Insolvency Testing, Extreme Value Statistics and Resampling

Steve Craighead

Abstract

By the use of resampling and extreme value statistics we will develop a method to reduce the time and costs of testing insurance company insolvency. Most ruin models require assumptions about the surplus distribution and/or assumptions about the claims count and severity distributions. The actuary may not be comfortable in making these ruin assumptions. Instead, he or she may generate such distributions by running extensive computer simulations of corporate models. We introduce a method to reduce the number of simulations necessary to estimate the lower 5 % tail of the surplus distribution. This approach ´cuts off´ the tail at approximately the tenth percentile of the sample distribution and generates new tails for the distribution by resampling. The median and average of the new order statistics approximate the lower 5 % tail.

Résumé

A l´aide du rééchantillonnage et des statistiques de valeurs extrêmes, nous développerons une méthode pur réduire le temps et les frais qu´il faut pour tester les risques d´insolvabilité des compagnies d´assurance. La plupart des modèles de ruine demandent des suppositions en ce qui concerne le distribution de surplus ou en ce qui concerne des distributions de compte et de sévérité des demandes. Il se peut que l´actuaire ne soit pas à l´aise à choisir ces suppositions de ruine. Plutôt, on pourrat produire tels distributions par tourner des simulations nombreuses des modèles actifs/passifs. Nous introduisons une méthode qui sert à réduire le nombre de simulations qu´il faut pour estimer le plus petit 5 % de la queue de la distribution de surplus. Cette procédure "coupe" la queue à la dixième centile de la distribution des échantillons et produit des nouvelles queues pour la distribution par rééchantillonnage. La médiane et la moyenne des nouvelles statistiques d´ordre convergent vers le dernier 5 % de la queue.

Keywords

Extreme value distributions, bootstrapping, insolvency, surplus analysis, normalized distances, resampling, Kolmogorov-Smirnov test.

Nationwide Life Insurance Company, One Nationwide Plaza 1-27-01, Columbus, Ohio 43215-2220 (U.S.A.); Tel: + 1-614-249 8775, Fax: + 1-614-249 0725, E-mail: usnatl4t@ibmmail.com

Introduction

In basic ruin analysis, Bowers et al. [3] set up a stochastic process with the following assumptions:

1. Claims count distribution,
2. Claims amount distribution,
3. No interest or asset performance,
4. Constant premiums, and
5. Constant expense loadings.

Even with these simple assumptions, there is still no closed formula for the probability of ruin (i.e., when surplus drops below zero), except when the process is compound Poisson with an exponential claims amount distribution. In the life insurance industry, however, we are required by regulations or professional standards to conduct computer simulation analysis upon different lines of business to observe when business performs poorly. We model our business as accurately as possible, allowing for interest and asset performance, changing premium and expense loadings. We do not make assumptions on the claims count or amount distributions; however, we do make many other assumptions such as the term structure of interest rates, relationship of our decrements to the level of interest rates, and asset default probabilities. Computer simulations reveal our business's behavior relative to our assumption of the term structure of interest rates. Optimally, we want to calculate the probability of ruin within the accuracy of these computer models.

There are three modeling techniques to consider: parametric, nonparametric, and semiparametric. Parametric methods like those in [12] require running n simulations and fitting the results to the best parametric distribution. This method has six problems:

1. Finding the proper size of n,
2. Finding the proper distribution,
3. Finding the proper fitting algorithm,
4. Determining whether or not the fitting algorithm is data dependent,
5. Estimating the support of the fitted distribution (i.e., the range of possible values of surplus), and

6. Measuring extraneous information which parametric models introduce relative to nonparametric models. Currently there are no diagnostics that can measure this information. See [14] for a brief discussion of this problem.

Nonparametric models [16] for fitting unknown distributions have these problems:

1. The determination of the proper algorithm,
2. The misestimation of the support,
3. The need of a stopping rule (i.e., when to stop augmenting the model),
4. The distortion in the extreme tails, and
5. The increase in the size of the confidence intervals.

The increase in the size of the confidence intervals is most likely related to the inability to measure extraneous information.

Semiparametric methods force a structure upon the model, but usually do not require estimates of specific parameters. The best example of this concept in Actuarial Science is the Whittaker-Henderson graduation technique, wherein Whittaker chose his distribution of true values "by analogy to the normal frequency law." However, he did not assume the distribution to be normal [11, 19]. Semiparametric methods have problems similar to parametric ones, such as:

1. Determining the proper size of *n*,
2. Determining the choice of the best distributional structure for the model,
3. Determining the choices of the best fitting algorithm,
4. Determining whether the fitting algorithm is data dependent or not,
5. Determining the support, and
6. Measuring extraneous information.

The resampling technique described in Section 3 is best described as semiparametric. However, it addresses the support misestimation problem.

In addition to the issue of choosing an appropriate modeling technique, there remains the challenge of properly selecting the level of confidence. That is, if estimating an extreme percentile like 1% or 0.1%, should the confidence level be at 90% or 95%, or should it be comparable to the level of the percentile, like 99.5% or 99.95%?

In my company research, I have been restrained to a nonparametric setting with a high level of confidence in the left tail. Note the following formula from [9]:

$$Pr(Y_k < \xi_p) = \sum_{w=k}^{n} \frac{n!}{w!(n-w)!} p^{w}(1-p)^{n-w},$$

where Y_k is the k'th ascending order statistic (i.e., $Y_1, Y_2, \ldots, Y_n$, *where* $Y_1 \leq Y_2 \leq \ldots \leq Y_n$), and ξ_p is the p'th percentile. Using this formula, we found that estimating ruin at the 0.1% level at a 99.95% left tail confidence level required n = 10,000. See Appendix I in [5] for the program that implements the above formula.

