Empirical Assessment of Household Susceptibility to Hazards-Induced Prolonged Power Outages

Amir Esmalian[1]; Shangjia Dong[2]; and Ali Mostafavi, A.M.ASCE[3]

[1]Ph.D. Student, Urban Resilience, Networks, and Informatics Lab, Zachry Dept. of Civil and Environmental Engineering, Texas A&M Univ., College Station, USA. E-mail: amiresmalian@tamu.edu
[2]Post-Doctoral Associate, Urban Resilience, Networks, and Informatics Lab, Zachry Dept. of Civil and Environmental Engineering, Texas A&M Univ., College Station, USA. E-mail: shangjia.dong@tamu.edu
[3]Assistant Professor, Urban Resilience, Networks, and Informatics Lab, Zachry Dept. of Civil Engineering, Texas A&M Univ., College Station, USA. E-mail: amostafavi@civil.tamu.edu

ABSTRACT

The objective of this study is to empirically assess household susceptibility to the power disruptions during disasters. In this study, a service gap model is utilized to characterize household susceptibility to infrastructure service disruptions. The empirical household survey data collected from Harris County, Texas, in the aftermath of Hurricane Harvey was employed in developing an appropriate empirical model to specify the significance of various factors influencing household susceptibility. Various factors influencing households' susceptibility were implemented in developing the models. The step-wise algorithm was used to choose the best subset of variables, and availability of substitutes, previous hazards experience, level of need, access to reliable information, race, service expectations, social capital, and residence duration were selected to be included in the models. Among three classes of models, accelerated failure time (AFT)-loglogistic model yielded the best model fitness for estimating households' susceptibility to disaster-induced power disruption. The model showed that having a substitute, households' need for the service, race, and access to reliable information are the most significant factors influencing household susceptibility to the power disruptions. Understanding households' susceptibility to infrastructure service disruptions provides useful insights for prioritizing infrastructure resilience improvements in order to reduce societal impacts.

INTRODUCTION

Infrastructure service disruptions are one of the impacts of natural disasters that threaten the communities' well-being. Electricity, water, communication, and transportation are among the critical services that households rely on prior, during, and after the hazards. In the infrastructure resilience literature, many research studies have been focusing on assessing physical performance and systems' exposure to the disaster-induced service disruption (Dong et al. 2019a, b; Rasoulkhani and Mostafavi 2018). These studies, however, assume the sub-populations of a community are equally susceptible and experience equal hardship in the face of natural hazards-induced service disruptions. Different households experience different levels of hardship under the same exposure due to their different level of susceptibility. This is because people do not hold equal needs and expectations of the infrastructure services (Tabandeh et al. 2018). Previous research (Coleman et al. 2019) has shown that there is a societal inequity in the experienced risks of households within an affected community. The socially vulnerable population is shown to suffer more hardship from infrastructure service disruptions. Hence, it is important to understand the determinants of social susceptibility and incorporate it into the assessment of societal risks due to infrastructure service disruptions.

Determining the social susceptibility to the service disruptions during natural disasters is challenging as households do not encounter prolonged service losses in their day-to-day lives. As a result, households do not have a clear perception of the potential threats and their ability to tolerate these disruptions before they actually experience one. Recognizing this, recent reports by the National Institute of Standards and Technology (Applied Technology Council 2016), and the National Infrastructure Advisory Council (Berkeley III and Wallace 2010) concluded that the current body of knowledge lacks fundamental empirical information about societal susceptibility and risks due to infrastructure disruptions in disasters. Addressing this important gap, we developed an empirical model to determine the household susceptibility to disaster-induced power outages based on survey data.

THEORETICAL BACKGROUND

We use a service gap model developed by Esmalian et al. (2019) to characterize the social susceptibility to the threats of infrastructure services disruptions. This framework suggests that households' experienced hardship in infrastructure service disruptions depends on the duration of the outages (service disruption exposure level) and their tolerance to withstand the negative impacts. Households have a varying zone of tolerance for the service disruptions and would experience disproportionate hardship when the service disruptions surpass their tolerance for the service. The service gap model explains why the same level of disruption exposure results in different levels of experienced hardship (Esmalian et al. 2019b). Thus, in the current study, the zone of tolerance was used to measure household susceptibility to infrastructure service disruptions.

