Casualty Actuarial Society E-Forum, Winter 2013

The CAS *E-Forum*, Winter 2013

The Winter 2013 Edition of the CAS *E-Forum* is a cooperative effort between the CAS *E-Forum* Committee and various other CAS committees, task forces, or working parties.

This *E-Forum* includes two papers written under the new heading of "Independent Research." Replacing the term "Additional Papers," Independent Research more accurately reflects the type of E-Forum submissions that are written independent of a call paper program.

E-Forum Committee

CAS *E-Forum,* Winter 2013

Table of Contents

Independent Research

For information on submitting a paper to the *E-Forum*, visit
http://www.casact.org/pubs/forum/.

A Note on the Upper-Truncated Pareto Distribution

David R. Clark, FCAS

Abstract

The Pareto distribution is widely used in modeling losses in Property and Casualty insurance. The thick-tailed nature of the distribution allows for inclusion of large events. However, in practice it may be necessary to apply an upper truncation point so as to eliminate unreasonably large loss amounts and to ensure that the first and second moments of the distribution exist.

This paper provides some background on the characteristics of the upper-truncated Pareto distribution, and suggests some diagnostics, based on order statistics, to assist in selecting the upper truncation point.

Keywords. Enterprise Risk Management, Pareto, Truncation, Order Statistics

1. INTRODUCTION

The Pareto distribution is useful as a model for losses in Property and Casualty insurance. It has a heavy right tail behavior, making it appropriate for including large events in applications such as excess-of-loss pricing and Enterprise Risk Management (ERM).

For applications in Enterprise Risk Management, however, there may be practical problems with the Pareto distribution because non-remote probabilities can still be assigned to loss amounts that are unreasonably large or even physically impossible. Further, a Pareto distribution with shape parameter $\alpha < 2$ will not have a finite variance, meaning we cannot calculate a correlation matrix between lines of business. In practice, an upper truncation point (T) is introduced and losses above that point are not included in the model. This upper truncation point may be considered the "Maximum Possible Loss" (MPL).

The difficulty for setting the upper truncation point is that the true Maximum Possible Loss for a given risk portfolio may not be easily determined. Analysts may hold different opinions as to what is possible.

In Enterprise Risk Management models, one goal is to evaluate the "tail" of the distribution, which can be very sensitive to the selection of the upper truncation point.

The goal of this paper will be to describe the characteristics of the upper-truncated Pareto and to offer some measures that may be useful in selecting the upper truncation point based on the sample of historical loss data. Some of these measures are results taken from the field of order statistics. We will not eliminate the need for the analyst to make an informed judgment when selecting the upper truncation, but we can give some objective measures to assist in making that judgment more informed.

1.1 Research Context

The literature on the Pareto distribution is vast. The text by Johnson et al. [5] provides the standard overview including historical genesis of the mathematical form, key characteristics and a comprehensive bibliography. Within the Casualty Actuarial Society literature, the 1985 paper by Stephen Philbrick is a recommended introduction and includes a brief discussion of upper truncation.

Our primary focus will be those characteristics of the Pareto distribution, particularly order statistics, that will be most useful for the Enterprise Risk Management application.

Order statistics is a branch of statistics that has grown over recent decades. It is concerned with inferences from an ordered sample of observations. In the CAS literature, an introduction to this topic related to estimating Probable Maximum Loss (PML, as distinguished from MPL) is given by Wilkinson (1982).

Extreme Value Theory (EVT) has developed as a branch from order statistics, with attention given to the distribution of the largest value of a sample. Much of EVT deals with approximations to the distribution of the largest value assuming the original distribution form is unknown.

1.2 Objective

The objective of this paper is entirely practical: given that the upper-truncated Pareto is widely used in insurance applications, we wish to supply analysts with additional information for selecting the upper truncation point.

1.3 Outline

The remainder of the paper proceeds as follows:

Section 2 will discuss the characteristics of the upper-truncated Pareto distribution itself.

Section 3 will review the maximum likelihood method for estimating the model distribution parameter.

Section 4 will introduce order statistics related to the upper-truncated Pareto and how they can be useful for selecting the upper truncation point.

Section 5 will present two brief examples to illustrate the technique of estimating the upper truncation based on the order statistics for the largest loss.