Using 10,000 interest rate scenarios for multiple lines of business, even though very accurate, is expensive, time consuming, hard to organize, and not as succinct as parametric ruin theory. In this paper, I introduce a method to analyze surplus distributions with fewer (n=1,000) simulations, giving rules of thumb that help determine whether the 1,000 trials are sufficient.

Klein [10] noted that changing the underlying distribution of the term structure produces a dramatically different surplus distribution. We will not address this complex issue in this paper, other than to note that the term structure which generates the surplus results in Section 5 is consistent between the different lines of business in a given year of study.

The following empirical analysis will be conducted on 107 different surplus distributions and 15 statistical distributions. The structure of this data, and the parameters for the 15 statistical distributions are described in Section 4, and 5. There are 10,000 samples for each of the distributions used.

Section 1 briefly describes resampling and discuss extreme value statistics in detail. Section 2 describes visual and statistical tests which allow the analyst to estimate the specific extreme value distribution of the surplus distribution. Section 3 discusses the semiparametric resampled extreme value bootstrap (REV) technique. Section 4 the REV technique is examined on some known parametric distributions. Section 5 studies fiftyone different lines of business and compares the REV estimates with the 10,000 trial empirical distributions for each of these lines of business. This section also develops three rules of thumb for the proper use of REV. In Section 6, we conclude with a discussion of results, other applications, and further research.

Section 1. Resampling and Extreme Value Statistics

The use of bootstrap resampling is very flexible and can be applied to a broad class of problems. See [6] for a good introduction to its use. The bootstrap allows you to sample a random variable without specifying a parametric model for the distribution. One constraint on this resampling is that the original samples must be independent and identically distributed (denoted iid) samples.

All statistical analysis below is constructed to analyze the right tail of a distribution. In our insolvency studies, we desire to study the left tail. In Section 5, we convert the left tail problem into right tail by multiplying the surplus values by -1.

This section is of a theoretical nature. The key points that you should gain from this section if you do not wish to study it in depth are the following:

1. Understand the basic definition of order statistics, and their relationship to quantile estimation..

2. Understand the basic properties of normalized distances where these distances are iid exponential distribution. See Theorem 4 and its surrounding discussion.. Also note the selection of the choice of k, where I decided where the tail begins.

3. Realize that there are only three types of tails for statistical distributions. These types are exponential, heavier than exponential, and lighter than exponential. We will use a nomenclature D(G) to indicate these types of distributions. Here D(G) = 3 for exponential, D(G)=2 for lighter, and D(G)=1 for heavier.

Theorem 1 through Theorem 3 is mostly involved in the discussion of some of the underlying behavior of the tails of distributions and Theorem 3 is actually a form of an existence proof that allows the use of the REV technique that is developed in Section 2.

Now, let $X_{1n} \geq X_{2n} \geq \ldots \geq X_{nn}$ denote the descending order statistics of a sample of size n from a population with CDF F(x). The CDF F(x) is said to be in the domain of attraction of the distribution G(x), denoted D(G), if the following theorem holds for F(x):

Theorem 1. (Central Limit Theorem for Extreme Value Statistics) If there exist real numbers $a_n > 0, b_n, \; n = 1, 2, \ldots$ such that for all real x

$$\lim_{n \to \infty} F^n((x - b_n)/a_n) = G(x)$$

and if the above limit is nondegenerate then G(x) takes on one of three functional forms (location and scale parameters aside):

$$G_{1,\alpha}(x) = \begin{cases} 0, & x \le 0 \\ exp(-x^{-\alpha}), & x > 0 \end{cases}$$

$$G_{2,\alpha}(x) = \begin{cases} exp(-(-x)^{\alpha}), & x \le 0 \\ 1, & x > 0 \end{cases}$$

$$G_3(x) = exp(-exp(-x)), -\infty < x < \infty.$$

where $\alpha > 0$.

Frechet found $G_{1,\alpha}$ in 1927, and Fisher and Tippet found the other two in 1928. Gnedenko proved the above Central Limit Theorem in 1943. After Gumbel used the three distributions [8], later literature has tended to refer to them as the Gumbel Type II, III, I distributions, respectively. Our naming convention follows that of Falk [7]. See [4, 7, 8, 15 , 21] for further discussion. Restated, the theorem says that as the number of samples approaches infinity, the distribution of the largest order statistic is either degenerate or one of the three above distributions. Gnedenko stated that $b_n = F^{-1}(1 - 1/n)$ and $a_n = F^{-1}(1 - e^{-1}/n) - b_n$ (See [15]).

Using the notation from Falk [7], below we use the relation x =:y to indicate that the symbol y is defined to be the expression x on the left-hand side.

The generalized Pareto distribution is defined as

$$\begin{cases} 1-x^{-\alpha}=:W_{1,\alpha}(x), & x\geq 1 \\ 1-(-x)^{\alpha}=:W_{2,\alpha}(x), & -1\leq x\leq 0 \\ 1-\exp(-x)=:W_3(x), & x\geq 0 \end{cases}$$

Note that $W_{1,\alpha}(x)$ is the standard Pareto distribution, $W_{2,1}(x)$ is the uniform distribution on $[-1,0]$, and $W_3(x)$ is the standard exponential distribution. Denote the three distributions collectively as $W(x)$.