Zone of tolerance is determined by different demographic characteristics of households, as well as household capabilities and resources they possess to mitigate the risk. Previous research related to the determinants of the zone of tolerance in the case of power outages identified the following influencing variables (Esmalian et al. 2019a):

1) *Need*: The more important it is for households to have access to electricity, the lower their zone of tolerance is to the service outages.
2) *Preparedness*: The greater households are prepared for threats of the disaster, the higher household' zone of tolerance will be to the power outages.
3) *Substitute*: Households which have access to a substitute for power will have a higher zone of tolerance to the disruptions.
4) *Social capital*: Households having a social capital as friends and family to rely on their help during the disaster have a higher zone of tolerance.
5) *Experience*: Households with previous experience with natural disasters have a higher zone of tolerance to the service disruptions.
6) *Service expectations*: Households with a higher expectation of the service losses have a higher zone of tolerance to the service outages.
7) *Information*: Households having access to more reliable information about the service disruptions have a higher zone of tolerance.
8) *Sociodemographic characteristics*: lower-income households, racial minorities, households with kids (age 10 or less), households with an elderly, renters, households who have lived in their residents shorter, and households living in multiple unit housings have a lower zone of tolerance to power outages.

In this study, we examined the significance of each of these factors in developing an empirical model of household susceptibility to service disruptions.

DATA

A household survey was distributed across Harris County, Texas in the aftermath of Hurricane Harvey. The online panel, Qualtrics, was used for distributing the survey. This study focused on sheltered-in-place households since those who had evacuated before the event might not experience service losses. After removing incomplete responses and those who evacuated, a total sample of 574 responses were used for examining the households' zone of tolerance for power outages. The following table displays the measurement of variables considered in the survey for modeling the zone of tolerance.

Table 1. Variable Description of the Models

Variable description	Variable description
Zone of tolerance *(number of days)*	Ownership *(1: yes; 0: no)*
Substitute *(1: yes; 0: no)*	Residence duration *(number of years)*
Experience *(1: yes; 0: no)*	Service expectation *(number of days)*
Need *(5: important always to 1: not at all important)*	Race minority *(1: yes; 0: no)*
Education *(1: less than high school; 2: high school graduate or GED; 3: trade/technical/vocational training; 4: some college; 5: 2-year degree; 6: 4-year degree; 7: graduate Level)*	Income *(1: less than \$25,000; 2: \$25,000-\$49,999; 3: \$50,000-\$74,999; 4: \$75,000-\$99,999;5: \$100,000-\$124,999; 6: \$125,000-\$149,999; 7: more than \$150,000)*
Preparedness *(1: over-prepared to 5: not at all prepared)*	Mobility/ disability *(1: yes; 0: no)*
Social capital *(1: yes; 0: no)*	Age ten *(1: yes; 0: no)*
Information reliability *(5: almost always to 1: never)*	Elderly *(1: yes; 0: no)*
Residence Type *(1: single-family housing; 0: multiple units)*	

METHODOLOGY

The focus of the current paper is to specify a statistical model with an appropriate fit and an efficient number of predictor for determining the households' zone of tolerance to power outages based on consideration of different influencing factors. First, the models were fitted to data, including all the available predictors, and then various algorithms were applied on models to find the best subset of the variables for estimating the zone of tolerance. The proper model for each class and the efficient subset of variables were selected, and then the selected models were compared based on their prediction accuracy using cross-validation.

Models Implemented

The zone of tolerance is measured based on the number of days that household could tolerate the disruption. Three modeling classes (Poisson regression models, Accelerated Failure Time (AFT) models, and ensemble learning methods (random forest and boosting)) were evaluated to find a suitable method for modeling household susceptibility measure by the zone of tolerance. The remainder of this section discusses each class of the models and their results.

