Cyber risk: An analysis of self-protection and the prediction of claims

Alana K. Azevedo[1], Agnieszka I. Bergel[2], Alfredo D. Egídio dos Reis[2],
[1]*ISEG-School of Economics and Management, Universidade de Lisboa; REM - Research in Economics and Mathematics, CEMAPRE and Universidade Federal do Ceará*
alanakna@phd.iseg.ulisboa.pt
[2]*ISEG-School of Economics and Management, Universidade de Lisboa; REM - Research in Economics and Mathematics, CEMAPRE*
alfredo@iseg.ulisboa.pt, agnieszka@iseg.ulisboa.pt

Abstract

For a set of Brazilian companies, we study the occurrence of cyber risk claims by analyzing the impact of self protection and the prediction of their occurrence. We bring a new perspective to the study of cyber risk analyzing the probabilities of acquiring protection against this type of risk by using propensity scores. We consider the problem of whether acquiring cyber protection improves network security using a matching method that allows a fair comparison among companies with similar characteristics. Our analysis, assisted with Brazilian data, shows that despite informal arguments that favor self-protection against cyber risks as a tool to improve network security, we observed that in the presence of self-protection against cyber risks, the incidence of claims is higher than if there were no protection. Regarding the prediction of the occurrence of a claim, a system considering a feedforward multilayer perceptron neural network was created, and its performance was measured. Our results show that, when applied to the relevant information of the companies under study, it presents a very good performance, reaching an efficiency in general classification above 85%. The fact is that the use of neural networks can be quite opportune to help in solving the problem presented.

Keywords: Cyber risk; cybersecurity; propensity score; neural network; multilayer perceptron.

1 Introduction

It is clear that the evolution of business relationships through digital platforms raises concerns about cyber security. The damage caused by security breach generates increasingly higher economic losses, but the search for protection and identification of the risks to which companies are subject is still moderate. Several ways to mitigate this risk were already widespread in this virtual environment, such as antivirus, firewall, data protection, authentication technology, secure communication, access restriction. It cannot be forgotten that the

losses are not restricted to the economic sense. A cyber attack can compromise a company both operationally and generate third-party issues.

Because it is a dynamic risk, with incidents that normally do not have a single pattern and that evolves quickly, assessing and estimating the cost and likelihood of a cyber attack is challenging for both companies and insurers. It is of fundamental importance that companies do not underestimate but manage the cyber risk to which they are exposed, even if they do not fully recognize its nature.

Many countries are already concerned with defining policies that involve cybersecurity standards, including Brazil, a country with substantial geographic and population dimension, consequently, exposed to great vulnerability. With this scenario, the Brazilian government has been looking for alternatives to value data privacy and security. The first step was the law that represents a civil landmark of the internet, approved in 2014, followed by the approval of the General Data Protection Law (LGPD) in 2018, which began taking effect in 2020.

Despite the existence of several forms of security measures against cyber attacks, the risk of both financial and operational losses is still considerable. According to Muggah and Thompson (2017), in 2014 Brazil was ranked as number one in the world for banking malware attacks, with nearly 300,000 compromised users. Furthermore, according to Diniz *et al.* (2014), at least 75 per cent of Brazilian Internet users claim to have been victims of some form of cyber crime. Institute (2019) allocates Brazil in the first position of the ranking of probability of data leakage. This probability of data leakage is 43%.

This paper proposes an analysis focusing on the occurence of claims, in the light of individual characteristics of Brazilian companies in relation to the use of technologies and cyber risk management. The main objective is to get information if self protection helps preventing cyber attacks by comparing the probability of occurrence of claims in companies using a propensity score matching method. We introduce a neural network model framework to study the differences, by using the company characteristics, in order to predict the occurrence of claims and to generate information about companies that may have high cyber risks attacks.

The manuscript is organized as follows. In next section we survey related work. Section 3 is devoted to the presentation of model frameworks, including definitions, assumptions, scenarios and procedures. It also includes information about the database. Section 4 and Section 5 present result discussion. Last section finishes our study by some concluding remarks.

2 Background and motivation

The key to an efficient mitigation of cyber risks is the assessment of all basic knowledge about the Information Technology (IT), the structuring of digital assets, the interconnected environment and the level of existing outsourcing.

Insurance companies try to lower the problem by considering that only people subject to large risks will buy their products minimizing the problem of adverse selection (refers to a situation where insureds have information that insurers do not have), so they will measure the risk and adjust the price they want charge for that risk. Pricing cyber insurance contracts has become particularly challenging due to the damage it could cause to businesses. Marotta *et al.* (2017) summarised the basic knowledge about cyber insurance available so far from both market and scientific perspectives, explaining basic terms and formulation of the area.

These authors discussed the issues which make this type of insurance unique and show how different technologies are affected by these issues.

Ogut *et al.* (2005) analyzed firm's IT security risk management strategies when their risks are interdependent. Only the interdependency that arises due to of interconnections of computers in different firms was modeled.

In case of cyber risk insurance Gordon *et al.* (2003) described a generic framework for using a cyber risk insurance for helping to manage information security risk. Mukhopadhyay *et al.* (2005), proposed the use of insurance as a supplementary tool to reduce the financial losses suffered due to *e*-risk using utility theory. The authors defined, classified and presented *e*-risk products available globally. In the same year, Böhme (2005) referred to an indemnity insurance model to evaluate the conditions under which coverage for cyber risks can be granted despite monocultures in installed platforms. A cost advantage for users of less widespread platforms could foster a more balanced market structure.

