Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

SMOTE-LASSO MODEL OF BUSINESS RECOVERY OVER TIME - CASE STUDY OF THE 2011 TOHOKU EARTHQUAKE

Rodrigo Costa, Jack W. Baker

Abstract

A methodology is presented to combine the synthetic minority over-sampling technique and the least absolute shrinkage and selection operator to analyze survey data and identify business characteristics correlated with recovery within selected time windows. The methodology addresses challenges that arise when data is imbalanced, and predictors are collinear. A case study using data from a survey of business recovery conducted one year after the 2011 Tohoku earthquake is presented to demonstrate the methodology's application. The survey collected data on 30 predictors describing the physical damage and utility disruptions experienced by the businesses and their sector, size, disaster preparedness, and recovery financing alternatives. The methodology identifies a strong correlation between physical damage and business recovery within 30 days. Industry sector, size, disaster preparedness, and disaster financing become statistically significant when recovery over longer periods is considered.

1 Introduction

Although a community of practice around modeling disaster loss and recovery has arisen in the last decade (Miles *et al.*, 2019), our understanding of the impact of a disaster on businesses remains limited (Brown *et al.*, 2019). On the one hand, larger economic cycles exert a strong influence on the well-being of individual firms. This makes it difficult to disaggregate macroeconomic and disaster-related effects. On the other hand, disaster recovery data collection is seldom performed systematically, and few disasters have been investigated from the perspective of business recovery.

Comprehensive studies of business recovery in the US demonstrate that direct physical damage is only one of the many factors influencing business loss and recovery (Dahlhamer and D'Souza, 1995; Dahlhamer and Tierney, 1998; Webb *et al.*, 2000; Alesch *et al.*, 2001). Often, physical damage plays a secondary role. This is because businesses may be affected by factors such as interruptions in supply chains (Kay *et al.*, 2019), demand changes (Sampson *et al.*, 2018), and the need to temporarily relocate after a disaster (Morrish and Jones, 2020). Disruptions to employees' livelihoods or commuting routes may also impact businesses. These external factors are dependent on the type of hazard, the extent of the damage, the community affected, the local economy and require extensive contextual knowledge to be understood.

To gain insights on the factors that make businesses more resilient to disasters, scholars have often relied on field studies and surveys of the affected organizations. Surveys collect data on businesses' characteristics and one or more metrics of recovery, e.g., current productivity or profitability, number of employees, and

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

financial status. Treating businesses' characteristics as independent variables and the recovery metric as the dependent variable, survey data are analyzed using statistical methods, e.g., ANOVA (Corey and Deitch, 2011), correlations tests (Brown *et al.*, 2015), statistical difference tests (Chang and Falit-Baiamonte, 2002), and logistic regression (Webb *et al.*, 2002). These analyses help to identify correlations between the business characteristics and the recovery metric, which in combination with empirical knowledge of the community and disaster contexts help scholars understand the bottlenecks for the recovery of businesses.

These studies have contributed to the understanding of the correlations between business characteristics and their recovery capacity. However, survey data are often collected several months after the disaster. In many cases, survey questions assess the recovery metric using scales that are independent of time, e.g., recovered or not recovered (Dahlhamer and D'Souza, 1995), or doing better, the same, or worse (Corey and Deitch, 2011). Consequently, if a business characteristic is strongly correlated with recovery within one month, but recovery is measured after six months, this information is lost. We argue that valuable insights can be obtained if correlations between businesses' characteristics and their capacity to recover within selected time windows are identified.

Identifying the business characteristics correlated with recovery at different time windows requires transforming the variable representing the recovery metric. This results in a class imbalance in the dependent variable. Furthermore, because we are interested in identifying the business characteristics strongly correlated with recovery for each time window, a statistical model with variable selection capabilities is desirable. This paper introduces a methodology for conducting logistic regression in combination with minority over-sampling and variable selection. The methodology allows logistic regression to be applied to imbalanced data sets containing predictors linearly correlated and identify among these the most significant. This methodology is applied to a case study involving survey data on business recovery after the 2011 Tohoku Earthquake. It is demonstrated how the proposed methodology can identify business characteristics correlated with recovery within a few days and those that become more statistically significant when recovery is extended over months.

2 SMOTE-LASSO Methodology

Consider a survey that collected data on business characteristics and time until recovery, T, independently of how recovery is being measured. The dark columns in Figure 1 are a illustration of the distribution of the number of businesses recovered at several time windows, e.g., $t_1, \ldots, t_n$. Identifying correlations between the business characteristics and their ability to recover *at* specific times may be misleading. Consider that the businesses characteristics are aggregated in a vector $\mathbf{X}$. If X_i is positively correlated with recovery at t_2, it will likely be negatively correlated with recovery at t_3. This may lead to erroneous interpretations of the statistical effect of X_i on recovery. For this reason, we consider the number of business that recovered *within* a given t as our dependent variable. This variable is indicated in Figure 1 with light gray bars. Thus, a predictive model is developed for the probability of $T \leq t$ (where T is time to recovery), given a vector $\mathbf{X}$ of business characteristics: $\mathrm{P}(T \leq t|\mathbf{X})$. Since the independent variable, $T \leq t$, has multiple discrete levels, logistic regression is an appealing alternative for fitting the model. Because these levels are not mutually exclusive, using a multinomial logistic regression approach is not an alternative. For this reason, a binary logistic regression is fitted for for $\mathrm{P}(T \leq t|\mathbf{X})$ for each t of interest.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

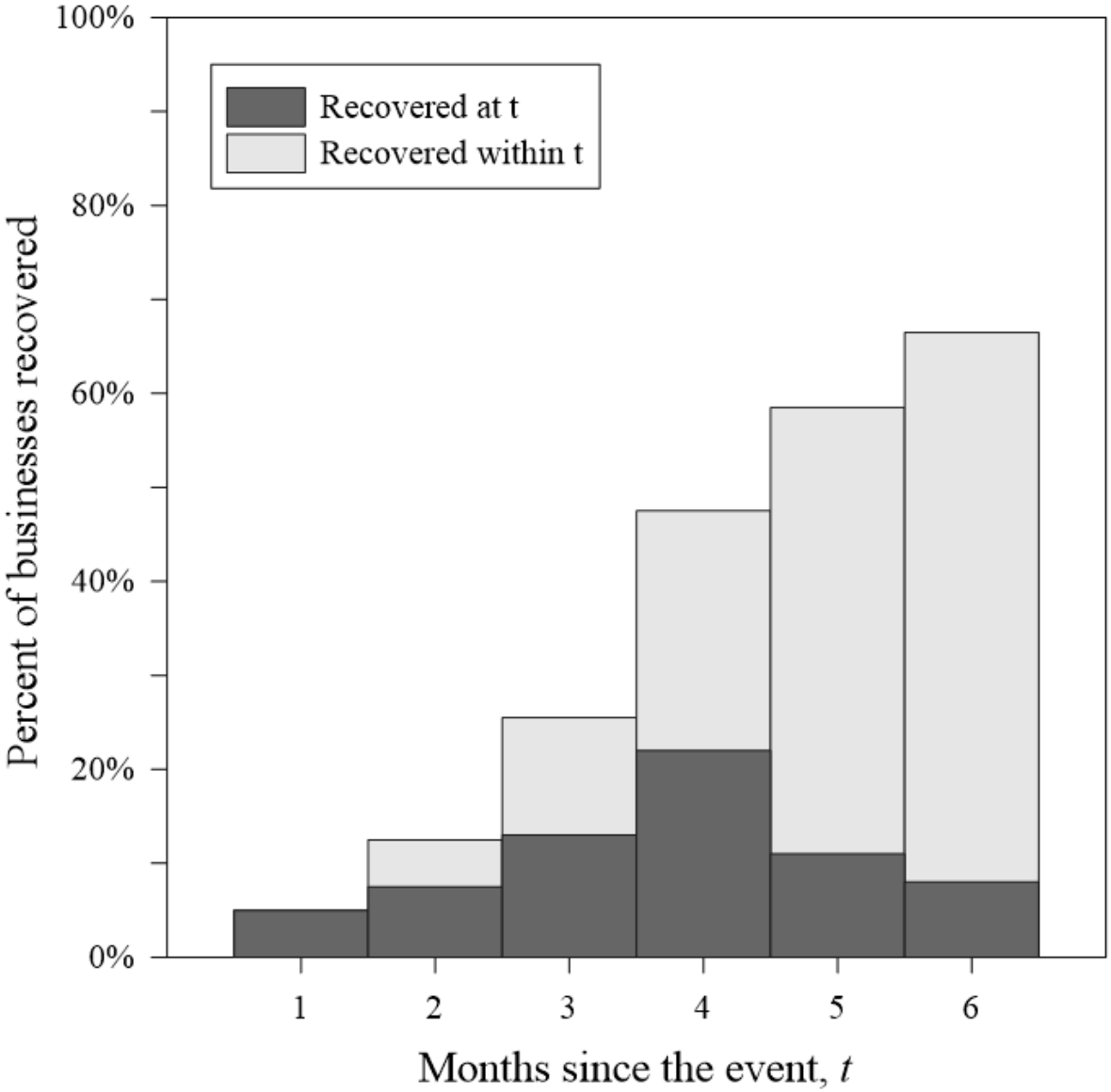

Figure 1. Illustrative representation of the percent of businesses recovered as a function of time after a disaster.

