SORT 41 (2) July-December 2017, 373-392 DOI: 10.2436/20.8080.02.64

Statistical and machine learning approaches for the minimization of trigger errors in parametric earthquake catastrophe bonds

Laura Calvet[1,2], Madeleine Lopeman[3], Jésica de Armas[4], Guillermo Franco[3] and Angel A. Juan[1,2]

Abstract

Catastrophe bonds are financial instruments designed to transfer risk of monetary losses arising from earthquakes, hurricanes, or floods to the capital markets. The insurance and reinsurance industry, governments, and private entities employ them frequently to obtain coverage. Parametric catastrophe bonds base their payments on physical features. For instance, given parameters such as magnitude of the earthquake and the location of its epicenter, the bond may pay a fixed amount or not pay at all. This paper reviews statistical and machine learning techniques for designing trigger mechanisms and includes a computational experiment. Several lines of future research are discussed.

MSC: 62-07.

Keywords: Catastrophe bonds, risk of natural hazards, classification techniques, earthquakes, insurance.

1. Introduction

Catastrophe (CAT) bonds are financial instruments that package catastrophe risk in a tradeable security. These tools are in effect responsible for the existence of a new market for trading risk at the frontier between finance and insurance, the so-called convergence market (Cummins and Weiss, 2009), which promises an enormous supply of capital for CAT risk transfer as long as pricing remains attractive for all parties involved. By purchasing a CAT bond, investors take the risk from a sponsor (risk ceding party) in exchange for some interest or spread. This spread constitutes the premium that compensates the risk-taking party.

[1] IN3-Department of Computer Science, Open University of Catalonia, 08018 Barcelona, Spain.

[2] Euncet Business School, 08225 Terrassa, Spain. {lcalvetl, ajuanp}@uoc.edu

[3] Guy Carpenter & Company, LLC, Dublin, Ireland, {madeleine.lopeman, guillermo.e.franco}@guycarp.com

[4] Faculty of Economics and Business Sciences, Pompeu Fabra University, 08005 Barcelona, Spain. jesica.dearmas@upf.edu

Received: December 2016
Accepted: August 2017

CAT bonds can be of different types depending on how their payment behaviour is structured. Earthquake CAT bonds in particular can base their payments on a variety of proxies (Wald and Franco, 2017). While some base payments on actual, experienced losses (indemnity), others (parametric) base them on the observable and measurable parameters that describe an event. Strategies to provide coverage for large losses ensuing after earthquakes through these parametric tools have been in use since the 1990s (Franco, 2014). Their popularity in the market is due to historically lower prices relative to traditional (re)insurance and their appeal among investment and hedge funds is due to their transparency. Lately, as traditional reinsurance pricing has decreased significantly, the price differential between traditional and alternative risk transfer (sometimes referred to as ART) is very small and is no longer the driving rationale for seeking parametric coverage. Rather, sponsors now look to parametric risk transfer for the flexibility and the ease of payment it provides.

Parametric earthquake CAT bonds employ a kind of trigger mechanism, typically a numerical check of some sort, to determine the payment that should take place when an earthquake occurs. These trigger mechanisms rely on obtainable physical characteristics of the event via respected third parties, often public agencies (Cummins, 2007; Croson and Kunreuther, 1999).

Since neither the investor nor the sponsor has the ability to manipulate this information, the risk transfer process is without moral hazard (the risk that the parties involved influence the payment outcome). Earthquakes around the world cause enormous losses, of which only about 30% have insurance coverage (Guy Carpenter, 2014). These financial impacts often disrupt individual livelihoods and national economies. Therefore, the possibility of expanding the coverage of insurance to minimize these impacts is very appealing. Making earthquake insurance more accessible, however, is difficult for traditional providers since their operations are typically resource- and time-consuming. Parametric risk transfer, in contrast, can be seamless, fast and cheap but in order to be viable, parametric solutions need to be accurate. They also need to be designed and customized without much effort so they can be easily industrialized.

Despite the popularity of parametric CAT bonds in the reinsurance market, the number of scientific works discussing these financial tools is rather limited. Thus, one of the main contributions of this paper is to help to fill this gap by proposing simple, fast, and automatic approaches able to design accurate trigger mechanisms. While a few authors have proposed ad hoc complex approaches based on the use of genetic algorithms, we explore here the use of more general tools based on statistical and machine learning methods. In particular, we review eight techniques to classify events as to whether they should trigger a payment or not, following a binary payment scheme often used in the industry. Events are classified using the fundamental parameters of focal location (i.e., longitude, latitude and depth) and the moment magnitude. Note that all the approaches need to be trained with a given dataset. These data need to constitute a large sample of events and have to include a monetary loss for each earthquake. Therefore, we turn to an earthquake CAT model to obtain a viable training dataset since historical catalogues

usually do not contain a large enough sample of this type of information. In order to test the performance of these techniques against a known benchmark, we recover the analysis presented in Franco (2010) and solve the same problem using the same dataset. We then compare across methods on such issues as accuracy, computational effort, and spatial correlation of the classifier results. According to the obtained results, the techniques employed can produce trigger mechanisms of equal or better accuracy than the technique described in the aforementioned paper. Moreover, several techniques provide huge efficiency gains in terms of decreasing classification time. Additionally, they provide scalability, being easily adapted to a larger parameter space and larger catalogues without losing much efficiency. Finally, they are relatively easy to implement using modern programming languages and software such as Python and R.

The rest of this work is structured as follows. While Section 2 contextualizes this work and describes the characteristics of the trigger mechanism, Section 3 reviews the related work. Afterwards, the techniques considered are introduced in Section 4. Section 5 explains the computational experiments. Finally, our conclusions and suggestions for future research are collected in Section 6.

