Cyber-Insurance: Copula Pricing Framework and Implications for Risk Management

Hemantha S. B. Herath
Associate Professor, Department of Accounting, Faculty of Business,
240 Taro Hall, 500 Glenridge Avenue, St. Catharines, Ontario, Canada L2S 3A1,
hemantha.herath@brocku.ca

Tejaswini C. Herath
Ph.D. Candidate, Department of Management Science and Systems, Jacobs Management Center,
University at Buffalo, Amherst, NY 14260,
tcherath@buffalo.edu

ABSTRACT

In recent years there has been a growing stream of research focusing on cyber-insurance. Risk transference with insurance has been suggested by both practitioners and academics to absorb losses caused by security breaches as well as to supplement the existing set of security tools to manage IT security residual risk after IT security investments are made. In this paper, we investigate pricing of cyber-insurance products using the emerging copula methodology for modeling dependent risks from an *actuarial approach* which is different to the process approaches of Bohme and Kataria (2006) and Mukhopadhyay et. al. (2006). We discuss a framework for assessing the empirical dollar loss distribution from the empirical distribution of the number of infected computers. We develop a cyber-insurance model and demonstrate the Gumbel copula to price insurance premiums using a numerical example with ICSA data.

Key Words: Cyber-Insurance; Copula; Correlated Risk, Copula Dependency, Information Security Risk Management.

Acknowledgements: Dr. Hemantha Herath acknowledges financial support from the Social Sciences and Humanities Research Council (SSHRC) of Canada.

1. Introduction

The survey results of financial losses due to information security breaches give us an overall glimpse of the severity of the problem of virus, hacker and denial of service attacks. Recent incidents (2005; 2006) of security breaches have resulted organizations millions of dollars in losses due to lost revenues not withstanding the intangible losses such as lost productivity, loss of customer goodwill/reputation and lost business opportunities . The recent CSI/FBI survey report notes that majority of the organizations

use or have security tools in place with almost 99% using Antivirus software, 98% having firewalls, 71% having proxy servers, and 68% having intrusion detection systems. However, in spite of the large use of security measures the losses due to breaches are still extremely high. Anderson (2001) explains this phenomenon clearly. It is difficult for security managers in any organization to know about and eliminate all the points of vulnerability in a system (i.e. create a fool proof system), when it only takes a hacker to exploit just one these points of vulnerability. Similarly, as Schneier (2002) point out, a new virus can easily compromise the perimeter security devices before a signature is available and implemented for anti-virus tools to track it down.

Recognizing that the elimination of security breach risk is close to impossible, NIST recommends several risk mitigation techniques that are based on technical as well as other non-technical controls. These include (Stoneburner et al. 2002):

1. Risk Assumption: To accept the potential risk and continue operating the IT system or to implement controls to lower the risk to an acceptable level
2. Risk Avoidance: To avoid the risk by eliminating the risk cause and/or consequence (e.g., forgo certain functions of the system or shut down the system when risks are identified)
3. Risk Limitation: To limit the risk by implementing controls that minimize the adverse impact of a threat's exercising a vulnerability (e.g., use of supporting, preventive, detective controls)
4. Risk Planning: To manage risk by developing a risk mitigation plan that prioritizes, implements, and maintains controls
5. Research and Acknowledgment: To lower the risk of loss by acknowledging the vulnerabilities or flaws and researching controls to correct them
6. Risk Transference: To transfer the risk by using other options to compensate for the loss, such as purchasing insurance.

In this paper, we focus on risk transference as a tool to minimize some of the financial losses to firms. Risk transference with insurance has been suggested by both practitioners and academics to absorb loss caused by security breaches as well as to supplement the existing set of security tools to manage IT security residual risk after IT security investments are made (see (Böhme 2005; Böhme and Kataria 2006; Gordon et al. 2003; Mukhopadhyay et al. 2006; Ogut et al. 2005) among others). Typically an individual or organization employs a combination of these risk management options simultaneously – retaining some of the risk, mitigating some and insuring the rest(Schneier 2002). While there are some similarities in cyber-insurance that are generic to other insurance products, there are certain characteristics unique to IT security. These unique characteristics, makes the pricing of cyber insurance challenging. First, internet related risks are unique, in terms of location, degree, visibility and traditional policies do not comprehensively address the additional risks that firms face as a result of being part of the digital economy (Gordon et al. 2003); and second, although issues related to pricing, adverse selection and moral hazard are common to all forms of insurance; an understanding of these issues related to cyber risks is warranted in designing insurance products.