One challenge that arises from defining the dependent variable as binary with classes $T \leq t$ and $T > t$ is that class imbalance becomes unavoidable for small and large t values. Consider the model to be fitted for $t = 2$ in Figure 1. The numbers of businesses before and after $t = 2$ may be significantly different. This class imbalance may lead to predictors that are skewed towards the over-represented class. Another challenge is identifying the business characteristics strongly correlated with $\mathrm{P}(T \leq t)$. Surveys of business recovery often collect several business characteristics. Building a model for $(T \leq t|\mathbf{X})$ using a large number of variables as predictors can cause two problems. First, a model with several predictors may suffer from multi-collinearity,

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

that is, correlations among predictors reducing the explanatory power of any individual variable when all the others are in the model (Agresti, 2003, p. 212). Second, there is often a trade-off between a model's ability to fit the training data and its ability to predict data on which it was not trained. This over-fitting is more prevalent when more predictors are included in the model (Friedman *et al.*, 2001, Section 7.2).

Figure 2 illustrates the proposed methodology to select the most significant predictors of P($T \leq t$). The first step is to randomly split the data into training and testing sets. The training set is subjected to a minority over-sampling procedure to reduce class imbalance. Then, the multi-split algorithm is used to select the significant predictors (Meinshausen *et al.*, 2009). The significant predictors are used on the balanced training data to obtain the final models, which are used for prediction on the test data set. Details of these steps are provided in the following.

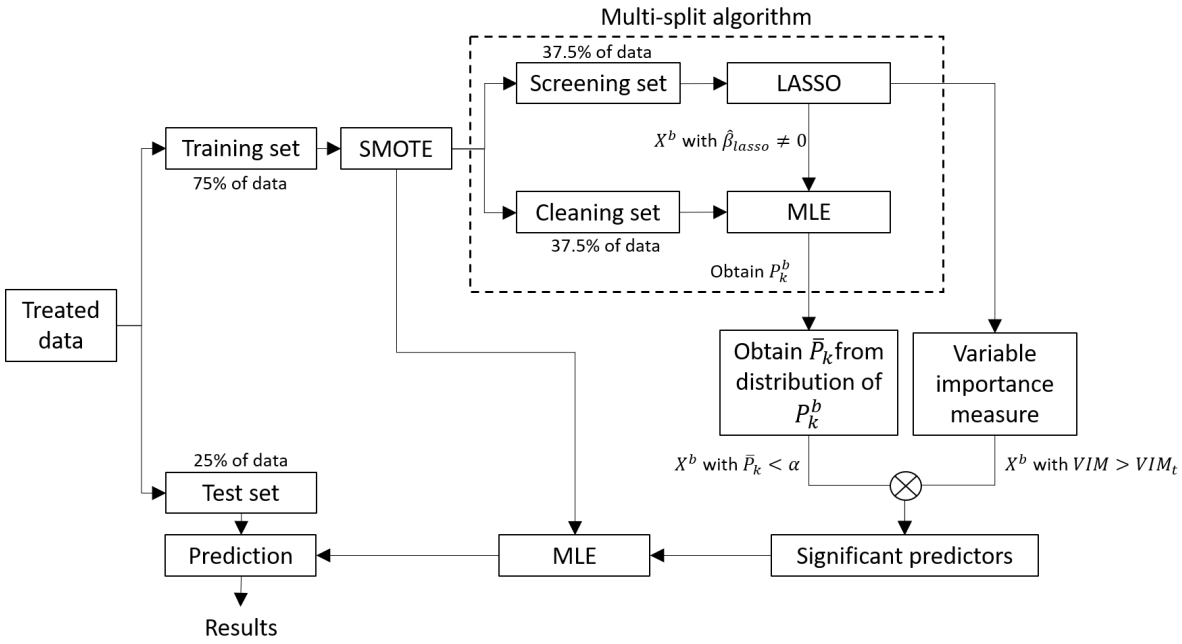

Figure 2. SMOTE-LASSO methodology for variable selection.

2.1 Synthetic Minority Over-sampling Technique

The 'class-imbalance' that arises when developing a model for P($T \leq t|\mathbf{X}$) has important consequences for the predictive model: it usually leads to classifiers that have poor predictive accuracy for the minority class, and that tend to classify most new samples in the majority class (Blagus and Lusa, 2013). Put simply, if 90% of the samples are of one class, a classifier that always predicts that class has a low error and may be selected as the best model.

To address this problem, the Synthetic Minority Over-sampling TEchnique (SMOTE) creates synthetic samples of the minority class based on its nearest K minority neighbors (Chawla *et al.*, 2002). For a variable with minority class represented by $\boldsymbol{m}$, the SMOTE samples, $\mathbf{s}$, are linear combinations of two similar samples of this class, say $\mathbf{m_I}$ and $\mathbf{m_R}$, as

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

$$\mathbf{s} = \mathbf{m_I} + u(0,1) \cdot (\mathbf{m_R} - \mathbf{m_I}) \tag{1}$$

where $u(0,1)$ is a uniformly distributed number between 0 and 1, and $\mathbf{m_R}$ is randomly chosen among the K minority class nearest neighbors of $\mathbf{m_I}$.

2.2 Logistic Regression

Logistic regression is employed to estimate $\mathrm{P}(T \leq t|\mathbf{X})$, the probability that a business recovered within t days. Logistic regression is named after its logistic functional form, an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1

$$\mathrm{P}(T \leq t|\mathbf{X}) = \frac{\exp(\boldsymbol{\beta}(t)^T \boldsymbol{X})}{1 + \exp(\boldsymbol{\beta}(t)^T \boldsymbol{X})} \tag{2}$$

where $\mathbf{X} = 1, X_1, \ldots, X_p$ are the p predictor variables, and $\boldsymbol{\beta} = \beta_0, \ldots, \beta_p$ are the coefficients of the model, which can be fitted using maximum likelihood. Note that the vector $\boldsymbol{\beta}$ is a function of the selected t. This functional dependence on t is noted in Equation 2, but for brevity is omitted in the following. The log-likelihood for N observations is (Friedman *et al.*, 2001)

$$\ell(\boldsymbol{\beta}) = \boldsymbol{y}\boldsymbol{\beta}^T \boldsymbol{x} - \log(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x})) \tag{3}$$

where $\boldsymbol{y}$ is an $N \times 1$ vector of zeroes and ones, $\boldsymbol{x}$ is an $N \times p$ matrix of observations, and the vector $\boldsymbol{\beta}$ that maximizes the log-likelihood estimator is chosen as the model coefficients, that is

$$\boldsymbol{\beta}_{mle} = \arg\max_{\boldsymbol{\beta}} \left\{ \boldsymbol{y}\boldsymbol{\beta}^T \boldsymbol{x} - \log(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x})) \right\} \tag{4}$$

Everything else being equal, the coefficient β_i represents the increase in the odds-ratio of $T \leq t$ to $T > t$ when X_i is increased by one unit.

2.3 Least Absolute Shrinkage and Selection Operator

As previously discussed, a model with many predictors will incur problems with multi-collinearity and over-fitting. The former occurs when many correlated variables are included in a regression model. A large positive coefficient on one variable can be canceled by a similarly large negative coefficient on its correlated cousin (Agresti, 2003), causing their coefficients to become poorly determined and exhibit high variance. Over-fitting occurs when the regression model is over-trained on the data used to construct it but has low predictive capacity when applied to new data. Both issues are mitigated with simpler regression models having fewer predictors. This paper uses the Least Absolute Shrinkage and Selection Operator (LASSO) to select a smaller set of significant predictors from a large set of available ones. The LASSO imposes a penalty on the size of the coefficients $\boldsymbol{\beta}$, forcing predictors with low predictive power to be selected out of

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

the final model (Tibshirani, 1996). A similar penalization is used in neural networks, where it is known as weight decay. Using the LASSO, the model coefficients are estimated as

$$\boldsymbol{\beta}_{lasso} = \arg\max_{\beta} \left\{ \boldsymbol{y}\boldsymbol{\beta}^T \boldsymbol{x} - \log(1 + \exp(\boldsymbol{\beta}^T \boldsymbol{x})) - \lambda|\boldsymbol{\beta}| \right\} \tag{5}$$

where $\lambda \geq 0$ is the Lagrangian multiplier that controls the amount of regularization and is usually chosen via cross-validation. As $\lambda \to 0$ the $\boldsymbol{\beta}_{lasso}$ converge to $\boldsymbol{\beta}_{mle}$ given in Eq. 4. Conversely, as $\lambda \to \infty$ the $\boldsymbol{\beta}_{lasso} \to 0$. Due to the regularization, certain coefficients are shrunk to zero. Thus, the LASSO prefers sparse models, i.e., with few predictors having non-zero coefficients.