Böhme and Kataria (2006) attempted to separately identify the internal (within one firm) and global (across multiple firms) correlation of cyber risks and to estimate their combine effect on the presence of cyber-insurance market. Bolot and Lelarge (2008) analyzed the impact of interdependent risks on the security investiments of the users, by using simple models based on the classical expected utility model that combine ideas from risk theory and network modelling. The key result was that using insurance would increase the security in the internet.

Schwartz and Sastry (2014) derived the expression of breach probability from standard assumptions. They provided a solution for user optimal security in environments with and without cyber insurance. The analysis confirms a discrepancy in informal arguments that favor cyber-insurance as a tool to improve network security, rather than merely manage risks. Fahrenwaldt *et al.* (2018) introduced and analyzed a new polynomial approximation of claims together with a mean-field approach that allows to compute aggregate expected losses and prices of cyber insurance.

Betterley (2018) presented an annual review and evaluation of insurance products designed to protect against the unique risk of data security for organizations. Xu and Hua (2019) develop a framework for modelling and pricing cybersecurity risk based on both Markov and non-Markov models. The authors proposed a simulation approach to compute the premium for cybersecurity risk and studied the effects of difference infection distributions and the dependence among infection processes on the losses. Macedo *et al.* (2019) proposed an innovative methodology for the design of insurance polices based on deterministic and stochastic service life prediction models. Single-parameter and multiparameter models were applied in the calculation of insurance premia.

One important approach used in some cases of cyber risk analysis is the copula model. Copulas are functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. It has great value for modeling dependent risks. In the case of insurance, one could model the non-linear dependencies in the pricing variables and use simulation to determine premiums.

Some authors have worked with copula models in the study of the cyber risk insurance. Mukhopadhyay *et al.* (2006) developed a framework, based on copula aided Bayesian Belief Network model, a graphical relationship between causal variables, to quantify the risk associated with online business transactions, arising out of a security breach, and thereby help designing *e*-insurance products. Herath and Herath (2011) developed a cyber-insurance

model using the emerging copula methodology. The authors estimated the premiums for the first part losses due to virus intrusions using three types of insurance policy models.

In the field of networks, Barracchini *et al.* (2014) suggested a possible alternative, original and supplementary solution to the issue of network security. The paper formulated the actuarial premises, following an actuarial multistate approach, for coverage in case of cyber damage. Ionită and Patriciu (2014) set a feedforward backward-propagating neural network in order to correlate threat data from agents installed on remote protected hosts. For the authors, the neural network assesses the risk of a cyber attack taking place and bringing the defense systems to an alarmed state in a timely manner.

Subroto and Apriyana (2019) presented an algorithmic model that uses social media big data analytics and statistical machine learning to predict cyber risks. Authors used Rweka package to carry out machine learning (ML) experimentation and artificial neural network (ANN) to build a confusion matrix to show how it is possible understand and predict vulnerabilities to threats.

Concerning propensity score, Rosenbaum and Rubin (1983) presented the central role of the propensity score in observational studies for causal effects. Shipman *et al.* (2017) discussed the usefulness and limitations of propensity score matching relative to more traditional multiple regression (MR) analysis.

Dehejia and Wahba (2002) discussed the use of propensity score-matching methods, pairing the experimental treated units with nonexperimental comparison units, and compared the estimates of the treatment effect obtained using the methods to the benchmark results from the experiment. Remarkably, we could not find previous publications addressing the use of propensity score for cyber risk studies.

Our objective is to shed new light on cyber risk, by measuring the difference on the number of claims of similar companies with and without self-protection against this type of risk. To do this, we undertake a sample of companies and apply a propensity score-matching method which involves pairing treatment and comparison units that are similar in terms of their observable characteristics, see Dehejia and Wahba (2002). Using these same characteristics, we develop a neural network system to predict the occurrence of claims and offer an innovative way as a support tool for better identification of the specific condition of the company about cyber risk.

The particular strengths and differential of the current study, compared to previous cyber risk studies, are: (i) It is based on a real data set of companies from Brazil with information of IT; (ii) Risk factors that are rarely studied, such as if the company has specialized IT staff, are analyzed here, and (iii) The use of propensity scores and neural network provide a meaningful comparison of companies' cyber risks. In addiction, the first can be understandable and persuasive to any audience and the last has the power of a universal approximator.

3 Methods and materials

In this section we present the models and how they were developed as well as the database considered.

3.1 The data

The Brazilian Institute of Geography and Statistics (IBGE) is the main provider of data and information in the country, which meets the needs of the most diverse segments of civil society, as well as federal, state and municipal government agencies.

In the year 2010, the research on the use of Information and Communication Technologies (ICT) investigated aspects of the use of these technologies by the Brazilian business segment. Among its themes, the research brought information about the use of computers and the internet in the activities of these organizations and the reasons given to explain their non-use. Information on adopted ICT security policies and the skills of the personnel employed in relation to these technologies were also presented.

The data used from the research was formatted in two stages. All information were anonymized and de-identified by IBGE prior to analysis. The first stage involved selecting only the information of companies that used computers and the internet. A total of 16,725 companies were considered. Once the selected companies had been established, the second stage involved the design of the sample so that we could develop the proposed study.

The variables, whose names are in parentheses, chosen from the research about each company were if own IT department (*depart*), if provided IT qualification (*quali*), if had IT security policy (*security*), if had wired local network (*wired*), if had wireless local network (*wireless*), if had intranet (*intra*), if had extranet (*extra*), if used cloud computing (*cloud*), if used out-of-the-box software ($readysoft$), if used free software ($freesoft$), if used software developed by another company ($otherssoft$), if own homepage (*homepage*), if used fixed broadband internet connection ($fixed$), if used mobile internet broadband connection (*mobile*), if made purchases of goods or services (*purchase*), if interacted with government agencies (*gov*), if used security measures, if had IT security related incidents.

