LEVERAGING DATA-DRIVEN APPROACHES TO EXPLORE THE EFFECT OF VARIOUS DISASTER POLICIES ON POST-EARTHQUAKE HOUSEHOLD RELOCATION DECISION-MAKING

Chenbo WANG¹, Gemma CREMEN² and Carmine GALASSO³

Abstract: Earthquake events can cause affected households to relocate. Post-earthquake relocation disrupts displaced households’ social ties as well as their access to affordable services. Simulation models that capture post-earthquake relocation decision-making can be useful tools for supporting the development of related disaster risk reduction policies that aim at mitigating disaster-induced relocation. Yet, existing versions of these models focus particularly on housing-related factors (e.g., housing repair costs), which are not the sole driver of post-earthquake relocation. In this paper, we integrate data-driven approaches and local perspectives into an existing simulation-based framework to holistically capture various context-specific factors perceived as being important to household relocation decision-making. The enhanced framework is used to quantitatively assess the effectiveness of various disaster risk reduction policies - both ‘soft’ (e.g., post-earthquake livelihood assistance funds) and ‘hard’ (e.g., upgrading existing infrastructure facilities to higher building codes) - in reducing post-earthquake household relocation, with an explicit focus on low-income households. We demonstrate it using a possible future (50-year) projection of “Tomorrowville”, a synthetic expanding urban extent that imitates a Global South setting. Our analyses suggest that livelihood assistance funds are more successful and pro-poor when it comes to mitigating positive post-earthquake relocation decision-making than hard policies focused on strengthening buildings (at least in the context of the examined case study).

Introduction

Devastating earthquake events can cause extensive damage to people’s homes, workplaces, and the infrastructure systems and networks that they depend on. As a result, affected households may decide to relocate following earthquake events (He et al., 2018). Post-disaster relocation has long-lasting negative impacts on relocated households’ social ties and can deprive them of access to affordable housing, healthcare, education, and employment (Badri et al., 2006). Therefore, it is crucial for policy makers to devise strategic disaster risk reduction (DRR) policies for mitigating positive post-earthquake relocation decision-making.

Simulation models that capture post-earthquake relocation decision-making can support the design of such DRR policies (Costa and Baker, 2022). Miles and Chang (2011) developed the ResilUS computational model to simulate community-based post-disaster housing recovery. ResilUS models households’ decisions to leave or to stay, accounting predominantly for factors related to housing reconstruction (e.g., the debt incurred by housing repairs). Nejat and Damnjanovic proposed an agent-based model using game theory to predict homeowners’ decision-making (i.e., stay and repair or sell and leave) based on the neighbourhood’s average reconstruction value and the predicted future value of reconstruction. Costa et al. (2022a) proposed an agent-based model for assessing temporary displacement and permanent relocation decision-making of households, accounting for a multitude of factors predominantly related to the immediate built environment, e.g., availability of water and electricity, neighbourhood conditions, repair progress, and socioeconomic factors.

Thus, most existing simulation models for household relocation decision-making focus particularly on housing-related factors. This means that they neglect or do not give sufficient attention to alternative factors that can motivate or discourage households from relocating, e.g., earthquake-induced livelihood impact. Many of these models are also not validated with empirical data or are only partially calibrated using highly aggregated relocation patterns observed after historical earthquake events (Nejat and Damnjanovic, 2012; Costa et al., 2022a; Miles and Chang, 2011).

¹ PhD Student, University College London, London, United Kingdom, chenbo.wang@ucl.ac.uk
² Lecturer, University College London, London, United Kingdom
³ Professor, University College London, London, United Kingdom
Therefore, further research to improve the understanding and modelling of households’ post-earthquake relocation decision-making is needed.

We aim to address this challenge using a data-driven modelling approach that holistically integrates various context-specific factors, allowing for a rich prediction of household-level post-earthquake relocation decision-making. Data-driven approaches (e.g., logistic regression, random forest) have been previously used in the literature to develop models for assessing or identifying factors related to post-disaster household behaviour (Loos et al., 2023; Costa et al., 2022b; Myers et al., 2008). However, these studies either: (1) did not explicitly focus on relocation; or (2) considered data at more aggregated resolutions (i.e., neighbourhood-level) than individual households; and (3) focused on hurricane rather than earthquake disasters. The data-driven model is integrated into an existing framework for policy-related risk-sensitive decision support on future urban development (Wang et al., 2023). This adapted framework can then be used to quantify the effectiveness of various DRR policies in mitigating positive post-earthquake relocation decision-making, with an explicit focus on the extent to which low-income households are disproportionately impacted. We demonstrate the enhanced framework using the “Tomorrowville” virtual urban testbed.