These unique challenges create a plethora of research questions that need to be addressed from the point of view of insurance companies (supply side) as well as the insured (demand side). For instance, pricing insurance products traditionally relies on actuarial tables constructed from historical records. The internet is relatively new, and as such data about security breaches and losses does not exist or does so only in small quantities. This is further exacerbated with the reluctance of organizations to reveal details of security breaches due to loss of market share, reputation etc.(Gordon et al. 2006). Although insurance companies currently provide cyber-insurance products, the accuracy of the pricing and whether or not insurance providers are charging the right premiums is still an open question (Gordon et al. 2003; Radcliff 2001).

In recent years there has been a growing stream of research focusing on cyber-insurance. Gordon and Loeb (2003) in their article nicely lay out a framework for using cyber-insurance as a risk management technique. In addition, they discuss the unique features of cyber-insurance as well as similarities in respect of pricing, adverse selection and moral hazard. Ogut et. al (2005), investigates the cyber-insurance problem from the issue of moral hazard and adverse selection. They show that the interdependency of IT security risk of different firms impact a firm's incentive to invest in cyber-insurance products. Böhme (2005) provides an intuitive discussion of the issue of correlated risk in cyber-insurance and argues whether the structural characteristics of the internet itself (i.e., the monopoly of a dominant platform(s)) will restrict the creation of a proper market for cyber-insurance. His arguments are through and clearly valid, and have implication for pricing. Using an indemnity insurance model, he evaluates the conditions under which insurance can be provided and model different premiums for users of dominant and alternate platforms. Extending Böhme (2005) work further, Böhme and Kataria (2006) investigate the correlated cyber-risks in a two-step risk arrival process, that is, within the firm (intra-firm risk correlation) and external to the firm (global risk correlation). In order to capture the global risk correlations the authors use the t-copula which is used to model correlation of extreme events.

Mukhopadhyay et. al. (2006), use a copula approach with the Bayesian Belief Networks (BBN) technique to quantify the e-risk associated with online transactions that would be affected by security breaches. They use the mutivariate normal copula to describe the joint distribution which is thereafter used to compute the conditional distribution at each node on the BBN. Using the software FULLBNT they are able to identify the probabilities associated with the specific causes of the security breaches. Using these probabilities for the risk of a breach and making a strong simplifying assumption that the dollar losses at each node in the network are distributed binomially with assumed specific values, they price the cyber-insurance premiums as a function of the expected value of the claim severity and its standard deviation. The claim severity is the product of the expected dollar loss amount and the probability of the loss. All of the above papers make a very significant contribution to understanding the issues of cyber-risk insurance.

In this paper, we investigate cyber-insurance pricing using the emerging copula methodology for modeling dependent risks from an actuarial approach which is different

to the approaches of Böhme and Kataria (2006) and Mukhopadhyay et. al. (2006). Mukhopadhyay et. al. (2006) uses a process view to quantify the operational risk that is based on modeling the chain of activities that constitute an operation and estimating the exact risk of each process. Böhme and Kataria (2006) work can also be categoried as a process approach that focuses on connectivity and system dynamics. Both these methods are useful for the actuarial approach that we develop in this paper. The use of copula methodology is unique in each of the three papers, Böhme and Kataria (2006) uses the t-copula, Mukhopadhyay et. al. (2006) uses the multi-variate normal copula and we use two Archimedean copulas - Clayton and Gumbel. Thus our paper makes the following contributions to the literature. First, we use the actuarial approach based on empirical loss distributions. In particular, we develop a framework for assessing the empirical dollar loss distribution from the empirical distribution of the number of infected computers. More specifically, we use the ICSA data of the actual virus incidents and the number of computers (modified at a firm level) to assess the empirical loss distribution. The empirical loss distribution is dependent on the security risk posture of a firm. Second, we provide a detailed survey of copulas (Clayton and Gumbel), discuss the fit test, how to simulate bivariate data from a known copula and develop a cyber-insurance model. As noted above, use of copulas is relatively new to cyber-security insurance industry. Third, we illustrate the model using a numerical example with ICSA data. This paper is organized as follows. In Section 2, we present a framework for assessing cyber risk and estimating the empirical distribution for dollar losses. Sections 3 provide a survey of copulas and a cyber-insurance model. In Section 4, we illustrate the methodology with the help of case study. Section 5 concludes the paper.