2. CAT bonds and the trigger mechanism

CAT bonds have allowed insurers, reinsurers, governments, private entities and catastrophe pools to cede risks of earthquake losses to the capital markets via transparent mechanisms associated with physical event features. Since they bypass the claims adjusting process, these tools provide a very fast recovery of funds to their sponsor after an event. Within the realm of parametric earthquake CAT bonds there are also several classes of tools. Some first-generation parametric CAT bonds, or so-called "CAT-in-a-box" triggers, rely on the main physical descriptors of an earthquake event (see, for instance, Cardenas et al., 2007; Franco, 2010; Franco, 2013). Others, second-generation indexes, rely on spatially-distributed features such as ground motions recorded at sensors located throughout a region (see for instance Goda, 2013; Goda, 2014; Pucciano, Franco and Bazzurro, under review). From here, this paper focuses on first-generation triggers.

Consider a set of l earthquake events in a geographic region of interest A. An earthquake event i is characterized by a magnitude m_i, a hypocenter depth d_i, and epicenter coordinates (x_i,y_i) within A. A binary trigger will determine whether a payment should be disbursed due to event i. This response is represented by the variable B', whose values $1/0$ indicate trigger/no-trigger (payment/no-payment). Two situations may arise: (1) at least one earthquake i triggers the bond ($B'_i = 1$) during its contract life, which means that the entire bond principal has to be disbursed and, as a consequence, the buyers of the bond lose their investment (and the bond sponsors receive compensation), or (2) no earthquake triggers the bond during its life, in which case the principal is returned to the investors with interest.

Since the payment of a large sum of money is at stake, it is important that the trigger performs as desired, i.e. that the trigger responds positively to events that cause a large loss beyond a design threshold and that it does not respond to events that cause a loss below this threshold. The accuracy of the trigger determines its success in the market. Triggers that behave erratically erode the confidence of the markets in these tools and therefore jeopardize the risk transfer process. It is crucial to design triggers that behave as they should.

To describe the accuracy of the trigger, first consider a reference variable B that represents its idealized behaviour and that depends on a measure based on the losses (typically monetary). For an earthquake event i, this variable can be described as follows:

$$B_i = \begin{cases} 0 & \text{if } L_i \leq L \\ 1 & \text{otherwise} \end{cases}$$

where L_i is the actual loss caused and L is a loss threshold specified by the sponsor, usually expressed in terms of a specific return period. In this idealized scenario, events trigger this CAT bond only if the corresponding loss is above a given pre-specified threshold L.

The objective of parametric trigger mechanism design is to develop a classification mechanism that uses physical parameters of events to determine the trigger behaviour B'. Discrepancies between variables B and B' or the sum of errors $(E = \sum_{i=1}^{I} I(B_i \neq B'_i))$, represent lack of correlation between the output of the trigger and the ideal trigger. Effective parametric trigger mechanism design aims to minimize these discrepancies.

A database including a set of events, their characteristics and the variable B can be used to calculate trigger errors for this specific set of events. A measure of the loss has to be obtained or estimated to compute B. It is preferable to have a reliable historical dataset including a high number of events but in earthquake research, this is not possible due to the low frequency of earthquakes, and the great uncertainty surrounding their associated losses, and the evolution of insured portfolios over time. For this reason, the design of triggers for seismic risk relies on simulated CAT model output.

According to the description offered in this section, the development of a trigger mechanism can be labelled as a binary classification problem, allowing us to employ a wide range of techniques to address it. In the following sections, some of them are introduced and tested, and their use is illustrated.

3. Related work

The literature related to CAT bonds has increased during the last few years due to their growing popularity. Combining instruments in finance and insurance fields with engineering seismic risk assessment, Tao and Tao (2005) propose a method to set the rate

for earthquake property and personal insurances with two kinds of deductibles. Moreover, the authors present a framework to set the annual coupon rate for earthquake CAT bonds, which considers the probability of a catastrophe occurrence from seismic risk assessment, the yields of reinvestment, the principal protected ratio and the issuance fee ratio. An illustrative example focused on an urban area of China is described. Zimbidis, Frangos and Pantelous (2007) produce a model for the risk dynamics of the magnitude of the earthquakes by using advanced techniques from the extreme value theory. The model is tested on historical data of earthquakes in Greece. Moreover, the theory of incomplete markets and price CAT bonds is discussed. Tao, Tao and Li (2009) builds a pricing model, which employs the probability of an earthquake, estimated by a seismic risk assessment method. The cash flows of the insurance in complete and incomplete markets are described by Geometric Brownian Motion and Jump-Diffusion processes, respectively. Wu and Zhou (2010) reviews the state-of-the-art approaches in modelling losses for CAT bonds' modelling and pricing. They are compared by using a catalogue of earthquakes in China from 1966 to 2008. The double exponential Jump-Diffusion model fits better. Damnjanovic, Aslan, and Mander (2010) propose an integrative model linking engineering design parameters with financial indicators. The authors explain a framework based on a four-step structural loss model and a transformed survival model, which estimates excess returns. Härdle and Cabrera (2010) study the calibration of a CAT bond for Mexican earthquakes, which proves that a hybrid strategy combining traditional reinsurance and CAT bonds presents a better performance in the sense that provides coverage for a lower cost and lower exposure in comparison with a strategy without CAT bonds. Goda (2013) compares the effectiveness of two trigger mechanisms for parametric earthquake CAT bonds: scenario-based and station intensity-based approaches. The results indicate that the latter method performs at least as well as the former. Additionally, different spatial correlation models of peak ground motions are studied. Later, Goda (2014) extends the station intensity-based trigger method, which uses direct observation of ground motions at recording stations, by promoting a flexible multiple-discrete payment structure. Gunardi and Setiawan (2015) present a study case for Indonesia, in which formulas are proposed for pricing three types of CAT bonds. A generalized extreme value distribution is used to model the probability of maximum magnitude for Indonesian earthquakes. Shao, Pantelous and Papaioannou (2015) investigate the pricing process for CAT bonds considering financial and catastrophe-independent risks. An application for earthquakes is considered employing extreme value theory, and a numerical example based on California is detailed. Finally, Cummins (2007) reviews the status of the market for CAT bonds and other risk-linked securities. It discusses the complementarity between CAT bonds and the reinsurance market. In addition, the role of other modern financing mechanisms such as risk swaps, industry loss warranties, and sidecars is explained.