Simulation-based framework
We adapt the framework proposed in Wang et al. (2023) to explicitly account for post-earthquake household decision-making, as shown in Figure 1. The adapted framework encompasses seven modules: (1) Policy Bundles; (2) Urban Planning; (3) Local Perspectives; (4) Seismic Hazard; (5) Physical Infrastructure Impact; (6) Social Impact; and (7) Computed Impact Metrics. Modules (1), (2), (4), (5), (6), and (7) are modified versions of the same modules within the original framework. Consideration of post-earthquake household relocation decision-making is facilitated by the Contextual Social Knowledge component, consisting of the Local Perspectives module and the Data-driven Model, which concurrently sits within the Social Impact module. Policy makers first design DRR policies (in the Policy Bundles module) and apply these policies to an urban plan associated with a specific time instance (in the Urban Planning module), both of which collectively produce a Visioning Scenario. The information stored in the Visioning Scenario and Contextual Social Knowledge informs the calculations in modules (4) to (6), which collectively comprise the Computational Model. Modules (4) to (6) produce seismic hazard calculations, physical infrastructure impacts, and social impacts, respectively. The form (i.e., inclusion of appropriate predictors) and parameterization or the selection of an appropriate Data-driven Model is guided by the Local Perspectives module, which provides relevant context-specific information on household relocation decision-making. The Data-driven model is used within the Social Impact module to predict whether households decide to relocate or stay. The results of these predictions are then translated into a Poverty Bias Indicator (PBI), which measures the extent to which low-income households are disproportionately affected in terms of earthquake-induced relocation. Each iteration of the framework produces an assessment of impacts for one specific Visioning Scenario. The optimal Visioning Scenario is the one that produces the lowest PBI. We use Monte Carlo sampling to capture uncertainties within modules (4) to (6), in line with Cremen et al. (2022). Modules introduced in Wang et al. (2023) are only briefly discussed. Described in detail are the newly introduced Local Perspectives module, the enriched Social Impact module, and the Computed Impact Metrics that depend on the Social Impact module.

Brief descriptions of existing modules
The Urban Planning module contains an urban plan that provides detailed information on land use, buildings, households and individuals associated with a specific urban area at a prescribed time (Mentese et al., 2023). Within the context of the proposed adapted framework, the Policy Bundles module encompasses one or more DRR policies that aim at mitigating positive post-earthquake relocation decision-making. These policies could be ‘soft’ (e.g., post-earthquake livelihood assistance funds), as well as ‘hard’ (e.g., upgrading existing infrastructure facilities to higher building codes). The Seismic Hazard module stores the seismic source and rupture features of a specific earthquake event and estimates the resulting ground motion intensity measures (IMs) at the locations of buildings. The Physical Infrastructure Impact module uses the ground motion fields (GMFs) output from the Seismic Hazard module in combination with fragility and/or vulnerability relationships to estimate physical damage and/or asset loss associated with buildings. The reader is referred to Sections 2.1 to 2.4 in Wang et al. (2023) for more details on these modules.
Contextual social knowledge

We integrate Contextual Social Knowledge to allow for context-specific and (local) people-centred characterization of post-earthquake household relocation decision-making.

The Local Perspectives module contains information on what local stakeholders (e.g., community representatives and disaster planning authorities) perceive as being important to the post-earthquake relocation decision-making of households within specific contexts. This information is used either to guide the form (i.e., the inclusion of appropriate predictors) and parameterization of the Data-driven Model or as criteria for selecting the appropriate Data-driven Model among a list of pre-existing ones. For example, if stakeholders consider earthquake impact on livelihood to be an influential factor in post-earthquake relocation decision-making of local households, then the Data-driven Model (whether pre-existing or not) should include earthquake impact on livelihood as a predictor.

The Data-driven Model makes predictions related to post-earthquake household relocation decision-making. It is developed by applying statistical learning methods (e.g., logistic regression, random forests) to data containing post-earthquake household relocation information and other appropriate context-specific details it depends on (e.g., age of head of household, household income group). The Data-driven Model is therefore inherently location-specific, allowing for more accurate characterization of post-earthquake household relocation decision-making compared to generic, heuristic models.

Social impact

The Social Impact module uses outputs from the Physical Infrastructure Impact module and leverages the Data-driven Model to assess post-earthquake relocation decision-making of each household, considering the policies within the Policy Bundles module. The post-earthquake relocation decision for the $h_{ih}$ household in the $j_{th}$ Monte Carlo sample, $I_{ih,j}$, is binary. $I_{ih,j} = 1$ means the $h_{ih}$ household decides to relocate and $I_{ih,j} = 0$ means otherwise.

