12th National Conference on Earthquake Engineering
Salt Lake City, Utah
27 June - 1 July 2022
Hosted by the Earthquake Engineering Research Institute

# Enhancing Post-Disaster Recovery Modeling Through High-Fidelity Household Income Estimation

J. Zhang[1], R. Costa[2], Á. Zsarnóczay[3], and J. W. Baker[4]

## ABSTRACT

Recent disasters have shown that income plays a central role in determining the capacity of impacted households to cope with the shock and recover from it. Researchers often rely on random sampling to generate synthetic household income from aggregated Census data. This conventional approach imposes limitations towards proposing realistic policies. A method is introduced to deduce minimum household income for single-family homeowners using publicly available data via the tax assessor and current population survey. Post-earthquake housing recovery simulations are used to evaluate the advantages of the proposed income estimation method relative to the germane random sampling approach. Preliminary results with the proposed method show a reduction in the number of low-income households assigned to high-valued dwellings. Results also suggest that the random sampling approach leads to inflated recovery delays and overestimates the vulnerability of select low-income households.

## Introduction

Household income has been a centerpiece in the studies of disaster impact and recovery [1-4]. Not only is income an important proxy for estimating economic well-being, but it is also the primary factor in assessing eligibility for post-disaster funding [5,6]. Public data on household income (e.g., from the U.S. Census) is usually available in aggregated form that does not provide information about individual households. Approximate approaches, such as randomly sampling incomes, are often used to disaggregate this data and assign income to individual households. A random sampling approach may assign low-income households to high-value buildings. If a simulation estimates that these buildings sustain even light damage, the assigned low-income homeowners may not have the means to afford repairing it. Therefore, it is possible for such models to overestimate the impacts that disasters could have on low-income homeowners.

In this paper, we introduce a method that considers the correlation between household income and building value via mortgage eligibility at time of purchase. We use San Francisco as an example to illustrate the advantages of assigning household income with the proposed method. Transaction dates and values are

[1] Graduate Student Researcher, Dept. of Civil & Environmental Engineering, Stanford University, Stanford, CA 94305 (jzhang01@stanford.edu)

[2] Postdoctoral Scholar, Stanford Urban Resilience Initiative, Stanford University, Stanford, CA 94305

[3] Research Engineer, Dept. of Civil & Environmental Engineering, Stanford University, Stanford, CA 94305

[4] Professor, Dept. of Civil & Environmental Engineering, Stanford University, Stanford, CA 94305

Zhang J, Costa R, Zsarnóczay Á, & Baker JW. Enhancing Post-Disaster Recovery Modeling Through High-Fidelity Household Income Estimation. *Proceedings of the 12th National Conference in Earthquake Engineering*, Earthquake Engineering Research Institute, Salt Lake City, UT. 2022.

extracted for buildings in San Francisco from the City's tax assessor database [7]. The proposed method accounts for correlations between household income and building value to assign income for the households that own the 95,387 single-family detached residences registered in the tax assessor database. We then demonstrate the implications of accounting for such correlations in a post-earthquake housing recovery simulation for the City.

## Estimating Household Income

The proposed method uses the transaction date and price of a property to estimate the minimum income required to afford the monthly payments of a primary mortgage associated with the property. Monthly mortgage payments are calculated as:

$$M = P \cdot [\ i(1 + i)^n\ ] \,/\, [\ (1 + i)^n - 1] \qquad (1)$$

where $M$ is the monthly payment amount, $P$ is the principal loan amount, $i$ is the monthly interest rate, and $n$ is the number of months to repay the loan. Assuming a down payment of 20%, the principal loan amount is 80% of the transaction value. An amortization period of 30 years is allotted to each mortgagor. The interest rate is extracted from historical mortgage records [8] to reflect the lending conditions at the time of transaction. We assume a gross debt-service ratio of 28% commonly used in the lending industry, such that the total annual payments calculated via Eq. 1. do not exceed 28% of the household annual income. The validity of this assumption will be investigated in future iterations of the model.