2. Framework for Assessing Cyber Risk

In pricing the premium, it is essential to identify the likelihood of a potential disaster as well as its impact. Risk is defined as a function of the ***likelihood*** of a given ***threat-source's*** exercising a particular potential ***vulnerability***, and the resulting ***impact*** of that adverse event on the organization (NIST) (emphasis in original).

Likelihood of threat source's exploiting vulnerability is random and is similar to situations normally faced by insurance companies such as expectation of a natural disaster. This likelihood can be assumed to follow Poisson distribution (Conrad 2005). Although the occurrence of virus is random, the possibility of breach may very well depend on the number of vulnerabilities. Number of vulnerabilities can be expected to depend on the security precautions taken by the firm (security posture coefficient - s). For example, daily monitoring and update of virus signatures will be less risky than if the firm updates the virus signatures weekly. Similarly, if the firm has resources and can afford hourly monitoring and update, the firm is likely to be safer than daily or weekly monitoring.

The dollar losses will depend on the type of computer affected and its relative losses. Security breaches pose various types of losses: (1) lost productivity (2) lost revenue, (3) clean up costs and (4) financial performance impact, to name a few (Farahmand et al.

2004). These costs can be expected to be dependent on the type of computer that is breached. For example, if the virus has crippled the administrative PCs, it will impose productivity loss of the persons as well as clean up costs related to the attack. If the computer involved is a web-server that is mostly used in e-business type of activity, in addition to clean up costs it is likely to result in lost revenues. In general, total losses due to a particular risk (e.g. virus attack) can be expressed as a function of: type of computer affected; losses related to type of computer and security posture of the firm. Thus the loss function based on the number of computers can be we written as.

$L = \quad s_1 \alpha_1 f_1 + s_2 \alpha_2 f_2 + s_3 \alpha_3 f_3 \ \ldots$

where $\alpha_1 + \alpha_2 + \alpha_3 \ldots = 1$

α_n : fraction of computers of type n

f_n : dollar impact for computer type n

s_n: security posture coefficient

3. Copula Methodology

"Copulas" are functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distributions whose one-dimensional margins are uniform on the interval [0,1] (Nelsen 1995, Frees and Valdez 1998). Copulas have been studied for over forty years. Sklar (1959) coined the term *copula* to describe functions which join together one-dimensional distribution functions to form multivariate distribution functions. Copulas are of interest to statisticians for two main reasons: Firstly, as a way of studying scale-free measures of dependence; and secondly as a starting point for constructing families of bivariate distributions, sometimes with a view to simulation (Fisher 1997).

Definition: Sklar (1959) Theorem

Let X and Y denote continuous random variables (lower case x, y represent their values) with bivariate distribution function $H(x, y)$ and marginal distribution function $F(x)$ and $F(y)$. Let $F^{-1}(.)$ and $G^{-1}(.)$ be the inverse of F and G. Then for any uniform random variables U and V with values $u, v \in [0,1]$ (i.e. make the probability transformation of each variate $U = F(X)$ and $V = G(Y)$ to get a new pair of variates $U \sim U(0,1)$ and $V \sim U(0,1)$), exist a copula C such that for all $x, y \in R$

$$H(x, y) = C(F(x), G(y)) = C(u, v) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

If F and G are continuous then C is unique. An important feature of copulas is that any choice of marginal distributions can be used. Hence copulas are constructed based on the assumption that marginal distribution functions are known.

Copulas are important because they allow us to study the dependence or association between random variables. There are several ways to measure dependence. The most

widely used measures are the Spearman's Rho and Kendall's Tau. Copulas precisely account for the dependence between random variables. For example between two random variables X and Y the dependence properties of the joint distribution (the manner in which X and Y move together) is precisely captured by the copula for strictly increasing functions of each variable. The two standard non-parametric dependence measure expressed in copula form are as follows:

Kendall's Tau is given by:

$$\tau = 4\iint_{I^2} C(u,v)\mathrm{d}C(u,v) - 1 \qquad \text{.....(2)}$$

and Spearman's Rho is given by:

$$\rho = 12\iint_{I^2} C(u,v)\mathrm{d}u\mathrm{d}v - 3 \qquad \text{.....(3)}$$

The expressions for Kendall's Tau and Spearman's Rho for some known families of copulas are presented in Section 3.0 .

