Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

INTEGRATING PLACE ATTACHMENT INTO HOUSING RECOVERY SIMULATIONS TO ESTIMATE POPULATION LOSSES

Rodrigo Costa, Chenbo Wang, and Jack W. Baker

Abstract

Following a disaster, residents of a community may be displaced from their damaged homes, leading to expensive and lengthy disruption, with many choosing to move away permanently. Population losses may hinder recovery and exacerbate inequalities across neighborhoods. This study considers household place attachment and identifies groups with low place attachment along with expensive and slow post-disaster recovery. We develop a framework to integrate place attachment considerations into housing recovery simulations. We use data from the American Housing Survey to develop housing and neighborhood satisfaction models and identify the neighborhoods with the least attached residents. A computational simulation framework is used to simulate post-earthquake housing recovery for a community and assess expected costs and time frames. We use the triad of low place attachment, high cost, and slow recovery to identify households prone to permanently moving away from their communities. A case study of housing recovery after a hypothetical earthquake near San Francisco is presented to demonstrate the application of the methodology. We identify that about 10% of the population in some neighborhoods are prone to moving away after a large earthquake. Households with low-income, renters, and older buildings are most likely to have low place attachment and experience costly and slow recovery. While existing approaches rely on heuristics, the approach and results in this paper provide quantitative means to assess potential population losses and inform efforts to reduce them. The framework to integrate place attachment into housing recovery simulations is versatile and employs publicly available information making it transferable to other communities.

1 Introduction

The movement of individuals and households between homes is called residential mobility. Post-disaster residential mobility may result in population replacement and loss. The former occurs when long-term residents move away and are replaced with new residents. This process may accelerate gentrification if the post-disaster area is reconstructed to higher standards [**?**], or concentrate socioeconomically disadvantaged persons if the disaster deteriorates local property values and public infrastructures, thereby decreasing local resettlement by more advantaged residents [**?**]. Residential mobility may also lead to a permanent decrease in the population. Population losses are 'one of the most destructive ills of post-disaster cities' [**?**, p. 87]. Large-scale residential mobility reduces community cohesion and hinders recovery [**??**]. Populations losses as low as 5% to 10% are accompanied by significant economic impacts such as the reduction of the taxpayer base and reduced demand from local businesses [**?**]. Displaced persons experience higher unemployment rates [?], limited participation in community recovery decisions [?], and the cultural trauma of forcibly moving to a new community [?]. Thus, there are multiple benefits of mitigating post-disaster residential mobility.

As communities continue work to understand and mitigate their disaster risks computer simulations become a valuable tool to inform planning, for example, as in the HayWired Scenario study in the San Francisco Bay Area [?]. However, these are complex problems and many of models employed in these simulations are still being developed and refined by the engineers, planners, and social scientists leading these studies. Due to infancy of this field, there exist limitations in our ability to simulate of certain processes. Post-disaster residential mobility is among them. The existing models often assume that residents will wait long periods to return home and repair as long as they can finance it [e.g., ??], or that residents are perfectly rational decision makers who will maximize their monetary gains [e.g., ??]. However, scholars have demonstrated that economic concerns alone cannot explain post-disaster return decisions [??]. There are non-material aspects of the decision to migrate [?], such as the sense of loss associated with a change in the environment in which one lives, i.e., solastalgia [??]. The aforementioned HayWired Scenario study acknowledges the effect of one's physical and social ties to a place, often called place attachment, on post-disaster decisions [e.g., ?, p. 11]. Due to the lack of more sophisticated models, the study assumes that a portion of the young, high-income renters have low place attachment and are the most likely to migrate out of the Bay Area after an earthquake. While the assumption is justifiable, the insight that the areas with high concentration of young, high-income renters are the most prone to population losses is a direct consequence of the model assumptions.

The challenges in the simulation of post-disaster decisions identified in the HayWired Scenario study provide the inspiration and practical foundation for this study. Here, the goal is to develop a methodology to simulate the post-disaster decision of the residents with less strict assumptions. This is done in two steps. First, we review the relevant literature on post-disaster decisions and place attachment to identify suitable ways to estimate the place attachment of the residents of a community. Then, we use place attachment as the lenses through which residents of single-family homes evaluate the benefits of staying and repairing against the option to move away. This approach by-passes the need to define *a priori* the demographic groups most prone to leave the community during recovery. Three contributions are offered in this study. First, we develop a methodology to assess the place attachment of households. The methodology employs publicly available data and is transferable to other metropolitan regions in the US. Second, we describe a workflow to assess expected earthquake-induced losses and housing recovery time for urban communities. The workflow is applied to a case study of San Francisco, and impacts of moment magnitude (M_w) 6.5, 7.2, and 7.9 earthquakes are examined. Lastly, we contrast the expected losses and recovery times for individual households with the results for place attachment. The goal is to identify the neighborhoods whose residents will jointly experience high losses, long times to regain a sense of normalcy and low place attachment. We argue that the combined pressure from these three factors is a better predictor of population loss than financial considerations alone. We identify the neighborhoods and socioeconomic groups with the highest potential for population losses. This information may help a city target neighborhoods and demographic groups that need help and foster a healthier post-disaster recovery.