The categorization of these variables was as follows. In the case of the "if the company used security measures", we categorized into companies without protection (≤ 2 measures) and companies with protection (> 2 measures). The variable "if the company had IT security incidents" was defined to be 0, 1, 2, 3, 4 or 5, since there were five types of possible incidents in the questionnaire and the company could have suffered from none to all. All other variables were classified as dummy variables.

There were only four variables in the questionnaire that did not present statistical significance by the logit model developed in this study. They were: if the company had an IT specialist; if the company used its own software; if the company used narrow band, and; if the company made sales through the internet.

Table 3.1 provides the characteristics of the sample we use. The table highlights the role of the use of internet.

Table 3.1: Sample Means of Covariates

Research on the Use of Information and Communication Technologies in Companies	
Variable	Sample
Proportion that had IT department	0.583
Proportion that provided IT qualification	0.445
Proportion that had IT security policy	0.364
Proportion that had wired local network	0.835
Proportion that had wireless local network	0.553
Proportion that had intranet	0.339
Proportion that had extranet	0.271
Proportion that used cloud computing	0.209
Proportion that used out-of-the-box software	0.954
Proportion that used free software	0.585
Proportion that used software developed by another company	0.753
Proportion that had homepage	0.661
Proportion that used fixed broadband internet connection	0.961
Proportion that used mobile broadband internet connection	0.418
Proportion that made purchases of goods or services	0.615
Proportion that interacted with government agencies	0.787

3.2 Propensity scores and the analysis of self-protection

The propensity score is a balancing score, a function of observed covariates x such that the conditional distribution of x given $b(x)$ is the same for treated ($z = 1$) and control ($z = 0$) units. Because units exposed to a certain treatment typically differ systematically from control units, balancing scores make direct comparisons between the two groups much more meaningly in nonrandomized experiments.

Rosenbaum and Rubin (1983) defines propensity score as the conditional probability of assignment to a particular treatment given a vector of observed covariates. Let x denote the vector of covariates for a particular company, and let the binary variable z indicate whether the company was exposed ($z = 1$) or unexposed ($z = 0$). The propensity score, $e(x)$, is the conditional probability of exposure given the covariates; that is,

$$e(x) = pr(z = 1|x), \tag{3.1}$$

presuming that

$$pr(z_1, ..., z_n|x_1, ..., x_n) = \prod_{i=0}^{n} e(x_i)^{z_i}(1 - e(x_i))^{1-z_i}. \tag{3.2}$$

a strict independence assumption is considered to simplify notation and discussion, even if it is not essential.

To begin the analysis, the first step is to calculate propensity scores, which are in this case individual probabilities of acquiring protection against cyber risk. These probabilities are obtained by estimating a Logit model given by:

$$e(x) = pr(z = 1|x) = \frac{e^{x'\beta}}{1 + e^{x'\beta}}. \tag{3.3}$$

where $z = 1$ if the company has protection against cyber risk and $z = 0$ otherwise, where x is a vector of covariates that affect the occurrence of claims. The vector x' contains the explanatory variables and β is the regression coefficients vector. Logistic distribution tends to give higher probabilities for $z = 0$ when $x'\beta$ is extremely small (and lower probabilities for $z = 0$ when $x'\beta$ is very large) relative to the normal distribution used in the Probit model. These two distributions tend to provide similar probabilities for intermediate values of $x'\beta$.

A major advantage of the propensity score method is that it attempts to minimize the information contained in the variables x, which will affect the decision of purchase protection. This is done by estimating, linked to these variables, the likelihood of the company acquire protection against cyber risk. Thus, the variables x will not be used directly, but the probabilities of participation derived from them.

By estimating propensity scores it is possible to calculate matching estimator that will be done considering matching from intervals. Matching based on propensity score is a correction strategy that adjusts the estimates generated for the selected trends. The method employs a predicted probability of a group member based on observed predictions, usually obtained by logistic regression to create a counterfactual group. Propensity score can be used for matching or as covariates. A guidance for the implementation of propensity score matching can be seen in the work of Caliendo and Kopeinig (2008).

In order to have a good matching based on the propensity score, large samples are needed, treatment groups and comparison with a substantial dimension must be identified, if possible, focus on variables that are precisely measured and stable and use a composite variable that minimize differences in groups between multiple points.

More generally, the estimate in this counterfactual analysis will be represented by the difference between the following portions:

1. Average number of claims by company with protection;

2. Average number of claims by company without protection.

Symbolically:

$$\Delta = E(y^1/b(x), z = 1) - E(y^0/b(x), z = 0), \tag{3.4}$$

where 1 denotes the existence of cyber protection and 0 no protection. E is the expectation operator. y^1 and y^0 are the number of claims of the company with and without protection, respectively. $b(x)$ is the probability of protection against cyber risk conditional on the company attribute set (X). The binary variable z, in this case, indicate whether the company has protection $(z = 1)$ or not $(z = 0)$. We consider the difference between the means of the variables of interest for individuals with identical observable characteristics.

Matching from intervals considers the comparison between the averages of the variable of interest number of claims (Y) of companies with and without cyber risk protection, which have on average the same propensity score estimate. Thus, the probability of participation

is estimated, that is, the propensity score for the companies. These companies are then grouped according to their probabilities. The final result represents the weighted sum of the mean differences of the variables of interest (Y) for each group, with the weights given by the participation of each company in each group.