Copulas provide a way to study scale free measures of dependence. In empirical applications where data is available we can use the dependence measure to specify the form of copula. Genest and Rivest (1993) provides a procedure for identifying a copula when bivarate data is available. One the correct copula is identified it can be used to simulate random outcomes from dependent variables.

3. 1 Survey of Gumbel and Clayton Copula

In this paper we survey two one parameter bivariate Archimedean copulas adopted from Frees and Valdez (1998) and Nelsen (1999) (Nelsen, 1990, pg 94-97 lists 22 one parameter families). Archimedean copulas are **easy** to apply and have nice properties. The parameter θ in each case measures the degree of dependence and controls the association between the two variables. For instance when $\theta \to 0$ there is no dependence and if $\theta \to \infty$ there is perfect dependence. Schweizer and Wolff (1981) show that the dependence parameter θ which characterizes each family of Archimedean copulas can be related to Kendall's Tau. This property can be used to empirically determine the applicable copula form.

(i) Clayton Copula (1978)

(a) Generator: $\varphi_\theta(t) = (t^{-\theta} - 1)$

(b) Bivariate Copula: $C_\theta(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}$

(c) Laplace Transform: $\varphi(t) = \varphi_\theta^{-1}(t) = (1-t)^{\frac{1}{\theta}}$

(d) Kendall's Tau $\tau_\theta = \frac{\theta}{\theta + 2}$

(ii) Gumbel Copula (1960)

(a) Generator: $\varphi_\theta(t) = (-\ln(t))^\theta$

(b) Bivariate Copula: $C_\theta(u,v) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{\frac{1}{\theta}}\right\}$

(c) Laplace Transform: $\varphi(t) = \varphi_\theta^{-1}(t) = \exp(-t^{\frac{1}{\theta}})$

(d) Kendall's Tau $\tau_\theta = 1 - \theta^{-1}$

3.2 Identifying a Copula Form

The first step in modeling and simulation is identifying the appropriate copula form. Genest and Rivest (1993) provide the following procedure (fit test) to identify an Archimedean copula. The method assumes that a random sample of bivariate data (X_i, Y_i) for $i = 1,2.....,n$ is available. Assume that the joint distribution function H has an associated Archimedean copula C_θ, and then the fit allows us to select the appropriate generator φ. The procedure involves verifying how close different copulas fit the data by comparing the closeness of the copula (parametric version) with the empirical (non-parametric) version.

The steps are follows:

Step 1: Estimate the Kendall's correlation using the non-parametric or distribution free measure

$$\tau_E = \binom{n}{2}^{-1} \sum_{i<j} \text{Sign}\left[(X_i - X_j)(Y_i - Y)_i\right]$$

Step 2: Identify an intermediate variable $Z_i = F(X_i, Y_i)$ having a distribution function $K(z) = \Pr(Z_i \le z)$. Construct an empirical (non parametric) estimate of this distribution as follows:

$$Z_i = \frac{\text{number}\left\{(X_i, Y_j) \text{ such that } X_j < X_i \text{ and } Y_j < Y_i\right\}}{n-1}$$

The empirical version of the distribution function $K(z)$ is $K_E(z) =$ proportion of $Z_i \le z$

Step 3: The next step is to construct the parametric estimate of $K(z)$. The relationship between this distribution function and the generator is given by $K(z) = z - \frac{\varphi(z)}{\varphi'(z)}$, where

$\varphi'(z)$ is the derivative of the generator and $0 \le z \le 1$. We show below the specific form of $K(z)$ for the three Archimedean copulas surveyed in this paper.

(i) Clayton Copula $$K(z) = \frac{z(1+\theta - z^{\theta})}{\theta} \qquad \text{.....(4)}$$

(ii) Gumbel Copula $$K(z) = \frac{z(\theta - \ln z)}{\theta} \qquad \text{.....(5)}$$

Repeat Step 3 for several different families of copulas**, i.e.,** several choices of $\varphi(.)$. By visually examining the graph of $K(z)$ vs z or using statistical measures such as minimum square error analysis, **one** can choose the *best* copula. This copula can be used in modeling dependencies and simulation.

