Exploring Policyholder Behavior in the Extreme Tail

A Process of Modeling and Discovery with Extreme Value Theory

By Yuhong (Jason) Xue[1]

ABSTRACT

Policyholder behavior risk should be treated as a long term strategic risk not only because its trend emerges slowly and its impact can only be recognized far in the future. More importantly, failing to recognize a change in trend can lead to inadequate capitalization or missed opportunities of strategic importance. Actuaries and risk officers have an obligation to employ more forward looking techniques to better understand and mitigate this risk. This paper demonstrates that Extreme Value Theory (EVT) can be used as such a tool to model policyholder behavior in the extreme tail, an area where judgment has been constantly applied. EVT is a mathematical theory that explores the relationship of random variables in the extremes. It is capable of predicting behavior in the extremes based on data in the not so extreme portions of the distribution. This paper applies EVT to the study of variable annuity dynamic lapse behavior in the extreme tail as an example. It illustrates a process whereby a company can fully model policyholder behavior including the extreme tail with existing data. No judgment about the extremal behavior is necessary. To this end, a variety of copulas are fitted to find the best model that describes the extremal dependency between ITM and lapse rate. The actual data combined with data simulated by the model in the extreme tail is then used to fit a dynamic lapse formula which displays different characteristics compared to the traditional methods.

1 Introduction

Firms face both short term and long term risks. During a time of uncertainty, business leaders will naturally focus more on the immediate threats: perhaps the loss of certain investments in the euro zone, the weakened market demand in a recession, or the urgent capital need in a volatile environment. Whatever the challenges may be, thanks to sound risk management, companies are able to adapt quickly and steer away from the many clear and present dangers. But is merely tactically managing short term risks enough to ensure long term sustainable success? The answer is clearly no. Of the companies that ultimately failed, the root cause was often not a failure to mitigate a short term threat. It was failing to identify a long term strategic risk, which might be a change in customers' behavior or a margin squeeze that threatens the whole industry.

[1] Yuhong (Jason) Xue, FSA, MAAA, Retirement Solutions, Guardian Life Insurance Company of America, New York, NY 10004
Email: yuhong_xue@glic.com

According to an online survey conducted by The Economist Intelligence Unit in 2011 on how companies are managing long term strategic risks, although more companies are now recognizing the importance of long term risk management, few claimed to be good at it. This is not at all surprising. Long term risk is inherently unknown, slow emerging, and does not easily lend itself to available analytical tools. For life insurance companies, one such long term risk is how policyholders will ultimately behave in extreme market conditions.

For years, actuaries have observed that policyholders behave differently in different markets. Holders of variable annuity policies, for example, tend to surrender less frequently in a prolonged down market than they would in a normal market. To account for this, dynamic behavior is assumed in a range of applications such as pricing, valuation, hedging and capital determination, where an inverse link is established between a market indicator, i.e. In The Moneyness (ITM), and a behavior indicator, i.e. full lapse rate. However, this link is only established based on past experience; in many cases, a block of business either has never been in a severe market or there is not enough data in a severe market to draw any credible conclusions. In other words, how policyholders would behave under extreme market conditions is largely unknown at this point. Nevertheless, assumptions about it are made everywhere, from pricing and reserving to economic and regulatory capital.

Assumptions or opinions differ widely from company to company in terms of how efficient policyholders would eventually be in severe markets. The price of new products and the amount of reserve and capital companies are holding are, to some extent, reflections of their opinions about the behavior in the extremes. Needless to say, getting it right in the long run is of strategic importance to companies, regulators and the industry as a whole. According to the same survey, one of the most important roles that the risk function should play is to question strategic assumptions and challenge management's entrenched view about the future. For actuaries and risk officers, what can we do to fulfill this important obligation with respect to setting policyholder behavior assumptions?

Thanks to the recent market volatility, some behavior data under stressed market conditions is beginning to emerge. In terms of existing analytical tools, however, we still don't have enough data to conclude anything about the behavior under extreme conditions. But is the newly collected data in the past few years dropping hints towards the extremes? Can we apply other tools to gain more insights? The answer is a definite yes. In this paper, we will illustrate how to apply Extreme Value Theory to shed some light onto policyholder behavior in the extremes.

