EXTREME CYBER RISKS AND THE NON-DIVERSIFICATION TRAP

Martin Eling

Werner Schnell

This Version: August 2017

Preliminary version – Please do not cite or distribute

ABSTRACT

As research shows heavy tailedness and tail dependencies are two distinct stochastic properties of cyber risk. These characteristics change classical diversification results when building up portfolios form cyber risks. Our results illustrate the occurrence of and the requirements for situations where diversification is suboptimal and insurers will not diversify their underwriting cyber portfolios. This situations, also called the diversification trap (Ibragimov et al., 2009), has the potential to explain the sluggish development of cyber insurance markets and why it lags far behind expectations. Our analysis also gives clues about how the diversification trap might be overcome.

INTRODUCTION

Cyber risks are characterized by heavy tailed marginal distributions (Maillart and Sornette, 2010; Edwards, Hofmeyr, and Forrest, 2015; Eling and Wirfs, 2016) and potential tail dependencies (Böhme and Kataria, 2006; Herath and Herath, 2011; Mukhopadhyay et al., 2013). In such cases diversification of risks might not lead to the benefit one typically hopes for. While under the classical expected utility optimization and normally distributed risk, utility increases and risk decreases as a portfolio gets more diversified (as long as the risks are not perfectly correlated), Ibragimov, Jaffee, and Walden (2009), Ibragimov (2004), and Ibragimov and Walden (2007) show that in the presence of heavy tails it can be optimal not to diversify at all. Moreover, diversification not only depends on the marginal distributions

Both authors are with the Institute of Insurance Economics, School of Finance, University of St.Gallen. Martin Eling can be contacted via martin.eling@unisg.ch or +41 71 224 79 80 and Werner Schnell via werner.schnell@unisg.ch or +41 71 224 79 92.

but also on the dependency structure (Ibragimov and Prokhorov, 2016), i.e. nonlinear (tail) dependencies significantly affect the diversification outcome. From a regulatory point of view the benefit from diversification is assessed by risk measures such as value at risk (VaR) or tail value at risk (TVaR). This paper empirically shows that heavy tails and tail dependencies in cyber risks cause risk measure to increase as the portfolio gets more diversified. Besides risk measures, which are of primary interest for regulators, we also consider an expected utility framework which might more appropriately capture the incentives and behavior of decision makers in a firm.[1]

The remainder of the paper is organized as follows. Firstly, we introduce the models used to analyze diversification in the Section "Methodology" and the cyber risk data used in the Section "Data". Then the Section "Results" uses data on cyber risk in order to calibrate the models and our findings are presented. Finally, we conclude by summarizing and discussing the results and providing recommendations.

METHODOLOGY

In order to analyze diversification, we derive a model for portfolio formation that is calibrated using cyber losses. We follow Ibragimov et al. (2009) and formulate the loss of a portfolio or a single insurer X_s as the sum of the portfolio's iid components Z_i:

$$X_s = n^{-1} \sum_{i=1}^{n} Z_i \tag{1}$$

This model considers an individual insurance company and implies equal weights on each risk Z_i and constant "size of the portfolio" as n increases. We use different marginal distribution for Z_i. Since in our case some distributions are not stable[2] we cannot rely on closed form solutions as Ibragimov et al. (2009) do. Instead, we simulate samples of the form $\mathbf{z} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_m\}$ where $\mathbf{z}$ is a matrix of dimension n for the portfolio size and m for the number of simulations (here 10 million).

In order apply the utility framework, the overall loss is modeled as a degenerated mixture distribution combining a Pareto distributed Z and a scalar μ (Ibragimov et al., 2009):

$$\tilde{Z}_i = \mu(1 - I) + I \cdot Z_i, \tag{2}$$

where I is Bernoulli distributed meaning it is one with a probability q and 0 with a probability $(1 - q)$. μ denotes the premium the insurer earns and is calculated as fair premium $\mu = q \cdot E(Z)$. The premium is thus determined exogenously and not by a model of market equilibrium. Now we expand the consid-

[1] Research has shown that the standard expected utility framework does not always well describe the behavior of people (Benartzi and Thaler, 1995). We thus do not only consider classical models, but also more recent models based on prospect theory (see, e.g., Kahneman and Tversky, 1979 and 1992).

[2] If Z_i's distribution is stable, it means that a portfolio of Z_i follows the same distribution up to a linear transformation. Special cases are the Normal for a Pareto index $\alpha = 2$, Cauchy for $\alpha = 1$, and Lévy distribution for $\alpha = 0.5$ (see, e.g., Ibragimov and Walden, 2007).

erations to several insurers cooperating in a risk pool. Similar but more general as a reinsurance arrangement, in a risk pool the risk each company takes on is shared with the other pool members. While the individual insures face some capacity restriction and the number of risk they can take on is limited, it might be that each insurer does not reach the critical underwriting size required to benefit from diversification as will be shown later. However, sharing risk in a pool could help to attain the required diversification. Put differently, business not beneficial from the perspective of a single insurer might be beneficial for several insurers working together. We aggregate the individual risks $\tilde{Z}_i$ for different portfolio sizes and denote the overall pool risk as X_p:

$$X_p = s^{-1} \sum_{i=1}^{n} \tilde{Z}_i, \tag{3}$$

where s is the number of insurers in the risk pool and n is the number of risks in the risk pool. Note that the expected utility framework is not applicable if the first order moments do not exist (Ibragimov and Walden, 2007), as is the case for the data we use here (the VaR is the only applicable measure then). However, limited liabilities, a rather realistic assumption, solves this problem:

