSET of the full sample)

Dependent Variable: Log of Claims ($) paid conditional on a claim being paid

	(1)	(2)	(3)	(4)	(5)
Log of Black Percentage	-0.302**	-0.317**	-0.259**	-0.419	-0.479**
	(0.124)	(0.121)	(0.129)	(0.224)	(0.163)
Log of Native and American Indian Percentage	0.062	-0.017	0.173	0.075	-0.007
	(0.137)	(0.101)	(0.254)	(0.158)	(0.087)
Log of Asian Percentage	-0.141	-0.262	-0.035	-0.390**	-0.444*
	(0.115)	(0.139)	(0.225)	(0.117)	(0.213)
Log of Hispanic Percentage	0.132	0.187	0.115	0.161	0.091
	(0.263)	(0.194)	(0.179)	(0.095)	(0.066)
Event Fixed/Random Effects	YES	YES	YES	YES	YES
Event FE*Distance to Epicenter	YES	YES	YES	YES	YES
Policy characteristics		YES	YES		YES
Other zip-code demographic				YES	YES
Zip-code Random Effects			YES		
Observations	212	212	212	212	212
Number of different zip-codes	81	81	81	81	81

Note: All regressions are estimated using a subset of claims filed from zip-codes within the USGS-defined "damage possible" zone. All models include earthquake event fix-effects and distance-to-epicenter by earthquake controls. Columns (1)-(3) are estimated without other zip-code characteristics controls; columns (4)-(5) are estimated with other zip-codes characteristics controls. Columns (2)(3)(5) are estimated with insurance contract characteristics covariates, including column (5) that include all of the insurance contract characteristics and zip-code characteristics controls. Column (3) is estimated assuming a multi-level structure of data and include EQ event random effects and zip-code random effects in lieu of other zip-code level variables. Since multi-level models assume that cross-level variables (i.e., zip-code level and policy level) are independent, and since our data suggests that policy limit has some strong correlation with zip-code median income, median home value, and college graduate percentage, we do not include additional zip-code level variables while estimating columns (3).

Robust standard errors clustered at the earthquake event level for fixed-effect models. Standard errors are naturally clustered for multi-level models.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

areas are more or less likely to file claims. If discrimination is self-fulfilling, then areas with higher percentages of minorities may have a lower tendency to file claims, expecting less successful outcomes. However, it could also be that those areas have a higher tendency to file claims. An alternative hypothesis that would be consistent with our finding on claims adjudication outcomes is that more "questionable" claims arise from zip-codes with higher percentages of minority population. In anticipation of such, claims adjusters end up being more suspicious when processing claims filed from certain areas, irrespective of the legitimacy of the individual claim.

We reorganize our dataset to be at the zip-code level, and aggregate the number of claims by earthquake event, if any, from all zip-code areas (including those with no claims). For each earthquake, we divide all zip-code areas into two groups: those inside the "damage possible" zones defined by the USGS after each earthquake event and those outside of such zones. We focus on the "outside" group to test if zip-codes with higher minority percentages are filing a greater number or fewer "questionable" claims, all else equal. For each earthquake, we use information on the distance from each zip-code centroid to the earthquake epicenter to control for the residual difference in tendency to file claims due to proximity to an earthquake. The following equation is estimated:

$$y_{i,j} = X_j^{'}\beta_1 + z_{i,j}^{'}\alpha_{1i} + \alpha_{0i} + \varepsilon_{i,j} \tag{6}$$

$y_{i,j}$ is a zip-code level outcome, i represents an earthquake event and j represents a zip-code area. We model the outcome variable as a binary variable (whether or not any claim is filed from zip-code j after event i) or as a count variable (number of claims filed from zip-code j after event i).