3.3 Pricing Cyber-Insurance

Copula methodology can be effectively used for forecasting the dollar value of losses from cyber attacks and pricing cyber-insurance. In this section, we discuss an application example that uses copula dependency. A key component of insurance pricing is understanding and modeling multivariate relationships. While linear regression may provide a basis to explain the relationship among two (or more) variables, the model is based on normality assumptions and linear dependence. Linear regression would work if the marginal distribution are normal. However, as ICSA data indicate the marginal distribution for the **number of computers affected** (X) and the **dollar value of losses** (Y) are not normal. In the case of the variable, **number of computers affected** (X) the marginal distribution is likely to be of the type Pareto or Exponential or Weibull, since a fewer number of computer virus (15%-25%) account for (85% - 75%) of the number of computes affected. Since the marginal distributions are non-normal the widely used classical Pearson's product moment correlation (ρ) cannot be used to model the dependency among the two variables. Correlation (ρ) measures the straight line association and thus the dependency is linear. Thus in forecasting the dollar value of losses and pricing cyber-insurance, the copula dependency is more appropriate.

In the copula approach for pricing cyber insurance, the first step is the use the procedures laid out in Section 3 above to identify the "**appropriate copula"** for modeling the non-linear dependence that explain the relationship between the two variables of interest, the number of computers affected (X) and the dollar value of losses (Y). That is, we identify the joint distribution of (X, Y) given by the specific copula function say $g(X, Y)$. Notice that now we can examine the distribution of any known function of X and Y. Consider, the expected insurance policy premium for a firm that has N computers (standalone PCs and servers). Let L and M be the lower and higher limits of the number of likely computers affected to determine the premium. Assuming the dollar value of the likely losses can be prorated based on the number of computers affected (or exposed), we can write the following function for pricing the cyber-insurance:

$$g(X,Y)=\begin{cases} a_1 & \text{if } X < L \\ a_2 + \left(\dfrac{X-L}{X}\right)\left(\dfrac{Y}{10}\right) & \text{if } L \le X < M \\ a_3 + \left(\dfrac{X-M}{X}\right)\left(\dfrac{Y}{10}\right) & \text{if } X \ge M \end{cases} \tag{6}$$

where, $a_i, \quad i = 1,2,3$ are constants. The expected insurance premium could be computed using Monte Carlo simulation. The steps are as follows:

Step 1: Generate a sequence of bi-variate data (X_i, Y_i) using the fitted copula (ie, Clayton or Gumbel or any other copula). The procedures for generating data from Clayton and Gumbel are well summarized in Frees and Valdez (1998), Nelson (1999) among others.

Step 2: For a given (L, M) compute the expected value of the cyber-insurance premium and the standard deviation as

$$E[g(X,Y)] = \frac{1}{S}\sum_{i=1}^{S} g(X_i, Y_i) \tag{7}$$

$$\sigma[g(X,Y)] = \sqrt{\frac{\frac{1}{S}\sum_{i}^{S} g(X_i,Y_i)^2 - [E(g(X,Y))]^2}{S}} \tag{8}$$

where S is the number of simulation runs.

4. Case Illustration

In this section, we illustrate the copula approach for pricing cyber-insurance using data from the ICSA. Consider Firm A[1] that has the following data pertaining to the **number of computers affected** (X) for each major computer viruses in 2003. Using the methodological framework discussed in Section 2, we estimate the **dollar value of losses** (Y) as given in Table 1. Notice, that both the number of computers affected as well as the dollar value of losses are random events. That is, the number of computers affected will depend on the severity of the virus, the company's security posture, and security policies in place. Similarly, the dollar value of losses will be random in the sense, that in rare instance of the same number of computers affected for two distinct viruses, the degree of the losses will not be identical because it will depend on each viruses ability to penetrate and harm the computers. Also, it will depend on the computer type affected (i.e., the proportion of stand alone computers, servers, network computers etc.). The data for Firm A is given in Table 1.

[1] We use the ICSA data for 2003, on actual compute virus incidences and the actual number of computers affected, scaled down one hundred times to firm level.

Table 1: Computer Virus Data for Firm A.