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

2 Place Attachment and Disasters

Place attachment describes the deeply rooted bonds that individuals develop with their communities. Definitions of place attachment are vast and discipline-dependent [??], often being intertwined with the definitions of place identity, place dependence, sense of place, and rootedness [?]. ? provide a comprehensive and up-to-date review of the literature on place attachment. The popular ? framework is adopted in this study. It defines place attachment as the tripartite of person (individually or collectively determined use and meanings), psychological process (affective, cognitive, and behavioral components), and place dimensions (the symbolic aspects, whether social environment and social meanings, and the physical environment, whether natural or built) [?]. Here, we focus on the place dimension of place attachment which is argued to be the most important by ?. For brevity, we use the term place attachment to refer to its place dimension in the following. Place attachment describes the qualities and specificity of the location to which one is attached and can be divided into social and physical place attachment [??]. Social place attachment has been defined as one's social ties and sense of belonging to a location, e.g., neighborhood [?]. Physical place attachment is related to one's dependence on the amenities or resources provided by a location to support one's goals. Thus, place attachment may be related to houses, streets, parks, and other outdoor settings.

Place attachment has been shown to affect people's risk perception, disaster preparedness, and how they respond to disasters. Attached persons tend to minimize risks that they are exposed to making them reluctant to landscape changes or moving out of risky areas such flood plains or wildland-urban interface zones [??]. Conversely, place attachment may positively influence disaster preparedness. It increases the likelihood that one will take action to prevent harm to the places they are attached to [???].

In post-disaster scenarios, persons forcibly separated from their usual living place may experience grief, similar to a situation where people lose an important social relationship [?]. Residents with weak or no place attachment are more likely to move away in the face of environmental change such as a disaster [?]. Conversely, residents who perceive their neighborhoods as an excellent place to live have been shown to be two to three times more likely to stay (or return) after a disaster [?]. Scholars have demonstrated that place attachment is a better predictor of willingness to move away than whether or not the residents had been born and raised in the region [?]. The influence of place attachment is stronger for homeowners who tend to report a larger social and emotional place connection than renters [?].

Hurricane Katrina has been extensively investigated from the perspective of disaster-induced out-migration. Cross-sectional studies of the recovery after Hurricane Katrina show that New Orleans' poorest permanently out-migrated [????]. However, ? argues that economic factors alone cannot explain the decisions of residents of New Orleans as many displaced New Orleanians returned to the city even if that entailed paying an economic price. ? identified that among African American and Vietnamese communities social capital and place attachment synergistically contributed to their decision to return. Among those who returned to the Ninth Ward after Hurricane Katrina, ? identified that they insisted that New Orleans provided a sense of place that cannot be found or replicated elsewhere.

3 Integrating Place Attachment into Housing Recovery Simulations

Post-earthquake housing recovery models are varied in detail and scope and often tailored to specific applications. Figure 1 illustrates some common features in such models. On the far left is the simulation of the ground motion. The output from these simulations is intensity maps of ground accelerations and displacements at the locations of each building of interest. The next step is damage and loss assessment. For portfolios, due to lack of detailed data, damage and losses are often estimated as a function of the ground motions using fragility curves and estimates of the building replacement cost per square foot [?]. More sophisticated approaches may split the damage and loss assessments into multiple tasks. Important outputs from the damage and loss assessment step are estimates of repair time (i.e., worker-hours needed to repair the building) and repair costs for each building. The next step is to assess the homeowners' access to financing. Financing is tied to demographic characteristics, e.g., high-income persons may have easier access to private loans but not qualify for public grants. The mismatch between repair costs and financing available can be used to estimate the indebtedness for each homeowner. If repairs would incur high debt, the homeowners may opt to sell the property. Homeowners who can finance repairs compete for the limited available recovery resources, such as construction workers. Because resources are scarce, some homeowners experience delays in their recovery processes. Thus, the recovery time for each building may substantially exceed the estimated repair time.

The two shaded boxes in Figure 1 indicate the contributions from this study. First, we use demographic data to estimate the strength of place attachment for each homeowner in our study. Second, we use the estimated indebtedness, recovery time, and place attachment to assess the likelihood of a homeowner engaging in repairs. Based on the literature discussed above, we assume that households with low place attachment are less willing to take on debt and wait long periods to return home after a disaster. That is, homeowners in these conditions are prone to moving away. Thus, to assess the potential population losses, we have two main tasks. The first is to assess place attachment for our population of interest. The second is to simulate housing recovery to predict household debt and displacement time of displacement. Thus we can assess, for each household, (i) the probability that its place attachment is low, $P(A = low)$; (ii) the probability that the debt incurred from the repair exceeds a given value, $P(D > d)$; and (iii) the probability that the housing recovery time exceeds some threshold, $P(T > t)$. Some of these probabilities will change based on the demographics of the household ($\boldsymbol{X}$) and the impact of the earthquake (E). Thus, the probability of residential mobility after an earthquake for a household, $P(M|\boldsymbol{X}, E)$, is defined here as

$$P(M|\boldsymbol{X}, E, A, t, d) = P(A = low|\boldsymbol{X}) \times P(D > d|\boldsymbol{X}, E) \times P(T > t|\boldsymbol{X}, E) \tag{1}$$

Note that Eq. 1 assumes the conditional statistical independence between place attachment, losses, and repair time, given demographics and impact. That is, A, D, and T are independent given $\boldsymbol{X}$ and E, but they will be dependent overall because they all depend on $\boldsymbol{X}$ and E. This formulation allows us to build predictive models that maintain overall dependence among variables, while simplifying the treatment of model prediction residuals. In the following sections, we introduce the models needed to assess the three terms on the right-hand side of Eq. 1.

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/RecoveryModel.png

Figure 1. Schematic representation of a housing recovery simulation model. The numbers in each box indicate the section in which the models are discussed. The shaded boxes indicate the contributions of this study.