Once household income at the time of purchase is established, the Integrated Public Use Microdata Series (IPUMS) is used to deduce the current income of households. The IPUMS is the largest collection of individual and household level Census data in the world [9]. Weighted household income metrics provide information to create annual snapshots of household income percentile rankings in the Bay Area between 1967 and 2019. These metrics are available from the Annual Social and Economic Component of the Current Population Survey in IPUMS. We assume that households maintain their income percentile ranking over time and project their income in 2021 from their historical ranking.

## Preliminary Results

### High-Resolution Spatial Distribution of Household Income

The spatial distribution of modeled annual household income is plotted in Fig. 1.a. The mean and median household income of homeowners for the study region are approximately 10% higher than the corresponding values from the Current Population Survey. This observation is within reasonable expectations as the study region is limited to San Francisco while the Current Population Survey extends into the entire metropolitan area including Oakland and Fremont - which demonstrates lower household income [9].

To enable comparison, a second set of household incomes is randomly sampled using income distributions from the SimplyAnalytics portal [10]. Replacement costs for each building is estimated by applying RSMeans building construction costs [11] to respective building areas. The ratios of building replacement cost to annual homeowner income is plotted in Fig. 1.b. It is evident from the histogram that the random sampling approach assigned over 5,000 owners to buildings with replacement costs exceeding 20 times their annual income while our proposed method produced a more reasonable distribution.

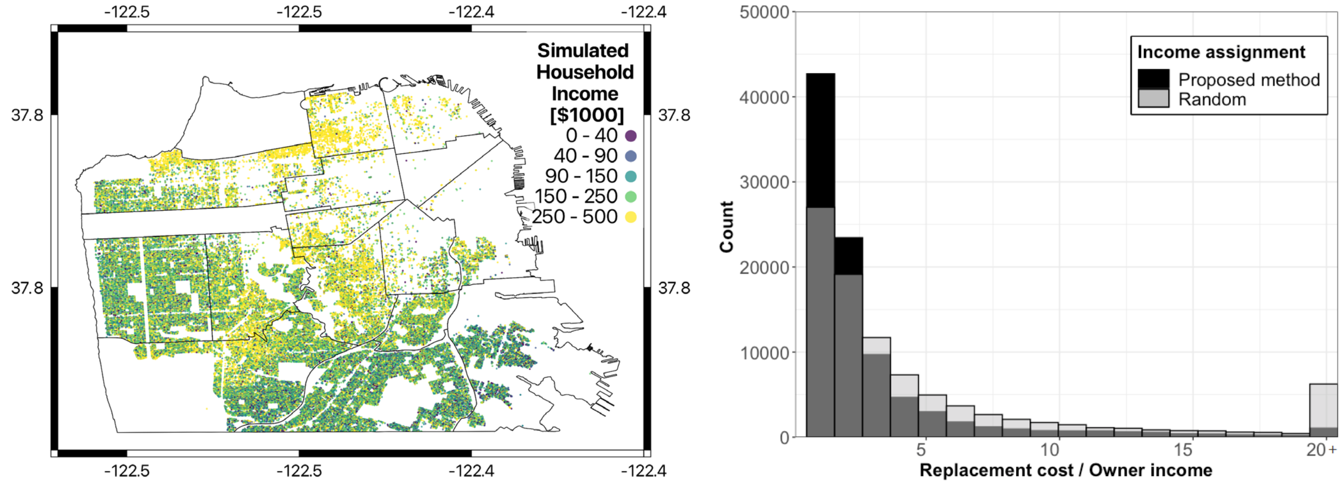


Figure 1. (a) Spatial distribution of simulated single family residence homeowner annual income in San Francisco; (b) Histograms comparing building replacement cost to household income generated from random sampling and our proposed method.