The variable selection included in the LASSO incurs in the problem that the p-values can no longer be trusted since the selected variables will tend to be the ones that are significant (Lee *et al.*, 2016). The multi-split algorithm described in the next section is deployed to address this problem.

2.4 Multi-split algorithm

Meinshausen *et al.* (2009) propose a multi-split algorithm for refining the variable selection using the LASSO and obtaining p-values. This approach can be implemented using free and extensively validated statistical packages (R Core Team, 2013), and it has been previously employed to study disaster recovery (Nejat and Ghosh, 2016). The algorithm splits the training data into a screening set and a cleaning set. The LASSO is applied to the screening set to identify the subset of predictors with $\boldsymbol{\beta_{lasso}} \neq 0$, denoted $\mathbf{X}_{lasso}$. Then, using only $\mathbf{X}_{lasso}$ as predictors, a maximum likelihood model is fitted to the cleaning set. Unlike the LASSO, the maximum likelihood provides estimated p-values for each predictor. Because the LASSO results rely on the data split, this process is repeated for B random splits. Thus, B sets $\mathbf{X}^{(b)}_{lasso}$, as well as B estimates of the p-value for each predictor, are available. From the B sets of $\mathbf{X}^{(b)}_{lasso}$, the number of times each predictor is significant is counted. Defining this number η_{X_i}, a measure called the Variable Importance Metric, *VIM*, can be calculated as (Nejat and Ghosh, 2016)

$$VIM = \frac{\eta_{X_i}}{B} \tag{6}$$

For each predictor, a summary p-value is calculated from the B estimates. The predictors for which $VIM > 0.75$ and p-value < 0.05 are selected to be included in the final model. The implications of these thresholds are discussed in the Appendix. The multi-split algorithm's steps are described in Algorithm 1.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

Algorithm 1 Multi-split algorithm, adapted from (Nejat and Ghosh, 2016)

1. Define a vector $\boldsymbol{\eta} = 0$ with $|\boldsymbol{\eta}| = |\mathbf{X}|$

for $b = 1, 2, \ldots, B$ **do**

2. Randomly split the data set into $D^{(b)}_{screen}$ and $D^{(b)}_{clean}$.
3. On $D^{(b)}_{screen}$, use LASSO to find the set of predictors $\mathbf{X}^{(b)}_{lasso}$ with $\hat{\boldsymbol{\beta}}_{lasso} \neq 0$.
   3.1. η_i=η_i+1 at the position of each $\mathbf{X}^{(b)}_{lasso}$.
4. Obtain the maximum likelihood estimate of $\boldsymbol{\beta}$ for $\mathbf{X}^{(b)}_{lasso}$ using $D^{(b)}_{clean}$.
   4.1. Obtain the raw p-values, $\tilde{P}^{(b)}_k$, for the regression coefficients associated with the set of predictors $\mathbf{X}^{(b)}_{lasso}$.
   4.2. Set $\tilde{P}^{(b)}_k = 1$ for regression coefficients corresponding to the predictors not included in $\mathbf{X}^{(b)}_{lasso}$.
   4.3. The final p-value is given by $P^{(b)}_k = \min\left(\tilde{P}^{(b)}_k \times |\mathbf{X}^{(b)}_{lasso}|, 1\right)$.

end for

5. Obtain the empirical quantile function q_δ, $\delta \in [0.05,1]$ for distribution of p-values obtained from the multi-split algorithm.
   5.1. Find δ^* that minimizes q_δ/δ.
   5.2. The quantile that yields the summary p-value is then given by $\min(4 \cdot q_{\delta^*}/\delta^*, 1)$.
6. Calculate the variable importance measure for each variable as $VIM_i = \eta_i/B$.
7. The set of predictors for which $VIM_i > 0.75$ and the summary p-value < 0.05 are considered the significant predictors.

3 The 2011 Tohoku Earthquake

The SMOTE-LASSO methodology proposed in this paper is used to investigate the characteristics of businesses affected by the 2011 Tohoku earthquake associated with recovery within different time windows. On March 11, 2011, a 9.1 M_w earthquake with an epicenter 70 kilometers east of the Oshika Peninsula of Tohoku struck Japan (Duputel *et al.*, 2012). The earthquake was followed by a tsunami with wave heights up to 40 meters, which caused damage to the Fukushima Daiichi power plant and a subsequent nuclear disaster. The compound result of these three events is the costliest disaster on record. Direct damage is officially estimated at ¥16.9 trillion (US$211 billion), including the value of damage to buildings, infrastructure, and other capital stocks (Kajitani *et al.*, 2013). For brevity, this compound event is referred to as the 'Tohoku earthquake' in this paper.

The Tohoku earthquake caused significant damage to the affected regions. Approximately 196,000 homes were damaged, of which nearly 45,000 were destroyed (Nanto, 2011, p. 1). More than 335,000 persons were displaced from the affected regions, and many lacked water and food for several days (Norio *et al.*, 2011). The Tokyo Electric Power Company (TEPCO) reduced its output by 21 GW, affecting 4.4 million homes (Norio *et al.*, 2011). Disruptions to two-thirds of the oil refineries between Tohoku and Kanto lead to widespread fuel shortages (Maruya, 2013). The earthquake and tsunami damaged 15 ports, 70 railway lines, and 23 railway stations, leading to severe transportation system disruptions (Kajitani *et al.*, 2013).

Many upstream industries were located in the affected area and suffered damage or experienced utility shortages, leading to supply chain interruptions. These interruptions caused a scarcity of products and

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

economic impacts across damaged and non-damaged areas (Norio *et al.*, 2011). Manufacturing production fell nearly 40% in the Tohoku region for the 12 months following the event (Matsushita *et al.*, 2017). Sales by department stores and supermarkets dropped by 20% in the Tohoku region, and 6% in Japan as a whole for March 2011 (Kajitani *et al.*, 2013). The number of international visitors to Japan dropped by 61% following the event, causing a significant impact on the tourism sector (Kajitani *et al.*, 2013).

Businesses confronting this devastation had limited ability to navigate the post-catastrophe environment. A survey of 736 large businesses conducted in 2009 by the Cabinet Office of the Japanese government identified that 55% had a disaster preparedness plan, and 25% were in the process of formulating one (Maruya, 2013). However, only 28% had a business continuity plan. Among medium-sized companies, only 36% had a disaster preparedness plan, 15% were formulating one, and only 13% had a business continuity plan. Cole *et al.* (2017) investigated business recovery after the 2011 Tohoku earthquake and identified that a continuity plan that allowed for the diversification of suppliers was beneficial for recovery. In another study, Matsushita *et al.* (2017) identified that businesses with continuity plans recovered at the same rate as those that did not suffer any physical damage. Conversely, businesses without a continuity plan lost 10% of their annual sales on average.

4 Data

This case study employs data obtained from surveys of 930 businesses conducted between October 2011 and February 2012. The data was collected by Sompo Risk Management, Inc. and provided to the authors. From the original data, incomplete responses are removed, and a total of 923 complete responses comprise the final data set. The survey asked about the businesses' physical damage and utility disruptions and the business size, industry sector, disaster preparedness, and source of recovery financing. The survey also asked how much time was needed to recover to pre-disaster levels of operation. This 'recovery time' is the variable of interest in this paper. The other survey responses are treated as potential predictors of recovery.

The survey also collected locations of main facilities from 386 businesses (42% of respondents). Figure 3 shows that most businesses were on the East Coast of the country, near the regions directly impacted by the event. Figure 3 includes contour lines of the peak ground acceleration to indicate the potential for direct shaking damage at the business location.