X_j is a vector of zip-code level demographic and socio-economic characteristics, which includes, the racial profile in terms of percentages of minority population (Black, Native American or American Indian, Asian, and Hispanic), population density, median household income, median home value, percentage of college graduates, percentage of owner-occupied homes, and percentage of homes built prior to 1949. X_j also includes imputed values of the number of insured housing units by zip-code, which are calculated as the product of the number of housing units and a predicted earthquake insurance take-up rate. Appendix B offers more details on how a set of predicted earthquake insurance take-up rates by zip-code is obtained based on a survey conducted by the University of Oklahoma Center for Risk and Crisis Management. $z_{i,j}^{'}\alpha_{1i}$ is a set of distance to epicenter controls that vary by earthquake event. α_{0i} is a set of event dummy variables, and $\varepsilon_{i,j}$ is the error term. Standard errors of the results are robust and clustered at the earthquake event level. We also estimate models with zip-code random-effects where errors are clustered at the zip-code level which are then clustered by earthquake event. Results are presented in Table 9.[19]

Table 9 Regression of zip-code level claim filing outcome on zip-code characteristics among areas outside of the USGS-defined "damage possible" zone.

Dependent Variable: zip-code level binary or count outcome (whether at least one claim has been filed or the number of claims filed)	(1)	(2)	(3)	(4)
Log of Black Percentage	0.184***	0.227	0.159***	0.193
	(0.067)	(0.177)	(0.061)	(0.120)
Log of Native and American Indian Percentage	-0.024	-0.023	-0.089	-0.075
	(0.183)	(0.245)	(0.210)	(0.180)
Log of Asian Percentage	0.001	-0.034	-0.083	-0.088
	(0.275)	(0.264)	(0.195)	(0.176)
Log of Hispanic Percentage	-0.503***	-0.582**	-0.434***	-0.518***
	(0.175)	(0.275)	(0.155)	(0.171)
Log of Insured Housing Units	1.164***	1.224***	1.069***	1.054***
	(0.156)	(0.286)	(0.125)	(0.109)
Log of Population Density	0.301**	0.375**	0.223***	0.257***
	(0.129)	(0.169)	(0.069)	(0.090)
Log of Median Household Income	-0.096	-0.119	-0.319	-0.216
	(1.227)	(1.115)	(1.057)	(0.804)
Log of Median Home Value	0.760	0.868	0.574	0.572
	(0.700)	(0.814)	(0.351)	(0.568)
Percentage of College Graduates	-0.030	-0.032	-0.026***	-0.026*
	(0.021)	(0.021)	(0.008)	(0.014)
Percentage of Owner-Occupied Home	2.799	3.232	1.668	1.735
	(2.314)	(2.098)	(1.850)	(1.304)
Percentage of Houses Built before 1949	0.016	0.017*	0.005	0.007
	(0.014)	(0.010)	(0.013)	(0.007)
EQ Event Fixed Effects	YES	YES	YES	YES
Event*Distance to Epicenter	YES	YES	YES	YES
Zip-Code Random Effects		YES		YES
Poisson Count Model			YES	YES
Observations	4303	4303	4303	4303

Note: All regressions are estimated using sample of zip-codes outside of the USGS-defined "damage possible" zone (different for each earthquake). Columns (1) and (2) estimate logistic regressions using whether or not any claim is filed from a zip-code as the dependent variable; columns (3) and (4) estimates Poisson regressions using the number of claims filed from a zip-code as the dependent variable. Columns (2) and (4) are estimated as multi-level models including EQ event and zip-code random effects.
Robust standard errors clustered at the earthquake event level.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

[19] We also tried regression models without the imputed insured housing units and simply use the number of housing units by zip-code instead, and regression models with a different set of predicted insurance take-up values from an alternative prediction equation. Results are broadly similar to Table 9.

Looking at places that the USGS states should not have suffered damage, there is some evidence that zip-codes with higher percentages of Black population tend to file more claims, all else equal. The sign of this variable is consistent, though its significance drops in random-effect models. From the logistic regression in model (1), we find that the probability that any claim is filed from a zip-code (outside the USGS-defined ”damage possible” zone) is higher for an area with a larger percentage Black population; and similarly for the Poisson regression in model (3).

Interestingly, the coefficient for the Percentage of Hispanic population is consistently both significant and negative, implying a significantly lower tendency to file claims. As intuition suggests, number of insured housing units and population density are positively associated with zip-code level claim filing frequencies. Other controls we include are generally not significant. These include measures of household income, home value, education attainment, home ownership, and age of construction.

