14th International Conference on Applications of Statistics and Probability in Civil Engineering, ICASP14
Dublin, Ireland, July 9-13, 2023

A Probabilistic Approach to Estimating Post-disaster Unmet Housing Needs Under Limited Information

Rodrigo Costa
Assistant Professor, Dept. of Systems Design Engineering, Univ. of Waterloo, Waterloo, Canada

Jack W. Baker
Professor, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, USA

ABSTRACT: This paper presents a methodology to estimate post-disaster unmet housing needs that is accurate and relies only on data obtained shortly after a disaster. Statistical models for aid distributed by the Federal Emergency Management Agency (FEMA) and the Small Business Administration (SBA) are developed and used to forecast funding provided by those agencies. With these forecasts, post-disaster unmet housing needs can be estimated shortly after a disaster, which can expedite the disbursal of financial aid. The approach can be used for multiple states and hazard types. As validation, the proposed methodology is used to estimate the unmet housing needs following disasters that struck California in 2017. California authorities suggest that the methodology employed by the Federal government underestimated the state's needs by a factor of 20. Conversely, the proposed methodology can replicate the estimates by the state authorities and provide accounts of losses, the amount of funding from FEMA and SBA, and the total unmet housing needs without requiring data unavailable shortly after a disaster. Thus, the proposed methodology assists with timely and accurate funding appropriations.

1. INTRODUCTION

In the US, the Department of Housing and Urban Development (HUD) has increasingly supported the recovery of uninsured, low-income, disaster-affected families. As HUD's role in supporting disaster recovery in communities increased over the last decades, criticisms have been raised regarding their limited and slow funding disbursement. In some cases, for reasons sometimes beyond their control, HUD's role in disaster recovery was deemed to have worsened socioeconomic inequalities (Gotham, 2014; McDonnell et al., 2018; Tafti and Tomlinson, 2019). However, obtaining better estimates of post-disaster unmet housing needs is challenging. HUD must combine information from FEMA with detailed data from other sources, including other federal agencies, private insurance claims payments, and possibly charitable assistance, to have a clear picture of unmet needs. Often, these data are unavailable for months after a disaster. For this reason, HUD grantees must produce an Action Plan for using the HUD funding, which includes re-calculating unmet needs conducted months after the disaster. Although the Action Plans still rely on often inaccurate or insufficient data, they provide a better representation of the unmet needs (Martin, 2018). A problem is raised if there is a significant discrepancy between the HUD-allocated funding and the grantee-estimated needs. For example, after the 2017 California Wildfires, a survey was conducted to assess homeowners' needs and define priorities resulting in delays in funding disbursement (California Department of Housing and Community Development, 2019). From this perspective, estimating unmet housing needs shortly after a disaster can help HUD allocate appropriate funding and communities to better plan their disaster recovery.

This paper presents a methodology that uses computational simulations to estimate post-disaster unmet housing needs using data that can be collected shortly after a disaster. Disaster financing data from the Federal Emergency Management Agency Individual Assistance Program (FEMA IAP) and the Small Business Administration Homeowner and Personal Property Loan (SBA HPPL) programs are used to build probabilistic models to estimate the expected approval rate and approved amount for these programs. The models account for hazard type (e.g., flood, wildfire, earthquake), state, and applicant demographics (e.g., income, insurance status, home ownership, and residence type). These models are combined with a methodology to estimate housing reconstruction costs based on FEMA guidelines and used to estimate post-disaster unmet housing needs. Thus, this paper provides two contributions. First, we develop statistical models to probabilistically estimate post-disaster assistance from the FEMA IAP and SBA HPPL programs. Second, we developed an improved methodology to estimate post-disaster housing needs using only data available shortly after a disaster. A case study application of the methodology to the 2017 California wildfire-related disasters shows that the proposed approach can replicate the losses, assistance, and unmet housing needs following these disasters.