Direct physical damage from the earthquake and tsunami was assessed through individual "yes or no" survey questions. Respondents reported damage to headquarter buildings, sales stores, production plants, or warehouses. A fifth question allowed respondents to report experiencing no damage to any of their facilities. Damage to headquarter buildings was strongly correlated to damage to sales stores. These two variables are thus coded as a single variable. For the same reason, damage to production plants or warehouses is coded as a single variable. Table 1 shows that the majority (55%) of all businesses did not experience damage to any of their facilities. Close to one-third of the businesses experienced damage to their headquarters or sales stores, and 19% reported damage to multiple facilities.

Businesses reported utility disruptions through individual "yes or no" questions. The length of the disruption was not captured. Table 2 shows that most businesses experienced some utility disruption. The most common disruption was to electrical power, with 76% of the businesses reporting experiencing it. Seventy-nine percent of the businesses experienced two or more utility disruptions.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

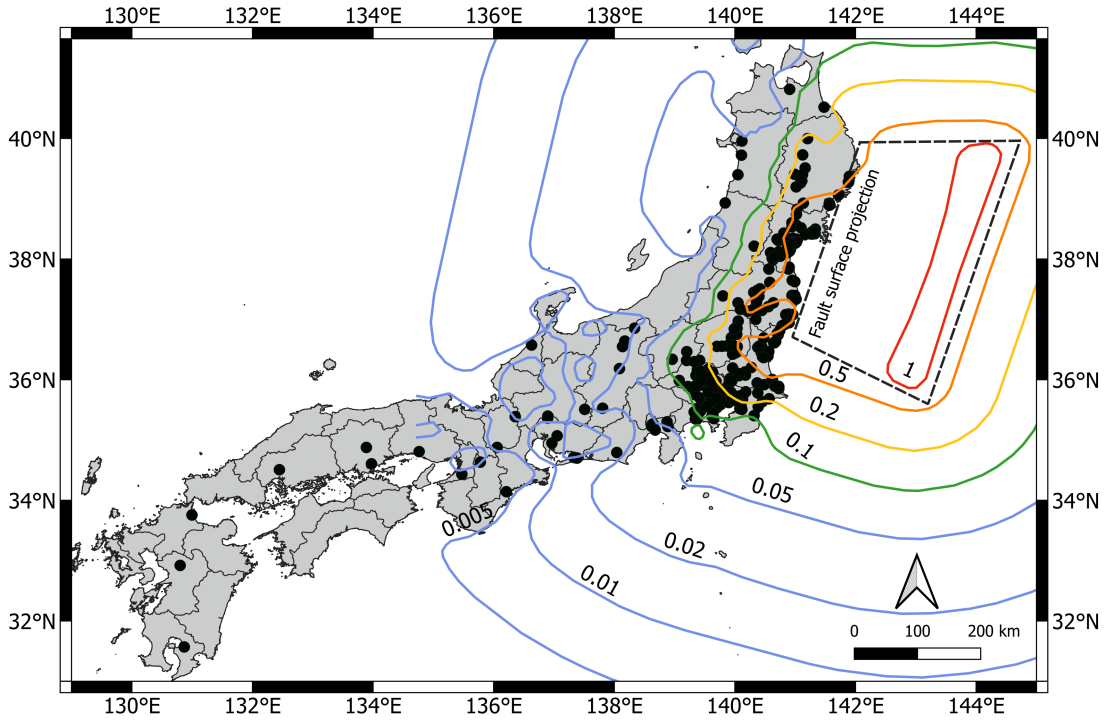

Figure 3. Location of businesses, along with fault surface projection (dashed lines) and contour lines of the peak ground acceleration in g (Worden *et al.*, 2020).

Table 1. Number of survey respondents reporting damage to various types of facilities. The number of reported cases exceeds 100% of respondents because some respondents had more than one type of damaged facility.

Facility damaged	Variable name	Count	%
None	D_1	504	55
Production plants or warehouses	D_2	374	41
Headquarters or sale stores	D_3	301	36

The survey included questions regarding the characteristics of the businesses. It identified a total of 27 different industry segments. These are grouped up into five sectors, as per Dahlhamer and Tierney (1998). Table 3 shows the prevalence of industry sectors among the businesses surveyed. The majority of the businesses, 56%, are in the manufacturing, construction, or contracting sectors. The finance, insurance, and real state sectors are the least represented, with only 7% of the total. The industry sector variable is coded so that the category "Other" is the baseline.

Businesses are also differentiated by size. Only 4% of businesses had fewer than 25 full-time employees,

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

Table 2. Number of survey respondents reporting disruptions to various types of utilities. The number of reported cases exceeds 100% of respondents because some respondents experienced multiple disruptions.

Type of utility disruption	Variable name	Count	%
Electrical power	U_1	701	76
Information and communication	U_2	659	71
Potable water	U_3	637	69
Gas	U_4	589	64
Sewage	U_5	555	60
Industrial water	U_6	524	57

Table 3. Number of survey respondents per industry sector.

Industry sector	Variable name	Count	%
Manufacturing, construction, or contracting	I_1	521	56
Services	I_2	131	14
Wholesale or retail	I_3	104	12
Finance, insurance, or real state	I_4	69	7
Other		98	11

which would categorize them as small businesses (Chang, 2010). Thus, all businesses with fewer than 300 employees are grouped, and two groups are defined. Table 4 shows that 60% of the businesses have ≥300 employees.

Table 4. Number of survey respondents per business size.

Industry size	Variable name	Count	%
≥300 employees	S_1	553	60
< 300 employees	S_2	370	40

Table 5 presents the prevalence of disaster preparedness measures taken by the survey respondents. Individual "yes or no" questions asked whether the business engaged in each preparedness activity. Data backup was the most common preparedness measure and creating a disaster action plan following closely. None of the measures were taken by the majority of the businesses. On average, 2.7 out of 11 disaster preparedness measures were adopted by the businesses investigated.

Data was collected on the sources of funding used for recovery. Only one business indicated using a public loan, which was not a sufficient sample to consider in the analysis below. The remaining businesses relied on insurance, private loans, internal reserves, or a combination of those. Two or three sources of financing were used by 9.6% businesses. Table 6 shows that the prevalent source of financing was insurance, with 36% of the businesses using it. It is also noted that the majority, 52%, reported not using any source of funding in particular. Table 1 showed that 55% of the businesses did not report any damage to their facilities, which

Table 5. Number of survey respondents reporting taking preparedness measures.

Preparedness measure	Variable name	Count	%
Data backup	P_1	456	49
Disaster action plan	P_2	379	41
Structural strengthening	P_3	318	34
Securing equipment	P_4	274	30
Disaster prevention office	P_5	255	28
Creation of continuity plan	P_6	253	27
Base isolation	P_7	145	16
Diversification of suppliers	P_8	139	15
Training on continuity plan	P_9	116	13
Cooperation among industry	P_{10}	102	11
Alternative customers	P_{11}	69	7

may explain the 52% that did not use any funding.

Table 6. Number of survey respondents reporting using each source of disaster financing. The number of reported cases exceeds 100% of respondents because some respondents used multiple sources of funding.

Funding source	Variable name	Count	%
None	F_1	482	52
Insurance	F_2	331	36
Internal reserves	F_3	179	19
Private loan	F_4	27	3

To determine the recovery time, T, the survey employed a multiple choice question with possible answers being t=0, 7, 14, 30, 90, 180, and more than 180 days. A time of zero days indicates that business operations were not impacted. Only 2% of the respondents indicated that recovery took longer than 180 days, indicating that they had not recovered at the time of the survey. Thus, the following sections study the characteristics of businesses that recovered within t=0, 7, 14, 30, 90, 180 days.

A third party obtained the data used in this study for private purposes. Thus, the data were not collected for this study's purpose, nor using a systematic approach (e.g. Stevenson *et al.*, 2018). Several potential drivers of business recovery were not surveyed, including the businesses' primary markets, how demand changes affected their recovery, and other describers of the community context. Although a question related to supply chain disruptions was included in the survey, less than 20% of the businesses answered the question. For this reason, supply chain interruptions were not included as a potential predictor.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

5 Results

To estimate the probability that a business recovered within t days, the variables in Tables 1 through 6 are used as predictors. The SMOTE-LASSO methodology is used to identify the predictors correlated with recovery within six intervals: 0, 7, 14, 30, 90, and 180 days. Recovery within 30 days, for example, indicates that the business recovered pre-disaster operation levels at any point during the first month of the event. Table 7 presents the estimated coefficients for the significant predictors at each t value. The standard errors of the coefficients are included in Table 8, in the Appendix. The variable names for the predictors are as listed in Tables 1-6. A positive coefficient value in Table 7 indicates that the predictor is correlated with an increased probability of recovery within t days. For example, D_1 indicates the absence of physical damage to the business' facilities, and the positive coefficient indicates that this is correlated with an increased probability of recovery within 30 days.