$$v(x) = \begin{cases} x & if\ x < k \\ k & if\ x \geq k \end{cases}. \tag{4}$$

Losses higher than k would lead to a default of the insurer and the claim would not be paid. On the truncated portfolio losses, we apply the power utility function:

$$u(v) = v(x)^{\beta}. \tag{5}$$

Since we assume that firms behave risk averse the utility function must be concave and therefore $\beta \in (0,1)$. Additionally, we also define the utility function based on the VaR as $u(X) = f(E(X), VaR(X))$ where the derivative with respect to the expectation is $f_{E(X)} > 0$ and with respect to VaR $f_{VaR(X)} < 0$ (see Ibragimov et al., 2009). This approach is similar to the classical portfolio approach but instead of minimizing the standard deviation while fixing μ, the VaR is minimized. However, both approaches would produce the same result if the risks were elliptical distributed.

Finally, the expected utility is approximated by the average of the number of simulations m (i.e. a Monte Carlo integration):

$$E(u) \approx m^{-1} \sum_{i=1}^{m} u(v). \tag{6}$$

The convexity of the utility function for large losses caused by the limited liability assumption is essential for the U-Shape utility curves and for the occurrence of diversification traps. However, we expect that an alternative assumption would do the same job. In this model we are going to replace the expected utility and limited liability assumptions by the prospect theory (see, e.g., Kahneman and Tversky, 1979 and 1992; and for the methodology see Prelec, 1998). We expect that the prospect theory will be able to imitate the convexity of the utility function and therefore will produce similar results. This would generalize and confirm the robustness of our findings. The aggregated value of the prospect theory

model is defined as:

$$V(x) = E_w\big(v(x)\big) = \int w(p(x))v(x)\,dx \approx \tfrac{1}{N}\textstyle\sum_{i=1}^{N}(w\big(p(x_i)\big) - p(x_i))v(x_i),$$

the value function is specified as:

$$v(x) = \begin{cases} x^{\beta};\ 0 < x < c \\ -\lambda(-x)^{\beta}; x > c \end{cases},$$

and the weighting function is specified as:

$$w(p) = \frac{p^{\gamma}}{p^{\gamma}+((1-p)^{\gamma})^{1/\gamma}},$$

where $w(p)$ is the probability weighting function, λ the loss aversion, $p(x)$ the distribution function here estimated by the empirical distribution, and β, as before, the risk aversion. The parameters are set according to Tversky and Kahneman (1992) to β=0.88, $\lambda = 2.25$, $\gamma = 0.61$. Moreover we set the reference point at the expected loss since the insurer would account for the expected value when calculating the permiums, $c = E(X)$. Since there is no closed form solution we simulate the model.

Dependent risks might further reduce the benefits of diversification. To model the dependency, we first used a Gaussian copula in order to simulate the portfolio distribution and thus assume that the dependency is linear. However, if the marginal distributions are non-normal (or more generally non-elliptical), the Person (linear) correlation is not an appropriate dependency measure since it does not capture the tail dependency (Embrechts, McNeil, and Straumann, 2002). Moreover, the correlation might not be applicable at all since the data suggest that the second order moments do not exist. Instead, we use copulas C to derive the joint distribution $F(.)$ (see, e.g., Wang, 1998):

$$F(Z_1, \dots, Z_n) = C\big(F_1(Z_1), \dots, F_n(Z_n)\big). \tag{7}$$

Different dependency structures are modeled with different copulas and parameters such as the Clayton copula (see Ibragimov and Prokhorov, 2016; Embrechts, Lambrigger, and Wüthrich, 2009; Embrechts, Nešlehová, and Wüthrich, 2009; Chen, Mao, Pan, and Hu, 2012). For the calibration of the copulas we orient ourselves at the empirical analysis conducted by Böhme and Kataria (2006).

DATA

The data considered in the main part of the paper are 1'553 cyber losses between 1995 and 2014 extracted from the SAS OpRisk database.[3] For detailed description of the data we refer to Biener et al. (2015) and Eling and Wirfs (2016). In order to analyze which distribution describes the data best we compare several goodness-of-fit statistics for several widely used distributions in Table 1 (a broader comparison with more distributions is provided in Eling and Wirfs, 2016).

Table 1
Goodness-of-Fit

	LogLik	AIC	BIC	KS	AD
Normal	-11'619.77	23'243.55	23'254.24	0.46	-
Lognormal	-4'498.01	9'000.03	9'010.72	0.08	17.26
Generalized Pareto	-4'461.65	8'929.30	**8'945.35**	0.07	7.24
Peak over threshold (PoT)	-4'456.43	**8'922.85**	8'949.59	0.06	11.05

Note: LogLik stands for the logarithmic likelihood of the maximum likelihood (ML) estimation, AIC for the Akaike information criterion, BIC for Bayesian information criterion, KS for the Kolmogorow-Smirnow test, and AD for the Anderson-Darling test. The POT approach slices a lognormal body and a Pareto distribution from the 80% quantile upwards.

Based on the goodness-of-fit-statistics we find that the generalized Pareto distribution and the POT approach fit the data best. The estimated Pareto index for the generalized Pareto distribution is 0.62 and for the POT approach it is 0.81.[4] We thus can confirm that cyber risks are indeed heavy tailed and the expectation and variance do not exist (see, e.g., Neslehová et al., 2006).