Since Y is the variable of interest (number of claims), the first procedure to do is to compute the differences in the number of claims between those who have (P) and have not (NP) cyber risk protection within each interval:

$$\Delta_e^S = \frac{\sum_{i \in S(e)} Y_i^P}{N_e^P} - \frac{\sum_{j \in S(e)} Y_j^{NP}}{N_e^{NP}}, \quad e = 1, 2, \ldots, m, \tag{3.5}$$

where $S(e)$ is the group of companies in interval e, Y_i^P and Y_j^{NP} correspond to the results observed for the companies i and j, respectively, of the groups with and without cyber risk protection in interval e. N_e^P and N_e^{NP} correspond to the respective numbers of companies in that same layer. The final result (denoted as Δ^S) is determined from the weighted average of the differences (Δ_e^S) obtained for the intervals:

$$\Delta^S = \sum_{e=1}^{m} \Delta_e^S \frac{N_e^P}{N^P}. \tag{3.6}$$

If the result returns a positive value, it means that the expected value of number of claims is higher for companies who purchase protection against cyber risk than for those that do not.

Even when comparing companies with approximate propensity scores in each group, matching from intervals may not use all available observations, since in certain cases it will be possible for companies without protection to be absent from some groups.

It is important to note that these types of estimators based on propensity score estimates, even solving the problem of matching between companies when the number of variables is large, may have some limitations when it comes to non-observable variables and its potential of participation. In addition, there can be no guarantee that there will be comparable companies.

3.3 A neural network model for claim prediction

The main objective of this model is to create a system for predicting the occurrence of a cyber risk attack. The system, based on relevant information, will help identify companies with a high risk of cyber attacks.

A neural network (NN) has the ability to learn from examples and to generalize the information learned. We designed an artificial neural network identifying the explanatory variables relevant to the problem under study.

To briefly define a neural network we considered the definition of Gallant and Gallant (1993) that stated that a NN model consists of a set of computational units and a set of one-way data connections. The author explains that at certain times a unit examines its inputs and computes a signed number called an activation as its output. Every connection has a weight with the purpose of determine the influence of an activation in the receiving

unit. This influence could produce a similar or a different activation depending of the sign of the weight.

According to Svozil *et al.* (1997), the main advantage of neural networks is the fact, that they are able to use some a *priori* unknown information hidden in data (but they are not able to extract it). The power to observe each aspect of the data set and how its units may or may not relate, gives neural network the ability to determine complex patterns across diverse volumes of data. Process of capturing the unknown information is called learning of neural network or training of neural network.

3.3.1 Multilayer perceptron and the use of neural network for classification

Before defining what would be a multilayer perceptron neural network, some other definitions are necessary.

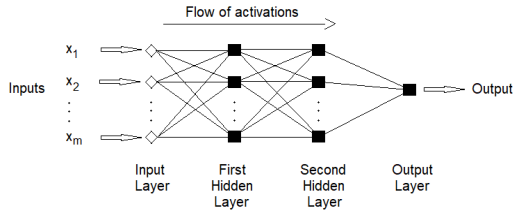
The first is the differentiation of groups of training algorithms. We have supervised and unsupervised algorithms. In a supervised learning the description of each training example also includes the class that the example represents. By the definition of Svozil *et al.* (1997), that neural network knows the desired output and adjusting of weight coefficients is done in such way that the calculated and desired outputs are as close as possible. In unsupervised learning, the neural network is trained without considering the class information associated with each training example.

Another important definition within neural networks is the direction of the flow of activations, feedforward or feedback. In a feedforward neural network, connections do not cycle, which can happen in a feedback network. In this way, when it comes to feedforward neural network, the neurons of a layer will be connected only with the neurons of the layer immediately after. Thus, communications will be unidirectional, with no connections between neurons in the same layer.

Several areas of knowledge can use neural networks in their analysis that are used to solve problems that involve control, prediction or classification. Thus, the types of data used in neural networks are diverse. The most common are categorical and quantitative data. Our study will work with categorical classes and, therefore, the learning of the neural network is called classification.

There are many types of neural networks studied in the literature on the subject. In this work, we considered a feedforward multilayer perceptron neural network (MLP) with supervised learning, see Demuth *et al.* (2014). A multilayer neural network typically has three layers. The first, called the input layer, connects the input variables. The last layer is the output layer that connects the output variables and where the solution to the problem is obtained. Layers in-between the input and output layers are called hidden layers. Figure 1 shows an example of the structure of a multilayer perceptron neural network.

Figure 1: Feedforward multilayer perceptron neural network composed of four layers



For Gallant and Gallant (1993) perceptron-based models are appealing because they are both fast and powerful in their ability to model data. Mathematically, according to definition of Nowosad and Campos Velho (2003), a perceptron network simply maps inputs vectors of real values into output vectors of real values. The connections have associated weights that are adjusted during the learning process, thus changing the performance of the network.

3.3.2 Model structure

For the implementation of a neural network we must determine the following variables: (a) The number of nodes in the input layer corresponding to the number of variables that will be used to feed the neural network; (b) The number of hidden layers; (c) The number of neurons to be placed in the hidden layers; and (d) The number of neurons in the output layer.

The proposed neural model is based on the selection of variables produced by a logit model. Thus, the information for each company to be processed by the neural network consists of sixteen explanatory variables identified as statistically significant by the logit model like the one presented in Table 4.1. The only neuron that leaves the network corresponds to the occurrence of a cyber attack, which has a binary character.

Each neuron in the input layer is connected with all neurons in the hidden layer. It is worth mentioning that the definition of the number of neurons in the hidden layer can reduce overfitting problems. Neural networks with few hidden neurons are preferred, as they tend to have a better generalization power. However, networks with few hidden neurons may not have the ability to model and learn data on complex problems, and underfitting can occur, that is, the network does not converge during training.