4 Place Attachment and Satisfaction

Place attachment can be challenging to measure. However, place attachment has been consistently demonstrated to be correlated to place (housing and neighborhood) satisfaction - even if the nature of this correlation is a debated topic [?]. Housing satisfaction has been defined as the contentment felt when housing aspirations are met in the actual housing inhabited [??]. Analogously, housing dissatisfaction has been suggested as metric of the gap between housing aspirations and current housing conditions [?]. Neighborhood satisfaction is broader, encompassing one's social networks. Social bonds take time to build, and the longer people live in an area, the more friends they are likely to have, and stronger their place attachment [??]. Thus, a household may be dissatisfied with a high-quality, well-maintained home because the housing costs are too high, (e.g., affordability issue) or because the family is being expanded (e.g., suitability issue). An affluent neighborhood may not satisfy a household if the commute to work is too long, or if their relatives do not live in it.

Here, we use housing and neighborhood satisfaction as proxies of household place attachment. Measures of housing and neighborhood satisfaction are publicly available from the American Housing Survey (AHS)[?]. To measure housing satisfaction, the AHS asked the respondents '[o]n a scale of 1 to 10, how would you rate your home as a place to live? (10 is best, 1 is worst).' An equivalent question was asked regarding neighborhood satisfaction. Previous studies have used AHS data on place satisfaction to gain insights on social capital building [?], demographic disparities [???], risk of housing problems [?], and to evaluate the success of subsidized housing programs [?].

In 2019, 1,883 occupants of single-family homes in San Francisco responded the survey answering both questions about place satisfaction and providing their demographic profile. We limit our scope to single-family buildings due to limitations that arise when investigating the recovery of multi-family buildings, which are discussed later. The responses from the 1,883 households are called 'samples' in the following. An overview of the AHS data employed in this study is shown in Table 1. These demographics are chosen because they have been correlated to socioeconomic vulnerability in past studies [?]. As described in the following, we use these data to build a model to estimate the housing and neighborhood satisfaction of households and use these as proxies for place attachment.

Figure 2 presents the prevalence of the neighborhood scores across selected demographic groups. Few households indicated scores below six, so values ≤ 6 are grouped. We note a significant difference in the level of housing and neighborhood satisfaction reported by renters and owners. Recent-immigrant, non-White, and renter households are the least likely to report high housing and neighborhood satisfaction.

We assume that a household has low place attachment when it jointly presents low housing and neighborhood satisfaction, that is

$$P(A = low|\boldsymbol{X}, E) = P(HS < s_h(\boldsymbol{X})|\boldsymbol{X}, E) \times P(NS < s_n(\boldsymbol{X})|\boldsymbol{X}, E, HS) \qquad (2)$$

where HS and NS are the housing and neighborhood satisfaction, and $s_h(\boldsymbol{X})$ and $s_n(\boldsymbol{X})$ are thresholds of housing and neighborhood scores that characterize 'low satisfaction.' The conditional dependence of NS on HS reflects a significant correlation between housing and neighborhood satisfaction that we identified while preparing the data for this study. The users define the thresholds $s_h(\boldsymbol{X})$ and $s_n(\boldsymbol{X})$. In this communication we employ the premise that all households have a consistent interpretation of satisfaction. However, the approach provides flexibility for the thresholds to be adjusted based on the household demographics. For example, disadvantaged households may indicate high satisfaction (e.g., 8) with a deteriorated home because

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Table 1. Demographic data available from the American Housing Survey. Demographics marked with a * are used as independent variables.

Demographic	Categories
Housing Unit	
Year built	real number
Building value	real number
Householder	
Immigrant	yes, no
Race	White, Black, Asian
Hispanic	yes, no
Bachelor degree	yes, no
Gender	female, male
Household	
Tenure	owner, renter
Income bracket	high, moderate, low
Income value	real number
Size	integer number
Year moved in	integer number
Has children	yes, no
Has elderly	yes, no
Has disable member	yes, no
Housing score*	integer number [1,10]
Neighborhood score*	integer number [1,10]

the alternative is homelessness. In these cases, the user may opt to use $s_h(Income = Low) = 9$. We next construct the models for $P(HS < s_h|\boldsymbol{X}, E)$ and $P(NS < s_n|\boldsymbol{X}, E)$ from the AHS data. One challenge that arises when building a model for $P(HS < s_h|\boldsymbol{X}, E)$ is that if $s_h = 6$, for example, the majority of the samples have scores above the threshold. A model fitted to this imbalanced data is prone to be biased towards the majority class. To mitigate this bias when predicting households' housing and neighborhood scores, we employ an approach that combines minority oversampling with an ensemble classifier.

4.1 SMOTEBoost Classifier

Several techniques exist to reduce the class imbalance. Undersampling consists of using only a subset of the samples from the majority class so that balance is achieved. A common drawback of undersampling is the loss of information from discarding many samples in the majority class. Conversely, oversampling consists of increasing the number of samples in the minority class to match the majority class samples. Oversampling is often achieved by drawing with replacement from the samples in the minority class. A potential problem in this approach is that oversampled sets may contain many copies of the same sample, leading to overfitting.

Another group of techniques focuses on creating synthetic samples from the minority class. In this study, we employ the Synthetic Minority Oversample TEchnique (SMOTE) [**?**]. The SMOTE creates synthetic

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_SatisfactionPerDemographic.png

Figure 2. Percent of respondents indicating a given housing or neighborhood score for selected demographic group. The vertical axis is normalized by the number of households in each group.

samples of the minority class based on its nearest K minority neighbors. Estimators fit using the SMOTE are less prone to overfitting and do not incur a loss of information.