Illustrating the tail dependencies is more difficult because of the lack of data and analyses. Many experts claim that cyber risks are correlated, e.g. because all companies are using the same software systems. But so far only little empirical evidence exists. A few papers from the IT domain discuss potential dependencies between cyber risk (Böhme and Kataria, 2006; Herath and Herath, 2011; Mukhopadhyay et al., 2013), but to our knowledge there is no study that empirically analyses the existence of dependence between potential cyber losses, and such a dependence exists, how it looks like. For this reason different potential dependency structures will be considered in our empirical part.[5]

[3] In the main body of the text we focus on the 1,553 cyber risk losses which are also considered by Eling and Wirfs (2016). As a robustness test we also we analyze in Appendix A a frequently considered data set on data breaches (e.g. Maillart and Sornette, 2010; Edwards et al., 2016) provided by the Privacy Rights Clearinghouse (PRC, 2017).

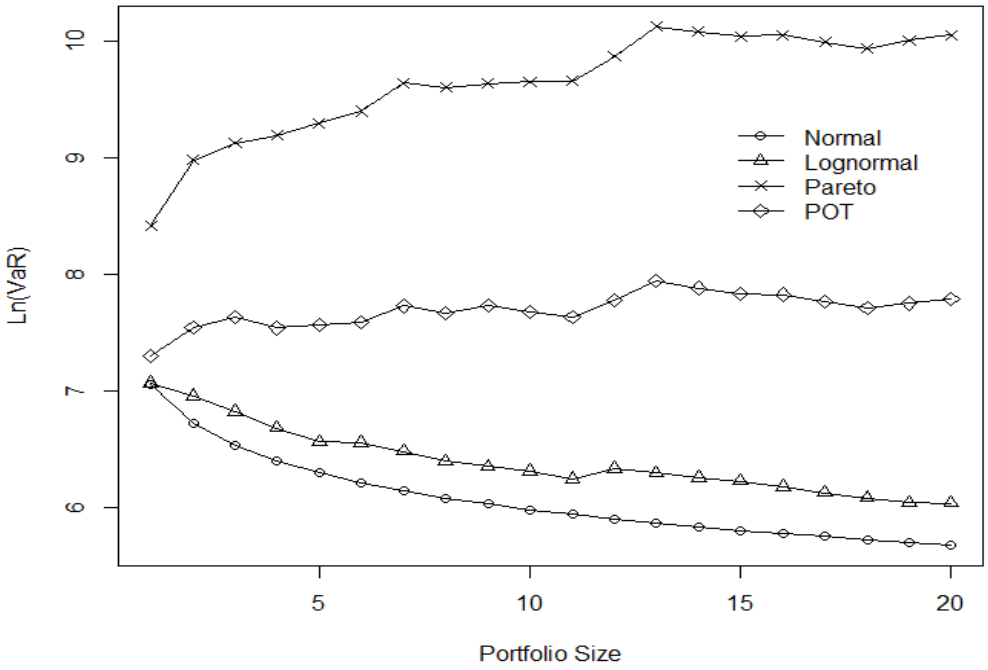
[4] The Pareto (or tail) index is the exponent α in $P(Z > z) = h(z)z^{-\alpha}$ where $h(z)$ is a slowly varying function (see, e.g., Neslehová et al., 2006). Here we define a distribution with a Pareto parameter $\alpha < 2$ as heavy tailed where the moments of order two and higher do not exists and if $\alpha < 1$ as extremely heavy tailed where moments of order one and higher do not exist (see Ibragimov and Prokhorov, 2016).

[5] Only Böhme and Kataria (2006) consider a data set (the number of potential attacks measured by honeypots), but they do not consider loss data and focus on the t-copula to capture potential tail dependencies. Herath and Herath (2011) model potential dependencies by Archimedean copulas (Clayton and Gumbel), while Mukhopadhyay et al. (2013) use Gaussian copula and linear correlations.

RESULTS

First we investigate the effect diversification has on the value at risk (VaR). Figure 1 shows the result for the $VaR_{0.995}(X)$ for different marginal distribution assumptions and depending on the degree of diversification n.

Figure 1
Diversification and VaR (independent)