Figure 1. Simulation-based framework for quantitatively assessing the effectiveness of DRR policies in mitigating positive post-earthquake relocation decision-making.
**Computed impact metrics**

The Computed Impact Metrics module uses the $I_{bh,j}$ outputs from the Social Impact module to determine collective relocation decisions made by low-income households ($I_{low,j}$) as well as households across all income groups ($I_{all,j}$). $I_{x,j}$ for the $j$th Monte Carlo sample is computed as:

$$I_{x,j} = \sum_{bh} \frac{I_{bh,x,j}}{n_x}$$

where $x$ refers to low- (low), middle- (mid), high- (high), or all- (all) income households, $I_{bh,x,j}$ is the $bh$th household’s relocation decision within group $x$ and $n_x$ is the total number of households associated with $x$. The module then translates $I_{low,j}$ and $I_{all,j}$ into a single-valued Poverty Bias Indicator $PBI_j$, which measures the extent to which low-income households disproportionately decide in favour of relocation. That is:

$$PBI_j = \frac{I_{low,j}}{I_{all,j}} - 1$$

A negative value of $PBI_j$ implies that the policies within the Policy Bundles module (and thus the associated Visioning Scenario) are pro-poor, i.e., the specific earthquake scenario considered does not result in a disproportionate number of positive relocation decisions among low-income households. See Section 2.6 in Want et al. (2023) for more details on $PBI$.

**Case-study descriptions**

We showcase the enhanced framework using the “Tomorrowville” virtual testbed. Tomorrowville imitates a Global South urban setting by means of its socioeconomic and physical characteristics (See Mentese et al., 2023, for details). We leverage the adapted framework to assess the effectiveness of four DRR policies in mitigating positive post-earthquake relocation decision-making, with an explicit focus on Tomorrowville’s low-income households. We concentrate our analysis on Tomorrowville’s uncertain future state, using a 50-year projection of its urban configuration (known as TV50_total), as shown on the left panel of Figure 2. We consider a $M_{7.0}$ earthquake scenario on a hypothetical fault near Tomorrowville, as shown on the right panel of Figure 2.

**Urban planning**

TV50_total includes 4,810 existing buildings in today’s Tomorrowville (TV0) and 5,346 new buildings anticipated to be built within the next 50 years (TV50_b2). TV50_total contains 8,713 residential buildings and 1,443 non-residential (e.g., commercial, industrial) buildings. New buildings to be built in TV50_b2 are much more earthquake resistant on average than those that exist in TV0 (see Wang et al., 2023, for details). There are three types of residential polygons (low-, middle-, and high-income polygons) where households within the same polygon belong to the same income group. TV50_total includes 6,766, 3,059, and 7,985 low-, middle-, and high-income households, respectively.

**Policy bundles**

We consider four DRR policies (see Table 1). Policy No.1, which provides livelihood assistance funds to households who have at least one member made unemployed by the earthquake, is a ‘soft’ (and compensatory) policy. The other policies, which involve upgrading the most vulnerable TV0 workplace and residential buildings to higher building codes, are ‘hard’ (corrective). Policy No.4 is also income-based, designed to explicitly facilitate pro-poor outcomes.

**Seismic hazard**

We consider a $M_{7.0}$ earthquake scenario on a hypothetical vertical strike-slip fault near Tomorrowville, which is identical to that considered in Wang et al. (2023), where more details can be found. We use Monte Carlo sampling to simulate 500 sets of GMFs for different IMs.

**Physical infrastructure impact**

We use fragility relationships associated with each building type (see Gentile et al., 2022, for details) to compute the damage state ($DS$) of each building for the simulated IM values. The exact fragility relationships used are influenced by the three hard policies included in the Policy Bundles module. The $DS$ damage classification is converted to a damage level $DL$ categorization to comply with the format of the Data-Driven Model (details to follow). $DS = 0$ ("no damage") is
mapped to $DL = 1$, representing "no damage", $DS = 1$ ("slight damage"), $DS = 2$ ("moderate damage"), and $DS = 3$ ("extensive damage") are mapped to $DL = 2$, representing "partial damage". $DS = 4$ ("complete damage") is mapped to $DL = 3$, representing "full damage" (FEMA, 2022; The Asia Foundation, 2019).