2. POST-DISASTER UNMET HOUSING NEEDS

HUD defines the post-disaster unmet housing needs, U, in a community as the difference between the total real estate losses (i.e., personal property losses are not accounted for) and the total amount of funding the community is expected to receive from insurance, FEMA, and SBA. That is,

$$U = L_T - F_{insurance} - F_{FEMA} - F_{SBA} \qquad (1)$$

thus, to estimate the unmet housing needs in a community struck by a disaster, one must collect data or estimate the losses, L_T, the total financing coming from insurance, $F_{insurance}$, the total funding coming from Federal Emergency Management Agency, F_{FEMA}, and the total financing coming from the Small Business Administration, F_{SBA}. However, in many cases, F_{FEMA} and F_{SBA} are not available for months after a disaster. Consequently, HUD often relies on an empirical approach where unmet housing needs are calculated as

$$U^* = \sum_{dc=3}^{5} H_{dc} \cdot M_{dc} \qquad (2)$$

where H_{dc} is the number of FEMA IAP applicants in each damage category dc, and the multiplier M_{dc} is the amount of unmet needs per home, empirically estimated by HUD. The damage categories are determined based on FEMA estimated losses. Only buildings at major low, major high, or severe damage categories are considered to have unmet housing needs. The multipliers M_{dc} are calculated using the median real property damage repair costs determined by the SBA for its disaster loan program for the subset of homes inspected by both SBA and FEMA for each eligible disaster (US Department of Housing and Urban Development, 2018). HUD's simplified method (Eq. 2) has been criticized by CDBG-DR grantees because it leads to an underrepresentation of disaster impacts for states with higher building costs (Florida Department of Economic Opportunity, 2018; California Department of Housing and Community Development, 2019).

3. PROPOSED METHODOLOGY TO ESTIMATE UNMET HOUSING NEEDS

This section introduces an alternative methodology to estimate post-disaster unmet housing needs that circumvents the need to wait for FEMA and SBA application results. To estimate the likelihood that a disaster-struck household would receive funding from FEMA, we employ data from the OpenFEMA (FEMA, 2022) and OpenSBA (Small Business Administration, 2022) portals. These portals concatenate data from all Presidential disaster declarations in recent years and make them publicly available. Note that local disasters where a Presidential declaration was not in place are not included. These data are presented as tables where most rows represent one individual applicant. The only exceptions are cases where few applicants are from the same zip code. In these cases, the data are aggregated. The OpenFEMA data provide information regarding the type of disaster, location, estimated losses and the amount received by individual applicants, as well as some demographic data, e.g., income. The OpenSBA data contain information regarding the type of disaster, location, and estimated losses, but no demographic characteristics. However, neither data set contains information regarding how much insurance funding each household received. Since FEMA and SBA only provide funding for losses not covered by insurance, the loss data are not a perfect predictor of the amount received because families with similar losses may have received different amounts from insurance.

3.1. Estimating funding from FEMA

To determine the total amount of FEMA funding coming to a community, F_{FEMA}, we must determine the likelihood that each individual household will be approved, $A_{FEMA,h}$, and the expected funding they may receive, $F_{FEMA,h}$. Using OpenFEMA data, we employ a model based on similarity. That is, we first determine the 'profile' for each household, P_h. The profile represents their losses L_h, the state where they live S_h, and their demographic characteristics X_h. Then, we estimate the likelihood that a household with profile P_h will be approved based on the likelihood of approval for all households with similar profiles, that is

$$A_{FEMA,h}(P_h) = \frac{\sum_{i=1}^{N} 1(P_i = P_h) \cdot 1(F_{FEMA,i} > 0)}{\sum_{i=1}^{N} 1(P_i = P_h)} \tag{3}$$

where $1()$ is an indicator function that returns the unit if the condition in parenthesis is true and zero otherwise. In a similar fashion, we estimate the expected funding from FEMA for a household with profile P_h as

$$F_{FEMA,h}(P_h) = \frac{\sum_{i=1}^{N} 1(P_i = P_h) \cdot F_{FEMA,i}}{\sum_{i=1}^{N} 1(P_i = P_h) \cdot 1(F_{FEMA,i} > 0)} \tag{4}$$

Finally, the expected total amount of FEMA assistance across all impacted households is

$$F_{FEMA} = \sum_{h=1}^{H} A_{FEMA,h}(P_h) \cdot F_{FEMA,h}(P_h) \tag{5}$$

where H is the total number of households that apply for FEMA in the community.