Poisson Regression Model

Poisson regression, which is a type of generalized linear models is useful when dealing with count data. Poisson distribution for the random component can take non-negative integer values and is a right-skewed distribution (Agresti 2007). As the zone of tolerance is measured in the number of days and has a positive skewness (Figure 1), Poisson regression could be an appropriate candidate for modeling zone of tolerance.

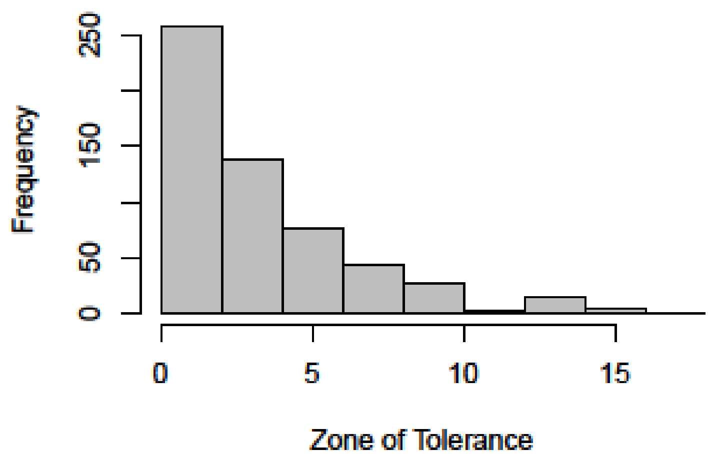

Figure 1. Histogram of the Zone of Tolerance

Poisson regression modeling, however, is based on an assumption that the mean and the variance of the distribution are equal. This assumption is not valid for the zone of tolerance, which has a mean of 3.83 and variance of 13.45 (based on the survey responses). Negative Binomial regression (Cameron and Trivedi 2013) is another method which can be used for such count data. This method accounts for the heterogeneity in the data by allowing the variance to exceed the mean. The Negative Binomial model with a log link function was tested, as shown in Equation 1.

$$\log \mu_i = x_i^T \beta + \varepsilon_i \tag{1}$$

where, μ_i is the mean tolerance, x_i^T is the vector of the predictors, β is the vector of parameters, ε_i is the error term, and $\exp(\varepsilon_i)$ being a random gamma variable with a mean of1 and a variance of α .

Accelerated Failure Models

AFT models are a type of survival analysis approaches for the time-to-event data. Survival data are generally positive and non-symmetrically distributed often with a positive skew. Such characteristics make survival analysis a proper model for the zone of tolerance. These models have shown to be successful choices in predicting the restoration time of the power outages (Liu et al. 2007). Adopting an AFT model would enable relating the zone of tolerance to the predictors with a linear relationship, as shown in Equation 2.

$$\log \mu_i = x_i^T \beta + \varepsilon_i \tag{2}$$

In this equation μ_i is the mean tolerance, x_i^T is the vector of predictors, β is the vector of parameters, and ε_i is error terms which are assumed to be independently distributed.

We used residual values of the models to investigate which distribution for the survival time would best fit the data. To this end, Kaplan-Meier estimator of residuals was used to determine whether the assumed distribution sufficiently fits the data (Hosmer and Lemeshow 1999). Exponential, Weibull, Log-logistic, and Lognormal distributions were considered for error terms ε_i. As shown in Figure 2, the logistic and Lognormal distributions yield a better fit compared to Exponential and Weibull.