Also, letting $\alpha > 0$, and $G = G_{1,\alpha}, G_{2,\alpha}, G_3$, then note that $1+\ln(G(x))=W(x)$. Also note the following relationship:

$$1+\ln(G(x)^{1/n})= \begin{cases} W_{1,\alpha}(n^{1/\alpha}x), & x\geq n^{-1/\alpha}, \\ W_{2,\alpha}(n^{-1/\alpha}x), & -n^{1/\alpha}\leq x\leq 0, \\ W_3(x+\ln(n)), & x\geq -\ln(n), \end{cases}$$

$$=:W_{(n)}(x).$$

We will call the $W_{(n)}(x)$ a shifted generalized Pareto distribution.

Theorem 2. (Distribution Tail Classification Theorem) $F\in D(G)$ iff its upper tail can be approximated in an appropriate way by a shifted generalized Pareto distribution, i.e., $F((x-\hat{b})/\hat{a}) \approx W_{(n)}(x)$ for some shift parameter $\hat{b}$ and scale parameter $\hat{a}$, as n approaches infinity.

Proof: See Falk [7].

To summarize, we can classify the tails of all distributions that satisfy Theorem 1 (up to scale and shift parameters) as either Pareto, exponential or $W_{2,\alpha}(x)$ with compact support.

See [5] for further discussion of various attempts to use the above to approximate the extreme quantiles of F(x).

The following technical theorem guarantees that the joint distribution of the tail order statistics converges to one of three forms. This joint distribution will be simulated in Section 3 using resampling. Dwass generalized the Central Limit Theorem in 1966, and Weissman proved it in 1977.

Theorem 3. (Generalized Central Limit Theorem) $(X_{1n}-b_n)/a_n$ converges in distribution to an extreme value distribution G iff for any $k \in N$ the ordered n'tuple,

$$((X_{1n}-b_n)/a_n,(X_{2n}-b_n)/a_n,\ldots,(X_{kn}-b_n)/a_n),$$

converges in distribution to $G^{(k)}$. Here $G^{(k)}$ has the k-dimensional Lebesgue-density

$$g^{(k)}(x_1,x_2,\ldots,x_k) = G(x_k)\prod_{i=1}^{k}G'(x_i)/G(x_i),$$

where $x_1 > x_2 > \ldots > x_k$ and is zero elsewhere.

So $G^k \in \{G_{1,\alpha}^{(k)}, G_{2,\alpha}^{(k)}, G_3^{(k)}\}$ is the only limit for the joint distribution of the k largest order statistics, standardized with the same location and scale parameters (the a_n and b_n don't have to be the same as given by Gnedenko). See [7] for a proof.

Weissman [18] proved the final theorem that we will need:

Theorem 4. (Normalized Spacing Theorem) For F in the domain of distribution $G_3(x)$, and for fixed k, as $n \to \infty$, the normalized spacings $ia_n^{-1}(X_{in}-X_{(i+1)n})$ $(i = 1,2,\ldots,k)$ are asymptotically jointly distributed as independent standard exponential random variables.

Reiss [13] discusses the appropriate value of k: "We see that in both cases there is a trade off between the following two requirements:

(a) k has to be large to gain efficiency,

(b) k has to be small enough to get asymptotic normality of the estimator."

This search for the proper k, which Boos [2] characterizes as where the distributional tail begins, is critical to models in [2,15, 21]. Several methods to determine k are outlined in [15, 17, 21]. Boos [2] makes an empirical estimate. My attempts to estimate the 1 and 0.1 percentiles by using several techniques outlined in [15, 21] produced inconsistent results. Therefore, I took a more empirical approach like Boos [2] and set $k = 0.1n + 2$ for all my simulations.

A discussion of applying two different Weissman estimators for a_n and b_n to determine k in [2, 15, 21] is further discussed in [5].

Section 2. Visual and Statistical Tests for Exponential Distributions

Boos [2] describes this visual test to determine the domain of attraction of a distribution's tail. Take the top n/5 sample order statistics, and plot them against $-\ln(i/(n+1))$ for $i=1,2,\ldots\left[\frac{n}{5}\right]$. If the graph appears to be a straight line, the distribution has an exponential tail attracted to $G_{(3)}^{(k)}(x)$. If the graph bends down (concave down) the distribution has compact support similar to the uniform distribution and is attracted to $G_{(2,\alpha)}^{(k)}(x)$. The tails of these types of distribution have less area than the exponential distribution and are said to be lighter than exponential tails. If the graph bends up (concave up) the distribution has a Pareto like tail and is attracted to $G_{(1,\alpha)}^{(k)}(x)$. These tails are considered to be heavier than exponential and will have more area under them than exponential.

I have found that the visual test is too subjective. The following statistical test (hereafter denoted CS Test) by Castillo et al. [4] reduces the confusion about the domain of attraction. Consider the following regions: $A_{12} = \{1,\ldots,[\sqrt{n}]\}$ and $A_{34} = \{[\sqrt{n}],\ldots,[2\sqrt{n}]\}$, where [x] denotes the greatest integer less than x. Let S_{12} and S_{34} be the slopes of the least square lines between $-\ln(-\ln((n+1-i-0.5)/n))$ and the order statistics $\{X_{in}\}$ for i in A_{12} and A_{34}, respectively. This is a correction of the formulation of the CS ratio in [5]. The quotient S_{34}/S_{12} will have large values for distributions attracted to $G_{(2,\alpha)}^{(k)}(x)$, small values (near zero) for those attracted to $G_{(1,\alpha)}^{(k)}(x)$, and midrange values for those attracted to $G_{(3)}^{(k)}(x)$. See [4] for further explanation.