Once the class imbalance in the classification problem is adjusted, we employ Adaptive Boosting (AdaBoost) to perform the model fitting [**?**]. AdaBoost has three main ideas. First, it uses many weak learners rather than a single more sophisticated learner. Weak learners are classification models that are intentionally simple and which do not have strong prediction capacity. In most applications, the weak learners employed in AdaBoost are decision trees with a single node, often called 'decision stumps.' The predictions of each stump are later combined to determine the most probable class for each sample. Second, classifications made by each stump are weighted by the errors it makes. Thus, the more incorrect one stump's prediction is, the less weight its vote has in the final classification. This property contrasts with the uniform weights used in a Random Forest, for example. Third, errors made by each stump are used to inform the creation of the next stump. This adaptive behavior once again contrasts with the independent trees in a Random Forest. Using these three ideas, an ensemble of weak learners trained via AdaBoost can make accurate predictions while being less susceptible to overfitting [**?**].

Combining the SMOTE and boosting algorithms is called 'SMOTEBoosting' and it improves prediction in imbalanced data sets [**?**, p. 107-119]. The data in Table 1 were used to fit the models using SMOTEBoosting. The model for housing satisfaction uses the modified housing scores (i.e., scores below six were grouped) as the dependent variable. With the exception of the neighborhood scores, all other variables in Table 1 are used as independent variables. We use the housing satisfaction model to predict the housing scores for the 1,883 samples. The model for neighborhood satisfaction is fitted using the modified neighborhood scores (i.e., scores below six were grouped) as the dependent variable. The demographics in Table 1 and the predicted housing scores are used as independent variables in the neighborhood model. The predicted housing scores are used in place of the surveyed housing scores to simulate the behavior of the SMOTEBoosting classifier when applied to a new data set for which surveyed scores scores are not available.

To test the SMOTEBoost classifier, we first split the 1,883 samples in the AHS into a training set (1318) and a testing set (565). The classifier is trained on the training set and used to predict the testing set. We tested the ability of the classifier to predict 'low satisfaction' using different scores as the thresholds. That is, for each household in the testing set we predicted if $HS < s_h$ and $NS < s_n$ for different s_h and s_n. To assess the quality of the classifier, Figure 3 shows the receiver operating characteristic (ROC) curves for each threshold. ROC curves summarize the trade-off between the true positive and false-positive rates for a predictive model using different probability thresholds. The ROC curve for a naive model that is correct 50% of the time is a straight line with a 45-degree slope. This line is shown in Figure 3. The area under the curve (AUC) for this model is equal to 0.5. Models with AUC>0.5 outperform the naive model, and AUCs closer to the unity are desirable. The models for neighborhood scores are slightly less accurate because they use the predicted housing scores as an independent variable. However, Figure 3 shows that all models fitted for all thresholds offer significant improvements over the naive model. These results demonstrate the quality of the SMOTEBoost classifier and its suitability to assess the probability that a household has low place attachment as per Eq. 2.

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_ROCs.png

Figure 3. Receiver operating characteristic curves for the predicted (left) housing and (right) neighborhood scores considering different thresholds.

5 Post-earthquake Housing Recovery Simulations

This section describes the steps used to simulate post-earthquake housing recovery and estimate repair cost and time for each household. The steps described in Figure 1 are employed for this purpose.

5.1 Simulation of Buildings and Households

This study employs the Federal Emergency Management Agency's Hazus methodology [**?**] to build a portfolio of all single-family buildings in San Francisco from 2019 SimplyAnalytics [**?**] projections based on the 2010 Census data. The methodology allows us to estimate the structural type, code design level, and replacement cost for each considered single-family building. We associate one household (defined by the demographics in Table 1) with each single-family home in the portfolio. The demographics of the households are sampled from the distributions in each census tract, and correlations between demographics are not directly simulated. For example, if in Census tract T 50% of the households have a Black householder, i.e., P($Race = Black|T$)=0.5, and 30% have a low income, i.e., P($Income = low|T$)=0.3, the probability that a household has a Black householder and low income is P($Race = Black|T$)P($Income = low|T$)=$0.5 \times 0.3 = 0.15$. San Francisco is comprised of 184 Census tracts and this approach partially captures the spatial correlation that exists between demographics. For the example above, the Pearson correlation coefficient is $r(Race = Black, Income = low)$=0.55.

5.2 Simulation of Ground Motion, Damage, and Losses

Shaking intensities are simulated at the centroid of each census block group. The ground-shaking simulations provide estimates of the peak-ground acceleration (PGA) and spectral acceleration (SA). The Open-Source Seismic Hazard Analysis (OpenSHA) Event Set Simulator is used to predict median values of PGA and SA [**?**]. The distributions of ground shaking at each location, and the correlations between spectral acceleration values at multiple periods and multiple locations, are predicted using empirical models [**???**]. Variability in predictions is captured by generating N realizations of ground shaking intensities associated with the given rupture of interest.

We consider buildings to potentially be on liquefiable soil, p_l, equal to the fraction of liquefiable soil within the census block group. For building on liquefiable soil, the probability of liquefaction is calculated as a function of the on-site PGA, the magnitude of the earthquake, the liquefaction susceptibility, which is assumed to be 'high,' and a five-feet ground-water level which is the default value for [**?**, Eq. 4-20]. Using the probability of liquefaction and the on-site PGA, we calculate permanent ground deformation considering the expected lateral spreading and ground settlement [**?**, Section 4.2.2.1.4]. The output from this assessment is a vector of permanent ground deformations. The estimated ground shaking and ground deformations are used to estimate damage using the methodology described in Sections 5.4 through 5.6.3 in **?**. Vectors of structural and non-structural damage states are output at this step, and then associated to vectors of repair costs ($\boldsymbol{C_r}$) and repair times ($\boldsymbol{T_r}$) for all buildings [**?**, Table 15.9].