Table 7. Regression coefficients estimates.

	Coefficient estimates for P($T \leq t|\mathbf{X}$)					
Predictor	t=0	t=7	t=14	t=30	t=90	t=180
Intercept	-0.17*	-0.45*	0.93***	0.86***	–	1.90***
D_1	3.06***	3.26***	2.32***	3.72***	–	–
D_2	-1.31***	-1.34***,†	-1.13***	-1.56***	-2.10***	-1.85***
D_3	-1.40***	-0.97***,†	-1.68***	-0.54**,†	-2.09***	-1.15***
U_1	–	–	–	–	0.66***	–
U_4	–	–	–	–	-0.61***	–
I_1	–	–	–	–	1.17***	–
I_2	–	–	–	–	1.18***	–
I_3	–	–	–	1.25***	2.19***	0.64**
I_4	–	–	-0.77**	-0.23***	0.91**,†	-1.42***
S_1	–	–	–	–	1.60***	–
P_4	–	–	–	–	–	0.51***
P_7	–	–	-1.64***	-0.64**	–	–
P_{10}	–	–	0.87***	–	–	–
F_2	–	–	–	–	–	-0.57***
F_3	–	–	–	–	–	0.64***

Significance codes: 0 '***' 0.001 '**' 1 '*'
†: the variable is included because it was selected for the two adjacent t values.

Table 7 shows that physical damage is the predictor more strongly correlated with recovery within 30 days - the presence of damage decreases the probability of recovery. Certain industry sectors and preparedness measures are also strongly correlated with recovery within 30 days. Surprisingly, businesses whose buildings were improved with base isolation were less likely to recover within this period. Three factors may help explain these results. First, much of the disruption caused by the event was due to factors other than ground shaking (which the base isolation should mitigate). Second, the base-isolation might have been less effective than anticipated against long-period ground motions such as those produced by the 2011 Tohoku earthquake

(Hayashi *et al.*, 2018; Ariga *et al.*, 2006). Third, businesses with base-isolation may be more likely to be located in high-seismic risk regions that were more impacted by the event; we do not have detailed location data for all buildings in order to confirm this, but among businesses that disclosed their locations, those with seismically isolated buildings were more prominently located on areas that experienced higher ground acceleration, as shown in the Appendix. These factors may explain the negative correlation with recovery within 30 days.

When recovery is not completed within 30 days, several additional predictors become significant. In terms of utilities, disruptions to gas, U_4, are negatively correlated with the probability of recovery, whereas disruption to electric power, U_1, are positively correlated. Belonging to the wholesale and retail sectors has a strong and positive correlation with recovery. A large portion of businesses in the wholesale and retail sectors indicated experiencing power shortages. This may explain the positive correlation between disruption to electric power, U_1, and recovery. Financing only becomes a significant predictor when recovery is extended to 180 days. Businesses that relied on their own funds, F_3, are more likely to recover within 180 days than those dependent on insurance payments, F_2. In Table 7, a few estimates of the predictors D_2, D_3, and I_4 are marked with daggers. This indicates that, at these t values, these predictors were not initially included in the models by the methodology. However, these predictors are significant for the t immediately before and after the t in question. They were thus included in the model based on this information. As expected, their inclusion influences the intercept value, causes negligible changes to other predictors' coefficients, and improves the models' predictive power.

In Figure 4, on each plot, the red dots indicate the predicted probability that a business is recovered by time t, $\mathrm{P}(T \leq t|\mathbf{X})$, comparing it to the data in the test set, in black. The abscissa indicates the possible values for the log-odds of $T \leq t$ to $T > t$ without including the intercept term, $\ell_i = \beta_i X_i$, with $i > 0$. Thus, different abscissa values indicate different business characteristics. The size of the black dots indicates how many businesses possess the same characteristics. For example, the top-left plot in Figure 4 shows that the majority of the businesses that were not recovered by t=0 had a $\ell_i < 0$. For the smaller t values, fewer predictors are significant, and fewer ℓ_i values are possible. For $t \geq 90$ days, the number of significant predictors increases, and the logistic curve is more apparent. As t increases, the correlation between the values of ℓ_i and the recovery state of businesses becomes less clear. This indicates relatively smaller predictive power for the fitted models.

To evaluate the fitted models' predictive power, receiver operating characteristic (ROC) curves are used. ROC curves summarize the trade-off between the true positive and false-positive rates for a predictive model using different probability thresholds. The area under the ROC curve, i.e., the AUC, is often used as a metric of the quality of a model compared to a naive model that is correct only 50% of the time. The AUC can take values between zero and unity. Figure 5, the ROC curve for the naive model is the straight gray line with a 45-degrees slope, and its AUC is 0.5. Models with AUC>0.5 have some predictive power, larger AUCs being more desirable. Figure 5 shows the ROC curves for the six models, one for each t investigated. The AUCs for the fitted models are indicated in the figure. As anticipated, the fitted models' predictive power decreases as t increases and the models become less sparse. However, in all cases, the AUC is considerably higher than 0.5, indicating that the fitted models significantly outperform a naive model.

To further investigate the factors correlated with timely recovery, the significant predictors in Table 7 are combined into three groups. The first refers to immediate impact and include predictors of physical damage

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

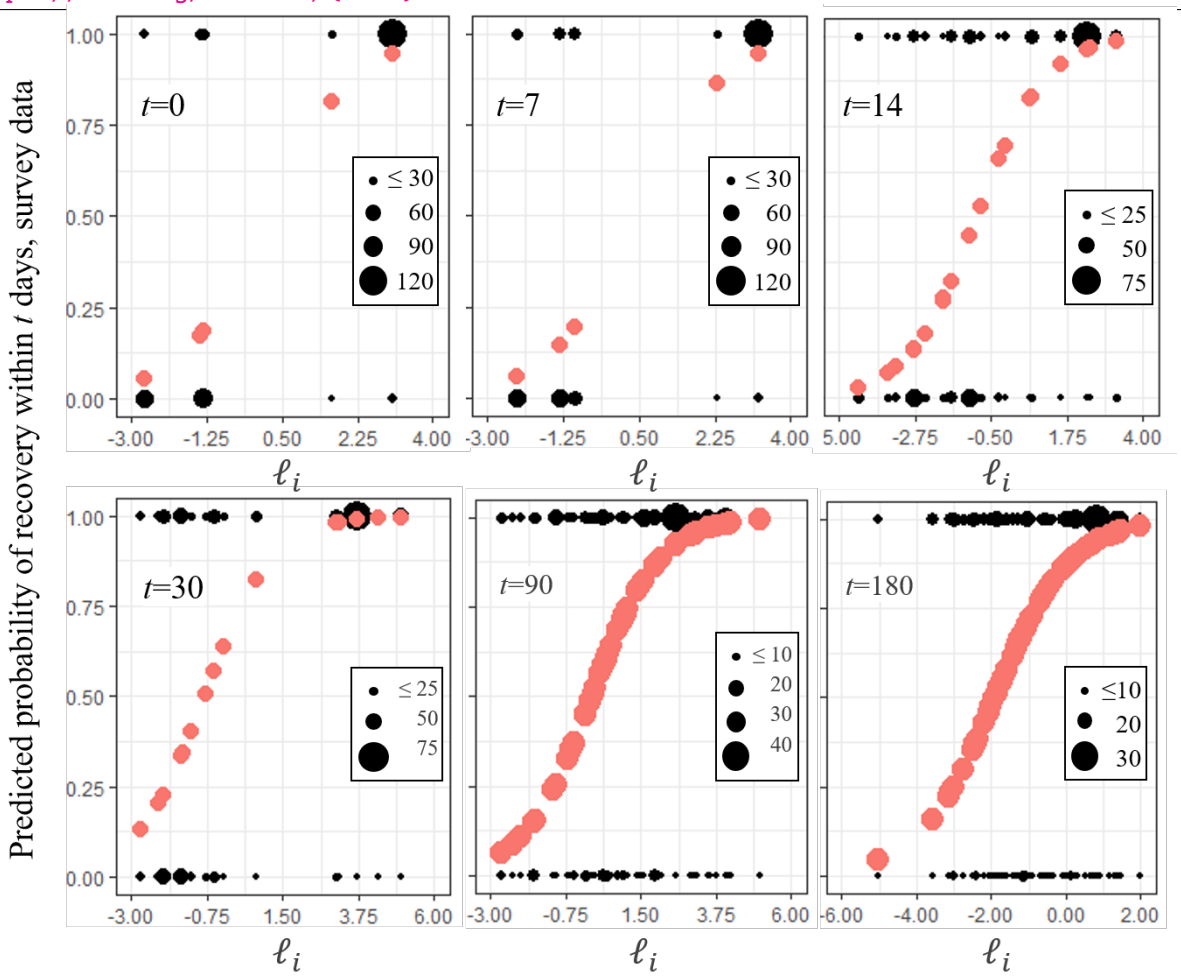

Figure 4. Predicted probabilities, $P(T \leq t|\mathbf{X})$, in red, compared to survey data, in black. One sub-figure is provided for each t of interest, with the t value labeled within the subfigure. The abscissa indicates the possible values for the log-odds of $T \leq t$ to $T > t$, i.e., $\ell_i = \beta_i X_i$. The ordinate axis indicates $P(T \leq t|\boldsymbol{X})$. The legend indicates the number of survey responses associated with which ℓ_i value.