4. Statistical and machine learning approaches

Classification techniques (Kotsiantis, 2007) constitute a set of procedures from statistics and machine learning (more specifically, supervised learning) to determine a category or class for a given observation. Having a dataset of l observations composed of explanatory or independent variables $(X_1, X_2, \ldots, X_n)$, and a response or dependent variable Y, these techniques attempt to explain the relationships between variables and/or classify new observations based on the explanatory variables. In the problem of CAT bond trigger design, the response of the mechanism is the dependent variable Y, while the characteristics of an earthquake event (i.e., the magnitude, the hypocenter depth, and the epicenter coordinates) represent the independent variables.

Nowadays, there are plenty of classification techniques. Some of the most employed, e.g., Linear Discriminant Analysis or Logistic Regression, have been applied for more than five decades. These are mainly linear methods. Boosted by the computing advances in the 1980s and 1990s, non-linear methods such as Classification Trees, Neural Networks and Support Vector Machines emerged and started to attract attention. This section introduces some well-known and powerful techniques that we propose to automatically design a trigger. The reader interested in comprehensive and practical descriptions is referred to the books written by Hastie, Tibshirani and Friedman (2009) and Lantz (2015).

4.1. The nearest neighbours classifier

The Nearest Neighbours classifier is a simple technique that assigns a new observation to the class of the most similar observations, so-called neighbours. Therefore, it is suitable when observations of the same class tend to be homogeneous. Its main weaknesses are: not producing a model (which hinders the exploration of relationships among variables), taking a relatively high amount of time, and consuming a large amount of memory. This classifier depends on a parameter k representing the number of neighbours. The neighbours are selected according to a distance function, usually Euclidean. This parameter allows the balance between overfitting and underfitting (also known as bias-variance trade-off): a large k reduces the variance caused by noisy data or outliers but may ignore small/local patterns; conversely, a small value may introduce too much bias.

4.2. The naïve Bayes classifier

The naïve Bayes classifier is based on Bayes' theorem. "Naïve" refers to the assumption that all variables are independent and equally important. Even if this condition is not usually met in real-life applications, this classifier frequently provides competitive results. The posterior probability for a given class y is computed as:

$$P(Y = y \mid X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n) =$$

$$\frac{P(X_1 = x_1 \mid Y = y)P(X_2 = x_2 \mid Y = y) \cdots P(X_n = x_n \mid Y = y)P(Y = y)}{P(X_1 = x_1)P(X_2 = x_2) \cdots P(X_n = x_n)}$$

The classification for a given observation is obtained by comparing the probabilities of each class given the values of the explanatory variables, and selecting the class associated to the highest probability. There are many classifiers differing in the assumption made regarding the distribution of $P(X_j = x_j \mid Y = y)$. Gaussian distributions constitute a typical choice. This technique employs frequency tables and, consequently, each variable must be categorical. Numeric variables are usually discretized.

4.3. Linear and quadratic discriminant analyses

In Linear Discriminant Analysis, the distribution of the explanatory variables is separately modelled in each of the classes, and then Bayes' theorem is used to flip these around into estimates for the probability of the response variable taking a specific value given the explanatory variables. Commonly, these distributions are assumed to be Gaussian. In this case, the resulting models are similar to those provided by Logistic Regression. Linear Discriminant Analysis is more commonly employed when there are more than two classes. While this technique assumes that observations are drawn from a distribution with a common covariance matrix in each class (which leads to linear decision boundaries), Quadratic Discriminant Analysis does not make assumptions on the covariance matrices (producing quadratic decision boundaries).

4.4. Classification trees

Contrary to global models (where a predictive formula is supposed to hold in the entire data space) such as those of Logistic Regression, Classification Trees try to partition the data space into small enough parts where a simple model can be applied. The results can be represented as a tree composed of internal and terminal (or leaf) nodes, and branches. Its non-leaf part is a procedure to determine for each observation which model (i.e., terminal node) will be used to classify it. At each internal node of the tree, the value of one explanatory variable is checked and, depending on the binary answer, the procedure continues to the left or to the right sub-branch. A classification is made when a leaf is reached.

The most relevant advantage of this classifier is the easiness to understand what trees represent. They mirror human decision-making more closely than other techniques. Furthermore, trees require little data preparation, are able to handle both numerical and categorical data, and perform well (i.e., use standard computing resources in reasonable time) with large datasets.

4.5. Logistic regression

Logistic Regression techniques are designed to model the posterior probabilities of each class by means of linear functions. These probabilities, such as the one shown below, must be non-negative and sum to one.

$$P(Y = y \mid X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n}}$$

These models are usually fitted by maximum likelihood employing Newton's method. The previous expression can be rewritten in terms of log-odds as follows:

$$\log\left(\frac{P(Y = y \mid X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n)}{1 - P(Y = y \mid X_1 = x_1 \cap X_2 = x_2 \cap \cdots \cap X_n = x_n)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

This technique is especially useful when the aim is to explain (i.e., not only classify) the outcome based on the explanatory variables. Non-linear functions can be considered including interactions and transformations of the original variables.

4.6. Clusterwise logistic regression

While Regression Analysis consists of fitting functions to analyse the relationship between variables, Clustering seeks subsets of similar observations (or variables) in a dataset. Thus, the aim of Clusterwise Regression is to combine both techniques in order to discover trends within data when more than one trend is likely to exist (DeSarbo and Cron, 1988). This technique is highly flexible because different functions can be estimated. It is considered a "white-box technique" in that its mathematical systems are not complex and its results are relatively easy-to-interpret.