2. CHARACTERISTICS OF THE UPPER-TRUNCATED PARETO

2.1 The [Untruncated] Single Parameter Pareto

The cumulative distribution function for the Pareto distribution is given below in formula (2.1). This form represents losses that are at least as large as some lower threshold, θ, following the notation in Klugman et al. This form is sometimes referred to as the "single parameter Pareto" with shape parameter α and a lower threshold used to define the range of loss amounts supported (θ is not considered a parameter). Sometimes this form of the distribution is referred to as a "European Pareto" (see Rytgaard, 1990) to distinguish it from the two-parameter form. An alternative form uses a shift $Y = X - \theta$, representing just the portion of the excess loss above the threshold and the theta θ treated as a scale parameter. For the remainder of this paper we will only consider the single parameter or "European" form of the distribution.

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha} \qquad \theta \leq x, \qquad \alpha > 0 \tag{2.1}$$

The moments of the unlimited Pareto distribution are given as follows.

$$E(X^k) = \frac{\alpha \cdot \theta^k}{\alpha - k} \qquad \alpha > k \tag{2.2}$$

We note that not all moments exist for the Pareto distribution. For example, when $\alpha \leq 1$ the expected value is undefined.

2.2 The Upper-Truncated Pareto

When we introduce an upper truncation point, T, the random variable for loss can only take on values between the lower threshold and the upper truncation point. It is also interesting to note that the shape parameter, α, can now be any real value and is no longer restricted to being strictly positive.

$$F(x) = \begin{cases} \dfrac{1-\left(\frac{\theta}{x}\right)^{\alpha}}{1-\left(\frac{\theta}{T}\right)^{\alpha}} & \theta \le x \le T, \qquad \alpha \ne 0 \\ \\ \dfrac{ln(x/\theta)}{ln(T/\theta)} & \theta \le x \le T, \qquad \alpha = 0 \end{cases} \tag{2.3}$$

For the special case of $\alpha = -1$, the distribution of losses is uniform between θ and T. This may be surprising given that most insurance applications are heavily skewed and restrict the shape parameter to positive values but it does show the flexibility of the truncated form. Negative alphas are theoretically valid but unusual in insurance applications; we will be concerned in this paper mainly with cases for $\alpha > 0$.

All moments for the upper-truncated Pareto will always exist.

$$E(X^k) = \frac{\alpha \cdot \theta^k}{\alpha - k} \cdot \frac{1-\left(\frac{\theta}{T}\right)^{\alpha-k}}{1-\left(\frac{\theta}{T}\right)^{\alpha}} \qquad \alpha \ne 0, k \tag{2.4}$$

We may note that for the values $\alpha = 0$ and $\alpha = k$, formula (2.4) does not hold directly but we can estimate the moments by making use of the limiting function below.

$$\lim_{\alpha \to 0} \frac{1-\left(\frac{\theta}{T}\right)^{\alpha}}{\alpha} = \lim_{k \to \alpha} \frac{1-\left(\frac{\theta}{T}\right)^{\alpha-k}}{\alpha - k} = ln(T/\theta) \tag{2.5}$$

To provide some additional insight into the shape of the upper-truncated form, we consider the expected values for some special cases. The value $\alpha = -1$ produces a uniform distribution for which the expected value is the mid-point or arithmetic average between θ and T. For the value $\alpha = 1/2$ the expected value is the square-root of the product of θ and T, also known as the geometric average. For the value $\alpha = 2$ the expected value is the harmonic average of θ and T, found by averaging their inverses.

Shape Parameter	$E(X\|\theta \le X \le T)$
$\alpha = -1$	$(T+\theta)/2$
$\alpha = 0$	$(T-\theta)/ln(T/\theta)$
$\alpha = 1/2$	$\sqrt{T \cdot \theta}$
$\alpha = 1$	$ln(T/\theta)/(\theta^{-1} - T^{-1})$
$\alpha = 2$	$2/(\theta^{-1} + T^{-1})$

Second moments are easily found by the recurrence relation:

$$E(X^2|\alpha) = E(X|\alpha) \cdot E(X|\alpha - 1) \tag{2.6}$$

2.3 Moment-Matching to Evaluate Upper Truncation

We can make use of the first and second moments to make an estimate of the upper truncation point $\hat{T}$. The moment-matched parameters are found by solving the equations below.

$$E\big(X|\hat{\alpha},\hat{T}\big) = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$E\big(X^2|\hat{\alpha},\hat{T}\big) - E\big(X|\hat{\alpha},\hat{T}\big)^2 = s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2.7}$$

In other words, we want to set an upper truncation point such that standard deviation of the fitted distribution is [at least] the standard deviation of the historical large losses.