The results of the visual test are displayed below in Figures 1 through 3. The parameters and slope ratios of these three examples are contained in Table 1 in Section 4.

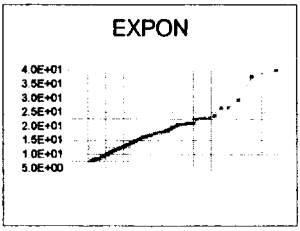

Figure 1

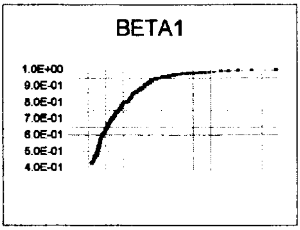

Figure 2

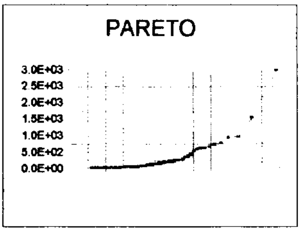

Figure 3

Section 3. **Semiparametric BootstrapTechnique**

This bootstrap technique approximates the joint distribution of the k largest order statistics, $G_3^{(k)}$, of Theorem 3 (i.e., an exponential tail). This technique essentially cuts the tail off and attaches additional tails with normalized distances based on the original tail.

We will denote this technique as Resampled Extreme Values (REV).

Here, define the vector of normalized distances, $\mathbf{d} = \{d_i\}$, of the k+1 largest order statistics to be

$$d_i = i(X_{in} - X_{i+1\,n}) \qquad i = 1, \ldots, k$$

Realize that this definition of normalized distances is different from Theorem 4. By Theorem 4, the $\{d_i\}$ are approximately iid exponential random variables when n is much larger than k. Because a_n is unknown and constant across all the distances, we can remove a_n to perform the simulation. Let $\{d_1^*, \ldots, d_k^*\}$ be a bootstrap resample of size k drawn with replacement from **d**. (Here just choose a sample of k discrete uniformly distributed random numbers between 1 and k inclusive, and use those random numbers as the subscripts for elements of the set $\{d_i\}$). Now, one defines the bootstrap order statistics $X_{1n}^* \geq X_{2n}^* \geq \ldots \geq X_{kn}^*$ of the k largest order statistics by

$$X_{jn}^* = X_{k+1\,n} + \sum_{i=j}^{k} i^{-1} d_i^* \qquad (j = 1, \ldots, k).$$

Here $X_{jn}^* = X_{jn}$ for $j \geq k+1$. The collective $\{X_{jk}^*\}$ simulates a sample from the joint distribution of the order statistics of the tail discussed in Theorem 3. Note also that each individual X_{jk}^* represents the $\frac{(n-j+1)}{n}$ quantile of the sample distribution. The collective $\{X_{jk}^*\}$ also represents a new tail being attached to the truncated distribution.

Notice here that we only need the normalized distances **d** to be iid. REV makes no assumption that the tail is approximately exponentially distributed. That is why REV is classified as semiparametric. See [15, 21, 22] for a further discussion of this resampling technique.

By ordering the results of our bootstrap samples, we are able to inexpensively collect information about the behavior of the tail of the underlying true distribution. Now we have ordered <u>ordered statistics</u>! Shades of recursion! The median or the mean of these ordered <u>ordered statistics</u>, in the bootstrap regime, is a good estimate of the underlying true tail. Also, the other ordered ordered statistics allows one to create nonparametric confidence intervals about the mean or median at each quantile.

In the data analysis of Sections 4 and 5, all sample populations have 1,000 different samples, and their tails start at k=102. We resample the sample distributions' tails 500 times each, and determine the median and the mean of the resampled X^*_{ik}, for i = 1..50. The choice of i = 1..50 simulates the upper 5% of the distribution.

Truncated distributions in general take on the form $F(x|y<X\leq z) = \frac{F(x)-F(y)}{F(z)-F(y)}$, where F(x) is the cumulative distribution. When examining a truncated left tail distribution, $y=-\infty$, and F(y)=0. When examining a truncated right tail distribution, $z=\infty$, and F(z) = 1.

By using a Kolmogorov-Smirnov (KS) test, the truncated resampled 5% tail is compared to the truncated 5% tail of the corresponding approximate population distribution. Note that truncated distributions fail the KS test more frequently because there are fewer points, and the maximum absolute deviation is magnified due to the division by the truncation percentage.

The approximate population distribution is a sample distribution of 10,000 samples instead of 1,000. If the median or mean tail estimates of the X^*_{ik} were a perfect match to the 10,000 trials, then each estimate would correspond to an order statistic $Y_{10i\ 10000}$ of the 10,000 trial populations. However, in reality, the median and the mean of $\{X^*_{ik}\}$ will be between $Y_{j\ 10000}$ and $Y_{(j+1)\ 10000}$ for some j. So to approximate the quantile of the 10,000 sample distribution to which the resampled quantile is compared, we let $\overline{X_{ik}}$ represent either the median or the mean of the $\{X^*_{ik}\}$ for i = 1..50. Using linear interpolation, the quantile of the median or the mean of the $\{X^*_{ik}\}$ is

$$\frac{j+\dfrac{\overline{X_{ik}}-Y_{j\ 10000}}{Y_{(j+1)\ 10000}-Y_{j\ 10000}}}{10000}$$

in the 10,000 sample population. These quantiles are then used in the KS comparison between the resampled tail and the approximate population distribution.