Distributions in the tail are of strong interest to people not only because of the significant economic consequences that tail events can inflict on financial institutions and society as a whole, but also because random events tend to correlate with each other very differently in the tail than in the normal range of distribution, which can greatly soften or exacerbate the financial impact. Understanding tail distribution seems like a daunting task since there is only limited data in the tail and even fewer or none in the extreme tail. However, mathematicians have discovered an interesting fact: for all random variables, as long as they satisfy certain conditions, the conditional distribution beyond a large threshold can be approximated by a family of parametric distributions. The dependence structure of two or more random variables once they are all above a large threshold can also be approximated by a family of copulas. This is called the Extreme Value Theory. It suggests that if you can define a large enough threshold and still have enough data to find a good fit for a distribution of the EVT family, then the extreme tail can be described by the fitted distribution. In other words, EVT has the power of predicting the extreme tail based on observations in the not so extreme range.

Researchers started to apply EVT to solving insurance problems two decades ago. Embrechts, Resnick, and Samorodnitsky (1999) introduced the basic EVT theory with examples of industrial fire insurance. Frees, Carrie`re, and Valdez (1996) studied dependence of human mortality between the two lives of a joint annuity policy. Dupuis and Jones (2006) explored the dependence structures in the extreme tail including large P&C losses and expenses they incur, hurricane losses in two different regions, and sharp declines of two stocks.

In this paper, we will illustrate a process of using EVT in studying the behavior data by exploring the relationship of ITM and lapse rate of a block of variable annuity business when ITM is extremely large, or in the extreme tail. By fitting a bivariate distribution only to the data that is large enough, we will take advantage of the predictive power of EVT and gain understanding of the lapse experience in the extremes.

The paper is organized as follows. Section 2 gives a brief introduction to copulas and EVT. Section 3 discusses the raw data used in this analysis. Section 4 presents the result of the EVT analysis including the fitted marginal distributions, the fitted copula that described the dependence structure, the simulated data, and the regression analysis. Finally, concluding remarks are made in section 5. All data analysis in this paper is performed using the R statistical language.

2 Introduction to Copulas and EVT

We present only the foundations of the theories here which are sufficient for the subsequent analysis. Readers can refer to Dupuis and Jones (2006) and the book: “An Introduction to Copulas” by Roger Nelson for more details.

2.1 Introduction to Copulas

A copula is defined to be a joint distribution function of standard uniform random variables:

$$C(u_1, \dots, u_p) = \Pr(U_1 \le u_1, \dots, U_p \le u_p)$$

where U_i, i = 1, . . . , p is uniformly distributed random variables and u_i is a real number between 0 and 1.

For any p random variables $X_1, \dots, X_p$ and their distribution functions $F_1(x_1), \dots, F_p(x_p)$, we can construct a multivariate distribution F by the following:

$$F(x_1, \dots, x_p) = C\left(F_1(x_1), \dots, F_p(x_p)\right)$$

Sklar (1959) showed that the inverse is true also. For any multivariate distribution function F, there exists a copula function C, such that the above formula holds.

Resulting from Sklar's theory, for any multivariate distribution, we now can write it in the form of their marginal distributions and a copula. In other words, the dependence structure of the multivariate distribution is fully captured in the copula and independent of the marginal distributions.

2.2 Extreme Value Theory

When studying the distribution of univariate large values, Pickands (1975) suggested using the Generalized Pareto (GP) distribution to approximate the conditional distribution of excesses above a sufficiently large threshold. That is, the distribution of Pr(X > u + y | X > u), where y > 0 and u is sufficiently large, can be modeled by

$$H(y) = (1 + \frac{\varepsilon(y-\mu)}{\sigma})^{-1/\varepsilon}$$

where μ is called the location parameter, σ is called the scale parameter and ε is called the shape parameter.

A similar result can be extended to the multivariate case whereby the joint excesses can be approximated by a combination of GP distributions for their marginals and a copula that belongs to the Extreme Value copula family.

One important consideration of applying EVT is the choice of the threshold. It should be sufficiently large so that the EVT distribution and copula converge to the real multivariate distribution. But it cannot be so large either that there is not enough data to provide a good fit. A reasonable choice of the threshold is a tradeoff between these two conflicting requirements.

In practice, three families of copulas are commonly used when studying extremal dependence. They are Gumbel (1960), Frank (1979) and Clayton (1978) copulas. Even though they don't belong to the EVT copula family, for dealing with very large but finite data, these copulas provide a good representation of the range of possible dependence structures. They are also easier to fit as they are represented by only one dependence parameter.