3.2. Estimating funding from SBA

To derive a relationship between the losses and loans received by households using the OpenSBA data, it is important to consider the distinction between the loss amount ($Loss_{T,h}$) in the OpenSBA data and the losses eligible for SBA loans ($Loss_{SBA,h}$) which is used to calculate the maximum loan amount. To avoid duplication of benefits, SBA calculates eligible losses as

$$Loss_{SBA,h} = Loss_{T,h} - F_{FEMA,h} - F_{insurance,h} \tag{6}$$

where $Loss_{T,h}$ are the total losses experienced by the household (i.e., present in the OpenSBA dataset), and $F_{FEMA,h}$ and $F_{insurance,h}$ are the funding received by the household. Consequently, in the OpenSBA data, two households with identical losses may have significantly different approved loans due to differences in their SBA-eligible losses. For this reason, we estimate the probability that a given loan is associated with multiple candidate losses, rather than estimating the expected loan given a loss. That is, the loan-to-loss ratio is approximated as a multinomial distribution given by

$$R_{LL} \sim f(r_1, \ldots, r_n; p_1, \ldots, p_n | L) \tag{7}$$

where the probability that the loan-to-loss ratio is in the interval r_* (e.g., r_1=[0,10%) and r_2=[10%, 20%)) is p_*. After conducting tests, we opted to split the losses to be used in Eq. 7 into $50,000 intervals up to $500,000, with one extra loss bracket covering losses above $500,000. Thus, a total of 11 models are developed for the loan-to-loss ratio conditioned by the losses. These models are presented in Fig. 1. Households experiencing losses smaller than the SBA cap of $200,000 are more likely to have higher loan-to-loss ratios. However, as the conditional losses increase and exceed the $200,000 cap, the loan-to-loss ratios tend to be smaller than the unity.

The models in Fig. 1 are easy to interpret and implement and we believe these traits will help with their adoption. As an example of the application of the models, say a household has $75,000 total losses (i.e., before FEMA or insurance funding is accounted for). The model in the second column in Fig. 1 should be employed for this household and it indicates that the household has 4% chance of having a loan-to-loss ratio in the interval [0%,10%), 9% chance of having a loan-to-loss ratio in the interval [10%,20%), and so on. Equation 7 is used to determine the loan-to-loss ratio for this household. The expected SBA loan is the maximum between its eligible losses ($Loss_{SBA,h}$) and the estimated loan-to-loss ratio, that is

$$F_{SBA,h} = max(R_{LL} \cdot \$75,000, Loss_{SBA,h}) \quad (8)$$

where $Loss_{SBA,h}$ is estimated by the proposed methodology using Eq. 6. Lastly, the expected total amount of SBA assistance across all impacted households is

$$F_{SBA} = \sum_{h=1}^{H} A_{SBA,h}(L_h) \cdot F_{SBA,h}(L_h) \quad (9)$$

4. CASE STUDY

In October 2017, a series of wildfires burned in Northern California. More than 200,000 acres burned, and 8,922 structures were destroyed. In December of the same year, another series of wildfires burned 308,383 acres across Southern California. The wildfires were followed by heavy rains, mudflows, and debris flows which compounded the devastation. In December 2017, FEMA issued Major Disaster Declarations DR-4344 and DR-4353 in response to these series of events, respectively. The disaster declarations led to a Presidential Disaster Declaration and the subsequent Congressional Appropriation of Funds, and on August 14th, 2018, HUD published the Federal Register allocating U_{HUD}=$124 million to California (US Department of Housing and Urban Development, 2018). Figure 2 presents a timeline of events following the 2017 Disasters in California. The highlighted period between the FEMA 4353 Disaster Declaration and the Notice of Appropriation could have been reduced if the HUD appropriation of funds was completed more quickly. The second highlighted period, between the approval of the State Action Plan and the Initial Program Awards, could have been reduced if the initial appropriation provided sufficient funding and the survey period was unnecessary.