Section 4. Known Parametric Distributions

Initially, we compare REV on fifteen separate parametric distributions. Table 1 lists the distributions, their parameters, their CS ratios, their domain of convergence D(G), and the KS probabilities for the median and mean. The S_{34}/S_{12} interval of acceptance for Region 1 (heavier than exponential tails) is between 0 and 0.70. The Region 2 (lighter than exponential) interval is greater than or equal to 2. The Region 3 (exponential tails) interval is between 0.7 and 1.99. The column D(G) reflects these three intervals. (Note: Beta1 through Beta4 are from the Beta distribution with different parameters chosen to test different Region 2 compact support distributions.)

Name	Parameters	$\frac{S_{34}}{S_{12}}$	D(G)	Median KS Prob	Mean KS Prob
BETA1	$\alpha_1=0.2$ $\alpha_2=0.8$	10.58	2	0.0120	0.0120
BETA2	$\alpha_1=3$ $\alpha_2=3$	2.76	2	0.2178	0.3362
BETA3	$\alpha_1=5$ $\alpha_2=1.5$	2.67	2	0.0000	0.0322
BETA4	$\alpha_1=2$ $\alpha_2=0.8$	7.52	2	0.0062	0.0062
CAUCHY	$\alpha=1$ $\beta=1$	0.03	1	0.0000	0.0000
CHISQ	$\nu=5$	0.63	1	0.9978	0.9996
EXPON	$\beta=5$	0.90	3	1.0000	1.0000
GAMMA	$\alpha=5$ $\beta=1$	1.60	3	0.9999	1.0000

Name	Parameters	$\frac{S_{34}}{S_{12}}$	D(G)	Median KS Prob	Mean KS Prob
LOGISTIC	$\alpha=5$ $\beta=1$	1.24	3	1.0000	1.0000
LOGNORM	$\mu=1$ $\sigma=0.5$	1.02	3	0.9992	0.9990
NORM	$\mu=0$ $\sigma=1$	1.85	3	0.8327	0.8551
PARETO	$\theta=1.1$ $a=5$	0.26	1	0.0000	0.0000
STUDENT	$\nu=5$	0.81	3	0.6960	0.7086
UNIFORM	$a=1$ $b=5$	6.56	2	0.0015	0.0025
WEIBULL	$\alpha=2$ $\beta=1$	1.34	3	0.9985	0.9979

Table 1.

Notice that the region of acceptance of Region 3 incorrectly excludes the Chi Square distribution. REV passes the truncated KS test whenever the KS probability is greater than 5%. Note that REV passes the KS test for both the Chi Square and Beta2 distributions. We will see how well REV estimates some Region 1 and Region 2 empirical distributions in the next section.

Section 5. <u>Empirical Surplus Distributions</u>

Following is an application to the projected surplus of 51 different lines of business, denoted "LOBm" for m = 1,...,51. The analysis includes three year-end projections (1992-4) for each LOB. The following analysis examines the right tail of a distribution. To study the surplus values below zero, all the surplus values were multiplied by -1 to turn the left tail analysis into a right tail analysis.

The given samples represent projected accumulated surplus at year 20, discounted to the present. This variable gives a good indication of long-term line of business solvency. Since this is not a parametric fit, the following score card will compare the 10,000 trial percentiles with the resampled 5% tail where n=1,000. The region of acceptance for the domain of convergence is the same as in Section 4. The truncated KS test is passed when the KS probabilities exceed 5%.

Name	1992 $\frac{S_{34}}{S_{12}}$	1992 D(G)	1992 KS Median	1992 KS Mean	1993 $\frac{S_{34}}{S_{12}}$	1993 D(G)	1993 KS Median	1993 KS Mean	1994 $\frac{S_{34}}{S_{12}}$	1994 D(G)	1994 KS Median	1994 KS Mean
LOB01	1.22	3	0.0007	0.0007	0.97	3	0.8749	0.8624	NA	NA	NA	NA
LOB02	1.21	3	0.9987	0.9980	2.30	2	0.9763	0.9696	1.89	3	0.1047	0.1119
LOB03	1.83	3	0.9999	0.9996	0.98	3	0.9993	0.9990	1.29	3	0.9863	0.9853
LOB04	1.18	3	0.3372	0.3501	0.62	1	0.9694	0.9639	1.16	3	0.9999	0.9997
LOB05	NA	NA	NA	NA	0.54	1	1.0000	0.9999	1.29	3	0.9994	0.9992
LOB06	1.75	3	0.9362	0.9135	0.55	1	1.0000	1.0000	0.81	3	0.9997	0.9992
LOB07	0.99	3	0.9980	0.9965	1.59	3	0.9884	0.9875	2.08	2	0.3083	0.2902
LOB08	NA	NA	NA	NA	NA	NA	NA	NA	1.26	3	0.3817	0.3584
LOB09	1.27	3	0.9728	0.9783	0.83	3	1.0000	1.0000	1.23	3	0.9538	0.9350
LOB10	2.11	2	0.9997	0.9996	1.35	3	0.8574	0.9481	0.90	3	0.8586	0.7877
LOB11	NA	NA	NA	NA	NA	NA	NA	NA	1.26	3	0.9999	0.9994
LOB12	NA	NA	NA	NA	NA	NA	NA	NA	1.27	3	0.9999	0.9994
LOB13	NA	NA	NA	NA	1.53	3	0.9831	0.9879	2.07	2	0.2484	0.2380
LOB14	NA	NA	NA	NA	0.77	3	1.0000	1.0000	NA	NA	NA	NA
LOB15	1.37	3	1.0000	1.0000	1.06	3	0.9889	0.9766	NA	NA	NA	NA
LOB16	1.35	3	0.3200	0.4361	1.41	3	0.9078	0.9020	NA	NA	NA	NA
LOB17	0.81	3	0.0936	0.0881	0.79	3	0.3164	0.2783	0.94	3	0.9999	0.9993