An interesting fact about extremal dependence is that it can disappear when the threshold goes to infinity. This is called asymptotic independence. The Frank and Clayton copulas both exhibit this characteristic but the Gumbel copula does not, meaning it still has strong dependence even when all random variables exceed huge numbers.

Table 1 summarizes the three copulas.

Family	Dependence parameter α	Mathematical representation	Asymptotic independence
Gumbel	α > 1	$(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$	No
Frank	α ≥ 1	$e^{[(-\ln(u))^{\alpha}+(-\ln(v))^{\alpha}]^{1/\alpha}}$	Yes
Clayton	-∞ < α < ∞	$\frac{1}{\alpha}\ln(1 + \frac{(e^{\alpha u}-1)(e^{\alpha v}-1)}{e^{\alpha}-1})$	Yes

Table 1: Summary of copulas commonly used in extremal dependence studies

3 Raw Data

The author constructed a hypothetical in-force block of lifetime GMWB business. Although it is not based on real company data, it resembles common product designs as well as general patterns of ITM and lapse experience observed by many industry surveys and studies.

The block is consists of mostly L share lifetime GMWB business issued from Jan 1999 to June 2006. The observation period is the 5 years from Jan 2006 to Dec 2011. For simplicity, we observe only the lapse behavior in the first year after the surrender charge period, or the shock lapse behavior.

There are two variations of the product designs. The first offers only annual ratchet on the lifetime withdrawal guarantee, while the second also has a roll up rate of 6% guaranteed for 10 years. ITM is defined to be: $\frac{\text{PV of future Payments}}{\text{Account Value}} - 1$

Based on the product type and their issue week, we sub-divide the block into 1,452 distinct cohorts. We observe the ITM and lapse rate in the shock lapse period for every cohort. The result is 1,452 distinct pairs of ITM and lapse rate data. In order to better analyze the data using EVT, we further process the data pair by converting the lapse rate to 1/lapse rate. For example, 10% lapse rate would be converted to 10. The ending data set is plotted in Figure 1.

The Pearson's correlation is 0.7 and the Kendall's Correlation is 0.5, both suggesting strong correlation between ITM and lapse rate. This is consistent with the visual in Figure 1.

However, our data is concentrated in the +/- 20% ITM range. It turns sparse when ITM goes beyond 20%. Perhaps it is warranted to look at data in the upper tail alone. We ranked the ITM and the reciprocal of lapse rate independently and plotted the observations when both variables are beyond their 85th and 90th percentile in figure 2. Interestingly, correlation is less obvious when it is beyond these thresholds.

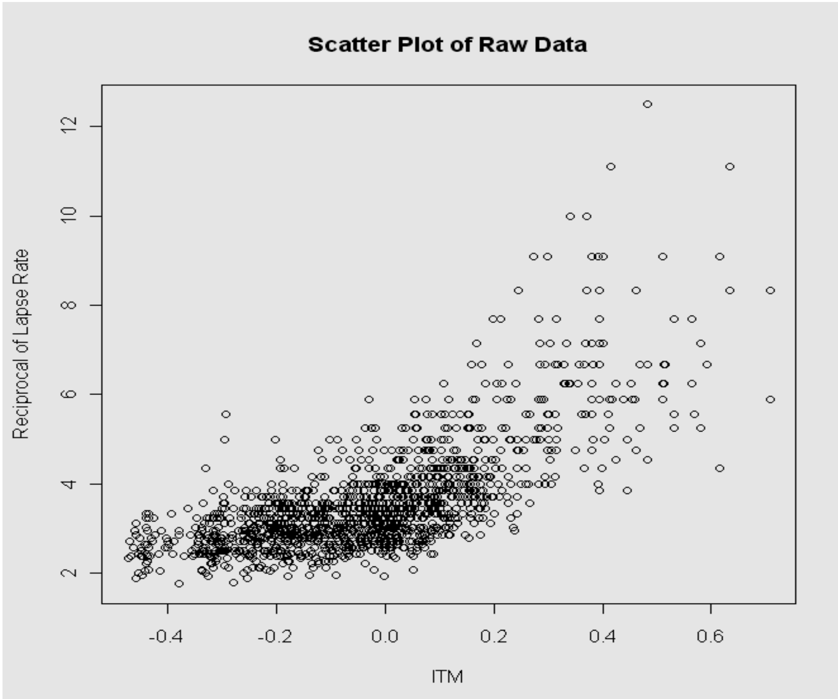
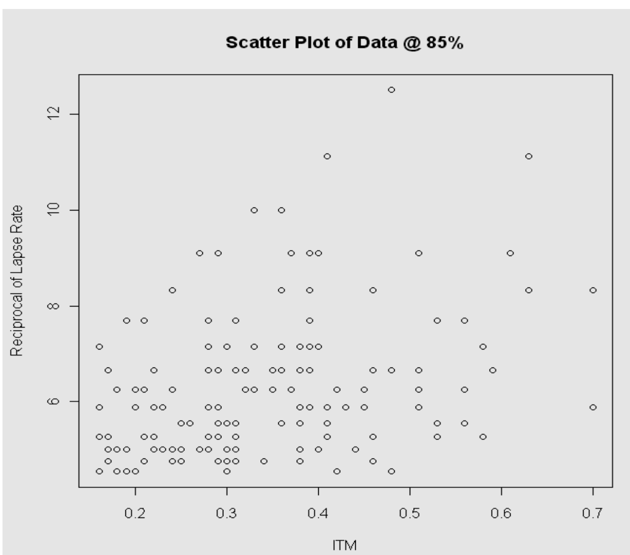