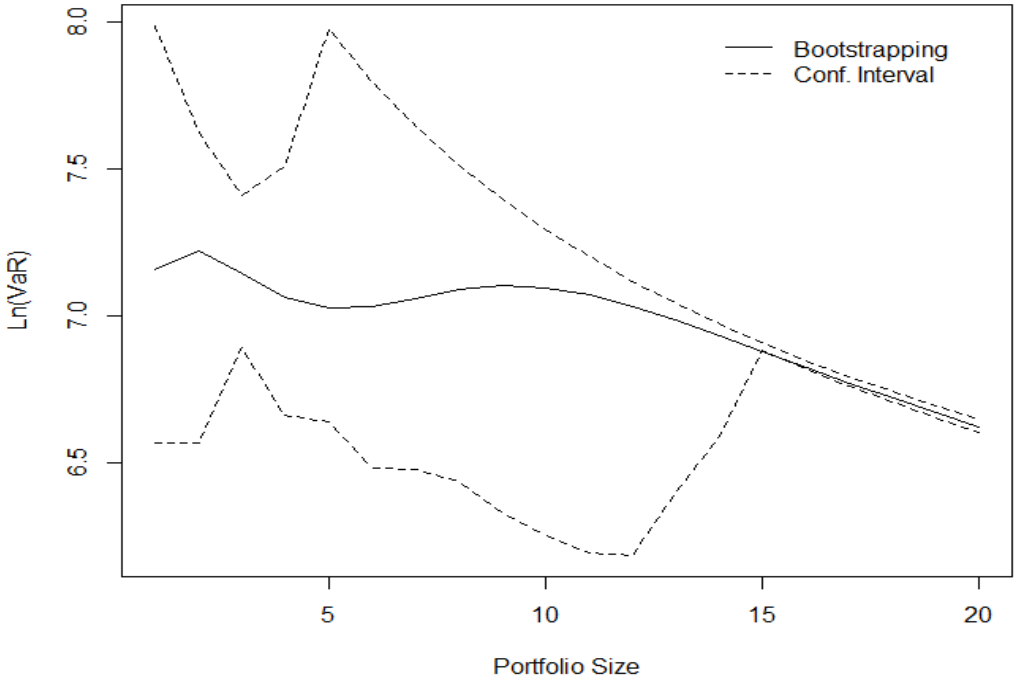


Note: Normal, Lognormal, Pareto, and POT stand for the distribution assumption used for Z_i.

The classical diversification result adapted to VaR is represented by the monotonically decreasing function of the normally distributed risks. Similarly, the VaR for lognormal risks also decreases but at a slower rate. Thus, the lognormal distribution, used in insurance practice and in regulatory models, also shows some degree of diversification.

We also use bootstrapping to simulate the VaR for different portfolio sizes. For the bootstrapping we draw directly from our original sample instead of the different distributions assumed above. The sample is drawn with replacement and is of equal size as the original data set (m=1'553 observations). Moreover, we calculate the confidence interval by repeating the bootstrapping itself. Figure 2 shows the bootstrapped VaR and its confidence interval.

Figure 2
Diversification and VaR (independent)



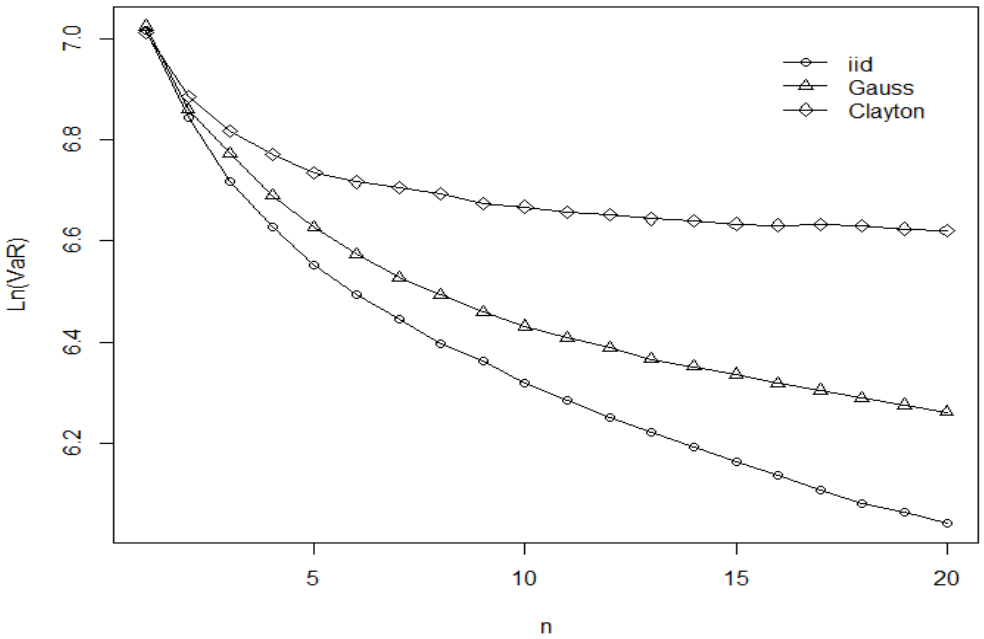
Note: The bootstrapping is based on the empirical distribution of Z_i. The dashed lines mark the 95% confidence interval of the bootstrapped VaR.

The bootstrapped VaR always lies above the lognormal VaR and the diversification benefit is much less prevalent than assumed by regulator. This result would be even more pronounced if the true distribution that generated our data is of a Pareto form as estimated above. In this case the VaR would actually increase and become superadditive[6] as the portfolio gets more diversified. As a consequence, not to diversify at all would be optimal from a risk management perspective. Note that the curves do not start for one risk at the same VaR. The reason for that is that the distributions are fitted to the data according to the maximum likelihood (ML) approach. Therefore, the VaR for one risk does not necessarily be the same for different distributions.

[6] The VaR of iid risk is superadditive (heavy tailed) if the Pareto index is $\alpha < 1$, additive if $\alpha = 1$ and subadditive if $\alpha > 1$ (see Neslehová et al. 2006).

Since the dependency affects the diversification results, we also simulate the VaR for different dependency structures. Figure 3 plots the VaR again as a function of the portfolio size for identical distributed risks and different copulas.[7]

Figure 3
Diversification and VaR (dependent)



Note: The identical (lognormal) distributed risk Z_i are aggregated assuming dependencies according to the Gauss and Clayton copulas. To compare the result the independent case (iid) is also plotted.