4.7. Neural networks

Neural Networks model the relationship between the explanatory variables and the response variable using a model inspired by how a biological brain responds to stimuli from sensory inputs. They extract linear combinations of the explanatory variables as derived variables and model the response variable as a non-linear function of these transformed variables. These models have several kinds of layers: the input layer, the output layer, and one or more hidden layers between them. Each layer contains neurons representing the variables. Increasing the number of hidden layers and/or neurons adds complexity and may improve computational capacity. With too few layers, the model may lack the flexibility to capture non-linearities in data. Neural Networks tend to have many weights, which can cause problems of overfitting. Weight decay is a method of regularization to prevent it. The "backpropagation" algorithm is a technique commonly employed for parameter estimation or training a Neural Network.

4.8. Support vector machines

A Support Vector Machine can be imagined as a surface that defines a boundary between various points of data that represent observations plotted in a multidimensional space. The goal is to create a flat boundary, called a hyperplane, which leads to fairly homogeneous partitions of data on either side. Among all potential hyperplanes, the one that creates the greatest separation between classes (a soft margin may be considered for the case on non-linearly separable data) is selected. The support vectors are the points from each class that are the closest to the hyperplane; each class must have at least one. In many real-life applications, the relationships between variables are non-linear. A key feature of this technique is its ability to efficiently map the observations into a higher dimension space by using the kernel trick. As a result, a non-linear relationship may be transformed into a linear one.

4.9. Discussion of classification techniques

Several techniques have been presented in the literature to design trigger mechanisms that determine – from an earthquake's physical characteristics – whether a principal bond should be paid (Franco, 2010, 2013). As mentioned, the aim of this work is to introduce and illustrate the application of simple, well-known, and efficient techniques that have heretofore not been explored in this context. Neural Networks and Support Vector Machines constitute two relatively modern and powerful techniques. Typically, they are able to reach high levels of accuracy by capturing nonlinear relationships between variables. However, this same characteristic makes them prone to overfitting. There are many procedures to avoid this problem such as the addition of a parameter to limit the growth of the weights or the introduction of randomness into the training data or the training algorithm. Sometimes it may be difficult to avoid overfitting and underfitting. Training Neural Networks often takes a long time, and both techniques require a non-trivial process of fine-tuning parameters. Furthermore, the resulting models are difficult if not impossible to interpret. For this reason, application of these techniques is almost always limited to classification/prediction purposes. Techniques such as Nearest Neighbours and Naïve Bayes Classifiers are easier to understand and implement and may provide relatively high accuracy. While the first is non-parametric and, consequently, flexible or unstable, the second relies on some assumptions that may be quite unrealistic in most cases.

Logistic Regression is a well-established technique, which enables the understanding of the effects of the explanatory variables on the response. Clusterwise Logistic Regression aims to incorporate the strengths of Logistic Regression while offering more flexibility, which should lead to a better understanding of the relationships among variables and higher accuracy. Classification Trees constitute an efficient technique that only uses the most important variables and results in a logic model. As other techniques studying non-linear relationships, these three techniques are particularly susceptible to overfitting or underfitting the model. Typically, small changes in training data may lead to signif-

icant modifications. In addition, Classification Trees may derive decisions that seem counterintuitive or are unexpected.

Closely related to Logistic Regression, the classic Linear/Quadratic Discriminant Analysis techniques search for the linear/quadratic combination of variables that explains the data the best. Logistic Regression is preferred if the assumption of normally-distributed explanatory variables does not hold. Otherwise, Discriminant Analysis can provide better results.

All these techniques have different features worthy of consideration when addressing a classification problem. Consequently, all are included in the following computational experiments.

5. Computational experiments

This section illustrates the application of the techniques introduced in Section 4 and compares the results with those obtained in Franco (2010). A framework for evaluation is presented such that the techniques can be compared to one another and to the reference methodology along the dimensions of accuracy, efficiency, and spatial correlation.

The dataset analysed is an earthquake catalogue representing 10,000 years of simulated seismic activity in and around Costa Rica. The catalogue contains a total of 24,957 earthquakes. These records should include the four main physical parameters enumerated before and the corresponding simulated loss. For each synthetic earthquake event in the catalogue, the model computes a ground motion footprint, which is in turn translated into estimated levels of damage to a user-defined portfolio of properties distributed in space. CAT Models have been discussed in previous studies (e.g. Grossi and Kunreuther, 2005) and we will not discuss the CAT modelling process here. The target of the classifier algorithms, in short, is to discriminate events based on their physical parameters to identify large loss-producing events. A more detailed description of the catalogue can be found in the aforementioned work. In this case study, the events in the database are assumed to be triggering events if their loss is equalled or exceeded with an annual probability of 1%.

5.1. Evaluation framework

In the case of parametric trigger design, it is difficult a priori to select the "best" classification technique for two main reasons. First, it is a multi-objective problem. Although from a statistical perspective, the sole objective may be to maximize accuracy, in real-life applications many other characteristics will likely play an important role. These may include ease of implementation, ease of explanation to non-experts, popularity, and existence of graphical representations or summaries of the outputs, among many others. The second reason is that, assuming we are only interested in the accuracy, the best technique will depend on the data at hand. Consequently, we present a general discussion of all techniques, and evaluate the trigger mechanisms they produce in three ways.