Figure 1: Scatter plot of raw data



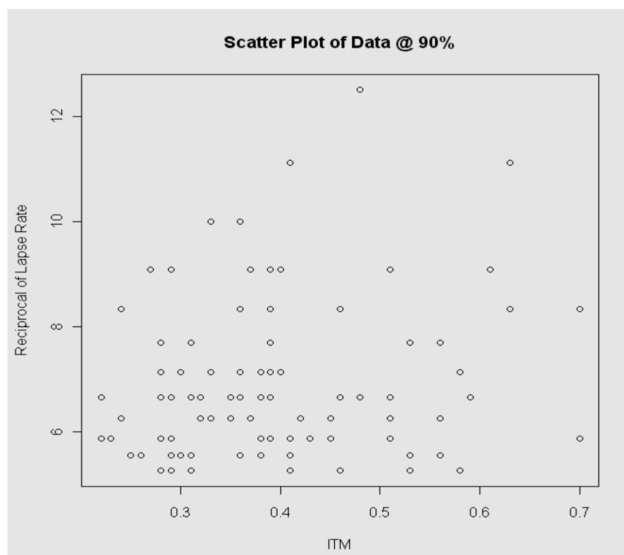


Figure 2: Scatter plot of reciprocal of lapse rate and ITM. Left: both exceed 85th percentile. Right: both exceed 90th percentile.

The distributions of ITM and lapse rate, or the marginal distributions, are also important. In figure 3 we plotted histograms of the two variables.