The next step in building a neural network is to determine the activation function. For Karlik and Olgac (2011) the most important unit in neural network structure is their net inputs by using a scalar-to-scalar function to transform the activation level of a unit (neuron) into an output signal. The most commonly used activation functions are the logistic function and the hyperbolic tangent function.

The activation function considered for the hidden and output layers was the hyperbolic tangent. After preliminary tests, this function showed a better convergence capacity. This function acts in the interval (-1,1). The hiperbolic tangent activation function takes the form

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{3.7}$$

Once the activation function has been defined, for the implementation of the neural network the database is divided. One part will be considered for the training phase and

another part for the test phase. The training phase adjusts parameters of the network model and trains it to be representative of the problem to be studied in order to minimize the mean square error between what is expected to be obtained with the input data and the result obtained with the output neuron. For this purpose the synaptic weights of the network are iteratively modified.

To define the training and test sets to be used in the design of the artificial neural network, 50% of the companies were considered. The data considered was randomly divided into two sets. The training set consisted of 4,180 companies, while the test set included 4,183 companies. The classification results will be given to the test set.

A neural network always learns from a training set. The learning is reflected in the generalization capacity that the network will display, when used in new situations and this happens using the test set.

Specified the architecture of the neural network, it will be necessary to define the network training algorithm. When it comes to estimating the model parameters, the most popularly used algorithm for this type of training is the backpropagation algorithm, one of the most important when it comes to neural networks. Backpropagation is a technique that calculates derivatives quickly. In this architecture the connection between the i-th and j-th neuron is characterized by a weight coefficient denoted as ω_{ij} and the i-th neuron by a threshold coefficient denoted as υ_i. The weight coefficient reflects the degree of importance of the given connection in the neural network, see Svozil *et al.* (1997). The output value of the ith neuron x_i is determined by:

$$x_i = \phi(\xi_i, \rho) \tag{3.8}$$

$$\xi_i = \upsilon_i + \sum \omega_{ij} x_j \tag{3.9}$$

where ξ_i is the potential of the ith neuron and $\phi(\xi_i, \rho)$ is the activation function. Svozil *et al.* (1997) define the threshold coefficient as a weight coefficient of the connection with formally added neuron j, where $x_j = 1$.

The following measures also must be chosen carefully since their influence can be decisive for the generalization of the network:

- Number of iterations of the algorithm;
- Stopping criterion;
- Starting weights, randomly selected;
- Learning rate.

To avoid overfitting, we used, as a stopping criterion, the estimation of the mean square error below the 0.01 threshold. Tetko *et al.* (1995) suggested the following formulation:

$$MSE = \frac{\sum (y_k - x_k)^2}{(\text{no. of compds.} \times \text{no. of output units})}. \tag{3.10}$$

where x_k is a calculated and y_k is a target value.

The learning rate controls how a neural network model learns a problem, whether slowly or quickly, and its choice depends on the function to be approximated.

For adjustment of the weight and threshold coefficients it holds that:

$$\omega_{ij}^{(k+1)} = \omega_{ij}^{(k)} - \lambda \left(\frac{\delta MSE}{\delta \omega_{ij}} \right)^{(k)} + \alpha \Delta \omega_{ij}^{(k)} \tag{3.11}$$

$$\upsilon_{i}^{(k+1)} = \upsilon_{i}^{(k)} - \lambda \left(\frac{\delta MSE^{(k)}}{\delta \upsilon_{i}} \right)^{(k)} + \alpha \Delta \upsilon_{i}^{(k)} \tag{3.12}$$

where λ is the learning rate ($\lambda > 0$) and α is the moment term ($\alpha \in (0, 1)$). A moment term was considered to increase the training speed of the network, that is, to accelerate the backpropagation algorithm and to avoid local minimums.

After the training process was completed, the trained networks were evaluated in relation to the classification of the companies belonging to the test set through the correct classification rate and sensitivity in the test set.

The neural network depends on the data considered to estimate the desired model. Therefore, the training phase must be rigorous, in order to avoid unrepresentative models. What is expected from a neural model is that its development will result in modeling with good generalization.

3.3.3 Performance measures

Because it is a two-class problem, one of the forms of representation for checking the performance of the neural network model is the confusion matrix. We classify a class as being positive (+) and another as being negative (-). The matrix model can be seen in Table 3.2, where:

- TP corresponds to the number of examples of the positive class correctly classified. In this case, the number of companies that have suffered cyber attacks and were classified as such;
- TN corresponds to the number of examples of the negative class correctly classified. In this case, the number of companies that did not suffer cyber attacks and were classified as such;
- FP corresponds to the number of examples of the positive class classified incorrectly. In this case, the number of companies that have suffered cyber attacks and were classified as companies without attacks;
- TN corresponds to the number of examples of the negative class classified incorrectly. In this case, the number of companies that did not suffer cyber attacks and were classified as companies that suffered attacks.

Table 3.2: Confusion matrix for two-class problems

	Predicted values	
Real values	+	-
+	TP	FN
-	FP	TN

Based on the confusion matrix, several other measures can be calculated to assess the effectiveness of the neural network model. In this work, the total error rate (err), total accuracy (ac), prevalence (p), sensitivity ($sens$) and specificity (esp) will be calculated.