**Post-Disaster Recovery**

We use the estimated incomes from the proposed method in housing recovery simulations following a hypothetical M7.2 earthquake along the San Andreas fault. This earthquake scenario is used for planning purposes by the City [12]. One hundred realizations of ground shaking intensity, corresponding structural damage, and subsequent repair costs are simulated using SimCenter's Regional Resilience Determination (R2D) tool [13]. The outputs from the R2D Tool are used as inputs through the Regional Risk and Recovery (R3) tool [14] to simulate the post-earthquake reconstruction of the housing stock in the City.

The estimated repair costs allow us to determine the loss-to-income ratio for each household. In each of the 100 simulations, we calculate the mean loss-to-income ratio for low-, moderate-, and high-income households considering incomes assigned through (1) our proposed method and (2) conventional random sampling. Low-income and high-income household thresholds are established at 80% and 200% of the San Francisco area-median-income (AMI) respectively. These thresholds were determined after conversations with local planners who indicated that the typical threshold of 120% AMI for high-income is not adequate for San Francisco [6]. Fig. 2.a. compares the loss-to-income ratios using each methodology, with each set of brown-teal-blue dots representing one recovery simulation for the 100 simulations. The 45-degree line indicates where the results would lie if the two income-assignment methods yielded identical results. Results below this line indicate higher loss-to-income ratios when using random income assignment, and that is observed in the low- and moderate-income groups. Conversely, the loss-to-income ratios for the high-income households are above the line, indicating that these are higher when the proposed method is used. The household incomes estimated with the proposed method are correlated to the value of the buildings occupied by these households. Thus, all else equal, high-income households are bound to experience higher losses.

Household income also affects the time to secure funding for repairs. High-income households are more likely to rely on bank loans with quick funding disbursement, whereas low-income households may rely on publicly-backed loans and grants which may take months to years to become available. Once funding is secured, some households may still be forced to wait for an available contractor crew to repair their homes due to limited reconstruction workforce. The later a household enters the competition for the workforce the longer it must wait. Because the wait for financing and the wait to hire a contractor postpone the start of the recovery process, we refer to this period as ‘recovery delay’.

We estimate the recovery delay in each simulation in a similar fashion to the loss-to-income ratio in 100 recovery scenarios. Fig. 2.b. shows that the recovery delay is largely overestimated for low-income households whereas it is significantly underestimated for high-income households when incomes are randomly sampled. The randomly sampled income produced high loss-to-income ratios for low-income households. Although the lower-income households may qualify for small private loans which are quick to be disbursed, the high loss-to-income ratios force them to rely on governmental aid for larger sums, which is slower to be disbursed. These results suggest that even though random income assignment has a relatively small impact on loss-to-income ratios, its influence on housing recovery is significant.