On March 15th, 2019, the California Department of Housing and Community Development (HCD) published an Action Plan for Disaster Recovery (California Department of Housing and Community Development, 2019) from the 2017 disasters. The HCD obtained data from the California Department of Forestry and Fire Protection (CAL FIRE), which identified that 7,640 homes were impacted: 137 were severely damaged and 7,503 were destroyed. The HCD estimated the average replacement cost for a home in California to be $300,000. The HCD assumed that the repair costs are a fraction of the replacement costs, between 50% and 75% for severely damaged homes and 100% for completely damaged homes. With this, the HCD estimated the total losses to be $2.283 billion. The HCD identified that many homeowners were not insured or held policies that did not cover the total building replacement costs, i.e., they were underinsured. Some fully insured homeowners had significant unmet needs due to increased materials and labor costs. Moreover, HCD advocated that HUD's criteria limiting the analysis to homes affordable to low-income households did not reflect the high-living-cost areas involved in these disasters. Thus, the HCD included all losses in calculating unmet housing needs. The HCD collected data from FEMA and SBA to estimate funding

Loan-to-loss ratio (%)	Real state losses [$1000] <50	50-100	100-150	150-200	200-250	250-300	300-350	350-400	400-450	450-500	>500
0-10	0.02	0.04	0.05	0.06	0.05	0.06	0.06	0.05	0.06	0.08	0.10
10-20	0.04	0.09	0.09	0.09	0.08	0.08	0.09	0.12	0.14	0.15	0.24
20-30	0.06	0.11	0.11	0.11	0.10	0.15	0.13	0.15	0.24	0.23	0.43
30-40	0.08	0.10	0.13	0.14	0.13	0.15	0.20	0.23	0.25	0.38	0.22
40-50	0.08	0.11	0.14	0.14	0.19	0.17	0.19	0.27	0.31	0.16	0.02
50-60	0.09	0.11	0.12	0.14	0.16	0.17	0.24	0.18	0.01		
60-70	0.11	0.12	0.12	0.12	0.13	0.15	0.09				
70-80	0.12	0.11	0.09	0.08	0.07	0.06					
80-90	0.16	0.11	0.08	0.08	0.06						
90-100	0.24	0.11	0.06	0.05	0.02						

Figure 1: Loan-to-loss ratio relationships conditioned on total real estate losses.

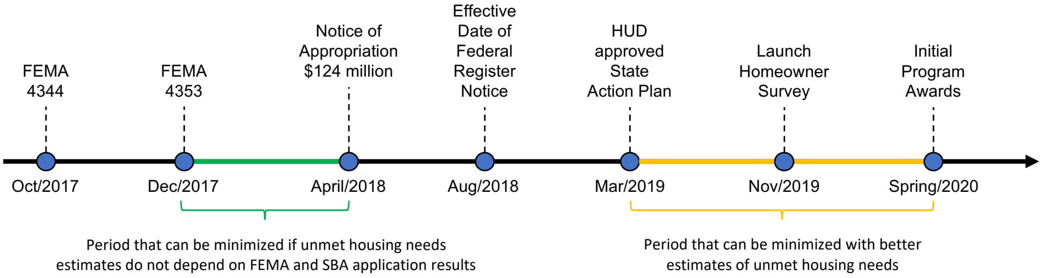

Figure 2: Timeline of CDBG-DR funding allocation following FEMA Disasters 4344 and 4353.

from these sources, i.e., F_{FEMA}=20.7 million and F_{SBA}=163.2 million. However, more than a year after the disasters, the HCD could not collect reliable data on insurance funding and so estimated the upper bound of the unmet housing needs using $F_{insurance} = 0$. Thus, Eq. 1 provided an unmet needs estimate of

$$U_{HCD} = \$2.283 \cdot 10^9 - \$0 - \$20.7 \cdot 10^6 - \$163.2 \cdot 10^6 = \$2,098 \times 10^9 \quad (10)$$

In the following, we discuss how the proposed approach is used to estimate funding from FEMA, SBA, and the unmet housing needs following these disasters.

4.1. Funding from Insurance

There are limited data regarding housing insurance for these disasters. The best information available was provided by the California Department of Insurance (CDI) to the HCD. We do not present this data here for brevity but it can be consulted in California Department of Housing and Community Development (2019). These data consist of the total sum of all insurance claims (i.e., residential, personal property, life, and automotive) and counts of the residential insurance claims per county. Thus, there is no available information on the amount of insurance funding provided for housing reconstruction. In the face of these limitations, we consider a scenario in terms of insurance that assumes that the number of insurance claims provided by the CDI represents the number of insured homes. This assumption overestimates the number of insured buildings since in some counties insurance claims exceed the number of damaged buildings. For this scenario, we assume that insurance will cover 50% of the repair costs. This assumption reflects the findings from the HCD regarding the issues with underinsurance.