The total error rate, Formula 3.13, is represented by the sum of the main diagonal of the confusion matrix, divided by the sum of all elements of the matrix. Accuracy is the measure that translates the precision of a test. This is Formula 3.14 below.

$$err = \frac{FP + FN}{TP + FP + TN + FN}. \tag{3.13}$$

$$ac = \frac{TP + TN}{TP + FP + TN + FN}. \tag{3.14}$$

Prevalence measures the proportion of companies that suffer claims. It allows to assess whether the condition under study is frequent. See Formula 3.15.

$$p = \frac{TP + FN}{TP + FP + TN + FN}. \tag{3.15}$$

Sensitivity, see Formula 3.16, translates the test's ability to identify a company that is experiencing cyber attack and specificity, see Formula 3.17, translates the test's ability to identify a company that does not suffer from cyber attack.

$$sens = \frac{TP}{TP + FN}. \tag{3.16}$$

$$esp = \frac{TN}{TN + FP}. \tag{3.17}$$

4 Propensity score matching for the impact of protecting against cyber risk on the number of claims

In the simulation procedure, summarized in Section 3.2, for the calculation of the propensity scores, results were obtained using the software STATA. In Table 4.1 we show how each significant variable affect the likelihood of acquiring cyber risk protection carried out by means of a logistic regression.

Table 4.1: Logit models for the probability of acquiring cyber risk protection

Variable	Effect
depart	-0.9762
	(6.47)+
quali	0.9574
	(10.03)+
security	1.4151
	(8.80)+
wired	0.0586
	(8.65)+
wireless	0.5079
	(6.68)+
intra	0.3363
	(3.38)+
extra	0.2646
	(1.97)**
cloud	0.2242
	(2.02)**
readysoft	0.6312
	(5.18)+
freesoft	0.3211
	(5.06)+
otherssoft	0.5918
	(9.39)+
homepage	0.2931
	(4.42)+
fixed	1.1239
	(9.20)+
mobile	0.8163
	(8.40)+
purchase	0.3924
	(5.92)+
gov	0.3976
	(6.32)+
constant	-1.0230
	(4.62)+
observations	16725

Absolute value of the z statistic in parentheses
*significant at 10%, **significant at 5%, +significant at 1%

Looking at the results of Table 4.1 we highlight that:

- Regarding the *depart* variable, for the values of the significant probabilities there is a negative effect, that is, companies that don't have an IT department are more likely to acquire protection from cyber risk;
- All the others variables presented significance and a positive effect.

From the estimates of the Logit model coefficients for each company that did and did not present cyber risk protection, estimates were obtained regarding the probability of acquire protection against cyber risk. To get an idea of the estimated probabilities, we show in Figure 2, the histogram of these probabilities. The left graph represent the distribution of the probabilities of acquiring protection against cyber risk for those companies without

protection, and the right graph represent the distribution of the probabilities of acquiring protection against cyber risk for those companies with protection. Estimates indicate that, in fact, a random sample of companies that did not have cyber risk protection implies lower probabilities of acquire for most of the group of companies.

Figure 2: Propensity score distribution

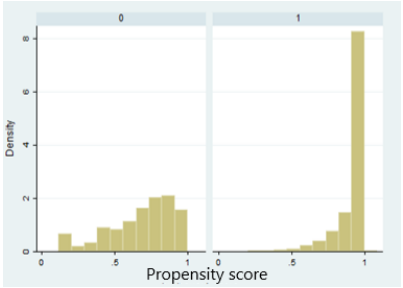


Table 4.2 shows the matching estimate for the impact of protecting against cyber risk on the number of claims.

Table 4.2: Impact of cyber risk protection on the number of claims

	Companies
Number of observations	
With protection	15122
Without protection	1603
Average claims	
With protection	0.6108
standard deviation	0.0071
Without protection	0.3693
standard deviation	0.0177
Propensity score (intervals)	0.0470
standard deviation	0.2040

With regard to the average of claims prior to the analysis of the propensity scores, companies that acquired protection against cyber risk suffered an average of 0.6108 attacks per year, while companies without protection suffered an average of 0.3693 attacks in the same period.

The propensity score point to the pattern that, on average, companies with cyber risk protection have a favorable difference when comparing the number of claims with those without protection, that is, it point to a positive impact of acquiring protection against risk in the occurrence of cyber risk attacks.

5 Performance analysis of neural networks in classifying the occurrence of cyber attacks

The neural network system designed in this work provides identification of the occurrence of a claim. It provides, when fed with new data, coming from the region under study, the

classification of the company (with or without claims) and the probability that the company will suffer a claim, identifying the specific condition of the company.

The architecture that presented the best results for the MLP was structured considering two hidden layers. The first containing 48 neurons and the second 16 neurons. Remembering that we consider 16 inputs and one output layer. We consider a constant learning rate equal to 0.01 to force the weight to be updated smoothly and slowly, avoiding big steps and unstable behavior. The value for the moment term was 0.5. The starting weights were randomly selected in the interval (-0.1, 0.1).

With such parameters, the NN reached the maximum number of iterations (in our case 6,000) and, therefore, the training was aborted. The results obtained can be seen in Tables 5.1 and 5.2.

The proportion of total agreement, that is, the proportion of companies in the test set that were classified as having a claim (non-occurrence), actually presenting a claim (non-occurrence), for the MLP was 86%.

Table 5.1: Confusion matrix

	Predicted values	
Real values	+	-
+	1847	316
-	273	1744

Table 5.2: Performance measures

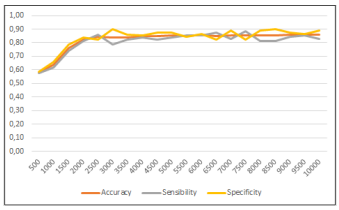
	Value
Total error rate	0.14
Total accuracy	0.86
Prevalence	0.52
Sensitivity	0.85
Specificity	0.86

According to the results presented in Table 5.2, we observed that the MLP presents good sensitivity, that is, it can efficiently classify the companies that declared to have suffered a claim. Both the sensitivity value of 0.85 and specificity of 0.86 show that the model proved to be quite efficient in classifying both positive and negative class.