One obvious caution on this estimate, of course, is that it does not guarantee that the indicated upper truncation point $\hat{T}$ is greater than the largest loss actually observed historically. We therefore take it as only one part of our evaluation.

3. MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Maximum Likelihood Estimation (MLE) is more commonly used than moment-matching for estimating parameters. When there is no upper truncation, the maximum likelihood estimator for the Pareto shape parameter α is found using a simple expression.

$$\hat{\alpha}_{MLE} = N \cdot \left(\sum_{i=1}^{N} ln\left(\frac{x_i}{\theta}\right) \right)^{-1} \tag{3.1}$$

When there is an upper truncation point, the maximum likelihood estimator for α is a bit more complicated and requires solving the equation below. We may note again that both the lower threshold θ and the upper truncation T are constraints supplied by the user and are not considered parameters to be estimated.

$$\hat{\alpha}_{MLE} = N \cdot \left(\sum_{i=1}^{N} ln\left(\frac{x_i}{\theta}\right) - \left\{ \frac{N \cdot ln\left(\frac{\theta}{T}\right) \cdot \left(\frac{\theta}{T}\right)^{\hat{\alpha}_{MLE}}}{1 - \left(\frac{\theta}{T}\right)^{\hat{\alpha}_{MLE}}} \right\} \right)^{-1} \tag{3.2}$$

If we do consider the lower and upper truncation points as parameters, then the MLE estimators are simply the smallest and largest observations respectively (see Aban et al., 2006); that is, the first and last order statistics from the sample.

$$\begin{aligned} \hat{\theta}_{MLE} &= MIN\{x_1, x_2, \cdots, x_N\} \\ \hat{T}_{MLE} &= MAX\{x_1, x_2, \cdots, x_N\} \end{aligned} \tag{3.3}$$

These MLE estimators are not as helpful for our purpose of selecting an upper truncation point. The goal of Maximum Likelihood Estimation is to find the model parameters that result in the highest probability assigned to events that we have actually observed. In the case of the upper-truncated Pareto, this goal is accomplished by assigning zero probability to values outside the range of what we have already observed. This is not helpful if we believe that worse events are possible.

However, we may note that the MLE in formulas (3.1) – (3.3) includes two statistics that summarize the sample of losses:

$$Mean\ of\ Logs\ =\ \frac{1}{N}\cdot\sum_{j=1}^{N} ln(x_i)$$

$$x_N = MAX\{x_1, x_2, \cdots, x_N\} \tag{3.4}$$

Together these represent sufficient statistics for the model parameters α and T, informally meaning that they contain all of the information available from the sample concerning these parameters.[1]

The MLE for $\hat{T}$ is referred to as "non-regular," meaning that we cannot estimate its variance through the regular procedure using the information matrix of second derivatives of the log-likelihood function. This is not, however, a great problem because we can estimate $Var(X_N)$ using the moment functions given in section 4.3.

Finally, it is important to remember that there is a relationship between the shape parameter α and the upper truncation T. A different alpha will be estimated depending upon the selected upper truncation point. To illustrate this relationship, the table below shows how the expected value of loss severity changes based on these parameters.

		Expected Pareto Severity Subject to Upper Truncation				
		Lower Threshold (Theta):		1,000,000		
		Maximum Possible Loss (Upper Truncation)				
		10,000,000	25,000,000	50,000,000	100,000,000	999,999,999
Alpha	0.75	2,839,841	4,072,455	5,257,028	6,698,663	13,948,679
	1.05	2,507,183	3,231,920	3,793,243	4,353,690	6,137,484
	1.35	2,234,010	2,641,165	2,890,943	3,093,714	3,513,688
	1.65	2,015,287	2,236,237	2,342,509	2,412,446	2,510,008
	1.95	1,843,001	1,959,873	2,003,684	2,027,046	2,049,735

[1] Chapter 7 of Arnold et al. provides a longer discussion on order statistics and sufficiency.

4. ORDER STATISTICS

For a sample of independent losses drawn from a continuous distribution, the order statistics are simply the sample put into ascending order, with x_1 being the smallest observation and x_N being the largest observation.