Table 1: *Structure of a confusion matrix. Note: This table summarizes the alignments and discrepancies between the behaviour of the designed trigger mechanism and the idealized trigger behaviour.*

		Predicted Class	
		$B' = 0$	$B' = 1$
Idealized Class	$B = 0$	True Positive (TP)	False Positive (FP)
	$B = 1$	False Negative (FN)	True Negative (TN)

First, the confusion matrix (Table 1) is obtained for each trigger mechanism. This table summarizes the alignments and discrepancies between the behaviour of the designed trigger mechanism and the idealized trigger behaviour (described in Section 2). In the context of parametric triggers, B' is a function representing the predicted trigger behaviour and B is a function representing the idealized trigger behaviour. In both cases, the function is equal to 1 if the bond triggers and is equal to 0 otherwise.

Next, several metrics are computed from the confusion matrix to quantify performance of each technique's trigger mechanism: error, sensitivity, and specificity. The formulas for computation thereof are shown below.

$$Error = \frac{FP + FN}{TP + FP + FN + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

Both false positive and false negative are equally penalized in this framework. In other words, we simply focus on minimizing the total number of errors. The Error metric above quantifies the rate at which the trigger mechanism misclassifies events[1]. Sensitivity characterizes how often the mechanism triggers when it should trigger, and specificity characterizes how often the mechanism does not trigger when it should not. The time required to design the trigger mechanism is also reported for each technique. The metrics described above constitute the numerical evaluation of the trigger mechanisms. Moreover, maps of the resulting trigger patterns are produced for a subset of techniques. This exercise is intended to assess whether classification techniques produce trigger mechanisms with realistic geospatial trigger patterns.

1. Note that error is equal to one minus accuracy.

5.2. Application of classification techniques

As mentioned in Section 2, the design of a parametric trigger mechanism is driven by the minimization of discrepancies between its outputs and those from a trigger with an idealized behaviour (one based directly on the losses). If the resulting trigger mechanism is expected to be useful for new or unseen observations, one should avoid employing the same observations for developing the mechanism and assessing its performance. This could lead to a problem of overfitting (i.e., obtaining complex models that capture specificities of the data but do not generalize well for other observations). An effective technique to avoid this problem is to split the dataset into three subsets: a training set used for constructing the triggers, a validation set employed to tune the parameters, and a test set required to assess their performance. We apply this approach using 50% of the observations for training, 25% for validation, and the remaining 25% for testing. z-score standardization has been applied for all techniques except Classification Trees, Logistic Regression and Clusterwise Logistic Regression. A confidence level of 95% has been considered for the statistical tests. Details of the application of each of the classification techniques are provided in the following paragraphs. The R program (R Core Team, 2012) has been used.

The nearest neighbours classifier. This technique requires a choice of the number of nearest neighbours to consider. Values ranging from 3 to 10 have been tested, and the corresponding accuracies associated to each value have been assessed using the validation set. Ultimately, 5 nearest neighbours are considered for construction of the trigger mechanism, since this provides the highest accuracy but is still small enough to reduce both the variance and the computational time required to make predictions.

Classification trees. Construction of a Classification Tree relies on the selection of the complexity parameter (a parameter that measures the tree cost-complexity). A total of 20 equidistant values from 0.01 up to 0.20 have been tested, and the corresponding accuracies associated with each value have been assessed using the validation set. The value 0.05 has been selected, since it provides the most accurate result. The representation of the tree is shown in Figure 1. Observations which satisfy the condition shown for each internal node terminate to the left; otherwise, they proceed to the right. The percentage shown at the bottom of each node indicates the proportion of observations that reach that node. The value above that percentage refers to the binary classification. For instance, the first condition is '$m < 6.15$', and it is evaluated for all observations (i.e., 100%). Approximately 70% of the earthquakes satisfy this condition and their prediction (i.e., B') is set to 0. The remaining earthquakes are further divided according to the condition '$y < 9.785$'. The same steps are iteratively repeated until a prediction is set for all earthquakes. Thus, eight conditions are considered and only 0.22% of the earthquakes are assigned a value of 1.

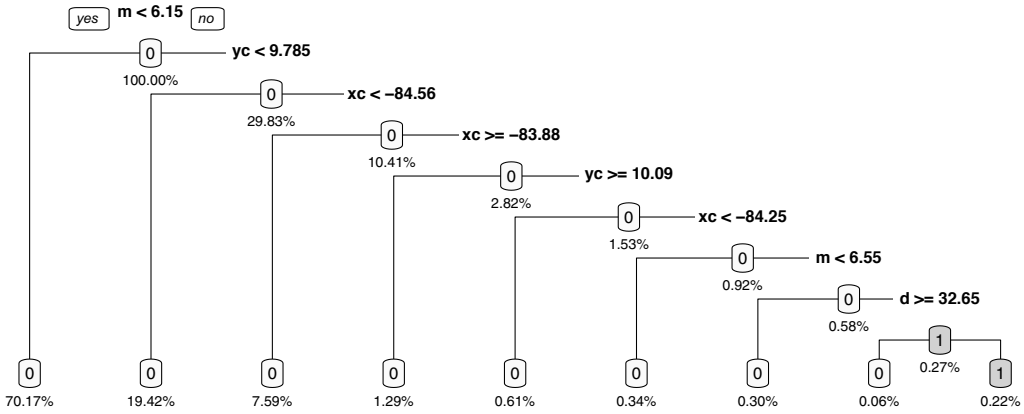

Figure 1: *Classification Tree. Note: Representation of the Tree. Observations which satisfy the condition shown for each internal node terminate to the left; otherwise, they proceed to the right. The percentage shown at the bottom of each node indicates the proportion of observations that reach that node. The value above that percentage refers to the binary classification.*

Neural Networks. Even if complex and powerful Neural Networks exist, we focus on a topology characterized by only one hidden layer. Despite its minimalism, this approach is commonly used, tends to provide good results and is conceptually simple. The number of units in the hidden layer (26) has been tuned by testing the set of values ranging from 10 to 40.