Remember that Table 9 represents outcomes for the subsample of zip-codes outside of the “damage possible zone”. There is some concern that limits our confidence in drawing conclusions from the results reported in Table 9. The definition of inside or outside the “damage possible zone” is somewhat arbitrary. The possibility exists that a home might have suffered damage even if it is outside the USGS damage zone. Therefore, we also include the continuous control variable – distance to epicenter – to capture the fact that places not far from the “damage possible zone” are still more likely to experience damage than places far from it.

7 Public Policy Implications and Conclusion

The current article studies racial discrimination in the adjudication of earthquake insurance claims in Oklahoma. Empirical evidence using these data suggests that claims arising from zip-code areas with a higher Black population are less likely to be paid and, if paid, are less likely to settle for as much as other claims, all else equal. These findings could be attributable to claimants' perceptions about the value the legal system places on claims (demand-side discrimination), insurers' perceptions about the value the legal system places on claims (supply-side discrimination), or, a difference in the underlying value of the claims that is unobservable in the data. Evidence that “questionable” claims are more commonly filed in zip-codes with a higher percentage of Black population runs counter to the demand-side discrimination explanation.

Racial discrimination in the settlement of insurance claims has received little scholarly attention, likely because of the paucity of data available to study. Our data, arising from earthquakes in the state of

Oklahoma, while very narrow, has the relatively unique advantage of letting us observe the settlements of claims of multiple people whose homes experienced the same loss event. Our primary findings, that Blacks are less likely to have their claims paid and, if paid, are on average compensated less, are not generalizable. Our data have a very specific focus. Further, our use of zip-code level data is a second-best approach. Superior data would allow a study at the individual household level. Data of this type, however, likely do not exist as insurers do not collect information on race in the underwriting of insurance policies.

Our findings suggest that additional research on whether racial discrimination exists in the settlement of insurance claims is warranted. Claims settlements that are based on the characteristics of loss events and the language of insuring agreements is expected by the parties to an insurance contract. Racial bias in claims settlement would both deprive individuals of value to which they are entitled and undermine confidence in insurance, one of the primary means in the economy for financially protecting oneself against loss.

References

Aigner, D. and Cain, G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 30(2):175–187.

Anbarci, N., Escaleras, M., and Register, C. A. (2005). Earthquake fatalities: the interaction of nature and political economy. *Journal of Public Economics*, 89(9-10):1907–1933.

Arrow, K. (1998). What has economics to say about racial discrimination. *Journal of Economic Perspectives*, 12(2):91–100.

Arrow, K. J. (1973). The theory of discrimination. In Ashenfelter, O. and Rees, A., editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, New Jersey.

Ashenfelter, O. and Rees, A. (1973). *Discrimination in Labor Markets*. Princeton University Press, New Jersey.

Ayres, I. and Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, 85(3):304–321.

Baker, T. and McElrath, K. (1996). Whose safety net? home insurance and inequality. *Law & Social Inquiry*, 21(2):229–264.

Becker, G. (1957). *The Economics of Discrimination*. University of Chicago Press, Chicago.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy*, 101(3):385–409.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Blanchflower, D. G., Levine, P. B., and Zimmerman, D. J. (2003). Discrimination in the small-business credit market. *The Review of Economics and Statistics*, 85(4):930–943.

Browne, M. J. and Puelz, R. (1999). The effect of legal rules on the value of economic and non-economic damages and the decision to file. *Journal of Risk and Uncertainty*, 18(2):189–213.

Cain, G. G. (1986). The economic analysis of labor market discrimination: A survey. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics, Volume I*, pages 693–785. Elsevier/North Holland.

Castillo, M., Petrie, R., Torero, M., and Vesterlund, L. (2013). Gender differences in bargaining outcomes: A field experiment on discrimination. *Journal of Public Economics*, 99:35–48.

Coate, S. and Loury, G. C. (1993). Will affirmative-action policies eliminate negative stereotypes? *The American Economic Review*, 83(5):1220–1240.

Cummins, J. D. and Tennyson, S. (1992). Controlling automobile insurance costs. *The Journal of Economic Perspectives*, 6(2):95–115.

Doerpinghaus, H., Schmit, J., and Yeh, J. J.-H. (2003). Personal bias in automobile claims settlement. *Journal of Risk and Insurance*, pages 185–205.