and utility disruptions (D_i and U_i). The second group represents organizational characteristics such as size, industry sector, and financing (S_i, I_i, and F_i). The last group comprises all preparedness measures (P_i). Figure 6 displays the prevalence of the significant predictor groups for all t. In most cases, predictors associated with the immediate impact are the most prevalent. Predictors associated with preparedness are the least common. For the shortest recovery time intervals, only the immediate impact variables are relevant. For longer time intervals, more predictors associated with organizational and preparedness measures are correlated with recovery.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

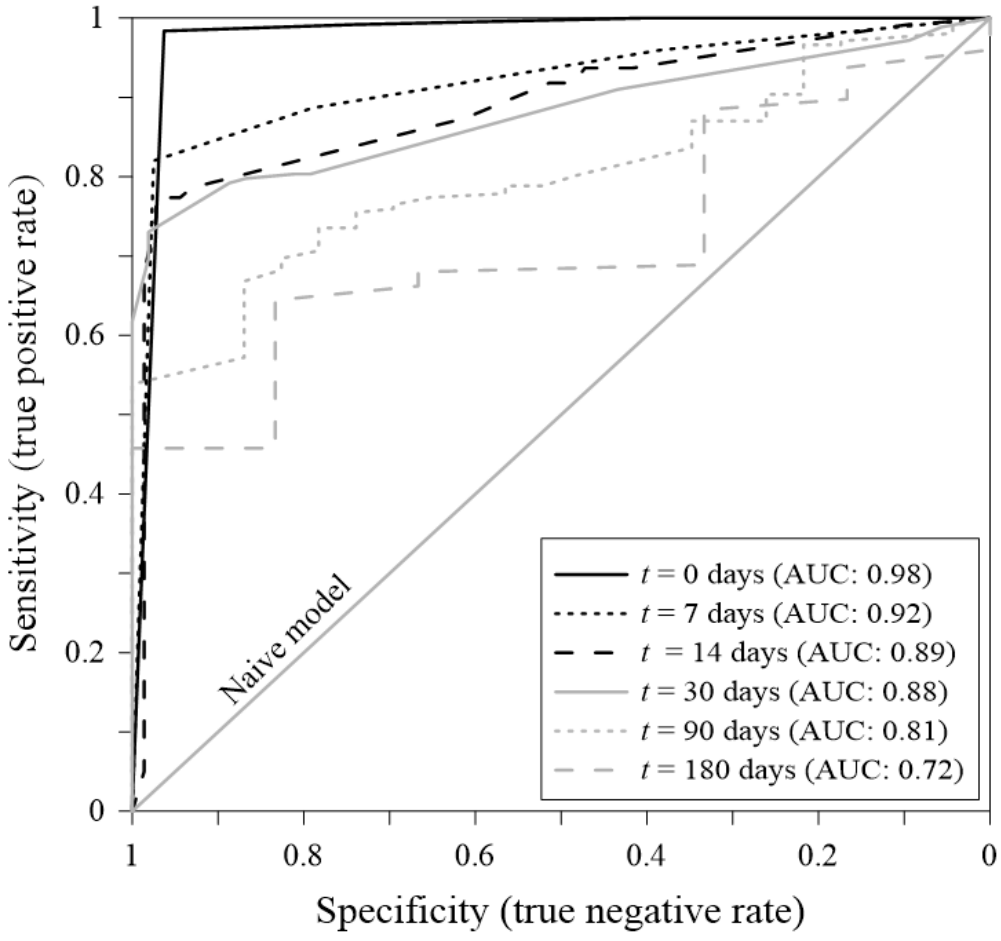

Figure 5. Receiver operating characteristic curves for different t values.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

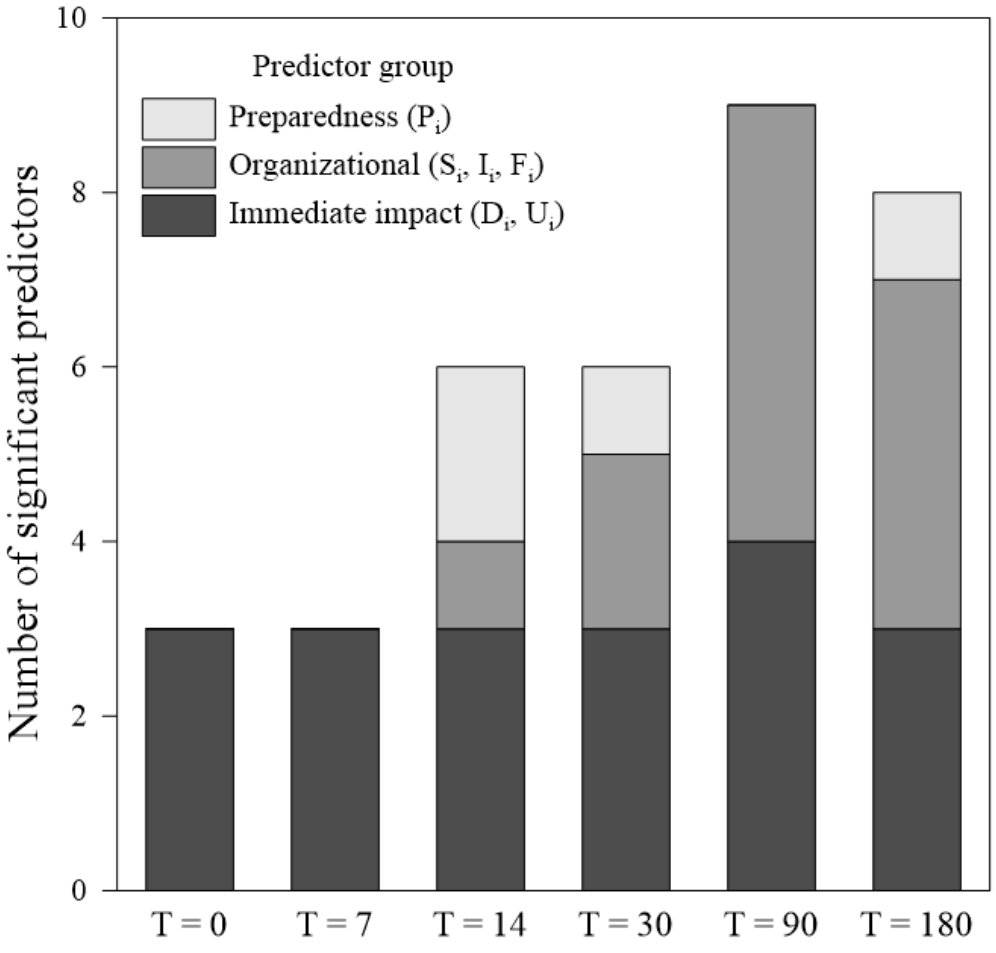

Figure 6. Numbers and types of significant predictor variables at each recovery time interval of interest.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

6 Conclusion

This paper introduces a methodology that can be applied to survey data to identify factors correlated with business recovery within selected time periods, t. These data are often imbalanced, in that for any t, the numbers of responses with $T \leq t$ and $T > t$ are significantly different. This imbalance can introduce biases in the modeling process and lead to a classifier with low predictive power for the minority class. Furthermore, a method is needed to identify the significant predictors from a larger set of candidates. The methodology in this paper combines minority over-sampling, the Least Absolute Shrinkage and Selection Operator, and a multi-split algorithm to address this problem. The methodology is applied to investigate the characteristics of businesses correlated with recovery within t after the 2011 Tohoku Earthquake. The results indicate that physical damage was strongly correlated with recovery in the first 30 days. When longer recovery times are considered, business characteristics become more important.