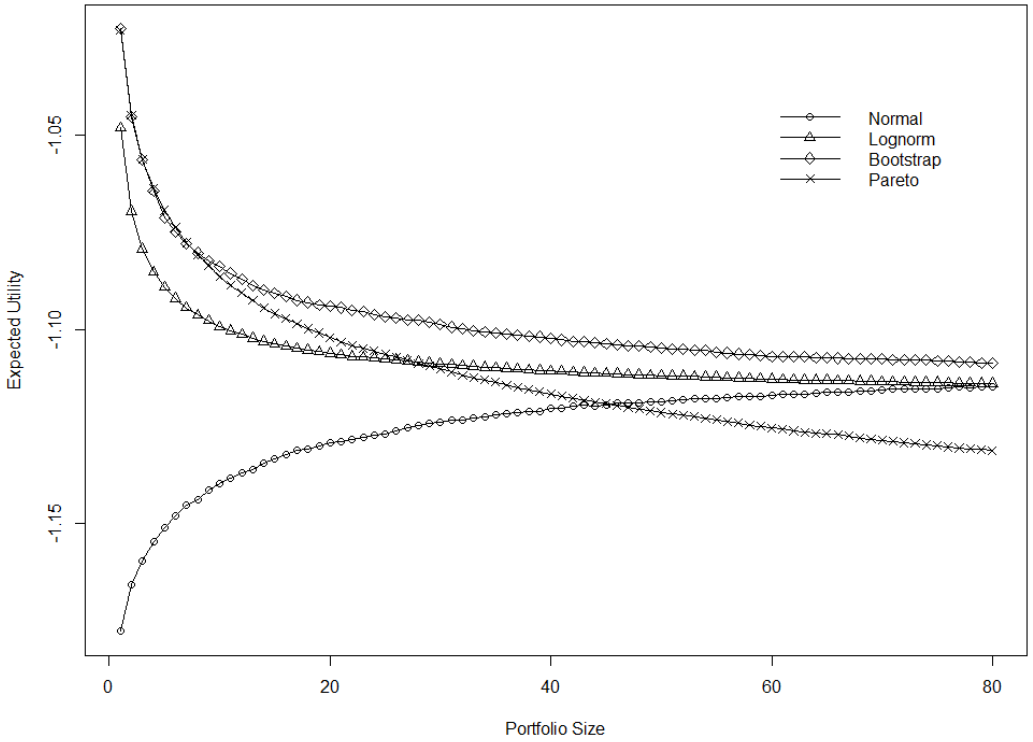
Figure 3 shows the lognormal marginal distributions combined with different dependency models. Since the VaR is decreasing for all copulas as the portfolio gets more diversified there is benefit from diversification. However, stronger dependency between the portfolio constituents would cause extreme losses to become more likely and the VaR to increases. Moreover, higher dependency in the tail as modeled by the Clayton copula increases the VaR even further.[8]

[7] The Clayton copula is calibrated to a correlation of 0.2.

[8] Note that the effect diversification has on VaR_q depends on the security level q (see Embrechts, Nešlehová, and Wüthrich, 2009). Generally, the lower q the better diversification works for VaR.

Figure 4 shows the expected utility based on a power utility function for iid risks (according to Equations 2-6 but 3 is replaced by 1).

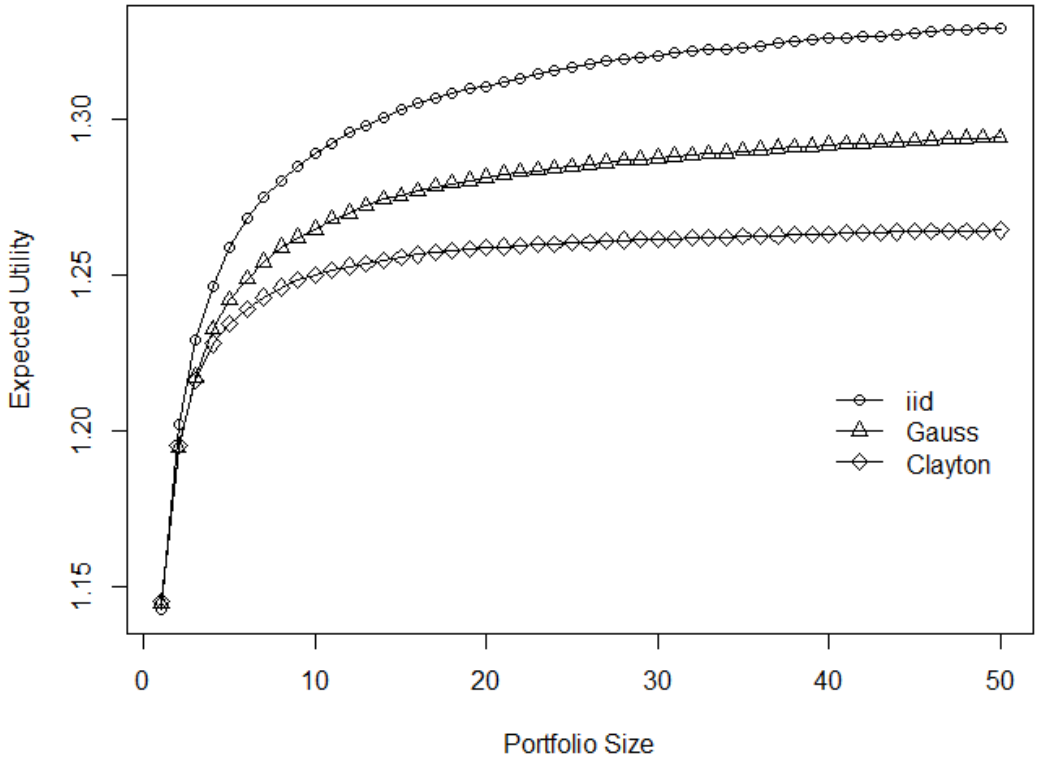
Figure 4
Diversification and Expected Utility (independent)



Note: The identical (lognormal) distributed risk Z_i are aggregated assuming dependencies according to the Gauss and Clayton copulas. To compare the result the independent case (iid) is also plotted.