Doerpinghaus, H. I., Schmit, J. T., and Yeh, J. J.-H. (2008). Age and gender effects on auto liability insurance payouts. *Journal of Risk and Insurance*, 75(3):527–550.

Harrington, S. E. and Niehaus, G. (1998). Race, redlining, and automobile insurance prices. *The Journal of Business*, 71(3):439–469.

Harrison, G. W. (1998). Mortgage lending in Boston: A reconsideration of the evidence. *Economic Inquiry*, 36(1):29–38.

Heckman, J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12(2):101–116.

Klein, R. (1989). Competition in private passenger automobile insurance: A report to the personal lines (c) committee. Technical report, Working paper. Kansas City, Mo.: National Association of Insurance Commissioners.

Klein, R. W. and Grace, M. F. (2001). Urban homeowners insurance markets in texas: A search for redlining. *Journal of Risk and Insurance*, pages 581–614.

Klumpp, T. and Su, X. (2013). Second-order statistical discrimination. *Journal of Public Economics*, 97:108–116.

Lin, X. (2016). Risk awareness and adverse selection in catastrophe insurance: Evidence from california's residential earthquake insurance market. *SSRN Working Paper*.

Liu, H., Ferreira, S., and Brewer, B. (2017). The housing market impacts of wastewater injection induced seismicity risk. *SSRN Working Paper*.

Loury, G. C. (1992). Incentive effects of affirmative action. *Annals of the American Academy of Political and Social Science*, 523:19–29.

Munnell, A. H., Tootell, G. M. B., Browne, L. E., and McEneaney, J. (1996). Mortgage lending in boston: Interpreting hmda data. *The American Economic Review*, 86(1):25–53.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661.

Pope, D. G. and Sydnor, J. R. (2011). What'ss in a picture? evidence of discrimination from prosper.com. *Journal of Human Resources*, 46(1):53–92.

Rubinstein, J. L. and Mahani, A. B. (2015). Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity. *Seismological Research Letters*, 86(4):1060–1067.

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40.

Wright, R. (2018). Yes, jury selection is as racist as you think. now we have proof. *New York Times*, Dec 4:online.