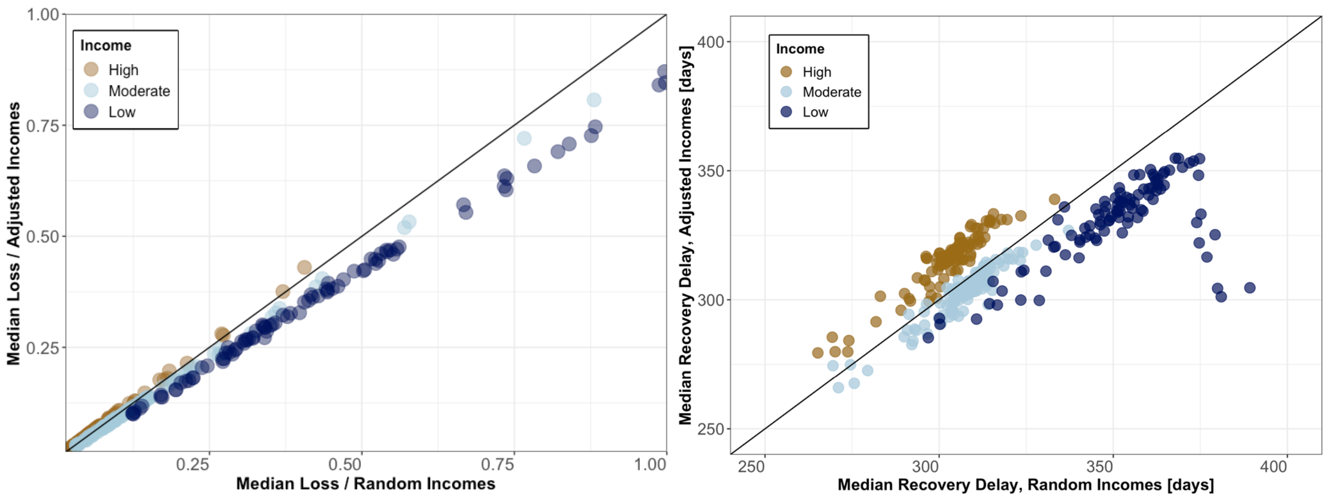


Figure 2. Marginal plots comparing the random sampling model and proposed method in terms of, (a) ratio of median loss to annual HH income and (b) median recovery delay for households, divided into three income categories.

## Conclusion

A novel method of assigning income to homeowners and its impact on post-disaster recovery modeling is explored in this study. Through the simulation of 100 post-earthquake recovery scenarios, we identified that the conventional random sampling process introduces biases that increase recovery delays and overestimate the vulnerability of low-income homeowners. The proposed method provides higher fidelity information when generating household income for regional disaster recovery simulations. Moving forward, our ongoing work further explores the validity of assumptions presented in this study, such as the income mobility and debt servicing for households.

## References


1. Abdulquadri B, Witt E, Lill I. Practice Framework for the Management of Post-Disaster Housing Reconstruction Programmes. *Sustainability* 2018; **10** (11), 3929.
2. Zhang Y, Peacock WG. Planning for Housing Recovery? Lessons Learned from Hurricane Andrew. *Journal of the American Planning Association* 2009; **76** (1), 5–24.
3. Peacock WG, Van Zandt S, Zhang Y, Highfield WE. *Inequities in Long-Term Housing Recovery after Disasters". In: Journal of the American Planning Association* 2014; **80** (4), 356–371.
4. Nejat A, Binder SB, Greer A, Jamali M. Demographics and the Dynamics of Recovery: A Latent Class Analysis of Disaster Recovery Priorities after the 2013 Moore, Oklahoma Tornado. *International Journal of Mass Emergencies & Disasters* 2018; **36** (1).
5. Moradi S, Nejat A. RecovUS: An Agent-based Model for Post-disaster Household Recovery. *Journal of Artificial Societies and Social Simulations* 2020; **23** (4) 13.
6. Alisjahbana I, Moura-Cook A, Costa R, Kiremedjian AS. An Agent-based Financing Model for Post-Earthquake Housing Recovery: Quantifying Recovery Inequalities Across Income Groups (in review).
7. DataSF. Assessor Historical Secured Property Tax Rolls 2021. Retrieved August 2021 https://data.sfgov.org/Housing-and-Buildings/Assessor-Historical-Secured-Property-Tax-Rolls/wv5m-vpq2.
8. FreddieMac. Primary Mortgage Market Survey 2021. Retrieved July 2021 http://www.freddiemac.com/pmms/.
9. Ruggles S, Flood S, Foster S, Goeken R, Pacas J, Schouweiler M, Sobek M. IPUMS USA: Version 11.0 [dataset]. Minneapolis, MN: IPUMS, 2021.
10. SimplyAnalytics. SimplyAnalytics & US Census Bureau American Community Survey Data 2020. Retrieved October 1st, 2021, from SimplyAnalytics database.
11. Building Construction Costs with RSMeans Data. Rockland, MA: Gordian RSMeans Data, Print, 2021.
12. Lee EM, Otellini P. Resilient San Francisco: Stronger Today, Stronger Tomorrow. Technical Report, 2016.
13. NEHRI SimCenter. SimCenter Testbeds. Retrieved October 2021 https://simcenter.designsafe-ci.org/testbeds/.
14. Costa R, Haukaas T, Chang SE. Agent-based Model for Post-earthquake Housing Recovery. *Earthquake Spectra* 2021; **(37)**, 46-72.