Table 2: *Kernels considered for Support Vector Machines. Note: This table presents some of the most popular kernels for Support Vector Machines in the literature, which are considered in the computational experiments.*

Linear	$k(a,b) = a^T b$
Polynomial [2]	$k(a,b) = (\alpha a^T b + c)^d$
Radial Basis [3]	$k(a,b) = exp(-\gamma \lvert a-b \rvert^2)$
Sigmoid [4]	$k(a,b) = tanh(\sigma a^T b + e)$

Support Vector Machines. In order to efficiently employ this technique, it is required to select a kernel and tune the corresponding parameters. The most popular kernels have been considered and are shown in Table 2. There is also a parameter cost related to the cost of a misclassification for which the following values have been considered: 0.01, 0.1, 1, 5, and 10. Using the validation set, each combination of cost and kernel (including

2. Values tested for α, c, and d, respectively: $\{0.1, 0.2, 0.3, 0.4\}$, $\{0, 0.2, 0.4, 0.6\}$, and $\{2, 3, 4, 5\}$.
3. Values tested for γ: $\{0.1, 0.2, 0.3, 0.4\}$.
4. Values tested for σ and e, respectively: $\{0.1, 0.2, 0.3, 0.4\}$ and $\{0, 0.2, 0.4, 0.6\}$.

the type and the corresponding set of parameters) has been tested by computing the associated accuracy level. According to the results, the best option is a polynomial kernel with the following parameters: $cost = 10$, $\alpha = 0.4$, $c = 0.4$, $d = 4$.

5.3. External validation

In order to validate the application of these techniques to the development of parametric triggers for earthquake catastrophe bonds, we compare our results with those provided by the methodology in Franco (2010). In this paper, the author proposes the construction of binary "cat-in-a-box" trigger mechanisms, where the geographical space is discretized in square boxes or sub-regions of the same size. Each sub-region belongs to a specific zone denoted as k. This approach relies on the concept of optimization and its aim is to determine the parameters of a trigger mechanism for each zone as well as the zone assignment of each sub-region such that the total trigger error is minimized. Concretely, the trigger mechanism for zone k has the following structure:

$$\forall (x_i, y_i) \in A_k, \qquad B_i' = \begin{cases} 0 & \text{if } m_i \leq M_k \quad \text{or } d_i \geq D_k \\ 1 & \text{if } m_i \geq M_k \quad \text{or } d_i \leq D_k \end{cases}$$

where M_k and D_k represent the parametric triggers for the zone, namely the magnitude and depth thresholds, respectively. All sub-regions belonging to zone k have the same trigger structure. An Evolutionary Algorithm (EA) is implemented to address this optimization problem and is executed for different combinations of geographic resolution and number of zones. Although the paper does not report computational times, these methods may consume several hours to perform the parameter optimization.

5.4. Performance

The performance of the trigger mechanisms designed using all nine statistical and machine learning techniques and using the EA employed in Franco (2010) is reported and discussed here. Performance measures are shown in Table 3. Total time takes into account the time to construct the trigger, fine-tune its parameters and test its performance.

A suitable trigger mechanism design should exhibit low error and high specificity and sensitivity and should require minimal computational effort. It can be concluded from the table that the non-linear and non-parametric techniques obtain the best performances of the statistical techniques in terms of accuracy, sensitivity and specificity. In particular, Nearest Neighbours classifier, Classification Trees, Neural Networks and Support Vector Machines are all consistently superior across the three metrics. The results reveal a high variability with respect to computational time, ranging from a few seconds to several minutes. There tends to be trade-off between accuracy and time-required, particularly in the cases of Neural Networks and Support Vector Machines, both of which require significantly more time than the other techniques.

Table 3: *Parametric trigger mechanism performance for ten design techniques. Note: This table shows the performance of the trigger mechanisms designed using all nine statistical and machine learning techniques and using the EA employed in Franco (2010). It considers the error, the sensitivity, the specificity, and the computational time.*

Technique	Error	Sensitivity	Specificity	Time (sec.)
Nearest Neighbours classifier	0.18%	99.84%	**94.44%**	7.22
Naïve Bayes classifier	0.77%	99.58%	4.35%	1.62
Linear Discriminant Analysis	0.64%	99.57%	0.00%	0.28
Quadratic Discriminant Analysis	0.42%	99.63%	57.14%	**0.12**
Classification Trees	0.24%	99.79%	87.50%	2.62
Logistic Regression	0.45%	99.58%	33.33%	0.87
Cluster-wise Logistic Regression	0.43%	99.57%	undefined	5.7
Neural Networks	**0.14%**	**99.94%**	82.14%	190.86
Support Vector Machines	0.27%	99.78%	81.25%	161.25
Evolutionary Algortihm (Franco, 2010)	0.34%	99.86%	55.56%	hours

Several techniques exhibit superior performances to the EA in terms of accuracy, sensitivity and specificity. While EA produces relatively low error rates, the time required is significantly longer than all of the statistical and machine learning techniques.

The triggering events in the idealized trigger mechanism (those for which $B = 1$) comprise less than 0.5% of the total test catalogue, while the other 99.5% of catalogue events do not trigger the idealized bond. Hence, a supposed "null" trigger mechanism in which no events ever trigger the bond would exhibit 99.5% accuracy (0.5% error), 100% sensitivity and 100% specificity. The burden in this case is therefore on any designed trigger mechanisms to outperform this null trigger mechanism benchmark. Eight out of the ten techniques produce trigger mechanisms superior to the null trigger mechanism in terms of accuracy, while the Naïve Bayes Classifier and Linear Discriminant Analysis perform worse by a small margin.

That so few events trigger the bond in the idealized scenario suggests that a larger catalogue might produce more informative and nuanced results using the statistical and machine learning techniques for parametric trigger mechanism design. With a larger catalogue to "learn" from, the techniques would have more triggering events from which to decipher patterns and connections. Reduction of the loss threshold used to construct the idealized trigger scenario would also generate more triggering events from which the statistical techniques could "learn", but since CAT bonds are typically constructed for relatively high return period losses (greater than 100 years), these solutions would not be relevant from a practical standpoint.