As expected, for normal distributed risk we attain the classical result for diversification. However, this is not true for heavy tailed distribution such as the Pareto distribution.

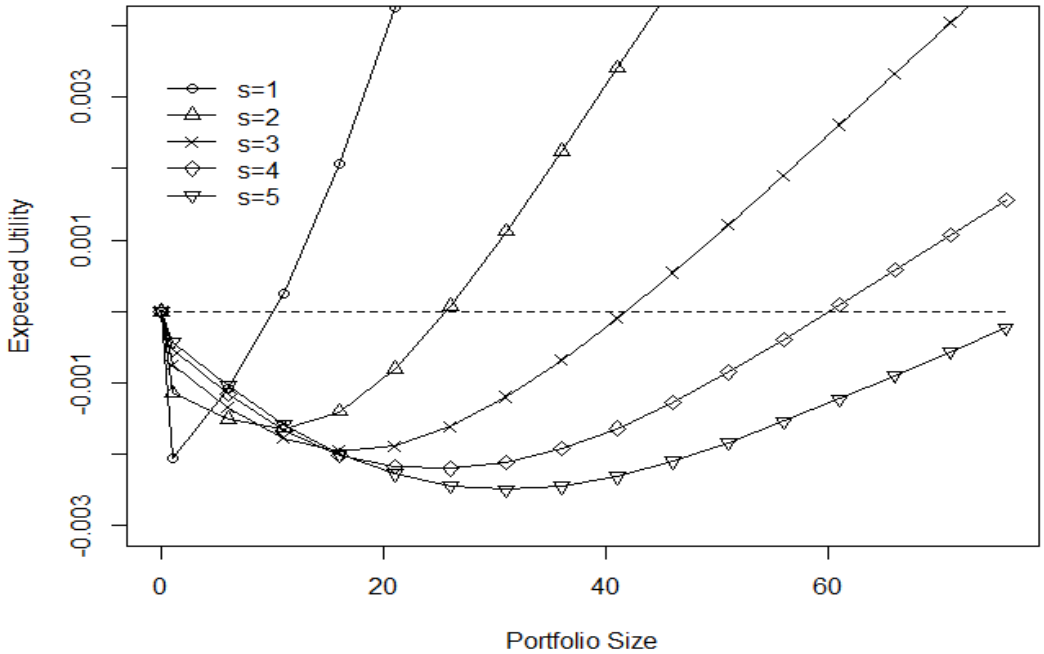
Figure 5
Diversification and Expected Utility (dependent)



Note: The identical (lognormal) distributed risk Z_i are aggregated assuming dependencies according to the Gauss and Clayton copulas. To compare the result the independent case (iid) is also plotted.

The result of the analysis described in Equations 2-6 is shown in Figure 6 for different risk in the insurer's portfolio.

Figure 6
Diversification and Expected Utility (Pareto Model)

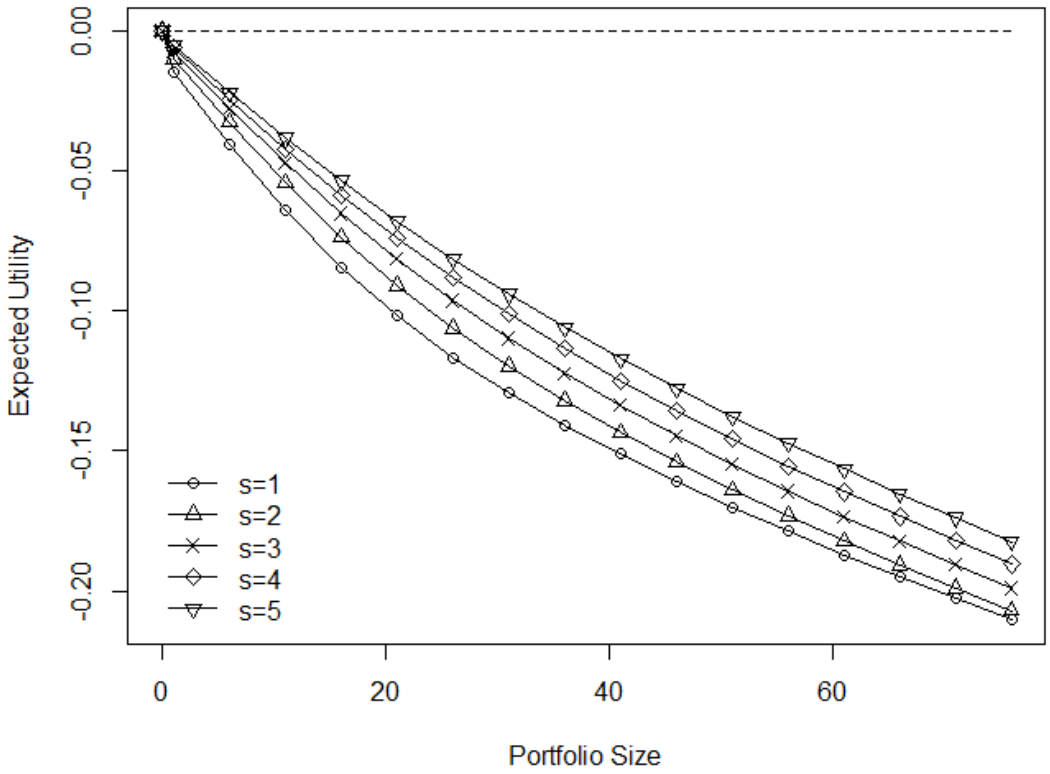


Note: For this analysis we use the following parameters: $k = 60$, $\mu = 6.6$, $q = 15\%$, $\beta = 0.0315$, and $Pareto\ index\ (\alpha) = 1$.