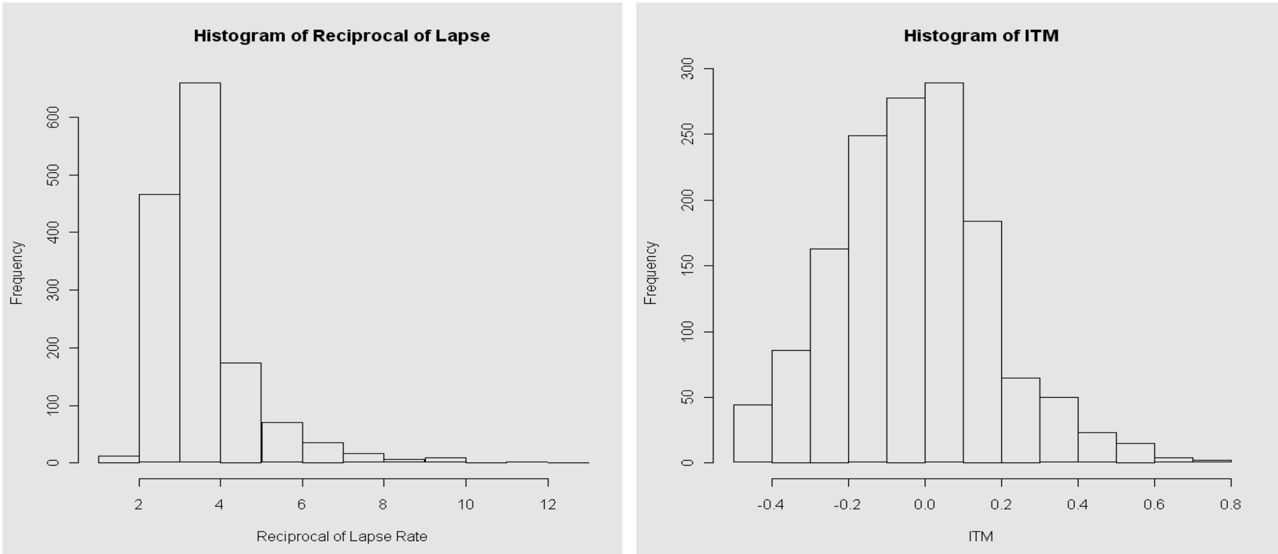


Figure 3: Left: histogram of reciprocal of lapse rate. Right: histogram of ITM

4 Result of EVT analysis

The purpose of this section is to illustrate the process of using EVT to model and explore the potential dependencies in the extremes between policyholder behavior and economic indicators such as ITM. We use shock lapse data from the hypothetical variable annuity block as an example. The goal is not to draw any definitive conclusions about the dependency. Rather, we will demonstrate the process by which one can model and discover insights into one's own data.

The process generally works as follows. We first analyze the marginal empirical distributions and define a threshold usually in terms of percentiles of the empirical marginals. Normally we would want to select a few such thresholds in order to find the largest that still leaves enough data for a good fit. We fit GP only to the marginal data that exceeds the threshold. Then using the same thresholds, we fit a number of copulas to the data exceeding the thresholds. Hopefully after this step, we can find a threshold and copula combination that provides a good fit to the tail. Now with the tail completely specified by the GP marginals and the copula, we can simulate the extreme tail data using the Monte-Carlo method. With the simulated data, we can perform analysis familiar to actuaries such as regression and stochastic calculation.

4.1 Marginal fitting

Here we fit GP to the excesses of ITM and the reciprocal of lapse rate data using the 55th, 85th and 90th percentile of their empirical distribution respectively as the threshold. Table 2 summarizes the result:

The fits were very good judging by the closeness to the empirical distributions. Figure 4 overlays the empirical density with the fitted GP density at 55th percentile threshold.

Threshold	Variable	Location	Scale	Shape
55th	ITM	-0.005	0.197	-0.193
	1/lapse	3.448	0.282	1.387
85th	ITM	0.161	0.259	-0.446
	1/lapse	4.545	1.986	-0.156
90th	ITM	0.223	0.245	-0.476
	1/lapse	5.000	2.222	-0.217