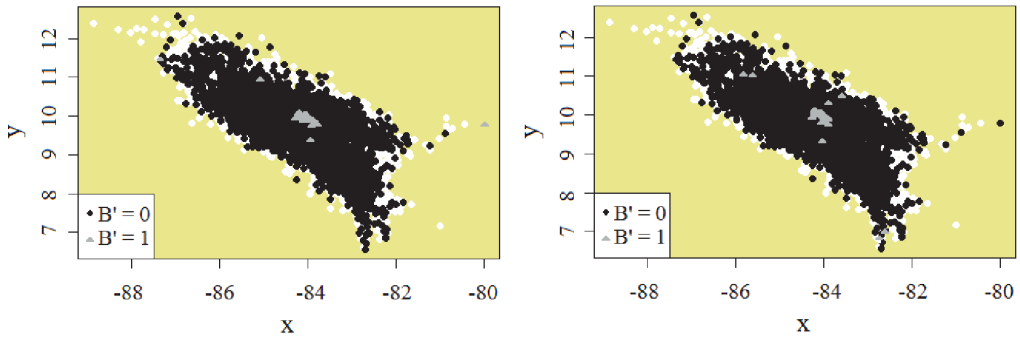

Figure 2: *Map of predictions obtained with Neural Networks (left) and the EA described in Franco (2010). Note: Evaluation of two techniques that produce trigger mechanisms: Neural Networks and the EA from Franco (2010). The first layer is composed of white points, where each point represents an earthquake of the catalogue. The second layer includes the black points, which identify those earthquakes belonging to the test set with* $B' = 0$. *Similarly, the third layer covers the gray triangles, i.e., earthquakes from the test set with* $B' = 1$, *respectively.*

While accuracy is certainly an indispensable feature of any suitable technique for design of parametric trigger mechanisms, a technique should also produce trigger behaviour that is meaningful from a physical perspective. Namely, a suitable technique for parametric trigger mechanism design should produce trigger behaviour that reflects the seismic hazard and/or development patterns in the region of study. For this reason, the physical performance of the techniques trigger mechanisms was evaluated representing earthquakes falling into the test set in maps. Figure 2 shows the evaluation of two techniques that produce trigger mechanisms that are suitable from a numerical perspective: Neural Networks and the EA from Franco (2010). The first layer is composed of white points, where each point represents an earthquake of the catalogue. The second layer includes the black points, which identify those earthquakes belonging to the test set with $B' = 0$. Similarly, the third layer covers the gray triangles, i.e., earthquakes from the test set with $B' = 1$. Note that the plot on the left (Neural Networks) gathers all gray triangles in the centre, while the plot on the right (EA) shows more dispersion.

6. Conclusions and future research

Natural catastrophes continue to cause enormous losses that remain largely uninsured, leaving populations vulnerable to severe financial impacts. The insurance and reinsurance industry, governments and catastrophe pools have started to employ financial instruments such as parametric CAT bonds to cede these catastrophic risks to the capital markets. Were these tools extended for more widespread usage at the retail level, we could progressively and massively reduce the “insurance gap” for earthquake risks. However, this requires the construction of accurate and unbiased parametric triggers

with extreme efficiency and automation, something that is not available in the industry today.

To address this problem, we have explored solving the trigger design challenge as a classification problem, employing well-known and powerful techniques from statistics and machine learning. From a numerical perspective, it has been shown that these techniques can produce trigger mechanisms of equal or better accuracy than previously published techniques (Franco, 2010). Furthermore, several statistical and machine learning methods provide huge efficiency gains in terms of decreasing classification time. Additionally, they provide scalability, being easily adapted to a larger parameter space and larger catalogues without losing much efficiency, and ease of implementation since there is a wide range of programs and programming languages that enable free and simple implementation of these statistical and machine learning techniques such as R (R Core Team, 2012), Octave (Eaton et al., 2014) and Scilab (Scilab Enterprises, 2012). Application of these statistical and machine learning techniques to the problem of parametric trigger design is not without complication, however, because while these methods provide accuracy and efficiency improvements, some of the examples shown in this paper produce trigger mechanisms with relatively low specificity values.

Several lines of future research emerge from the introduction of classification techniques to the development of trigger mechanisms for earthquake CAT bonds. First, it is apparent from the experiments in this paper that more meaningful insights as to the applicability of classification techniques to the development of trigger mechanisms could be gleaned from the use of a larger earthquake catalogue. It would also be worthwhile to examine the behaviour of the trigger mechanisms at multiple return periods, particularly lower ones. There is a natural imbalance in the data at high return periods since very few events trigger the bond. Consequently, there are two groups of events subjected to classification (depending on whether they should trigger a given CAT bond), but they greatly differ in size. Techniques may present low accuracy with respect to the minority (triggering) group and still have a good global accuracy. Analysis of the same simulated earthquake catalogue at lower return periods would reduce this classification group imbalance but would not produce a usable trigger mechanism, since CAT bonds are typically constructed to protect against high return period losses. Therefore, such an experiment could provide valuable insights into the different classification techniques but would not produce directly usable trigger mechanisms. A popular numerical alternative to this complication is to oversample events in the minority group, which would constitute an artificial expansion of the original earthquake catalogue.

Introduction of such a large number of alternative techniques for parametric trigger mechanism design motivates the development of a selection framework. From the standpoint of practical implementation, it would be interesting to identify the most desirable characteristics for a trigger mechanism and order them. For instance, if accuracy is supreme, one should explore the use of more modern and complex techniques such as Random Forests and Multi-Layer Neural Networks (provided a larger catalogue was

available). In contrast, if the interpretability plays the largest role, it would make sense to employ more classical techniques and study graphical tools.