Name	1992 $\frac{S_{34}}{S_{12}}$	1992 D(G)	1992 KS Median	1992 KS Mean	1993 $\frac{S_{34}}{S_{12}}$	1993 D(G)	1993 KS Median	1993 KS Mean	1994 $\frac{S_{34}}{S_{12}}$	1994 D(G)	1994 KS Median	1994 KS Mean
LOB18	NA	NA	NA	NA	NA	NA	NA	NA	1.24	3	0.2216	0.2552
LOB19	2.53	2	0.9998	0.9996	1.63	3	0.9999	0.9996	NA	NA	NA	NA
LOB20	NA	NA	NA	NA	0.74	3	1.0000	1.0000	NA	NA	NA	NA
LOB21	NA	NA	NA	NA	1.43	3	0.4646	0.4646	NA	NA	NA	NA
LOB22	NA	NA	NA	NA	1.43	3	0.4046	0.4370	NA	NA	NA	NA
LOB23	NA	NA	NA	NA	0.76	3	1.0000	0.9999	NA	NA	NA	NA
LOB24	NA	NA	NA	NA	1.78	3	0.1614	0.2182	NA	NA	NA	NA
LOB25	NA	NA	NA	NA	1.03	3	0.9999	1.0000	NA	NA	NA	NA
LOB26	1.42	3	0.9504	0.9285	0.77	3	0.9999	0.9999	0.94	3	0.9990	0.9982
LOB27	1.78	3	0.9921	0.9922	0.82	3	0.9998	0.9998	1.47	3	0.9931	0.9933
LOB28	1.20	3	0.9954	0.9958	0.83	3	1.0000	1.0000	1.19	3	0.9728	0.9729
LOB29	2.30	2	0.4052	0.4280	0.83	3	0.9943	0.9934	1.05	3	0.9431	0.9006
LOB30	NA	NA	NA	NA	1.35	3	0.6901	0.7270	NA	NA	NA	NA
LOB31	1.23	3	0.9334	0.9227	1.65	3	0.9942	0.9893	NA	NA	NA	NA
LOB32	1.33	3	0.5442	0.4753	1.52	3	0.9653	0.9232	NA	NA	NA	NA
LOB33	0.85	3	0.9998	1.0000	2.21	2	0.9943	0.9936	1.25	3	0.0106	0.0093
LOB34	1.09	3	0.9695	0.9897	1.15	3	0.7138	0.6612	1.33	3	0.5057	0.5108
LOB35	0.87	3	0.9914	0.9967	2.04	2	0.9991	0.9957	1.63	3	0.0805	0.0824
LOB36	0.84	3	1.0000	1.0000	1.92	3	1.0000	1.0000	1.63	3	0.9737	0.9652
LOB37	NA	NA	NA	NA	3.04	2	0.8869	0.8347	1.81	3	0.9549	0.9379
LOB38	NA	NA	NA	NA	0.75	3	0.6771	0.6775	1.05	3	0.9998	0.9993
LOB39	NA	NA	NA	NA	0.27	1	0.7308	0.6603	0.15	1	0.0000	0.0000
LOB40	NA	NA	NA	NA	NA	NA	NA	NA	0.57	1	0.9985	0.9974
LOB41	NA	NA	NA	NA	NA	NA	NA	NA	0.83	3	0.2876	0.2261
LOB42	0.74	3	1.0000	1.0000	0.73	3	0.7506	0.7466	NA	NA	NA	NA
LOB43	NA	NA	NA	NA	0.61	1	0.9561	0.9522	1.14	3	0.9996	0.9993
LOB44	1.80	3	0.9741	0.9696	0.80	3	1.0000	0.9998	1.11	3	0.9984	0.9985

Name	1992 $\frac{S_{34}}{S_{12}}$	1992 D(G)	1992 KS Median	1992 KS Mean	1993 $\frac{S_{34}}{S_{12}}$	1993 D(G)	1993 KS Median	1993 KS Mean	1994 $\frac{S_{34}}{S_{12}}$	1994 D(G)	1994 KS Median	1994 KS Mean
LOB45	2.14	2	0.9347	0.9356	0.93	3	0.9990	0.9985	NA	NA	NA	NA
LOB46	1.34	3	0.7499	0.7557	1.26	3	1.0000	1.0000	NA	NA	NA	NA
LOB47	2.05	2	0.6214	0.6121	2.55	2	0.9985	0.9943	NA	NA	NA	NA
LOB48	1.82	3	1.0000	1.0000	1.55	3	1.0000	1.0000	3.23	2	0.4459	0.3990
LOB49	0.22	1	0.0000	0.0000	0.85	3	0.9726	0.9653	0.70	3	0.9863	0.9646
LOB50	2.14	2	0.8341	0.9052	0.83	3	1.0000	1.0000	NA	NA	NA	NA
LOB51	1.67	3	0.9982	0.9964	1.34	3	1.0000	1.0000	NA	NA	NA	NA
D(G) SCORE CARD												
1 PASS	NA	NA	0.00	0	NA	NA	5	5	NA	NA	1	1
1 FAIL	NA	NA	1.00	1	NA	NA	0	0	NA	NA	1	1
2 PASS	NA	NA	6.00	6	NA	NA	5	5	NA	NA	3	3
2 FAIL	NA	NA	0.00	0	NA	NA	0	0	NA	NA	0	0
3 PASS	NA	NA	23.00	24	NA	NA	35	35	NA	NA	25	25
3 FAIL	NA	NA	2.00	1	NA	NA	0	0	NA	NA	1	1
Total			31.00	31			45	45			31	31

Table 2

Here, we see that Region 3 distributions are well estimated by REV.