Table 2: result of fitting GP to the excess of ITM and Reciprocal of Lapse

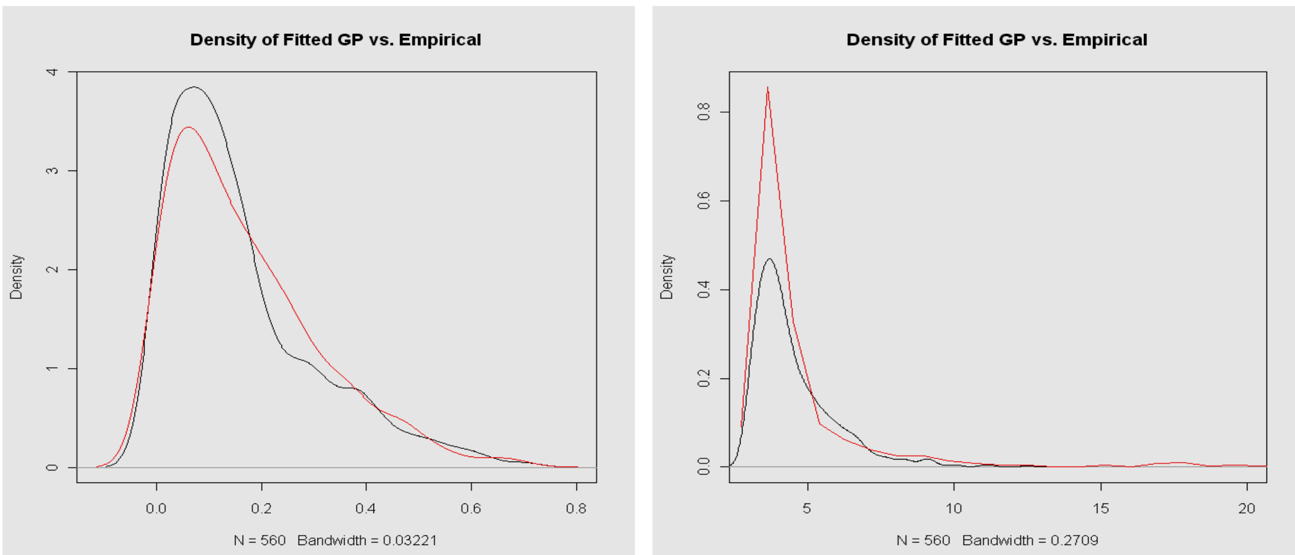


Figure 4: density functions of fitted GP (red) vs. empirical (black) with data exceeding 55th percentile. Left: ITM density. Right: reciprocal of lapse rate density.

4.2 Copula fitting

Now that the marginal distributions are fully specified with the three thresholds, we turn to model the dependence structure between ITM and lapse rate in the tail. We fit the Gumbel, Frank and Clayton copulas to data exceeding the thresholds. Other copulas are also possible. But these three capture a range of dependencies and are sufficient for our purpose.

The method of fitting is canonical maximum likelihood (CML). This approach uses the empirical marginal distributions instead of the fitted GP marginal distributions in the fitting process. It estimates the copula dependence parameter more consistently without depending on the fitted marginal distributions. Table 3 shows the results.

At the 55th percentile threshold, the Gumbel copula provides the best fit. However at the 85th percentile, the Clayton copula fits better. At the 90th percentile, there are less than 100 data points which is not enough to fit the copulas properly. Therefore, we chose the 85th percentile as our threshold and the Clayton copula to model the tail. The distribution of ITM and lapse is fully specified using the fitted Clayton copula and the GP fitted marginals for ITM and lapse exceeding the 85th percentile.

Threshold	55th		85th		90th	
Number of data pairs	560		145		95	
Copula	Parameter	Pseudo Max Loglikelihood	Parameter	Pseudo Max Loglikelihood	Parameter	Pseudo Max Loglikelihood
Gumbel	1.715	140.869	1.278	8.893	1.106	1.236
Frank	4.736	134.379	2.420	10.678	0.912	1.043
Clayton	0.801	69.952	0.601	10.881	0.148	0.531

Table 3: result of fitting copula to the empirical data.

4.3 Simulation

With the bivariate distribution fully specified above the 85th percentile, we performed a Monte-Carlo simulation and generated 500 ITM and lapse data pairs. We also did the same using the fitted Gumbel copula as a comparison. Figure 5 shows the scatter plots of the simulated data.