Some limitations in the case study results should be acknowledged. The majority of the surveyed businesses had more than 25 employees and were all clients of the same insurance company. Recovery is measured as regaining pre-disaster operation levels. Although this is a common metric of business recovery, it is arguably reductionist. Certain business characteristics that have been demonstrated to be correlated with recovery (e.g., supply chain interruptions, changes in demand, and impact on employees) were not surveyed. For this reason, it is not possible to determine the mechanisms that lead to the correlations identified in this study. Lastly, it is possible that some businesses closed before the survey was conducted, resulting in survivorship biases.

Independent of unique features in the case study data, the techniques presented in this paper for addressing the imbalanced data and selecting predictors from a large and collinear set of candidates are general. They should be useful for future studies of post-disaster recovery. The proposed methodology can help scholars and practitioners to gain insights that would otherwise be lost. This can help identify the characteristics of businesses that make them capable of recovering timely after a disaster.

7 Data Availability Statement

Some or all data, models, or code used during the study were provided by a third party. These include all survey data used in this paper. Direct request for these materials may be made to the provider as indicated in the Acknowledgments.

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies.

Costa, R., and Baker, J. (2020) "Multi-Split LASSO Algorithm v1.0" DOI:10.5281/zenodo.4072332 [Online, accessed on October 3, 2020]

8 Acknowledgments

The authors thank Sompo Group (especially Sompo Risk Management, Inc., Sompo Holdings, Inc. and SOMPO Digital Lab, Inc.) for providing the survey data used in this study. We also thank Chenbo Wang for helping translate and interpret the survey results.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

9 Funding

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

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

A Appendix

A.1 Standard Errors for the Regression Coefficients

Table 8 presents the standard errors for the coefficients presented in Table 7.

Table 8. Regression coefficients standard errors.

	Coefficient standard errors for P($T < t|\mathbf{X}$)					
Predictor	t=0	t=7	t=14	t=30	t=90	t=180
Intercept	0.28	0.31	0.22	0.22	–	0.14
D_1	0.32	0.35	0.26	0.41	–	–
D_2	0.27	0.28	0.20	0.19	0.14	0.12
D_3	0.27	0.29	0.19	0.54	0.14	0.11
U_1	–	–	–	–	0.18	–
U_4	–	–	–	–	0.16	–
I_1	–	–	–	–	0.18	–
I_2	–	–	–	–	0.24	–
I_3	–	–	–	0.26	0.29	0.20
I_4	–	–	0.25	0.32	0.31	0.16
S_1	–	–	–	–	0.23	–
P_4	–	–	–	–	–	0.12
P_7	–	–	0.22	0.20	–	–
P_{10}	–	–	0.22	–	–	–
F_2	–	–	–	–	–	0.11
F_3	–	–	–	–	–	0.15

A.2 Results from the SMOTE-LASSO methodology

The results in Figure 7 provide more context to the results presented in this paper. For all predictors, the figure plots on the abscissa axis the variable importance measures, VIM, and on the ordinate axis the p-values. It was assumed that predictors with p-values > 0.05 and VIM < 0.75 should be excluded from the final models. The threshold of 0.05 is commonly used for the p-values. The threshold of 0.75 for the *VIM* is arbitrary. The dashed red lines indicate these two thresholds. In Figure 7, each dot indicates one of the 30 predictors. The objective of the figure is to discuss the adequacy of the threshold for the VIM, rather than identify the VIM and p-values of each predictor. For this reason, and for clarity, the dots are not labeled. The red dots represent predictors that are selected to remain in the final models. Some dots overlap, and for this fewer dots maybe apparent than predictors in Table 7. The VIM threshold was relevant in only two cases. At t=7 days, the threshold included a predictor with VIM slightly larger than 0.75. At t=90 days, a predictor with VIM close 0.5 was excluded. In other cases, only significant changes to the VIM threshold would affect the significant predictors. Thus, the results in this paper are not strongly dependent on the VIM threshold and a sensitivity analysis is therefore judged not necessary.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

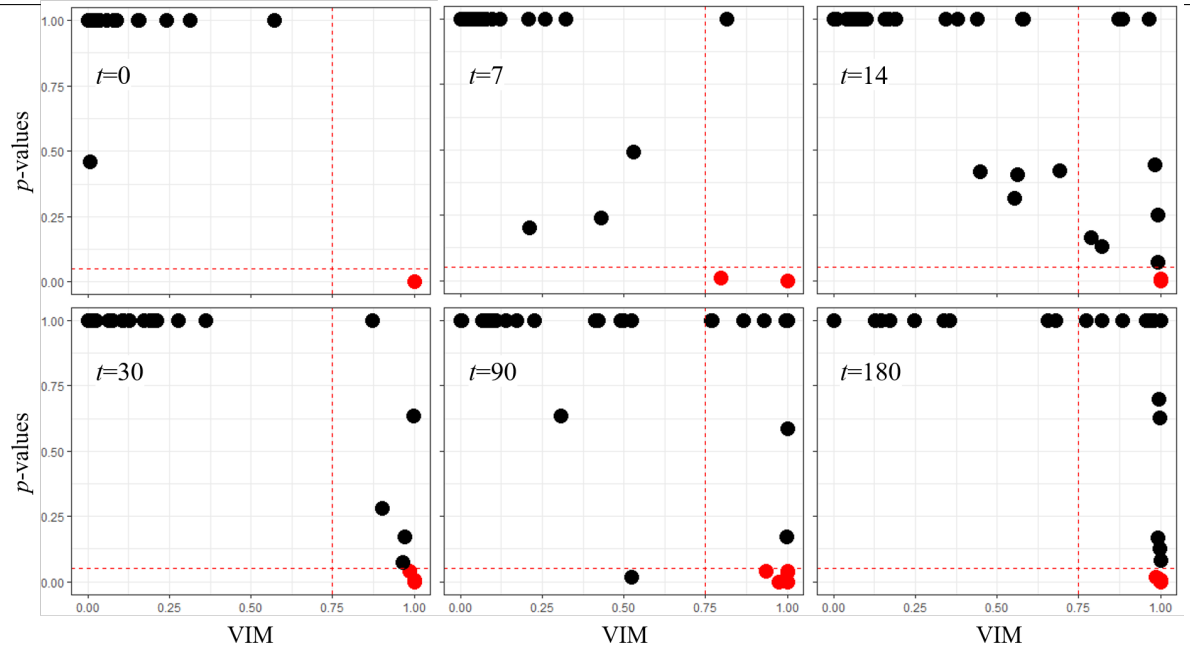

Figure 7. Impact of VIM and p-value thresholds on variable selection.

A.3 Location of businesses with seismically isolated buildings

As discussed in the Results section, the businesses with seismically isolated buildings performed worse than those without it. In Figure 8, the locations of the businesses with known locations are plotted, along with contours of peak ground acceleration, PGA. The businesses with seismically isolated buildings are highlighted in blue. A significant portion of the blue dots are within the $0.5g$-$1.0g$ zone. In Table 9 the counts of businesses within each PGA zone are shown, confirming that the percentage of seismically isolated buildings is higher within the $0.5g$ PGA zone than the percentage of non-isolated buildings. Thus, one possible explanation for the negative correlation of seismic isolation with recovery is the fact that seismic isolation is also correlated with the earthquake hazard.

Table 9. Number of businesses with seismically isolated buildings per estimated PGA zone

PGA	With seismic isolation		Without seismic isolation	
Zone	Count	%	Count	%
$0.5g$-$1g$	38	63	169	54
$0.2g$-$0.5g$	12	20	66	21
$0.1g$-$0.2g$	10	17	75	25