The figure shows that for risk with a Pareto index of 1 and limited liability, the expected utility for different portfolio sizes is U-shaped. Thus the benefit from diversification first decreases before it eventually increases again. The question is whether above a critical portfolio size the utility becomes bigger than underwriting not cyber risks at all. For three pool members the critical size would be 60 policies. Whether the market supplies zero or more than 180 polices is a question of strategic behavior described by game theory and whether there exists a reinsurance market.

The result of the analysis described in Equations 3-7 is shown in Figure 7 for different risk in the insurer's portfolio.

Figure 7
Diversification and Expected Utility (Pareto Model)

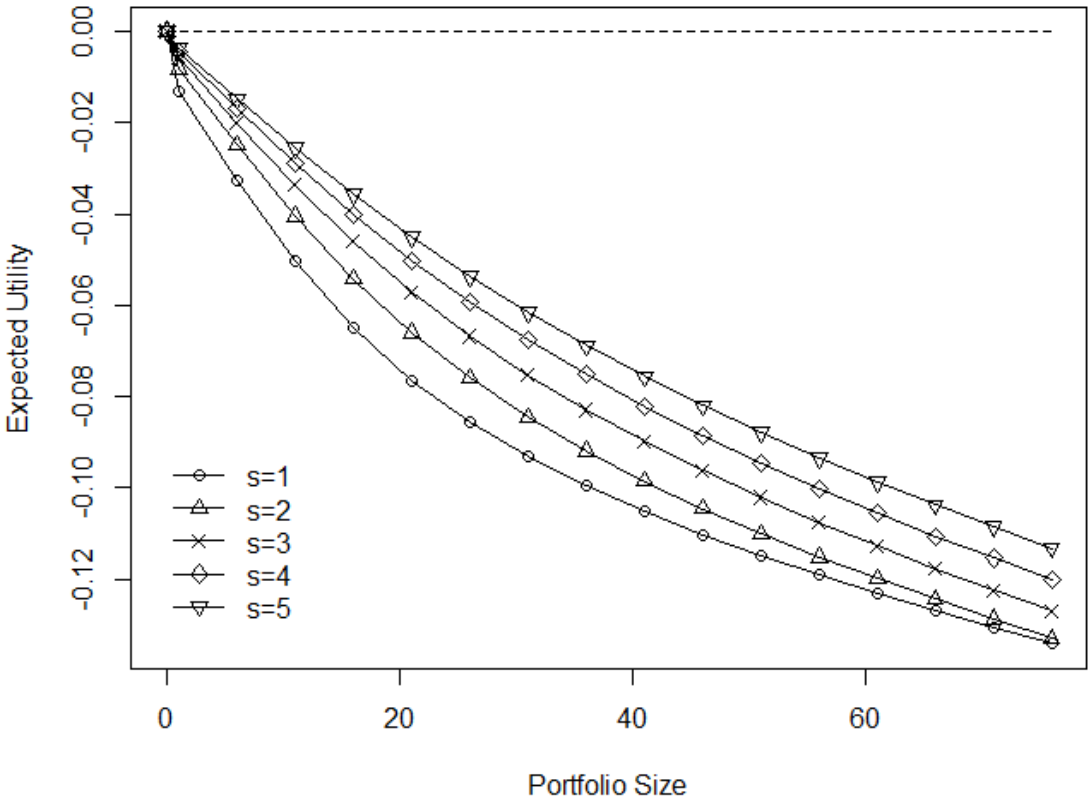


Note: For this analysis we use the following parameters: $k = 60$, $\mu = 6.6$, $q = 15\%$, $\beta = 0.0315$, and $Pareto\ index\ (\alpha) = 0.62$.

As shown, using a Pareto index of 0.62 (as estimated from the data) changes, ceteris paribus, the result completely. Since the expected utility decreases monotonically not providing any insurance would be optimal and the market would fail completely. A numerical analysis shows that the U-shape can only be observed if the tail index is in the range of (0.8, 1.2) that is similar to the findings of Ibragimov et al. (2009) for cat risk. While the situation in Figure 6 leaves room for sovereign intervention, the model in Figure 7 does not.

Figure 8 shows the same analysis for the POT model that combines the lognormal distribution for the body with the Pareto distribution for the tail.

Figure 8
Diversification and Expected Utility (POT Model)



Note: For this analysis we set a threshold at the 80% quantile, use lognormal for the body and a Pareto distribution for the tail. The parameters have been chosen as follows: $k = 60$, $\mu = 6.6$, $q = 15\%$, $\beta = 0.0315$, and $Pareto\ index\ (\alpha) = 0.81$.

Similar to the Pareto model in Figure 7 the expected utility monotonically decays for all pool sizes as the portfolio sizes increases. Therefore, it is not beneficial for insurers to supply any cyber insurance and the market fails.

Results for the prospect theory approach are yet to come.