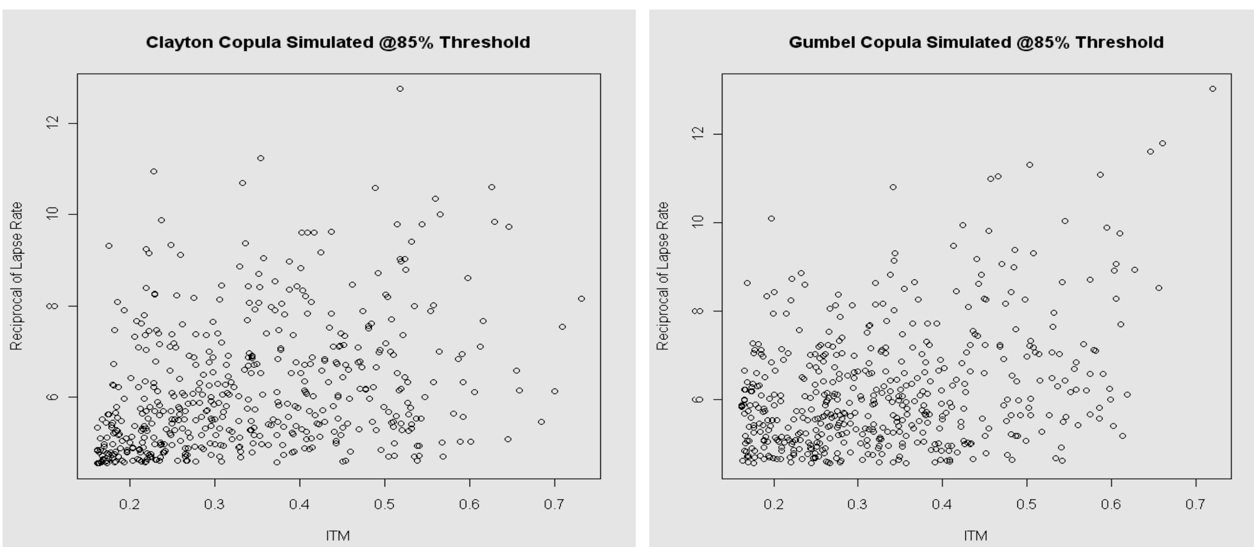


Figure 5: simulated data above joint threshold at 85th percentile by fitted bivariate distribution. Left: data simulated by fitted Clayton copula. Right: data simulated by fitted Gumbel copula

The plots on the left (Clayton) is more scattered than the one on the right (Gumbel). This is consistent with the asymptotic dependence characteristics of the two copulas. Recall that Clayton is asymptotically independent, which means the dependency of the two random variables will become weaker in the extreme tail. Gumbel, on the contrary, will show dependency in the extreme tail.

4.4 Regression analysis

Of all the statistical techniques that actuaries use, regression is one of the favorites. It explains dependence between variables reasonably well when enough data is distributed around the mean. It breaks down where data is sparse or randomness is too high. Nonetheless, the regressed function offers a very simple way of describing a dependence structure. In fact, it is commonly used in actuarial modeling such as in deriving dynamic lapse functions used to link a market variable to a multiplicative factor of the lapse rate.

We applied a common regression technique called Generalized Linear Models (GLM) to regress the combined simulated data above the 85th percentile threshold and the raw data. We did the same for the raw data alone as a comparison.

For the GLM regression, we chose the Poisson link function so that the resulting regressed function takes the form of $Multiplicative\ Factor = e^{\alpha * \left(\frac{AV}{Guar}\right) + \beta}$ where α and β are parameters to estimate, the multiplicative factor is used to multiply the base lapse rate, and the AV/Guar is the ratio of account value over the perceived guarantee value defined to be the present value of all future payments. The regressed functions are plotted in figure 6.

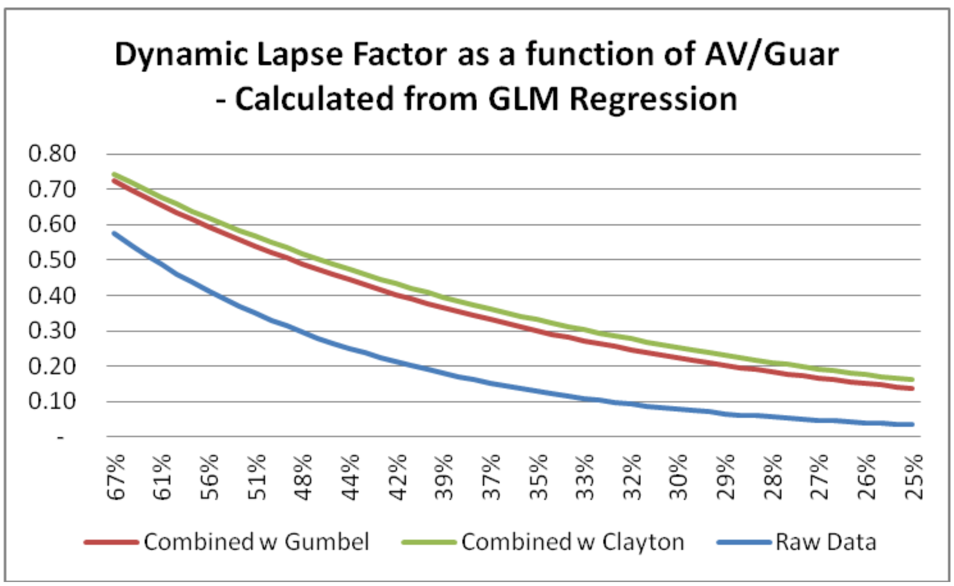


Figure 6: regression analysis using simulated data

When the raw data is combined with the simulated data generated by our chosen Clayton copula, the resulting dynamic lapse function has the highest value, suggesting it will reduce the base lapse rate the least. This implies, for example, when account value is only 25% of the guaranteed value, the lapse rate would be about 25% of the base. Repeating the same for the combined raw and Gumbel copula simulated data results in a slightly lower dynamic lapse function. This is consistent with the asymptotic dependence characteristic of the two copulas: the Clayton copula is predicting less dependence between ITM and lapse rate in the extreme tail.