The technological developments characterizing the era of Big Data and the Internet of Things have potentially fascinating implications in this field. These avenues open the possibility of designing triggers not only based on few physical characteristics of an earthquake but on much more information obtained through broad networks of sensors. Metaheuristics, simheuristics (i.e., algorithms combining metaheuristics and simulation techniques) and other classical instruments may be used to perform a feature selection or extraction. Finally, the capacity of simulators to create larger catalogues is ever-increasing, constantly being able to generate more and more data, more and more reliably. In this scenario, non-linear approaches such as Deep Learning would be worth exploration.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness, FEDER (TRA2013-48180-C3-P, TRA2015-71883-REDT) and the Erasmus+ program (2016-1-ES01-KA108-023465). The authors are also thankful to AIR Worldwide for permitting the usage of the AIR Earthquake Model for Central America version 1.1 for the reproduction of the example contained in Franco (2010).

References

Cardenas, V., Hochrainer, S., Mechler, R., Pflug, G. and Linnerooth-Bayer, J. (2007). Sovereign financial disaster risk management: the case of Mexico. *Environmental Hazards*, 7 , 40–53.

Croson, D.C. and Kunreuther, H. (1999). Customizing reinsurance and cat bonds for natural hazard risks. In Conference on Global Change and Catastrophic Risk Management.

Cummins, J.D. (2007). CAT bonds and other risk-linked securities: state of the market and recent developments. *Social Science Research Network*. http://ssrn.com/abtract=105740.

Cummins, J.D. and Weiss, M.A. (2009). Convergence of insurance and financial markets: hybrid and securitized risk-transfer solutions. *Journal of Risk and Insurance*, 76, 493–545.

Damnjanovic, I., Aslan, Z. and Mander, J. (2010). Market-implied spread for earthquake CAT bonds: financial implications of engineering decisions. *Risk Analysis*, 30, 1753–1770.

DeSarbo, W.S. and Cron, W.L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of classification*, 5, 249–282.

Eaton, J.W., Bateman, D., Hauberg, S. and Wehbring, R. (2014). GNU Octave version 3.8.1 manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform. ISBN 1441413006. http://www.gnu.org/software/octave/doc/interpreter.

Franco, G. (2010). Minimization of trigger error in cat-in-a-box parametric earthquake catastrophe bonds with an application to Costa Rica. *Earthquake Spectra*, 26, 983–998.

Franco, G. (2013). Construction of customized payment tables for cat-in-abox earthquake triggers as a basis risk reduction device. In *Proceedings of the International Conference on Structural Safety and Reliability (ICOSSAR)* (pp. 5455–5462). New York, NY.

Franco, G. (2014). Earthquake mitigation strategies through insurance. Chapter *Encyclopedia of Earthquake Engineering*. (pp. 1–18). Springer Berlin Heidelberg.

Goda, K. (2013). Basis risk for earthquake catastrophe bond trigger using scenario-based versus station intensity-based approaches: a case study for southwestern British Columbia. *Earthquake Spectra*, 29, 757–775.

Goda, K. (2014). Seismic risk management of insurance portfolio using catastrophe bonds. *Computer-Aided Civil and Infrastructure Engineering*, 30, 570–582.

Grossi, P. and Kunreuther, H. (2005). *Catastrophe Modeling: A New Approach to Managing Risk*. Springer.

Gunardi and Setiawan, E.P. (2015). Valuation of Indonesian catastrophic earthquake bonds with generalized extreme value (GEV) distribution and Cox-Ingersoll-Ross (CIR) interest rate model. In *AIP Conference Proceedings* (Vol. 1692, pp. 020024).

Guy Carpenter (2014). Chart: gap between economic and insured losses. http://www.gccapitalideas.com/2014/01/20/chart-gap-between-economic-and-insured-losses/.

Härdle, W.K. and Cabrera, B.L. (2010). Calibrating CAT bonds for Mexican earthquakes. *Journal of Risk and Insurance*, 77, 625–650.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning*. (2nd ed.). Springer.

Kotsiantis, S. (2007). Supervised machine learning: a review of classification techniques. *Informatica*, 31, 249–268.

Lantz, B. (2015). *Machine Learning with R*. Packt Publishing Ltd.

Pucciano, S., Franco, G. and Bazzurro, P. (under review). Loss predictive power for strong motion networks for usage in parametric risk transfer: Istanbul as a case study. *Earthquake Spectra*.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. ISBN 3- 900051-07-0. http://www.R-project.org/.

Scilab Enterprises (2012). *Scilab: Free and Open Source Software for Numerical Computation*. Scilab Enterprises Orsay, France.

Shao, J., Pantelous, A. and Papaioannou, A.D. (2015). Catastrophe risk bonds with applications to earthquakes. *European Actuarial Journal*, 5, 113–138.

Tao, X.X. and Tao, Z.R. (2005). Financial instruments integrated with engineering risk assessment for earthquake disaster reduction. *WIT Transactions on The Built Environment*, 82.

Tao, Z., Tao, X. and Li, P. (2009). Pricing Model for Earthquake CAT Bonds. In *International Conference on Business Intelligence and Financial Engineering* (pp. 740–744).

Wald, D. and Franco, G. (2017). Financial decision-making based on near-realtime earthquake information. In *Proceedings of the 16th World Conference on Earthquake Engineering*. Santiago de Chile. Paper: 3625.

Wu, D. and Zhou, Y. (2010). Catastrophe bond and risk modeling: a review and calibration using Chinese earthquake loss data. *Human and Ecological Risk Assessment*, 16, 510–523.

Zimbidis, A.A., Frangos, N.E. and Pantelous, A.A. (2007). Modeling earthquake risk via extreme value theory and pricing the respective catastrophe bonds. *ASTIN bulletin*, 37, 163–183.