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

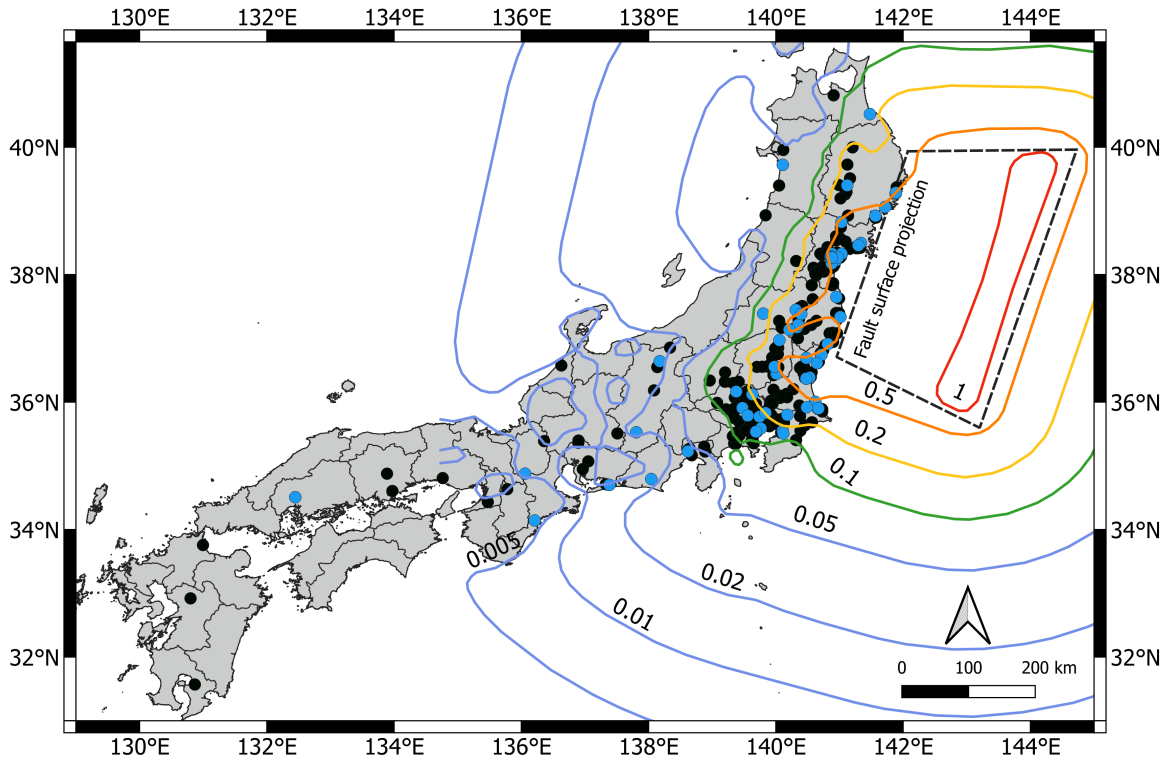

Figure 8. Location of businesses (black dots), fault surface projection (dashed lines), and contour lines of the peak ground acceleration in g (Worden *et al.*, 2020). Blue dots are businesses with seismically isolated buildings.

References

Agresti, A. (2003). *Categorical data analysis*, volume 482, John Wiley & Sons.

Alesch, D. J., J. N. Holly, E. Mittler, and R. Nagy (2001). Organizations at risk: What happens when small businesses and not-for-profits encounter natural disasters, Technical report, University of Wisconsin-Green Bay Center for Organizational Studies.

Ariga, T., Y. Kanno, and I. Takewaki (2006). Resonant behaviour of base-isolated high-rise buildings under long-period ground motions, *The Structural Design of Tall and Special Buildings* **15**(3), 325–338.

Blagus, R., and L. Lusa (2013). SMOTE for high-dimensional class-imbalanced data, *BMC Bioinformatics* **16**.

Brown, C., E. Seville, T. Hatton, J. Stevenson, N. Smith, and J. Vargo (2019). Accounting for business adaptations in economic disruption models, *Journal of Infrastructure Systems* **25**(1), 04019001.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038.
https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

Brown, C., J. Stevenson, S. Giovinazzi, E. Seville, and J. Vargo (2015). Factors influencing impacts on and recovery trends of organisations: evidence from the 2010/2011 Canterbury earthquakes, *International Journal of Disaster Risk Reduction* **14**, 56–72.

Chang, S. E. (2010). Urban disaster recovery: a measurement framework and its application to the 1995 Kobe earthquake, *Disasters* **34**(2), 303–327.

Chang, S. E., and A. Falit-Baiamonte (2002). Disaster vulnerability of businesses in the 2001 nisqually earthquake, *Global Environmental Change Part B: Environmental Hazards* **4**(2), 59–71.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer (2002). SMOTE: synthetic minority over-sampling technique, *Journal of artificial intelligence research* **16**, 321–357.

Cole, M. A., R. J. Elliott, T. Okubo, and E. Strobl (2017). Pre-disaster planning and post-disaster aid: examining the impact of the Great East Japan earthquake, *International journal of disaster risk reduction* **21**, 291–302.

Corey, C. M., and E. A. Deitch (2011). Factors affecting business recovery immediately after hurricane katrina, *Journal of Contingencies and crisis management* **19**(3), 169–181.

Dahlhamer, J. M., and M. J. D'Souza (1995). Determinants of business disaster preparedness in two US metropolitan areas, *International journal of mass emergencies and disasters Volume* .

Dahlhamer, J. M., and K. J. Tierney (1998). Rebounding from disruptive events: business recovery following the Northridge earthquake, *Sociological spectrum* **18**(2), 121–141.

Duputel, Z., L. Rivera, H. Kanamori, and G. Hayes (2012). W phase source inversion for moderate to large earthquakes (1990–2010), *Geophysical Journal International* **189**(2), 1125–1147.

Friedman, J., T. Hastie, and R. Tibshirani (2001). *The elements of statistical learning*, volume 1, Springer series in statistics New York.

Hayashi, K., K. Fujita, M. Tsuji, and I. Takewaki (2018). A simple response evaluation method for base-isolation building-connection hybrid structural system under long-period and long-duration ground motion, *Frontiers in Built Environment* **4**, 2.

Kajitani, Y., S. E. Chang, and H. Tatano (2013). Economic impacts of the 2011 Tohoku-Oki earthquake and tsunami, *Earthquake Spectra* **29**(1_suppl), 457–478.

Kay, E., C. Brown, T. Hatton, J. R. Stevenson, E. Seville, and J. Vargo (2019). Business recovery from disaster: A research update for practitioners, *The Australasian Journal of Disaster and Trauma Studies* **23**(2).

Lee, J. D., D. L. Sun, Y. Sun, J. E. Taylor, *et al.* (2016). Exact post-selection inference, with application to the LASSO, *The Annals of Statistics* **44**(3), 907–927.

Maruya, H. (2013). Proposal for improvement of business continuity management (BCM) based on lessons from the Great East Japan Earthquake, *Journal of JSCE* **1**(1), 12–21.

Costa, R., and Baker, J. W. (2021). "SMOTE-LASSO Model of business recovery over time - case study of the 2011 Tohoku earthquake." *Natural Hazards Review*, 22(4), 04021038. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000493

Matsushita, N., E. Hideshima, and H. Taniguchi (2017). The Mitigation Effect Of BCP On Financial Damage: An Empirical Study Of The Non-Manufacturing Industries In The Great East Japan Earthquake, *Journal of JSCE* **5**(1), 78–86.

Meinshausen, N., L. Meier, and P. Bühlmann (2009). P-values for high-dimensional regression, *Journal of the American Statistical Association* **104**(488), 1671–1681.

Miles, S. B., H. V. Burton, and H. Kang (2019). Community of practice for modeling disaster recovery, *Natural Hazards Review* **20**(1), 04018023.

Morrish, S. C., and R. Jones (2020). Post-disaster business recovery: An entrepreneurial marketing perspective, *Journal of Business Research* **113**, 83–92.

Nanto, D. K. (2011). *Japan's 2011 Earthquake and Tsunami: Economic Effects and Implications for the United States*, DIANE Publishing.

Nejat, A., and S. Ghosh (2016). LASSO Model of Postdisaster Housing Recovery: Case Study of Hurricane Sandy, *Natural Hazards Review* **17**(3), 1–13, ISSN 15276988, doi:10.1061/(ASCE)NH.1527-6996.0000223.

Norio, O., T. Ye, Y. Kajitani, P. Shi, and H. Tatano (2011). The 2011 eastern Japan great earthquake disaster: Overview and comments, *International Journal of Disaster Risk Science* **2**(1), 34–42.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org/.

Sampson, K., T. Hatton, and C. Brown (2018). The silent assassin: Business demand changes following disaster, *Journal of business continuity & emergency planning* **12**(1), 79–93.

Stevenson, J. R., C. Brown, E. Seville, and J. Vargo (2018). Business recovery: an assessment framework, *Disasters* **42**(3), 519–540.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society: Series B (Methodological)* **58**(1), 267–288.

Webb, G. R., K. J. Tierney, and J. M. Dahlhamer (2000). Businesses and disasters: Empirical patterns and unanswered questions, *Natural Hazards Review* **1**(2), 83–90.

Webb, G. R., K. J. Tierney, and J. M. Dahlhamer (2002). Predicting long-term business recovery from disaster: a comparison of the Loma Prieta earthquake and Hurricane Andrew, *Global Environmental Change Part B: Environmental Hazards* **4**(2), 45–58.

Worden, C. B., M. Thompson, M. Hearne, and D. J. Wald (2020). Shakemap manual online: technical manual, user's guide, and software guide, Technical report, U.S. Geological Survey, doi:https://doi.org/10.5066/F7D21VPQ, URL http://usgs.github.io/shakemap/.