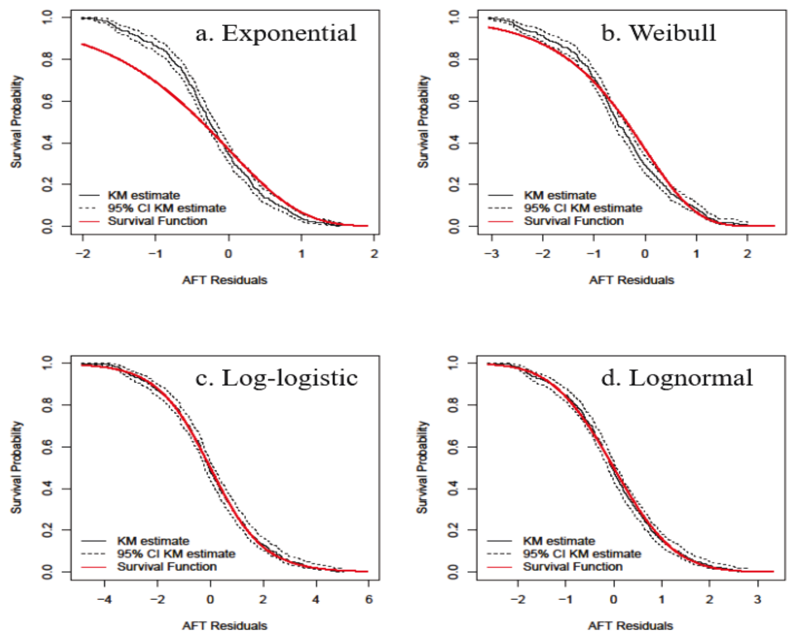

Figure 2. Checking the Fit of the Survival Models (a) Exponential (b) Weibull, (c) log-logistic, and (d) log-normal AFT models.

Ensemble Learning Methods

Random forest and boosting are also examined to model the zone of tolerance. This class of models does not use a specific probability distribution for the response variable (Nateghi et al. 2014). Random forest models generate a number of decision trees by drawing bootstrap resamples from the dataset and use their averages for the prediction. The regression tree for generating each of the decision trees is built by randomly selecting m variables out of the total number p predictors to decorrelate the trees. Boosting is based on a similar approach; however, the trees are generated sequentially. This means that each tree is not generated with the bootstrap sample; instead, the tree is fit on the modified form of the original dataset (James et al. 2014). In both methods, one needs to determine the tuning parameters. In this study, cross-validation was used to find the optimum tuning parameters based on the test error. Random forest model was generated using 250 trees and $m = 2$. Boosting was developed by 100 trees, shrinkage $\lambda = .04$, and depth =1.

Model Selection

The stepwise model selection algorithm utilized to choose the best subset of the influencing variables for the Negative Binomial model and AFT models. Here, the goal was to find the model which can properly fit the data and also control for the number of variables included in the model. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used for the stepwise model selection. AIC criteria may tend to choose more variables when the sample size is large; thus, we examined BIC as well, which imposes a higher penalty for having many variables as shown by the equations below.

$$AIC = -2 \times \ln(L) + 2 \times (p) \tag{3}$$

$$BIC = -2 \times \ln(L) + p \ln(n) \tag{4}$$

L refers to the likelihood of the fitted model, p is the number of parameters used in the model, and n is the sample size.

Table 2. Summary of the Results for the Zone of Tolerance Models

Variables	Negative Binomial		AFT-Lognormal		AFT-Loglogistic	
	AIC	BIC	AIC	BIC	AIC	BIC
Substitute	✓	✓	✓	✓	✓	✓
Experience	✓	✓	✓	✓	✓	✓
Need	✓	✓	✓	✓	✓	✓
Service expectation	✓		✓	✓	✓	
Preparedness						
Social capital	✓		✓			
Information	✓	✓	✓	✓	✓	✓
Residence Type	✓					
Ownership						
Residence duration			✓		✓	
income						
Race minority	✓		✓	✓	✓	✓
education						
Mobility/ disability					✓	
Age ten						
Elderly						

RESULTS

The summary of the results for selecting the best subset of the variables to be included in the models is presented in Table 2. Based on the results shown in Table 2, we ultimately selected a model containing substitute, experience, need, information, racially minority, service expectations, social capital, and residence duration. Substitute, experience, need, information, and racially minority are significant factors in almost all the models using AIC and BIC criteria (Table 2), and thus included in the model. Service expectation of a household was included in the model as it influences the general level of preparedness of the households. Social capital, which refers to the resources available in a household's social network, was chosen since it has been shown to be an important variable in household recovery in disasters (Aldrich 2011). Residence duration affects the protective actions of the households. Households which lived in their homes for a longer period of time are more probable to take effective risk mitigation actions such as buying a generator (Lindell and Perry 2000; Stein et al. 2014); therefore, residence duration is included in the model.