CONCLUSIONS

This analysis shows two important aspects from a regulatory point of view. With respect to VaR, we first show that diversification does not work sufficiently well for cyber risks as measured. The regulator thus must account for that. For example, since the risk does not decrease with diversification there should be no capital discount for diversification. Or the regulator could limit the amount of underwriting risks the insurers has in its books. Second, if the market for cyber risk is in a diversification trap according to the utility framework, we showed why the market for cyber insurance completely or partially fails. As a consequence, idiosyncratic risks cannot be diversified and therefore would be relevant for the pricing. The premium charged by insurer increases and might even become prohibitive high and as

a consequence the market fails. Market failure must be met by different response of the regulator. The regulator might incentivize the use of reinsurance market, risk pools and instruments alike. As there seems to be a game theoretical coordination problem, the government could also provide help so that the cyber risk market would achieve the critical size needed for harvesting the benefits of diversification.

The limitation of this paper lays with the quality of available data and whether the data represents cyber risk in general well. The availability of better data in the future would open up new research opportunity. Moreover, your analysis could be extended by modeling the premium endogenously as the market clearing price.

In the next version of the paper we are going to analyses what effect model risk has on our diversification result and extend the analysis by the prospect theory and the mean-VaR framework. Moreover, we will use a numeric approach to analyze how sensitive our result is to different parameter combinations (e.g. how strongly can the Pareto index deviate from our estimate so that there is still a U-shaped expected utility).

APPENDIX A

The analysis for the data breaches provided by Privacy Rights Clearinghouse (PRC, 2017) is under constructions.

REFERENCES

Benartzi, S., and R. H. Thaler, 1995, Myopic Loss Aversion and the Equity Premium Puzzle, *Quarterly Journal of Economics*, 110(1): 73-92.

Biener, C., M. Eling, and J. H. Wirfs, 2015, Insurability of Cyber Risk: An Empirical Analysis, *Geneva Papers*, 40(1): 131-158.

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

Chen, D., T. Mao, X. Pan, and T. Hu, 2012, Extreme Value Behavior of Aggregate Dependent Risks, *Insurance: Mathematics and Economics*, 50(1): 99-108.

Edwards, B., S. Hofmeyr, and S. Forrest, 2015, Hype and Heavy Tails: A Closer Look at Data Breaches, Working Paper, *14th Annual Workshop of the Economics of Information Security*.

Eling, M., and J. H. Wirfs, 2016, Cyber Risk is Different, Working Paper. http://www.ivw.unisg.ch/~/media/internet/content/dateien/instituteundcenters/ivw/wps/wp160.pdf.

Embrechts, P., A. J. McNeil, and D. Straumann, 2002, Correlation and Dependence in Risk Management: Properties and Pitfalls, *Risk Management: Value at Risk and Beyond*, 176-223.

Embrechts, P., D. D. Lambrigger, and M. V. Wüthrich, 2009, Multivariate Extremes and the Aggregation of Dependent Risks: Examples and Counter-Examples, *Extremes*, 12(2): 107-127.

Embrechts, P., J. Nešlehová, and M. V. Wüthrich, 2009, Additivity Properties for Value-at-Risk under Archimedean Dependence and Heavy-Tailedness, *Insurance: Mathematics and Economics*, 44(2): 164-169.

Herath, H., and T. Herath, 2011, Copula-Based Actuarial Model for Pricing Cyber-Insurance Policies, *Insurance markets and companies: Analyses and Actuarial Computations*, 2(1): 7-20.

Ibragimov, R., 2004, Portfolio Diversification and Value at Risk under Thick-Tailedness, *Harvard Institute of Economic Research Discussion Paper #2086*. http://post.economics.harvard.edu/hier/2005papers/HIER2086.pdf.

Ibragimov, R., 2005, New Majorization Theory in Economics and Martingale Convergence Results in Econometrics, Ph.D. Dissertation, Yale University.

Ibragimov, R., and A. Prokhorov, 2016, Heavy Tails and Copulas: Limits of Diversification Revisited, *Economics Letters*, 149: 102-107.

Ibragimov, R., and J. Walden, 2007, The Limits of Diversification when Losses may be Large, *Journal of Banking and Finance*, 31(8): 2551-69.

Ibragimov, R., D. Jaffee, and J. Walden, 2009, Nondiversification Traps in Catastrophe Insurance Markets, *Review of Financials Studies*, 22(3): 959-993.

Kahneman, D., and A. Tversky, 1979, Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, 47(2): 263-292.

Kahneman, D., and A. Tversky, 1992, Advance is Prospect Theory: Cumulative Representation of Uncertainty, *Journal of Risk and Uncertainty*, 5(4): 297-323.

Neslehová, J., P. Embrechts, and V. Chavez-Demoulin, 2006, Infinite Mean Models and the LDA for Operational Risk, *Journal of Operational Risk*, 1(1): 3-25.

Maillart, T., and D. Sornette, 2010, Heavy-Tailed Distribution of Cyber-Risks, *European Physical Journal B*, 75(3): 357-364.

Mukhopadhyay, A., S. Chatterjee, D. Saha, A. Mahanti, and S. Sadhukhan, 2013, Cyber-Risk Decision Models: To Insure IT or not?, *Decision Support Systems, 56*(1): 11-26.

Prelec, D., 1998, The Probability Weighting Function, *Econometrica*, 60: 497-528.

Privacy Rights Clearinghouse (PRC), 2017, Data Breaches. https://www.privacyrights.org/data-breaches.

Wang, S., 1998, Aggregation of Correlated Risk Portfolios: Models and Algorithms, *Proceedings of the Casualty Actuarial Society*, 85(163): 848-939.