Assessing the relationship between the parameters adopted and the results obtained, we will analyze three of these relationships to better understand the functioning of the neural network in question.

Figure 3 shows the evolution of the values of the performance measures in relation to the increase in the number of iterations. What has been observed is that such rates increase with the increase in the number of iterations, mainly the accuracy, reaching a level of stability after 6,000 iterations, not needing to perform a greater number of iterations since it would only mean more computational time. The rates of sensitivity and specificity show a greater oscillation around accuracy, but with an increasing trend.

Figure 3: Performance measures *versus* iterations



Concerning the variation of the learning rate and the value of the momentum, its influence in the rates of the performance measures was similar. As the learning rate and momentum values increase, the sensitivity rate also increases. Accuracy and specificity decrease. The latter with greater intensity.

Figure 4: Performance measures *versus* Learning rate

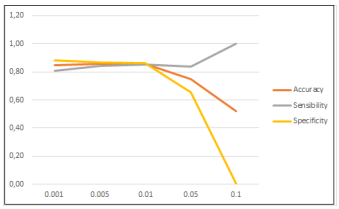
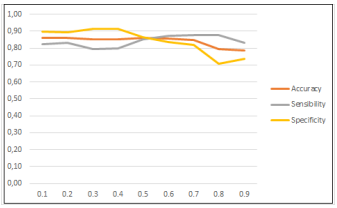


Figure 5: Performance measures *versus* Momentum



It is important to mention that several simulations were carried out until the best model was found to represent the problem under study. For an analysis of the performance of neural networks in the classification, a comparison of the results of MLPs obtained in this work was made. In Table 5.3 we can observe that there was a significant improvement in the results from the first simulation performed.

First, following the indication of Blum (1992) who explains that the number of neurons must be between the size of the final layer and the initial phase, which in our case, must be between 1 and 16, we started simulating a neural network with only one hidden layer with 8 neurons. All others parameters continued the same. The result showed a very inefficient neural network with an accuracy of only 0.57.

After several attempts, we identified that the best architecture would be a neural network with two hidden layers with 48 and 16 neurons, respectively.

Another MLP neural network was developed considering as activation function the logistic function. The greater accuracy rate achieved was 0.82 and, with the sensitivity rate equal to 0.73, the efficiency of this model was worse than the model selected as the best one.

Table 5.3: Comparison of the results obtained from the MLP simulations

Classifier	Accuracy	Sensitivity	Specificity
MLP with the best architecture	**0.86**	**0.85**	**0.86**
MLP with three layers and 8 neurons on the hidden layer	0.57	0.45	0.71
MLP similar to the best model but considering logistic activation function	0.82	0.73	0.91
MLP similar to the best model but without momentum	0.86	0.81	0.90

As can be seen, using the MLP neural network without considering momentum, the same accuracy was obtained, but with a greater discrepancy between the sensitivity and specificity rates. The sensitivity rate decreased to 0.81, while the specificity increased to 0.90. Such values indicate that the implemented NN is very good in classifying patterns belonging to the negative class, but has lost efficiency for data from the positive class.

The fact is that the use and study of neural networks can be quite opportune to help in solving the problem presented.

6 Remarks and conclusions

This paper presents an analysis of self protection through a propensity score-matching method, producing accurate estimates of the effect of a certain treatment in a non-experimental environment, and a performance analysis in classifying the occurrence of cyber attacks through feedforward multilayer perceptron neural networks.

Rubin (1973) explains that when combined with covariance adjustments for paired differences, multivariate paired sampling is known to be one of the most robust methods to reduce bias due to imbalances in the observed covariates. The correspondence with the estimated propensity score was quite successful in reducing bias. This method highlights the differences between the comparison and the treated units.

Feedforward multilayer perceptron neural networks have been used as a classification model in several fields of study. It is important to know that the database can be constantly updated, both to improve the statistical characterization of the occurrence of cyber attacks and to incorporate new information that can better describe cyber risks. The choice of significant variables in the implementation of neural networks is imperative. The inclusion of variables not relevant to the problem under study may disturb the performance of the neural network, as well as the classification error.

Our analysis of self protection shows that despite informal arguments that favor protection against cyber risks as a tool to improve network security, we observed that in the presence of protection against cyber risks, the incidence of claims is higher than if no protection existed. This type of analysis is only the beginning of many of the possibilities that could extend this study of self protection and cyber attacks. This result could represent a problem of signaling and screening. Companies have a better idea of the risks they face. Those who know that they face large risks are more likely to get self-protection or to buy insurance against cyber risk than those who face small risks. This is the problem of adverse selection. All of these assumptions can be the initial trigger for the improvement and extension of this study.

The classification results using a MLP neural network trained with a backpropagation algorithm were very good, with 86% global hits. It was also possible to conclude that, as the number of iterations increases, there is a trend to increase the accuracy of the classifications.

The neural model, proposed here, can be conducted in an innovative way as a supporting tool for the decision making of insurers, aiming at useful responses to risk management.

It is a fact that people and organizations are more connected and investing in technological innovations. On the one hand, such behavior is inevitable as the world is increasingly interconnected by digital relationships. On the other hand, cyber threats are increasingly frequent, generating significant organizational, financial and reputational impacts. For companies, deciding to invest in cybersecurity is directly linked to the cost of implementing and adapting to these technologies.

Xu and Hua (2019) lists that information about network configurations, network flows, cyber incidents and security protocols, among others, is used to develop statistical models to model and predict cyber security risks. Their approach facilitates risk assessments for a large-scale network.