The three classes of models were developed using the selected subset of the variables. Negative Binomial model, AFT- Loglogistic, AFT-Lognormal, Random forest, and Boosting were compared based on their test error to select a proper modeling class. Here we compared the out-of-sample modeling accuracy of the models by using cross-validation. The model performance was calculated using 10-fold cross-validation. Subsequently, the data were divided into 10 equally sized groups, the first group was held out as the validation dataset, and the model was created by the remaining data. Rotating the validation dataset, the final cross-validation error was computed by averaging each test error. Calculated Mean square error (MSE) of the models is presented in Table 3.

Table 3. Prediction Results for the Zone of Tolerance Models

Model	MSE
Negative Binomial	15.87
AFT-Lognormal	13.48
AFT-Loglogistic	12.29
Random forest	12.55
Boosting	12.39

The results of Table 3 suggest that the most accurate models are AFT with Loglogistic assumption of the errors, boosting, and random forest. We select the AFT model as the final model for measuring the zone of tolerance because of its simplicity over the random forest and boosting. The rest of this section presents the fit of the AFT model.

Table 4. Final Loglogistic AFT Model for Prediction the Zone of Tolerance

Variables	Mean	St. Error	P-value
Intercept	1.7825	0.1798	<0.0001
Substitute	0.2606	0.0773	0.0008
Experience	0.2363	0.0954	0.0132
Need	-0.1379	0.0326	<0.0001
Service expectation	0.0240	0.0102	0.0181
Social capital	0.0798	0.0642	0.2130
Information	0.0887	0.0328	0.0069
Residence duration	0.0040	0.0023	0.0820
Race minority	-0.2146	0.0702	0.0023
Log(scale)	-0.9373	0.0385	<0.0001

The results of the AFT model show that households which have a substitute for the zone of tolerance have a higher zone of tolerance for power outages. Previous experience of the households leads to a higher level of tolerance for power outages. These households have a better perception regarding the prolonged service outages and are better capable of withstanding the inverse impacts. The greater a household need for services, the lower their tolerance to the disruptions. For example, households which need the electricity for using a medical device would have a lower tolerance to the service losses. Households' expectation of the potential disruptions is associated with their tolerance. Their expectation affects their preparation and other protective actions that household would take to mitigate the risks, and therefore influence their zone of tolerance. Social capital was also included in the model, and it is shown to have a positive association with the zone of tolerance. Households with access to more reliable information have reported a greater zone of tolerance. Access to proper information enables households to better prepare and cope with service losses. Residence duration of households positively influences their zone of tolerance, which is mainly due to their higher protective actions. Finally, the model suggests that racial minorities have a lower zone of tolerance to the power outages. This result highlights a disparity in household susceptibility for vulnerable populations such as racial minorities.

CONCLUSION

This study examined the zone of tolerance for characterizing household susceptibility to power outages caused by natural disasters. Using survey data collected from the affected households in the aftermath of Hurricane Harvey, multiple empirical models were developed for determining households' tolerance to the power outages. Negative Binomial models, AFT models, and ensemble learning methods were examined for determining the most appropriate statistical model.

The contributions of this study are threefold. First, this study identified the influencing factors such as having a substitute, having previous experience with a disaster, the need to the service, access to reliable information, being racially minority, household's expectation of the disruptions, social capital, and duration of the residents as the best subset of variables for determining the zone of tolerance. Second, the analysis showed that the AFT model with Loglogistic distribution outperforms the other methods with its prediction accuracy. The empirical model of household susceptibility complements the existing models of power outage prediction in order to determine the societal risks of prolonged power outages in extreme weather events and other natural disasters. Third, and from a practical perspective, the developed empirical model enables estimation of households' susceptibility and risks due to infrastructure service losses and provides utility companies with a tool to strategically allocate the resources to minimize the societal impacts of service disruptions through better investment prioritization and restoration operation planning.