**Table 1. Policies considered for this case study.**

<table>
<thead>
<tr>
<th>Policy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provide livelihood assistance funds to households with at least one member made unemployed by the earthquake.</td>
</tr>
<tr>
<td>2</td>
<td>Replace non-RC workplace buildings with high-code RC buildings.</td>
</tr>
<tr>
<td>3</td>
<td>Replace non-RC residential buildings with high-code RC buildings.</td>
</tr>
<tr>
<td>4</td>
<td>Replace non-RC low-income residential buildings with high-code RC building</td>
</tr>
</tbody>
</table>

**Local perspectives**

We consult hypothetical local stakeholders of Tomorrowville (i.e., community representatives and emergency response authorities) on the factors considered to be important for local households' relocation decision-making after potential future earthquake disasters. This consultation identifies earthquake-induced livelihood impact, residential building damage, and the age of the household head as significant factors.

The inclusion of these predictors in the Data-driven Model is consistent with our literature review of past disasters worldwide. Myers et al. (2008) analysed county-level data from the U.S. Census Bureau and found that disaster-hit regions with more severe housing damage experienced greater outmigration in the wake of Hurricanes Katrina and Rita. Age plays a role in household...
relocation decision-making via its correlation with place attachment. Older people often have higher place attachment than younger people, due to the longevity of their presence within a location (Anton and Lawrence, 2014). People with high place attachment show reluctance to move away from places to which they are attached (Johnson et al., 2020). Earthquake impact on livelihood (e.g., lack of local jobs) also affects household's relocation decisions after earthquake events (Wang et al., 2015).

Data-driven model

A bespoke Data-driven Model is developed specifically for this case study, based on the information provided by Local Perspectives. The required input variables of the Data-driven Model (i.e., earthquake impact on livelihood, residential building damage, age of the household head) are outputs or encapsulated information of the Policy Bundles module, the Urban Planning module, and the Physical Infrastructure Impact module.

Due to the synthetic nature of Tomorrowville, it is not possible to use data collected in the same locality to develop the Data-driven Model. We therefore use information from a region with a comparable social, physical, and economic context. The data used to develop the Data-driven Model is derived from the Independent Impacts and Recovery Monitoring (IRM) project, a longitudinal study conducted to monitor disaster-induced social impacts, recovery patterns, and disaster-affected households' evolving needs after two devastating earthquakes struck Nepal in April and May 2015 (The Asia Foundation, 2019). The IRM project revisited the same disaster-affected households and asked them similar questions over a five-year duration following the disaster. Questions included e.g., “to what extent was your livelihood affected by the earthquake?” In this study, we adopt the fourth-round survey data (collected in April 2017) - as opposed to previous survey rounds conducted during the emergency response and the early recovery phase - to focus on long-term household relocation.

To construct the Data-driven Model, the survey data must include responses from households that indicate the earthquake impact on their livelihood, residential building damage level (DL), as well as the age of their household head and information on their relocation decisions. The residential building DL (i.e., no damage, partial damage, and full damage) and the age of the household head (provided in the form of age groups, i.e., 18-25, 26-35, 36-45, and 46 and above) are directly available from the IRM survey data. We assume a uniform distribution to sample a specific values of the ages of household heads, due to a lack of more detailed information. The earthquakes are assumed to have affected the livelihoods of households who indicated that their jobs were “completely affected” or “somewhat affected”. We assign positive relocation decisions to households that had at least one member planning to migrate when the survey was conducted.

The survey data used contain responses from 4,854 households affected by the earthquakes. Excluding households with “unknown” residential building damage leads to a total of 3,519 complete responses (samples), which are used for the model development. Among those, 157 households were associated with positive relocation decisions and 1,724 households experienced earthquake-induced livelihood impacts. 2,735, 439, and 345 of households with complete responses experienced full damage, partial damage, and no damage to their homes, respectively. 230, 649, 824, and 1,816 of households with complete responses had heads with ages of 18-25, 26-35, 36-45, and 46 and above, respectively.

The majority (95.5%) of complete data samples are not associated with a positive relocation decision, which renders the dataset imbalanced (He and Garcia, 2009). A model fit to imbalanced data is biased toward the majority classification (in this case, the decision not to relocate; He and Garcia, 2009). We adopt oversampling to obtain a balanced dataset, which involves adding more samples randomly drawn from the minority class (i.e., those with positive relocation decisions). In the adjusted dataset, 2,512 samples are labelled as having ‘positive relocation decisions’ and 3,362 samples are labelled otherwise.

We use random forest to fit the Data-driven Model with the balanced dataset. Random forest is a non-parametric statistical model that aggregates the predictions given by multiple (thousands or more) decorrelated decision trees (Breiman, 2001). The outcomes of the model are households' probabilities of having a positive relocation decision. We perform ten-fold cross validation, and evaluate the model performance by calculating the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for each test fold. The closer the AUC is to unity, the more predictive power the model has (Hastie et al., 2009). The highest mean AUC obtained across the ten folds through tuning the hyper-parameters of the random forest (e.g., number of trees) is 0.71,
which we consider to be satisfactory. The Data-driven model is finally refit on the entire dataset, using the tuned hyper-parameters.