If we analyze only the raw data, a dynamic lapse function that produces a lower multiplier would emerge. Because the data in the extreme tail is scarce, the regression relies on the data in the +/- 20% ITM range to define the influence of ITM over lapse rate. The low dynamic lapse function in the extreme tail is just an extension of that strong dependence observed in the +/- 20% ITM data range. However, our modeling using EVT suggests that the dependence is not necessarily that strong in the extreme tail. Perhaps it is consistent with the belief that no matter how great the perceived value of the guarantee, there will always be some policyholders who elect to surrender the policy due to other considerations.

Using the simulated data, a more accurate approach would be to model lapse stochastically. Interested readers can test if that will yield materially different results than the regression approach.

5 Summary and Conclusion

To the insurance industry, how policyholders behave under extreme market conditions has remained largely unknown. Yet companies have to make assumptions about this behavior in order to price products and determine reserves and capital. These assumptions vary widely from company to company depending on how efficient they believe their policyholders would eventually be in prolonged severe markets. In the end, only one future will emerge and getting it wrong can hurt companies' ability to execute their business strategies. A long term risk such as this one often emerges very slowly, which means if we just monitor it in the rear view mirror, by the time we noticed a changing trend, it may already be too late.

Actuaries and risk officers have an obligation to not only monitor the experience as it happens, but also to employ forward looking tools to gain insight into an emerging trend and more importantly to advise business leaders on how to mitigate the risk. This paper explores the extremal relationship of variable annuity lapse rate and ITM as an example to illustrate the process of applying EVT to model the extremal dependency between policyholder behavior and market variables. It also demonstrates the predictive power of EVT by simulating the dependency in the extreme tail. The example suggests that the dependency between ITM and lapse rate may not be as strong in the extreme tail as people might expect. Although the conclusion of this analysis should not be generalized as it can be highly data dependent, this paper shows EVT could reveal a different dynamic in the extreme tail than traditional techniques.

With the newly collected data through the recent market cycles, the industry is in a position to reexamine its understanding of extremal policyholder behavior. There could be profound implications to a wide range of applications ranging from capital determination to pricing. This process need not be limited to VA dynamic lapse study; it can be used in other behavior studies such as exploring VA withdrawal behavior. However, as illustrated in section 4, the choice of the threshold is an important consideration. One should exercise caution when applying EVT if the underlying data does not allow for a good threshold selection.

ACKNOWLEDGMENT

The author wishes to thank Elizabeth Rogalin and Michael Slipowitz for providing valuable comments for this paper.

REFERENCES

[1] Clayton, D.G. 1978. "A Model for Association in Bivariate Life Tables and its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence", *Biometrika* 65:141–51.

[2] Dupuis, D. J. and B. L. Jones, 2006. "Multivariate Extreme Value Theory and Its Usefulness in Understanding Risk", *North American Actuarial Journal* 10(4).

[3] Embrechts, P., S. I. Resnick, and G. Samorodnitsky. 1999. "Extreme Value Theory as a Risk Management Tool", *North American Actuarial Journal* 3(2): 30–41.

[4] Frank, M. J. 1979. "On the Simultaneous Associativity of $F(x, y)$ and $x \_ y \_ F(x, y)$", *Aequationes Mathematicae* 19:194–226.

[5] Frees, E. W., J. F. Carrie`re, and E. Valdez. 1996. *"*Annuity Valuation with Dependent Mortality*", Journal of Risk and Insurance 63: 229–61.*

[6] Gumbel, E. 1960*. "*Distributions des valeurs extremes en plusieurs dimensions", *Publications de l'Institut de statistique de l'Universite´ de Paris 9:171–73.*

[7] Nelson, R. 1999. "An Introduction to Copulas*" New York: Springer.*

[8] Pickands, J. 1975. "Statistical Inference Using Extreme Order Statistics", *Annals of Statistics* 3(1): 119–31.

[9] Sklar, A. 1959. "Fonctions de repartition a n dimensions et leurs marges", *Publ. Inst. Stat. Univ. Paris* 8:229–231.

[10]The Economist Intelligence Unit, 2011. "The Long View – Getting New Perspective on Strategic Risk*",* http://www.businessresearch.eiu.com/long-view.html