ACKNOWLEDGMENT

The authors would like to acknowledge the funding support from the National Science Foundation under grant number 1846069 and National Academies' Gulf Research Program Early-Career Research Fellowship. Any opinions, findings, conclusion, or recommendations expressed in this research are those of the authors and do not necessarily reflect the view of the funding agencies.

REFERENCES

Agresti, A. (2007). *An introduction to categorical Data Analysis. Annu. Rev. Sociol*.

Aldrich, D. P. (2011). "The power of people: Social capital's role in recovery from the 1995 Kobe earthquake." *Natural Hazards*, 56(3), 595–611.

Applied Technology Council. (2016). "Critical assessment of lifeline system performance: understanding societal needs in disaster recovery." *Prepared for U.S. Department of*

Commerce National Institute of Standards and Technology, Engineering Laboratory, Gaithersburg, MD., NIST CGR(16-917–39).

Berkeley III, A. R., and Wallace, M. (2010). *A Framework for Establishing Critical Infrastructure Resilience Goals. National Infrastructure Advisory Council*.

Cameron, A. C., and Trivedi, P. K. (2013). *Regression Analysis of Count Data*. Econometric Society Monographs, Cambridge University Press.

Coleman, N., Esmalian, A., and Mostafavi, A. (2019). "Equitable Resilience in Infrastructure Systems: Empirical Assessment of Disparities in Hardship Experiences of Vulnerable Populations during Service Disruptions." *Natural Hazards Review*.

Dong, S., Wang, H., Gao, J., and Mostafavi, A. (2019a). "Robust component: a robustness measure that incorporates access to critical facilities under disruptions." *Journal of Royal Society Interface*.

Dong, S., Wang, H., Mostafizi, A., Gao, J., and Li, X. (2019b). "Measuring the topological robustness of transportation networks to disaster-induced failures: A percolation approach." *Journal of Infrastructure System*.

Esmalian, A., Dong, S., Coleman, N., and Mostafavi, A. (2019a). "Determinants of risk disparity due to infrastructure service losses in disasters: a household service gap model." *Risk Analysis*.

Esmalian, A., Rasoulkhani, K., and Mostafavi, A. (2019b). "Agent-Based Modeling Framework for Simulation of Societal Impacts of Infrastructure Service Disruptions during Disasters." (June), 15–23.

Hosmer, D. W., and Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley & Sons, Inc., New York, NY, USA.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.

Lindell, M. K., and Perry, R. W. (2000). "Household adjustment to earthquake hazard. A review of research." *Environment and Behavior*, 32(4), 461–501.

Liu, H., Davidson, R. A., and Apanasovich, T. V. (2007). "Statistical forecasting of electric power restoration times in hurricanes and ice storms." *IEEE Transactions on Power Systems*, 22(4), 2270–2279.

Nateghi, R., Guikema, S., and Quiring, S. M. (2014). "Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models." *Risk Analysis*, 34(6), 1069–1078.

Rasoulkhani, K., and Mostafavi, A. (2018). "Resilience as an emergent property of human-infrastructure dynamics: A multi-agent simulation model for characterizing regime shifts and tipping point behaviors in infrastructure systems." *Plos One*, 13, e0207674.

Stein, R., Buzcu-Guven, B., and Subramanian, D. (2014). "The Private and Social Benefits of Preparing For Natural Disasters." *Private and Social Benefits of Preparing International Journal of Mass Emergencies and Disasters*, 32(3), 459–483.

Tabandeh, A., Gardoni, P., and Murphy, C. (2018). "A Reliability-Based Capability Approach." *Risk Analysis*, 38(2), 410–424.