In the course of this analysis, we pinpointed an error in the underlying model. For example, the misestimate of LOB01 in 1992 turns out to be an example of leverage points. See [1] for a discussion of how leverage points are handled within regression analysis. LOB01 was our first attempt at this type of corporate model, and existence of leverage points indicates an error in the computer modeling. This impact of leverage points on REV also indicates that the technique is not robust in the treatment of outliers.

The misestimate of LOB33, in 1994, indicates where REV does not represent the extreme tail well over the entire 5% tail. The test passes, however, when restricting the truncated KS test to the 2.5% tail. So REV only approximates the 2.5% tail well for LOB33.

The fact that all tests passed for all regions in 1993, I believe, was related to the fact that interest rates were very low in 1993. The ten year Treasury Bond values for 1992, 1993, and 1994 were respectively 6.77%, 5.83%, and 8.09%. In 1993, the dispersion of the surplus was tighter than in 1992 or 1994. The underlying interest rate scenarios for the corporate model were generated from a simple lognormal generator. Since the actual standard deviation of a lognormal distribution declines with the starting interest rates, the behavior of the standard deviation of the scenarios in 1993 narrowed from that of 1992 or 1994, and led to the narrowing of the overall surplus results.

When conducting the KS tests, I observed that the results varied by the choice of 1,000-samples for some specific LOBs. To measure this data dependency, I conducted multiple KS tests with 25 different 1,000-samples from each of the 122 sample distributions, totaling 3,050 tests. Table 3 shows the collective results of these tests for the 10%, 7.5%, 5%, and 2.5% resampled tails.

	2.5% median	2.5% mean	5% median	5% mean	7.5% median	7.5% mean	10% median	10% mean
Region 1 Pass	167 64.7%	167 64.7%	178 69.0%	175 67.8%	182 70.5%	178 69.0%	205 79.5%	192 74.4%
Region 1 Fail	91 35.3%	91 35.3%	80 31.0%	83 31.2%	76 29.5%	80 31.0%	53 20.5%	66 25.6%
Region 2 Pass	270 54.9%	268 54.5%	314 63.8%	319 64.8%	433 88.0%	438 89.0%	453 92.1%	457 92.9%
Region 2 Fail	222 45.1%	224 45.5%	178 36.2%	173 35.2%	59 12.0%	54 11.0%	39 7.9%	35 7.1%
Region 3 Pass	1927 83.8%	1935 84.1%	2020 87.8%	2033 88.4%	2134 92.8%	2144 93.2%	2148 93.4%	2163 94.0%
Region 3 Fail	373 16.2%	365 15.9%	280 12.2%	267 11.6%	166 7.2%	156 6.8%	152 6.6%	137 6.0%

Table 3

The smaller percentage tails influence the truncated KS test, so the overall results improve as one examines more of the simulated median or mean tail. Recall that the KS test mirrors the standard KS test except that the maximum absolute deviation is divided by the truncation percentile, thus magnifying the deviation. However, the choice of where the tail begins, as set by the variable k, has a counter influence and will create situations that may pass at the 2.5% or 5% level but will not pass at 10%. Refer to the above discussion for 1994's LOB33.

We could improve the Region 3 pass percentage by shifting its region of acceptance. A rough global maximum Region 3 interval is between 0.55 and 2.99. Using the 10% mean tail for the 3,050 KS truncated tests improved the pass percentage from 94.00% to 94.27% in the above tests. Looking at just the parametric distributions, we now correctly include the Chi Square distribution and incorrectly include Beta2 and Beta3. If the Region 3 interval is between 0.55 and 1.94, then thea local maximum pass percentage of 94.23% results. This region correctly includes the Chi Square distribution and properly excludes the two Beta distributions, which were included with the rough global maximum.

In the analysis of surplus, I have developed the following Rules of Thumb when using REV as a tool to reduce the number of trials:

Rules of Thumb.

1. All business with tails heavier than exponential (as measured by the CS ratio) should use 10,000 trials.
2. All expensive business, where overestimation would be critical, with tails Exponential or lighter, should use 10,000 trials.
3. If visual and statistical test behavior is maintained into subsequent years, additional runs may not be necessary.

Section 6. Conclusions and Further Research

REV is a very effective estimator when coupled with the CS test for Region 3 distributions. This is evidenced by the KS results at the end of the last section. The truncated KS test discussed above is a two-tailed test. We could improve the statistics by considering a one-tail test where we only test the maximum of the deviation of the resampled tail from the tail of the empirical population distribution (instead of the maximum of the absolute difference). This approach would allow the resampled tail to lie above the empirical population distribution. If an estimation process places greater area under the tail of the surplus distribution, the process is more conservative, hence permitting a one-tail test.

Strawderman and Zelterman [15] have developed a method that does not use resampling, but approximates the bootstrap CDF through a saddlepoint approximation found by Wood et. al. [20]. Further research could be conducted here.

Further analysis of the CS test should be conducted as well. One particular area of analysis is to test the sensitivity to changes in underlying parameters. For example, further research could improve the choice of *k*, replacing the simplified rule used above. Another promising method would be the technique developed in [17].