	Virus	X # of Computers	Y $ Losses
1	W32/Blaster	1291	$ 355,648.72
2	W32/Slammer	849	$ 339,832.66
3	W32/Sobig	238	$ 115,729.51
4	W32/Klez	140	$ 65,090.38
5	W32/Yaha	118	$ 45,402.25
6	W32/Swen	108	$ 66,053.73
7	W32/Dumaru	87	$ 39,182.88
8	W32/Mimail	70	$ 19,556.82
9	W32/Nachi	63	$ 20,087.13
10	W32/Fizzer	58	$ 20,465.35
11	W32/BugBear	50	$ 10,180.13
12	W32/Lirva	47	$ 11,769.29
13	W32/Sober	21	$ 6,944.48
14	W32/SirCam	21	$ 5,339.08
15	W32/Ganda	19	$ 7,547.77
Mean		212	$ 75,255
Standard Deviation		363	$ 114,702

In order to price the cyber-insurance premium for Firm A, based on the number of computers affected and the dollar value of losses pertaining to each virus incidence, we have to simulate bi-variate data (X_i, Y_i) with non linear dependence using the best fit copula. For the simulation, we need to identify the marginal distributions for the number of computers affected (X) and the dollar value of losses (Y) above in Table 1. We use ARENA software for identifying the marginal distributions. The marginal distribution for number of computers affected (X) and the dollar value of losses (Y) above both are Weibull distribution of the following form. The fitted marginal distributions are $X \sim 18 + \text{Weibull}(118, 0.586)$ and $Y \sim 5340 + \text{Weibull}(38900, 0.586)$.

Notice that both the distributions are shifted Weibull distributions. There are several properties, which pave way for using copula dependencies to simulate the pair of values (X_i, Y_i). First, since the marginal distributions are non-normal (Weibull), one cannot use linear regression for simulating; second, finding the joint distributions of the two variables is not easy since they are shifted Weibull distributions; and finally Pearson's product moment or linear correlation cannot be use since the marginals are non-normal. However, the approach as demonstrated below enables us to identify the appropriate copula to determine a joint distribution that can be used with any marginal distribution. Furthermore, copula allows modeling nonlinear dependency which is more appropriate for estimating premiums.

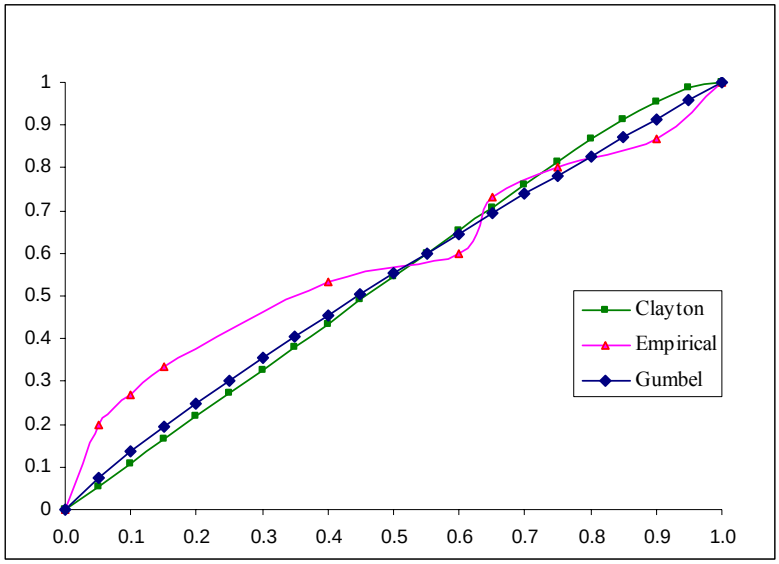

Figure1: Best Fit Copula

We use Kendall's Tau to measure the dependence between number of computers affected (X) and the dollar value of losses (Y) for the data in Table 1, which using the statistical software SPSS is computed as 0.848. Next, we obtain θ values of 11.15789 and 6.578947 respectively using equations in section 3.1 (i)-(d), and (ii)-(d) for Clayton and Gumbel copulas. In order to identify the appropriate copula we follow the procedure outlined in Section 3.2. The empirical distribution $K_E(z)$ and the $K(z)$ values for Clayton and Gumbel copulas based on equations in section 3.2 (i) and (ii) and (iii) are shown in Figure 1. Using a visual fit it is evident from Figure 1 that Gumbel copula provides the best fit (both are relatively close). However, minimizing the mean square error would be a more robust method to select the appropriate copula.