5.3 Simulation of Housing Recovery Debt

We adopt the model of **?**, with modifications, to simulate recovery financing. This model was developed considering post-earthquake housing recovery financing for a household in San Jose, California. Four funding sources are included: earthquake insurance, bank loans, Small Business Administration (SBA) loans, and Community Development Block Group for Disaster Recovery (CDBG-DR) grants. If the claims and applications are successful, insurance and loans are disbursed within weeks. The grants from CDBG-DR may take months to years to be disbursed since these funds have to be approved by Congress after each disaster. Funding from the Federal Emergency Management Agency Individuals and Households Program is not accounted for due to these grants being relatively small when compared to the expected losses [?]. For each funding source, the model provides the probability of receiving funding, the expected amount received, and the time to receive the funding.

The original model considers that homeowners who cannot obtain total financing cannot repair. Because we are interested in identifying the burden of repairing one's home, we assume that all homeowners will attempt to repair it. However, the gap between the financing needed to repair the home (R_c), and the financing that homeowners can obtain from insurance (F_i), bank loans (F_b), SBA loans (F_{sba}), and CDBG-DR grants (F_{cdbg}) is assessed and it is defined as the 'indebtedness' associated with the disaster, $D(\boldsymbol{X}, E)$, that is

$$D(\boldsymbol{X}, E) = R_c(E) - \Big(F_i(\boldsymbol{X}, E) + F_b(\boldsymbol{X}, E) + F_{sba}(\boldsymbol{X}, E) + F_{cdbg}(\boldsymbol{X}, E)\Big) \tag{3}$$

where $\boldsymbol{X}$ and E denote dependence on the household demographics and losses associated with the earthquake being considered. We assume homeowners will use their savings, sell non-liquid assets, or get high-interest loans to pay this debt. Thus, this amount is used to proxy the additional challenges households need to surpass to repair their homes. The probability that D exceeds a threshold d, $P(D > d|\boldsymbol{X})$, is needed for Eq. 1 and is given by

$$P_h(D > d|\boldsymbol{X}) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(D_i > d|\boldsymbol{X}, E) \tag{4}$$

where N represents the number of realizations of the earthquake of interest and $\mathbf{1}$ is an indicator function that returns 1 if $D_i > d$ and zero otherwise. Note that $P_h(D > d|\boldsymbol{X})$ is only calculated for owner households. Renters are not responsible for paying for repairs. Therefore, we do not consider debt as a factor influencing their decisions. The consequence of this assumption is that homeowners require one more adverse condition to be present for them to be susceptible to housing mobility. This assumption reflects findings from empirical studies that identified that homeowners are less likely to move away in the aftermath of disasters.

5.4 Simulation of Recovery Times

We estimate an available construction workforce of 1,000 crews in San Francisco, based on data from the Business Analyst Online [?]. The availability of other types of workers (e.g., inspectors or engineers) is not accounted for. The housing recovery of single-family homes is often 'bottlenecked' by the availability of contractors [?]. We assume that if the demand for contractor crews is higher than the local supply, workers come from nearby communities over time up to a limit. If the available supply exceeds this limit, workers leave the city also over time. The limit used in this study is 80% of the current demand, which yields housing recovery rates similar to those observed in previous large disasters [?]. Thus, during recovery the number of construction crews in the community is at least 1,000, but it can increase to 80% of the total demand if the demand is higher than 1,000 crews.

The recovery is simulated over discrete time steps. At each time step, households that have obtained funds (Section 5.3) 'request' a contractor crew. If the number of contractor crews available exceeds the number of requests in the current time step, all households that requested a contractor can start repairs. The contractors stay allocated to the households for a time equal to the repair time of each building, T_r. After that, they

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

become available for another household. When there are more requests than contractors, available contractors are allocated to the household that made the earliest request. This process produces the recovery trajectories indicated at the far right in Figure 1.

The recovery simulation allows us to estimate recovery time for each homeowner. The recovery time for a household with demographics $\boldsymbol{X}$ after earthquake E, $T(\boldsymbol{X}, E)$, is:

$$T(\boldsymbol{X}, E) = T_f(\boldsymbol{X}, E) + T_c(\boldsymbol{X}, E) + T_r(E) \tag{5}$$

where T_r is the repair time, T_f is the time needed to obtain financing, and T_c is the time need for a contractor to become available to work in the building. That is, T_c=0 if there are contractors immediately available in the community. The recovery trajectories for each earthquake simulation provide the probability that a household's recovery time exceeds t given its demographics, $P_h(T > t|\boldsymbol{X}, E)$. This probability is the last factor needed to assess Eq. 1. For each household, this probability is

$$P_h(T > t|\boldsymbol{X}, E) = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(T_i > t|\boldsymbol{X}, E) \tag{6}$$

where T_i is the recovery time after each earthquake simulation $i = 1, \ldots, N$. Thus, $P(T > t|\boldsymbol{X})$ is the probability that the recovery time for a household will exceed t after an earthquake.