A Supplemental Tables and Figures

Figure A.1 Distribution of zip-code level Black population percentage

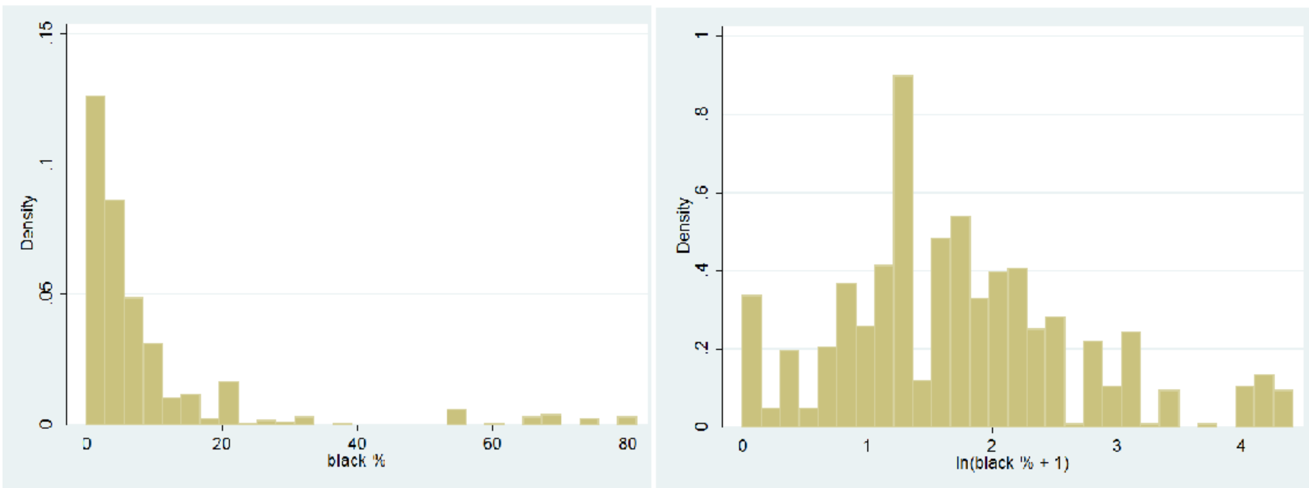

Table A.1 Pearson Correlation Table

	Deduc	Limit	Delay	Engin	PubAd	PopDe	MedInc	Medhv	Colleg	Black	Native	Asian
Deductible (%)	1											
Log of Policy Limit	-0.0691	1										
Log of Delay in Filling Claims	-0.076	0.0664	1									
Engineer Report (YES=1)	-0.2618	0.123	0.1414	1								
Public Adjuster (YES=1)	0.0157	-0.0042	-0.0214	0.0078	1							
Log of Population Density	-0.0273	0.126	0.0621	-0.0029	-0.0685	1						
Log of Median Income	-0.0555	0.326	-0.0014	-0.0013	-0.0036	0.0548	1					
Log of Median Home Value	-0.0473	0.3912	0.0513	0.025	-0.0089	0.2624	0.8243	1				
College Graduates %	-0.0207	0.394	0.0907	0.0343	-0.0612	0.4416	0.6168	0.8266	1			
Log of Black %	0.0448	0.0017	0.0757	-0.0025	-0.0443	0.6214	-0.3707	-0.1183	0.1174	1		
Log of Native or American Indian %	-0.041	-0.1841	-0.0634	0.059	0.0324	-0.5894	-0.2021	-0.383	-0.5306	-0.5285	1	
Log of Asian %	-0.0393	0.2496	0.0893	0.0075	-0.0504	0.7172	0.3171	0.524	0.6213	0.3349	-0.534	1
Log of Hispanic %	-0.0281	-0.1183	-0.0262	-0.0202	-0.0213	0.552	-0.2683	-0.2412	-0.1556	0.3665	-0.1431	0.2924

The number of observations is 846 claims filed. This correlation table is divided into four quadrams: upper-left shows correlation among policy-level characteristics; lower-right shows correlation among zip-code level characteristics; lower-left shows correlation between policy-level and zip-code level variables.

B Predicting Earthquake Insurance Take-Up

The University of Oklahoma Center for Risk and Crisis Management conducted a 20 wave survey between 2014 and 2019 to assess how residents perceived their exposure to natural hazards. Participants were randomly selected from across the state. They completed an internet-based survey that took about 30 minutes. Participants were surveyed quarterly for five years. They were compensated $10 for each round of the survey in which they participated.

In two waves of the surveys, there is a question asking whether or not the respondent has earthquake insurance. Information on the participants' zip-code location, race, age, gender, household income, education level, and the square footage of house are also reported. There are in total 2627 unique participants from these two waves of the survey (including 1808 participants that participated in both).

We focus on 2269 unique participants who answered the question "do you have earthquake insurance?", and who completed survey questions related to their race, income, and education. We wanted to know if any demographic and socioeconomic characteristic is predictive of the take-up of earthquake insurance. Table B.1 presents result from a logistic regression with the dependent variable being "1" if an individual answered "Yes" to the "do you have earthquake insurance?" question. We find that household income, education level, and the square footage of one's house are significantly positively associated with the probability of having earthquake insurance; different race variables are not statistically significant, though generally have negative signs, compared to the baseline of White.

Using the coefficient estimates from Table B.1 and the method of prediction at the means, we predict a take-up rate of earthquake insurance for all households in a zip-code area: $Pr(Y=1|X=\bar{x}) = \frac{exp(\bar{x}'\hat{\beta})}{1+exp(\bar{x}'\hat{\beta})}$, where $\bar{x}$ is a vector of mean values of the covariates in a zip-code.

Table B.1 Regression of having earthquake insurance (YES=1) on individual characteristics

	Dependent Variable: Have Earthquake Insurance (YES=1)
Log of household income	0.441***
	(0.089)
Education (College degree=1)	0.282**
	(0.114)
Log of square footage	0.390***
	(0.137)
Black	-0.062
	(0.331)
American Indian or Native	-0.021
	(0.265)
Asian	-0.359
	(0.795)
Two races or other (non-white)	-0.562*
	(0.337)
Hispanic	-0.506
	(0.499)
(Intercept)	-9.286***
	(1.087)
Observations	2269
LR chi-sq (8)	88.81
Prob > chi-sq	0.0000
Log likelihood	-1108

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$