In order to simulate bi-variate outcomes (X_i, Y_i) using the Gumbel copula we use the algorithm suggested by Marshall and Olkin (1988). The algorithm is as follows:

For the Gumbel copula the generator and Laplace transform are given in Section 3.1 (ii). The inverse generator is equal to the Laplace transform of a positive Stable variate $\gamma \sim \mathrm{St}(\hat{\alpha},1,\Theta,0)$ where $\Theta = \left(\cos\left(\frac{\Pi}{2\theta}\right)\right)^{\theta}$ and $\theta > 0$

Step 1: Simulate a positive Stable variate $\gamma \sim \mathrm{St}(\hat{\alpha},1,\Theta,0)$

Step 2: Simulate two independent uniform [0,1] random numbers u_1 and u_2

Step 3: Set $X = F^{-1}\left(u_1^*\right)$ and $Y = G^{-1}\left(u_2^*\right)$ where $u_i^* = \varphi\left(\frac{1}{\gamma}\ln u_i\right)$ and $\varphi(t) = \exp\left(-t^{\frac{1}{\theta}}\right)$ for $i \in [1,2]$

Nolan (2005) and Cherubini et. al (2004) suggest the following procedure to simulate a positive random variable $\gamma \sim \mathrm{St}(\hat{\alpha},1,\Theta,\delta)$

Step 1 (a): Simulate a uniform random variable $\upsilon = U\left(\frac{-\Pi}{2}, \frac{\Pi}{2}\right)$

Step 1 (b): Independently draw an exponential random variable (ε) with mean = 1

Step 1 (c): $\theta_0 = \arctan\left(\tan\left(\frac{\Pi\hat{\alpha}}{2}\right)\right)\Big/\hat{\alpha}$ and compute

$$z = \frac{\sin\hat{\alpha}(\theta_0 + \upsilon)}{(\cos\hat{\alpha}\theta_0 \cos\upsilon)^{\frac{1}{\hat{\alpha}}}}\left[\frac{\cos(\hat{\alpha}\theta_0 + (\hat{\alpha}-1)\upsilon)}{\varepsilon}\right]^{\frac{(1-\hat{\alpha})}{\hat{\alpha}}}$$

Step 1 (d) $\gamma = \Theta z + \delta$

In order to determine the cyber-insurance premium for a firm, we simulate a large number of bivariate data (X_i, Y_i) by repeating the algorithm $S = 10{,}000$ times. The annual insurance premium for different levels of L and M with the following values for $a_1 = 400$, $a_2 = 125$ and $a_3 = 300$ are given in Table 2.

Table 2: Computed Premiums

	L=25	
	Mean	Standard Deviation
M=50	$912	$608
M=100	$1007	$701
M=150	$1267	$831

5. Discussion

In this paper, we use a copula based model for pricing cyber-insurance. The cyber insurance pricing model explicitly considers the risk posture of the firms since both the valuation parameters - the number of computers affected and the dollar losses depend on the risk posture of the firm. The framework, proposed for assessing the risk posture and estimating the dollar losses considers computer product diversity, lost productivity, lost revenue, and clean up costs among others. Furthermore, as previously discussed, the model considers nonlinear dependencies for correlated risks and allows simulating from a copula model without explicitly having to determine the joint distribution for the two given marginals. Hence, the model is versatile in the sense that any type of marginal distributions can be used. Thus, our paper makes a significant contribution to the literature in cyber-insurance risk modeling and pricing since we address the problem from an actuarial approach .

One of the primary constraints in pricing cyber-insurance as identified by Gordon and Loeb (2003) is the unavailability of large amounts of historic data on e-crimes and related losses. As Gordon and Loeb (2003) point out, it is further exacerbated due to the fact that firms do not reveal details concerning security breaches. That is, since cyber-insurance products are new, its is not known whether the right premiums are being

charged. One of the primary limitations of the proposed model is that it uses historic data to determine the appropriate copula for pricing the premiums. Thus one can argue, that it suffers from the same lack of data problem as current insurance pricing models. However, the proposed copula based cyber-insurance model makes a methodological contribution since it provides a different modeling perspective using the actuarial approach. Finally, the framework provides awareness to managers, that it is important to collect data on e-crimes and security breaches for negotiating for lower premiums on cyber-insurance products. For future research, we hope to develop in more detail the premiums based on product diversity, and more specific security postures. In this paper we consider one aspect of risk (type of loss) to elaborate the methodology while drawing upon what is sparsely available data on virus losses (ICSA labs). Finally, as future research we hope to attempt to integrate other correlated risks using the process approaches with the actuarial approach for a comprehensive model.

References

"A Chronology of Data Breaches," 2005, available at: http://www.privacyrights.org/ar/ChronDataBreaches.htm.

"2006 Disclosures of U.S. Data Incidents," 2006, available at: www.privacyrights.org/ar/ChronDataBreaches.