6 Assessing Potential Population Loss

To combine the models described above and demonstrate the application of the proposed framework, we present a case study that investigates the impact of earthquakes on 124,563 single-family buildings in San Francisco. Three earthquakes occurring on the San Andreas fault with moment magnitudes of 6.5, 7.2, and 7.9 are selected. These earthquakes represent planning scenarios that the City of San Francisco has considered. One hundred simulations post-earthquake housing recovery for each earthquake are conducted to capture uncertainty. In each simulation, random variables representing the ground motion, damage, losses, repair time, repair financing, and recovery speed of each household take different values. Conversely, household demographics and physical place attachment are simulated once before the first simulation is run.

The case study only includes the recovery of single-family homes for several reasons. Single-family homes are less varied in terms of their structural features than multi-family homes. The financing mechanisms available to repair single-family homes are more straightforward than those available to multi-family homes. More importantly, multi-family homes are often owned or managed by companies or strata, and the processes involved in deciding to rebuild are not trivial to simulate. It is thus unclear if place attachment plays a pivotal role in the decision to repair multi-family buildings. Figure 4 shows the spatial distribution of single-family homes in San Francisco. These are concentrated on the west side of the city, in wealthier neighborhoods, which are also closer to the San Andreas fault, the source of the earthquakes considered in the following analyses.

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_SFResidences.png

Figure 4. Number of single-family residences per census tract in San Francisco.

6.1 Place Attachment

The first step in using Eq. 1 to estimate potential population loss is to assess the number of households with low place attachment. Once the demographics of each household are simulated from Census data, the SMOTEBoost classifier is used to estimate their housing and neighborhood satisfaction. Households with housing and neighborhood scores below seven are considered to have low place attachment. Figure 5 shows the probability of low place attachment for households in each Census tract. Lighter colors indicate areas whose residents are more prone to residential mobility. The areas in white contain fewer than 50 single-family residences (e.g., Golden Gate Park). The Northeast and Southeast parts of the city present the highest percentage of households with low pre-disaster place attachment. These areas have a significant number of households with low income and under-represented minorities. The earthquakes considered in this study come from the San Andreas fault to the West of the city. Thus, based on distance from the source, the neighborhoods with the lowest place attachment are on average exposed to lower ground motion intensities.

Figures/Fig_LowPlaceAttachment.png

Figure 5. Estimated percentage of households with low place attachment. Areas in white contain fewer than 50 single-family residences. Lighter colors indicate areas whose residents more prone to residential mobility.

6.2 Housing Recovery Simulations

The impact of the three earthquakes on the housing stock in Figure 4 is summarized in Table 2. Not surprisingly, the number of buildings with severe or complete damage, and the losses, increase with the earthquake magnitude. It is assumed that only buildings at severe or complete damage require major repairs

[?]. Repair time is a function of the damage state, hence it is constant. Repair delay measures the time from the day of the event to the moment when repairs start. Repair delay is bound by the ability of households to obtain financing and the competition for the skilled workforce in the community. There is significant variability in the repair delay (e.g., due to the funding sources available to each homeowner). The mean repair delay is slightly higher for completely damaged buildings, reflecting the longer period needed to finance the more expensive repairs to these buildings. The mean repair delays also increase slightly with earthquake magnitude, reflecting additional delays due to supply constraints when damage is more widespread.

Table 2. Expected impacts of the three earthquakes on the building portfolio.

Earthquake magnitude [M_w]	Structural damage state	Number of buildings	Mean loss per building [US$]	Repair time [days]	Mean repair delay [days]
7.9	Severe	22,269	131,495	90	496
	Complete	16,584	217,541	180	526
7.2	Severe	11,488	92,565	90	499
	Complete	5,694	161,159	180	528
6.5	Severe	1,003	60,954	90	489
	Complete	443	125,958	180	498

Figure 6 shows the spatial distribution of the impacts from the M_w 7.9 earthquake, the most damaging of the three. The results are aggregated by Census tract. The maps show the average repair costs for the buildings in the tract considering 100 realizations of the earthquake, calculated as

$$L_t = \frac{1}{100 \cdot Nb} \sum_{j=1}^{100} \sum_{i=1}^{Nb} R_{c,i,j} \tag{7}$$

where Nb is the number of buildings. Losses are affected by distance to the San Andreas fault, building value, soil conditions, and the age of the buildings.

Recovery is then simulated over eight years following the earthquake for each damage realization, using 14-days time steps. Each simulation takes about 20 minutes to run in a high-performance computer and results in one recovery curve. The average recovery curve for each earthquake is presented in Figure 7. Although not shown in the figure, there is significant variability in the immediate damage, i.e., the drop in the number of habitable residences at time $t = 0$. The recovery progresses quickly until about three years after the earthquake. Around this time, all households that are not dependent on public grants funding have repaired their homes. After this point, the constant slope of the recovery curves reflects the slow distribution of public grants funding in time.

The housing recovery simulations are used with Equations 4 and 6 to estimate the number of households experiencing long recovery time and high debt. The equations require the thresholds d and t to be defined. These thresholds can vary based on household demographics. For example, a young family with no children has fewer restrictions to moving away and may be less willing to wait long periods to repair their home. San

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_MapLoss79.png

Figure 6. Average repair cost per Census tract per building expected after the M_w 7.9 earthquake. Areas in white contain fewer than 50 single-family homes.