Examing the surplus data, the underlying assumptions and term structure remained constant throughout the 10,000 trials for each given valuation year. The business models have changed between valuation years. For example, LOB49 was modeled the same in years 1992 and 1993, but was changed in 1994. However the domain of convergence of the distribution went from one to three between years 1992 and 1993, and remained at three in year 1994. It appears that the term structure of interest rates affects the domain of convergence, and further research should be conducted on this effect.

Region 1 distributions, according to Gumbel [8] and Strawderman and Zelterman [15], can be transformed to Region 3 distributions using the natural logarithm. This transformation may modify the rules of thumb, further reducing the required number of simulations.

In conclusion, we have developed a method to reduce the time and costs of testing insurance company insolvency. We have removed several assumptions about the surplus distribution and have replaced them with results from extreme value theory. We have introduced the CS test, which will allow actuaries to determine the riskiness of a line of business. We hope that these methods and tests will be adopted to reduce risk and develop better and safer products in our industry.

POSTSCRIPT: A copy of the program and the fifteen parametric sample distributions implementing the REV technique is available by request.

Acknowledgments: I would like to thank Steve Sedlak and Steve Hodges for their direction, insight and advice. Also, I would like to thank Russ Osborn for the French translation of the Abstract and both Russ Osborn and Brian Ulery for their skillful rewording of the text. Especially, I would like to thank the department wizard, Martha Rohlck, who managed the nightmare of generating the 10,000 trials for the last three years. All my love and gratitude go to my wife Patricia and my daughter Michelle and my sons, Samuel, Bradley, Evan and Carl, for their patience while researching and writing this paper.

Bibliography

[1] Anonymous, S-PLUS Guide to Statistical and Mathematical Analysis Version 3.3. StatSci Division MathSoft, Inc. Seattle, Washington September 1995, p. 34 of chapter 7.

[2] Boos, D. D., "Using Extreme Value Theory to Estimate Large Percentiles." Technometrics. February 1984, Vol. 26, No. 1, pp. 33-39.

[3] Bowers, et al., Actuarial Mathematics. Society of Actuaries, Schaumburg, Ill. 1986, pp. 345-374.

[4] Castillo, E., Galambos, J., and Sarabia, J. M., "The Selection of the Domain of Attraction of an Extreme Value Distribution from a Set of Data." Lecture Notes in Statistics Vol. 51 Extreme Value Theory. Edited by J. Husler, R. Reiss. Springer-Verlag Berlin Heidelberg 1989, pp. 181-190.

[5] Craighead, S. "Extreme Value Statistics, Resampling, and Insolvency Testing." To appear in ARCH 1995.

[6] Efron, B. and Gong, G., "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." The American Statistician. February 1983, Vol. 37, No. 1, pp. 36-48.

[7] Falk, M., "Best Attainable Rate of Joint Convergence of Extremes." Lecture Notes in Statistics Vol. 51 Extreme Value Theory. Edited by J. Husler, R. Reiss. Springer-Verlag Berlin Heidelberg 1989, pp. 1-9.

[8] Gumbel, E. J., Statistics of Extremes. Columbia University Press, New York, 1958.

[9] Hogg, R. V. and Craig, A. T., Introduction to Mathematical Statistics, 2'nd ed. The Macmillan Company, New York, 1965, pp. 177-182.

[10] Klein, G. E., "The Sensitivity of Cash-Flow Analysis to the Choice of Statistical Model for Interest Rate Changes." TSA, 1993, Vol. XLV, pp. 79-186.

[11] London, R. L., Graduation--The Revision of Estimates. Actex, Winsted Conn., 1985, p. 58.

[12] Panjer, H. H. and Willmot, G. E., Insurance Risk Models, Society of Actuaries, Schaumburg, Ill. 1992.

[13] Reiss R.-D., "Extended Extreme Value Models and Adaptive Estimation of the Tail Index." Lecture Notes in Statistics Vol. 51 Extreme Value Theory. Edited by J. Husler, R. Reiss. Springer-Verlag Berlin Heidelberg 1989, p. 157.

[14] Soofi, E. S., "Capturing the Intangible Concept of Information." JASA, December 1994, Vol. 89, No. 428, pp. 1243-1254.

[15] Strawderman, R. L. and Zelterman, D., "A Semiparametric Bootstrap for Extreme Order Statistics." To be published.

[16] Tarter, M. E. and Lock, M. D., Model-Free Curve Estimation. Chapman & Hall, New York, NY.

[17] Wang, J. Z., "Selection of the k Largest Order Statistics for the Domain of Attraction of the Gumbel Distribution." JASA, June 1995, Vol. 90, No. 431, pp. 1055-1061.

[18] Weissman, I., "Estimation of Parameters and Large Quantiles Based on the k Largest Observations." JASA, December 1978, Vol.73, No. 364, pp. 812-815.

[19] Whittaker, E. T. and Robinson, G., The Calculus of Observations, 4'th ed. London and Glasgow: Blackie and Son, Ltd., 1944, p. 304.

[20] Wood, A. T. A., Booth, J. G. and Butler, R. W., "Saddlepoint Approximations to the CDF of Some Statistics With Nonnormal Limit Distributions." JASA, June 1993, Vol. 88, No. 422, pp. 680-686.

[21] Zelterman, D., "A Semiparametic Bootstrap Technique for Simulating Extreme Order Statistics." JASA, June 1993, Vol. 88, No. 422, pp. 477-485.

[22] Zelterman, D. and Lindgren, T., "Bootstrap Methods for Extreme Order Statistics," in INTERFACE '90 Proceedings, ed. C. Page and R. LePage, New York: Springer-Verlag 1992, pp. 580-582.