4.2. Funding from FEMA's Individual Assistance Program

The model described in Section 3.1 is used to generate 1,000 estimates of the FEMA IAP funding received by Californians following the 2017 Disasters. Figure 3 shows the results obtained. For reference, the vertical dashed line shows the empirical estimate of $15.3 million in assistance obtained from the OpenFEMA portal. The results show that the proposed methodology can replicate the empirical results, albeit with a tendency to underestimate FEMA funding. These results can be explained by the devastating nature of the wildfires which completely destroyed most homes. Consequently, most applicants would be expected to receive the maximum FEMA IAP funding, around $36,000. The proposed model includes data from disasters where complete damage was not as frequent, hence its tendency to underestimate the reality of the 2017 Disasters in California.

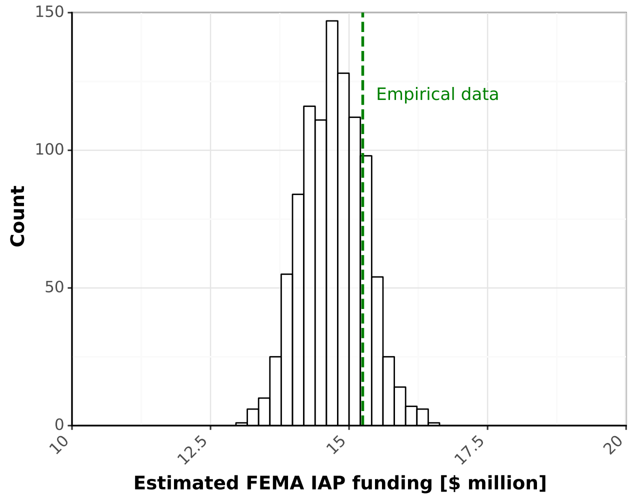

Figure 3: Results from 1,000 estimates of the total FEMA IAP funding received by California in response to FEMA Disasters 4344 and 4353 using the proposed approach. The vertical dashed line indicates empirical results after the event, as collected from the OpenFEMA portal (FEMA, 2022).

4.3. *Funding from SBA's Homeowner Personal Property Loan Program*

Figure 4 shows the results from 1,000 simulations of the SBA funding using the method described in Section 3.2. The dashed vertical line indicates the empirical value obtained after the event (2,371 successful SBA applicants from California received $152 million between 2017 and 2018, according to the OpenSBA portal (Small Business Administration, 2022)). The results in Fig. 4 show that the proposed methodology can replicate empirical results, however, with a tendency to overestimate the empirical values. The model to estimate SBA funding takes as input the estimated FEMA funding. Thus, the tendency of the proposed approach to underestimate FEMA funding is one explanation to its tendency to overestimate SBA funding. Nonetheless, the proposed models provide estimates that are similar to the empirical data in both cases.

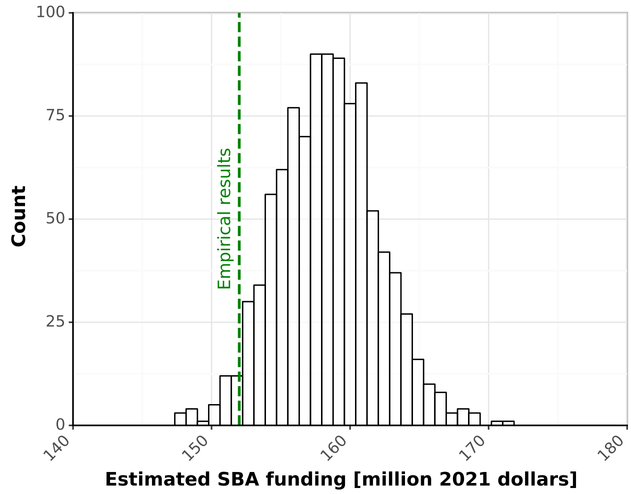

Figure 4: Results from 1,000 estimates of the total SBA HPPL funding received by California in response to FEMA Disasters 4344 and 4353 using the proposed approach. The vertical dashed line indicates empirical results collected from the OpenSBA portal (Small Business Administration, 2022).