Francisco also has a strong housing market, and the prospect of long-term gains may justify housing repair costs that would not be viable in other parts of the country. For exploratory studies, 'what-if' scenarios can be used to determine lower and upper bounds. Here, we consider two combinations of t and d. The first represents a household with strict thresholds to decide to stay; namely, recovery should cause a debt that is less than its annual income and be finished within one year. The second combination represents a household willing to incur a debt equal to twice their annual income and wait up to two years to return home. We adopt t=2 years as the upper bound because support to disaster-induced displaced persons typically lasts no longer 24 months [**?**].

The results from the two 'what-if' scenarios are presented in Figure 8. Lighter colors correspond to the more

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_RecoveryCurves.png

Figure 7. Post-earthquake housing recovery curves. The curves represent average results from the 100 simulations of each earthquake.

strict scenario. Two key insights are drawn from the figure. First, the choice of the thresholds t and d has a similar impact on the results as the choice of a different earthquake magnitude. These results highlight the need for community-specific quantitative research to assess the willingness of households to spend on and wait for recovery. Second, in the case study, there is a low probability ($< 3\%$) that a homeowner will not be able to afford repairs using funding from insurance, public and private loans, and grants after an earthquake. However, there is a non-negligible probability (30% for the M_w 7.9 earthquake) that the recovery process will take longer than two years, i.e., the upper bound assumed for t. Our residential mobility results are thus primarily controlled by the recovery time rather than costs.

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_Sensitivity.png

Figure 8. Percentage of households experiencing high repair costs and recovery time after an earthquake on the San Andreas fault. Symbols indicate the mean results. Dashed lines indicate the one-standard deviation confidence intervals.

6.3 Population Loss

The results from Figs. 8 and 5 allows us to use Eq. 1 to estimate the probability that a given household will move away after a earthquake, $P_h(M|\boldsymbol{X}, E)$. In turn, it allows us to estimate the potential population loss at different parts of the city as

$$L_c(E) = \frac{1}{H} \sum_{i=1}^{Nb} P_h(M|\boldsymbol{X}_h, E) \tag{8}$$

where $L_c(E)$ is the population loss in a census tract with Nb single-family homes. Equation 8 can be applied to any combination of earthquake and thresholds t and d to gain insights about potential population losses. For brevity, Figure 9 shows the results for the M_w 7.9M earthquake considering the strict scenario. The map is overlayed with the contour lines of the planning districts in the city. The Richmond and Outter Sunset districts are the most prone to losing population after any earthquake.

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_PopulationLoss79.png

Figure 9. Potential residential mobility as percentage of Census tract population considering the strict scenario for the M_w 7.9 earthquake scenario.

The socioeconomic characteristics of the at-risk households are as crucial as their locations. Figure 10 presents the demographics of the households most likely to leave the city during recovery under the assumptions in the case study. The total heights of the bars represent the results for the M_w 7.9M earthquake, and the dashed lines are the results for the M_w 7.2M earthquake. Expected population losses after the M_w 6.5M are minimal and not shown. Disparities are observed across building and household characteristics. Low-income renters who occupy the more physically vulnerable buildings (i.e., low-code buildings) are the most prone to residential mobility following an earthquake. Disparities across racial, ethnic, and immigrant groups are minor. Note that the earthquake intensity is higher on the west side of the city and that we only include single-family buildings in the analysis. Non-white immigrant households are more prevalent on the east side of the city, and these households tend to occupy multi-family buildings.

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Lastly, the disparities between owners and renters are higher for the M_w 7.2M earthquake.

Figures/Fig_PMDemographics.png

Figure 10. Probability of out-migration for households in different socioeconomic groups. The heights of the bars indicate the results for the M_w 7.9 earthquake. The dashed lines indicate the results for the M_w 7.2 earthquake.

To conclude the analyses in this case study, we compare the results obtained from the proposed approach (shown in Fig. 10 to two other approaches. In the first approach, indicated as 'None,' place attachment is not accounted for, e.g., the *PA* terms is not included in Eq. 1. In the second approach, indicated as 'HayWired', only young, high-income renters are assumed to move out, which is similar to the assumption in the HayWired Scenario study. The results considering the $M_{7.9}$ earthquake are shown in Figure 11. The probabilities of residential mobility across all demographics for each approach are $P_{None}(M|E)$=0.14, $P_{HayWired}(M|E)$=0.033, and $P_{Proposed}(M|E)$=0.034. Thus, the if place attachment is disregarded, near five times more households are considered at prone to moving away. To gain insights on how each approach distributes P($M|E$) across demographies, the results in Figure 11 are normalized. That is, the ordinate axis shows P($M|E, X_i$)/P($M|E$) and values above the red line at Y=1 indicates a demographic is over-represented among those expected to move away. The normalization eliminates the influence of the selection of thresholds in Eq. 1. The major disagreement between approaches is in regards to the influence of income on P($M|E$). If place attachment is not accounted for (i.e., 'None' approach), P($M|E$) is evenly distributed (i.e., all close to Y=1.) In the 'HayWired' approach, high-income households are the only ones at risk of moving away permanently. In the proposed approach, lower income residents are more prone to residential mobility. Thus, the proposed approach has similar data requirements, yields results similar to those from the HayWired Scenario approach albeit using a less strict assumption, and is more consistent with empirical findings which show that socioeconomically disadvantaged persons are more likely to out-migrate [**?????**].

Costa, R., Wang, C. and Baker, J. W. (2022). “Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses.” Natural Hazards Review, 23(4), 04022021. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

Figures/Fig_PMDemographics_2.png

Figure 11. Probability of out-migration for households in different socioeconomic groups using different approaches to simulate the effect of place attachment on the decision to stay: ’None’ considers place attachment has no effect; ’HayWired’ considers only young, high-income renters are prone to moving away and; ’Proposed’ is the approach described in this study. Values above the red horizontal line indicate a demographic group is overrepresented among those expected to move away.