**Social impact**

The Social Impact module uses inputs from the Physical Infrastructure Impact module (i.e., the DL of each building), the Urban Planning module (e.g., the workplace buildings where employed individuals work, the age of the head of each household), and the Policy Bundles module to calculate residential building DL and earthquake-induced household-level livelihood impact.

We assume that workplace buildings with at least extensive damage ($DS \geq 3$) cannot function and the livelihoods of people working in these buildings are impacted. A household's livelihood is classified as being impacted if the livelihood of one or more employed members is impacted. We assume that policy No.1, which provides livelihood assistance funds to households whose livelihoods are impacted, eradicates the effect of earthquake-induced livelihood impact on household relocation decision making. In other words, we assume that implementing policy No.1 removes earthquake-induced livelihood impact from all affected individuals within Tomorrowville. This module finally leverages the Data-driven Model to compute $I_{hh,j}$, adopting 0.50 as the classification cut-off threshold between $I_{hh,j} = 0$ and $I_{hh,j} = 1$.

**Results**

Figure 3 shows for all policies (and no policy) the empirical cumulative distribution functions (CDF) of $I_{all,j}$ (left panel), and $PBI_{j}$ (right panel), across the 500 GMFs. Policy No.1 is the most effective in mitigating positive post-earthquake relocation decision-making. Policy No.2 is slightly more effective than policy No.3 in lower-intensity ground-motion realizations (i.e., corresponding to the lower tail of $PBI_{j}$ values), whereas policy No.3 performs better in higher-intensity ground-motion realizations. These findings reflect the underlying relative significance of each predictor and how it changes as the ground-motion intensity varies. That is, earthquake impact on livelihood (related to policy No.2) has a larger marginal impact on overall household relocation decisions than residential building damage (related to policy No.3), when ground-motion intensity is lower. Policy No.1 is also the most pro-poor policy among those considered in this case study, slightly outperforming even policy No.4 that is explicitly designed to facilitate pro-poor outcomes by specifically targeting low-income households.

**Conclusion**

We present a forward-looking approach for assessing the effectiveness of DRR policies in mitigating positive post-earthquake relocation decision-making, which involves enriching an existing framework for risk-informed policy design (Wang et al., 2023) with local perspectives and an accompanying data-driven model. This enrichment is facilitated by the introduction of a so-called contextual social knowledge component, which allows for better context-specific characterization of post-earthquake household relocation decision-making.
We demonstrate the adapted framework by assessing the effects of multiple DRR policies (i.e., livelihood assistance funds, and residential and workplace building upgrading programs) implemented on an expanding virtual urban testbed, Tomorrowville, particularly focusing on the extent to which the policies mitigate positive post-earthquake relocation decision-making among low-income households. We find that a soft policy of post-disaster livelihood assistance provision (policy No.1) is more effective and more pro-poor in mitigating positive post-earthquake relocation decision-making than hard policies centred on the seismic strengthening of buildings (policies No.2, 3, and 4). This finding emphasises the fact that hard strategies only consisting of resource-intensive engineering interventions might not always be the most effective disaster risk reduction solution for urban areas. It also reveals that soft policies can offer a pro-poor means of mitigating disaster impacts without the need to explicitly restrict their remit based on politically sensitive income thresholds. These findings demonstrate that the framework can be used to inform disaster risk reduction policies and to support forward-looking risk-sensitive urban development practices in yet-to-be urbanised regions.

Acknowledgements

The authors acknowledge funding from UKRI GCRF under grant NE/S009000/1, Tomorrow’s Cities Hub, funding from the Earthquake Engineering Field Investigation Team (EEFIT) under EEFIT Research Grant, and the University College London Overseas Research Scholarship (UCL-ORS).

References

Anton CE and Lawrence C (2014), Home is where the heart is: The effect of place of residence on place attachment and community participation, *Journal of Environmental Psychology*, 40, 451–461.


Costa R and Baker JW (2022), Challenges and opportunities in post-disaster housing recovery simulations, in *Proceedings 12th National Conference on Earthquake Engineering*, EERI.


Costa R, Wang C and Baker JW (2022), Logistic models linking household recovery capacity to demographic characteristics, in *Proceedings 13th International Conference on Structural Safety and Reliability*, ICOSSAR.


He H and Garcia EA (2009), Learning from imbalanced data, *IEEE Transactions on knowledge and data engineering*, 21(9), 1263–1284.