4.4. *Estimated Unmet Housing Needs*

Finally, the proposed methodology is used to estimate the total unmet housing needs. We combine the loss estimates by the HCD (e.g., $2,098) and the results from the 1,000 simulations of FEMA and SBA funding to create 1,000 estimates of unmet housing needs using Eq. 1. Figure 5 presents the unmet housing needs estimates using the proposed approach and compares it to the $2.58 billion unmet housing needs calculated by the HCD. We note that the abscissa axis in Fig. 5 extends over a $0.1-billion range. Thus, the proposed methodology provides estimates that are within 0.1/2.4≈5% of the unmet needs estimated by the HCD using empirical data. Conversely, the HUD estimate (i.e., Eq. 2) is $0.124 billion, or 95% smaller than the HCD-estimated unmet needs.

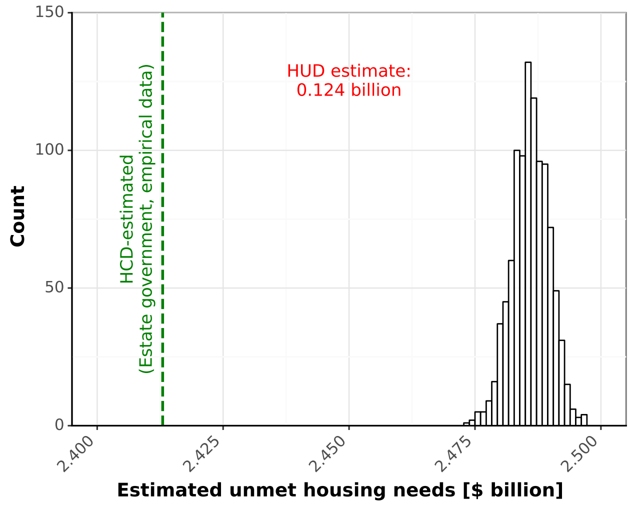

Figure 5: Results from 1,000 estimates of the unmet housing needs using the proposed approach. The vertical dashed line represents the estimate by the California Department of Housing and Community Development.

5. CONCLUSIONS

This study proposes a novel methodology that addresses some limitations in the state-of-the-art approaches to estimating post-disaster unmet housing needs for communities. The methodology uses data from the OpenFEMA and OpenSBA portals regarding assistance provided after major disasters in the U.S. in the last 20 years to build predictive models for the approval rate and approved amount from each agency. Thus, the proposed methodology can be used shortly after a disaster and provides accuracy equivalent to state-of-the-art approaches. A case study is presented where the methodology is used to estimate unmet housing needs after a combination of disasters that struck California in 2017 (i.e., FEMA DR-4344 and FEMA DR-4353). Unmet housing needs estimates provided by the California Department of Housing and Community Development (HCD) using data collected about one year after the disaster are used as benchmarks. The case study demonstrates that the proposed methodology can replicate the HCD estimates, using only data available much sooner after the Disasters. Thus, the proposed methodology can help communities better prepare and respond to a disaster by providing accurate and quick estimates of unmet housing needs.

6. REFERENCES

California Department of Housing and Community Development (2019). “State of California Action Plan for Disaster Recovery in Response to 2018 Disasters.” *Report no.*, Sacramento, California.

FEMA (2022). “Individual Assistance Housing Registrants for Large Disasters (v1).

Florida Department of Economic Opportunity (2018). “State of Floria Action Plan for Disaster Recovery in Response to Hurricane Irma.” *Report no.*, Tallahassee, California.

Gotham, K. F. (2014). “Reinforcing inequalities: The impact of the cdbg program on post-katrina rebuilding.” *Housing Policy Debate*, 24(1), 192–212.

Martin, C. (2018). “Community Development Block Grant-Disaster Recovery Program: Stakeholder Perspectives.” *Report no.*

McDonnell, S., Ghorbani, P., Desai, S., Wolf, C., and Burgy, D. M. (2018). “Potential challenges to targeting low-and moderate-income communities in a time of urgent need: the case of cdbg-dr in new york state after superstorm sandy.” *Housing Policy Debate*, 28(3), 466–487.

Small Business Administration (2022). “Disaster Loan Data.

Tafti, M. T. and Tomlinson, R. (2019). “Theorizing distributive justice and the practice of post-disaster housing recovery.” *Environmental Hazards*, 18(1), 7–25.

US Department of Housing and Urban Development (2018). “Federal Register Volume 83, Number 157.” *Report no.*, US Department of Housing and Urban Development.