7 Discussion

The results in this case study show an overlap between the demographics of the households with low place attachment and those with difficulty to promptly finance post-earthquake housing repairs. We emphasize that these results do not directly predict population loss. Rather, they identify households who have to spend more (relative to their incomes), have to wait longer, and have fewer reasons to stay. San Francisco’s long-term resilience plan includes measures to ’increase neighborhood quality of life, overall physical conditions and, to build community capacity" [?, p. 109]. These measures are expected to improve housing and neighborhood conditions and foster place attachment. The framework presented in this paper helps to assess the benefits of these actions.

Nonetheless, the problem tackled in this study is complex, and the framework has limitations. Per the **?** framework, we only tackle the place dimension of place attachment. The person and process dimensions are not accounted for, and our findings should be interpreted in light of this simplification. Furthermore, we use satisfaction as a proxy of place attachment. Although this has a basis in the literature, it is not a direct measure of place attachment. Moreover, the American Housing Survey satisfaction scale may not be consistently interpreted by all respondents. We envision that these limitations can be overcome if questions related to place attachment are included in the American Housing Survey, [e.g., **?**, p. 14]. That data could be employed in the proposed framework with minimal modifications.

Housing prices in San Francisco are much higher than the national average. Thus, the decision of homeowners to stay and repair or sell and leave may be significantly affected by their perception of the future monetary benefits from having a home in San Francisco. Moreover, we assume that each homeowner possess one home in San Francisco, and that homeowners wish to repair their homes as soon as possible. More detailed data regarding homeowners with multiple homes would allow this assumption to be refined. While our case study demonstrates the application of the proposed models, further investigation is needed to determine if the empirical findings regarding the role of place attachment on post-earthquake decisions observed in other communities are transferable to San Francisco.

Another limitation is with our model for financing the repair of rented housing. Owners of rented buildings do not have the same access to financing that is available for owner-occupied buildings. Thus, their financing processes are unclear. Here we optimistically assume that all landlords are high-income persons with sufficient insurance and private financing. With a more realistic model requiring longer waits for funding in some cases, a larger number of renter households would be prone to moving away. However, no such realistic models are available at present. Lastly, the model assumes that buildings will be repaired to a pre-disaster state. Given San Francisco's competitive real state market, it is likely that some homeowners and landlords will improve damaged buildings or replace them with higher-density units.

These limitations stem from the complexity of the problem and the challenges of anticipating human decisions. However, the proposed framework can be used in 'what-if' studies, and these assumptions are consistent across all scenarios considered. In that case, the impact of these limitations is minimized, and the comparable results yield meaningful comparisons.

8 Conclusion

In this study, we integrate place attachment considerations into housing recovery simulations. Place attachment is used as a surrogate for willingness to rebuild. We identify the households with low place attachment and whose housing recovery process is expected to be the most challenging. Our premise is that households with low place attachment are less willing to take on debt and wait extended periods to restore their livelihoods. We introduce a classification algorithm that combines the Synthetic Minority Oversampling TEchnique and Adaptive Boosting (SMOTEBoost) to estimate household place attachment

Costa, R., Wang, C. and Baker, J. W. (2022). "Integrating Place Attachment into Housing Recovery Simulations to Estimate Population Losses." Natural Hazards Review, 23(4), 04022021.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000571.

from data from the American Housing Survey (AHS). We use the place attachment estimates to study post-earthquake decision of households. We introduce a housing recovery simulation framework to estimate repair costs and housing recovery time for single-family buildings. We combine the place attachment, repair cost, and repair time results to estimate population losses. The place attachment assessment and the housing recovery simulations are decoupled. Thus, the place attachment assumptions can be revised without rerunning the computationally expensive housing recovery simulation. While we focus on post-earthquake decisions, the SMOTEBoost algorithm can be used to assess place attachment and investigate post-disaster decisions after other types of extreme events such as hurricanes and floods.

The application of framework is demonstrated in a case study of the potential population loss in San Francisco during the recovery from hypothetical earthquakes on the San Andreas fault. The case study quantifies housing repair costs (relative to household income), time to secure funding, and building repair time for 124,563 single-family households in San Francisco. The potential population loss is investigated under different scenarios. The results indicate that low-income renters occupying older buildings are the most prone to moving away after a disaster.

The framework presented in this study addresses the concern with the loss of populations with low place attachment which has recently emerged in studies of the regional impacts of earthquakes [**?**, p. 11]. Previous studies have ignored the influence of place attachment or assumed *a priori* which demographic groups are most prone to moving away after a disaster. In consequence, existing approaches provide limited insights on the demographic groups expected to struggle and perhaps move away during post-earthquake recovery. The framework in this paper is based on a review of studies of previous disasters. It employs data from the American Housing Survey which are publicly available for multiple locations in the US. It is empirically-based, can be employed to multiple regions, and is more nuanced in determining the demographic groups most prone to residential mobility. The framework can be incorporated in pre-disaster studies to estimate population losses using 'what-if' scenarios [**?**, e.g.] and to evaluate the benefits of taking actions to improve neighborhood cohesion [**?**, p. 109]. Some challenges remain in the application of the proposed framework, as highlighted in the Discussion section. Nonetheless, it offers a more robust procedure that can replace semi-heuristic approaches and help formalize the simulation of housing recovery.

9 Funding

Funding for this for this work was provided by the Stanford Urban Resilience Initiative.