Anderson, R. "Why information security is hard: An economic perspective,," 17th Annual Computer Security Applications Conference (ACSAC) ,, New Orleans, LA, 2001.

Böhme, R. "Cyberinsurance Revisited," Workshop on the Economics of Information Security (WEIS), Harvard University, 2005.

Böhme, R., and Kataria, G. "Models and Measures for Correlation in Cyber-Insurance," Workshop on the Economics of Information Security (WEIS), Cambridge University, UK, 2006.

Cherubini U., E. Luciano and W. Vecchiato 2004, *Copula Methods in Finance*, John Wiley and Sons, West Sussex.

Clayton, D. G., 1978, "A Model for Association in Bivariate Life Tables and its Applications in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence ", *Biometrika,* Vol. 65, pp. 141-151.

Conrad, J.R. "Analyzing the Risks of Information Security Investments with Monte-Carlo Simulations," Workshop on the Economics of Information Security (WEIS), Harvard University, 2005.

Farahmand, F., Navathe, S., Sharp, G., and Enslow, P. "Evaluating Damages Caused by Information Systems Security Incidents," in: *The Economics of Information Security,* J. Camp and R. Lewis (eds.), Kluwer, 2004, pp. 85-94.

Fisher, N. I., 1997, "Copulas," in *Encyclopedia of Statistical Sciences,* Updated Vol. 1, S. Kotz, C. B. Read, abd D. L. Banks, Editors, John Wiley and Sons, New York, pp. 159-163.

Frank M. J., 1979, "On the Simultaneous Associativity of F(x,y) and x + y –F(x,y)," *Aequationes Math.* , Vol. 19, pp.194-226.

Frees E. W., and E. Valdez, 1998, "Understanding Relationships Using Copulas," *North American Actuarial Journal,* Vol. 2, no. 1, pp. 1-25.

Genest C., 1987, "Frank's Family of Bivariate Distributions," *Biometrika,* Vol. 74, pp. 549-555.

Genest C., and L. Rivest, 1993, "Statistical Inference Procedures for Bivariate Archimedean Copulas,"*Journal of the American Statistical Association,* Vol. 88, pp. 1034-1043.

Gordon, L.A., Loeb, M.P., Lucyshyn, W., and Richardson, R. "2006 CSI/FBI Computer Crime and Security Survey ", Computer Security Institute, available at: http://www.gocsi.com/forms/fbi/csi_fbi_survey.jhtml.

Gordon, L.A., Loeb, M.P., and Sohail, T. "A Framework for Using Insurance for Cyber-Risk Management," *COMMUNICATIONS OF THE ACM* (46:3) 2003.

Gumbel E. J., 1960, "Distributions des valeurs extremes en plusiers dimensions," ***Publ. Inst. Statist. Univ. Paris***, Vol. 9, pp.171-173.

Marshall A. W., and I. Olkin, 1988, "Families of Multivariate Distributions,"*Journal of the American Statistical Association,* Vol. 83, pp. 834-841.

Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., and Sadhukhan, S.K. "e-Risk Management with Insurance : A framework using Copula aided Bayesian Belief Networks," 39th Hawaii International Conference on System Sciences, Hawaii, 2006.

Nelsen R. B., 1999, *An Introduction to Copulas*, Springer-Verlag New York, Inc.

Nelsen R. B. 1995, "Copulas, Characterization, Correlation and Counterexamples," *Mathematics Magazine*, Vol. 68, no. 3 (June), pp.193-198.

Nolan J. P., 2005, *Stable Distributions: Models for Heavy Tailed Data*, Forthcoming

Ogut, H., Menon, N., and Raghunathan, S. "Cyber Insurance and IT Security Investment: Impact of Interdependent Risk," Workshop on the Economics of Information Security (WEIS), Harvard University, 2005.

Radcliff, D. "Calculating e-risk," *ComputerWorld* (35:7), Feb 12 2001, p 34.

Schneier, B. "Computer Security: It's the Economics, Stupid," Workshop on Economics of Information Security (WEIS), 2002.

Sklar, A. 1959, "functions de repartition a n dimensions et leurs merges," ***Publ. Inst. Statist. Univ. Paris***, Vol. 8, pp.229-231

Stoneburner, G., Goguen, A., and Feringa, A. "Risk Management Guide for Information Technology Systems," National Institute of Standards and Technology (NIST), 2002.