One of the measures that is on the rise, in addition to self-protection, is cyber insurance. This line of business is evolving rapidly but faces challenges such as determining the severity of claims. This challenge makes pricing difficult as there is still insufficient data to properly characterize the risk and one may have to deal with extreme events. In this work, no premium calculation was made since the interest was to investigate the nature of the risk itself.

Acknowledgements

The data used in this paper are from the Survey on the use of information and communication technologies in companies, for the year 2010, conducted by "Instituto Brasileiro de Geografia e Estatística" (IBGE), and were obtained through authorized access to the institution's restricted data access room. The results, analyzes and interpretations presented are the sole responsibility of the authors, neither representing the official view of IBGE nor constituting official statistics.

The author was partially supported by the Project CEMAPRE/REM - UIDB/05069/2020 - financed by FCT/MCTES through national funds.

References

Barracchini, C., Addessi, M. E., *et al.* (2014). Cyber risk and insurance coverage: An actuarial multistate approach. *Review of Economics Finance*, 4:57–69.

Betterley, R. (2018). The betterley report. Cyber/privacy insurance market survey 2018. *The Betterley Reportt*. `https://www.irmi.com/docs/default-source/publication-tocs/betterley-report---cyber-risk-market-survey-june-2018-summary.pdf?`

Blum, A. (1992). *Neural networks in C++: An object-oriented framework for building connectionist systems*. John Wiley & Sons, Inc.

Böhme, R. (2005). Cyber-insurance revisited. In *WEIS*. `http://infosecon.net/workshop/pdf/15.pdf`.

Böhme, R. and Kataria, G. (2006). Models and measures for correlation in cyber-insurance. In *WEIS*, volume 2, page 3. `https://core.ac.uk/download/pdf/162458449.pdf`.

Bolot, J.-C. and Lelarge, M. (2008). A new perspective on internet security using insurance. In *IEEE INFOCOM 2008-The 27th Conference on Computer Communications*, pages 1948–1956. IEEE.

Caliendo, M. and Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1):31–72.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161.

Demuth, H. B., Beale, M. H., De Jess, O., and Hagan, M. T. (2014). *Neural network design*. Martin Hagan.

Diniz, G., Muggah, R., and Glenny, M. (2014). Deconstructing cyber security in brazil: Threats and responses. `https://igarape.org.br/wp-content/uploads/en/2014/11/Strategic-Paper-11-Cyber2.pdf`.

Fahrenwaldt, M. A., Weber, S., and Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA*, 48(3):1175–1218.

Gallant, S. I. and Gallant, S. I. (1993). *Neural network learning and expert systems*. MIT press.

Gordon, L. A., Loeb, M. P., and Sohail, T. (2003). A framework for using insurance for cyber-risk management. *Communications of the ACM*, 46(3):81–85.

Herath, H. and Herath, T. (2011). Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies: Analyses and Actuarial Computations*, 2(1):7–20.

Institute, P. (2019). IBM: Cost of a data breach report 2019. `https://www.all-about-security.de/fileadmin/micropages/Fachartikel_28/2019_Cost_of_a_Data_Breach_Report_final.pdf`.

Ionită, M.-G. and Patriciu, V.-V. (2014). Biologically inspired risk assessment in cyber security using neural networks. In *2014 10th International Conference on Communications (COMM)*, pages 1–4. IEEE.

Karlik, B. and Olgac, A. V. (2011). Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122.

Macedo, M., de Brito, J., Silva, A., and Oliveira Cruz, C. (2019). Design of an insurance policy model applied to natural stone facade claddings. *Buildings*, 9(5):111.

Marotta, A., Martinelli, F., Nanni, S., Orlando, A., and Yautsiukhin, A. (2017). Cyber-insurance survey. *Computer Science Review*, 24:35–61.

Muggah, R. and Thompson, N. B. (2017). Brazil struggles with effective cyber-crime response. `https://igarape.org.br/brazil-struggles-with-effective-cyber-crime-response/`.

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., and Sadhukhan, S. K. (2006). e-risk management with insurance: A framework using copula aided bayesian belief networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, volume 6, pages 126a–126a.

Mukhopadhyay, A., Saha, D., Chakrabarti, B., Mahanti, A., and Podder, A. (2005). Insurance for cyber-risk: A utility model. *Decision*, 32(1).

Nowosad, A. G. and Campos Velho, H. F. (2003). New learning scheme for multilayer perceptron neural network applied to meteorological data assimilation. In *Proceedings of the 24th Iberian Latin-America Congress on Computational Methods in Engineering*, pages 1–8. UFOP.

Ogut, H., Menon, N., and Raghunathan, S. (2005). Cyber insurance and IT security investment: Impact of interdependent risk. In *WEIS*. `http://infosecon.net/workshop/pdf/56.pdf`.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, pages 159–183.

Schwartz, G. A. and Sastry, S. S. (2014). Cyber-insurance framework for large scale interdependent networks. In *Proceedings of the 3rd International Conference on High Confidence Networked Systems*, pages 145–154.

Shipman, J. E., Swanquist, Q. T., and Whited, R. L. (2017). Propensity score matching in accounting research. *The Accounting Review*, 92(1):213–244.

Subroto, A. and Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning. *Journal of Big Data*, 6(1):50.

Svozil, D., Kvasnicka, V., and Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1):43–62.

Tetko, I. V., Livingstone, D. J., and Luik, A. I. (1995). Neural network studies. 1. comparison of overfitting and overtraining. *Journal of Chemical Information and Computer Sciences*, 35(5):826–833.

Xu, M. and Hua, L. (2019). Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal*, 23(2):220–249.