4.1 The Distribution of the Largest of N Losses

The upper-truncated Pareto, the distribution of the largest of a sample of size N losses, is given in formula (4.1).

$$F_N(x_N) = F(x_N)^N = \left(\frac{1-\left(\frac{\theta}{x_N}\right)^{\alpha}}{1-\left(\frac{\theta}{T}\right)^{\alpha}}\right)^N \tag{4.1}$$

This distribution is unimodal, with the mode defined below in formula (4.2). The mode is not directly dependent upon the value of the upper truncation point T except for the case in which T is set below where the mode would otherwise be calculated.

$$Mode_N = MIN\left\{\theta \cdot \left(\frac{\alpha \cdot N + 1}{\alpha + 1}\right)^{1/\alpha}, T\right\} \tag{4.2}$$

Percentiles from this distribution are easily calculated, and this provides us with a hypothesis test for the upper truncation point.

4.2 Hypothesis Test for Existence of Upper Truncation

One preliminary question is whether an upper truncation is indicated at all. A simple hypothesis test can help answer this question. The null hypothesis is that there is no upper truncation point (that is $T = \infty$). We then compare the actual largest loss x_N observed in the history and ask whether it is reasonable that a Pareto with no upper truncation would have generated that loss. If the largest observed loss is "significantly" less than would have been expected, then we reject the null hypothesis and conclude that an upper truncation point should be used.

The test statistic is a p-value:

$$p = F(x_N|T = \infty)^N = \left(1 - \left(\frac{\theta}{x_N}\right)^\alpha\right)^N \approx exp\left(-N \cdot \left(\frac{\theta}{x_N}\right)^\alpha\right) \tag{4.3}$$

The shape parameter α used in this test should be based on the MLE estimate with no upper truncation as given in formula (3.1). The test is only appropriate for a sample size N large enough that the largest observation x_N does not have a significant impact on the estimate of α.

The approximation on the right side of (4.3) is the Fréchet distribution, which is a limiting case for the sample maximum and a standard result from Extreme Value Theory[2]. The hypothesis test is given in this form in Aban et al. The idea is that if the p-value is small, say $p < .05$, then we reject the null hypothesis that a Pareto with no upper truncation point is appropriate. Unfortunately, this test does not tell us what the upper truncation point should be; in fact, it does not even tell us that an upper-truncated Pareto is correct but only that an untruncated Pareto is unlikely.

Conversely, if the p-value is large, say $p > .05$, that does not necessarily mean that an untruncated Pareto should be used – but only that our sample data is not sufficient to reject it. The usefulness of the test is therefore quite limited.

4.3 Evaluating Moments for the Largest Loss

The calculation for the moments of the distribution of the largest loss is not trivial but can be accomplished with a careful strategy. Using a binomial series expansion of the distribution of the largest loss, the moments can be written as follows.

$$E_N\left(X_N^k\right) = \theta^k \cdot \sum_{j=1}^{N} \binom{N}{j} \cdot (-1)^{j-1} \cdot \frac{\alpha \cdot j \cdot \theta^k}{\alpha \cdot j - k} \cdot \frac{\left(1 - \left(\frac{\theta}{T}\right)^{\alpha \cdot j - k}\right)}{\left(1 - \left(\frac{\theta}{T}\right)^\alpha\right)^N} \qquad \alpha \neq k \tag{4.4}$$

As discussed above, the terms when $\alpha j = k$ can be evaluated using the limit formula (2.5). The difficulty with this form is that for large sample sizes, say $N > 30$, the factorial functions become extremely large, making the calculation numerically unstable. An alternative form that works for

[2] This is a result of the Gnedenko, Fisher-Tippett Theorem. The Fréchet distribution is given in Loss Models by Klugman et al. as the "Inverse Weibull" distribution.

larger values of N makes use of the incomplete beta distribution.

$$E_N\left(X_N^k\right) = \theta^k \cdot \frac{\Gamma(N+1)\cdot\Gamma(1-k/\alpha)}{\Gamma(N+1-k/\alpha)} \cdot \frac{\beta\left(1-\left(\frac{\theta}{T}\right)^{\alpha}; N, 1-k/\alpha\right)}{\beta\left(1-\left(\frac{\theta}{T}\right)^{\alpha}; N, 1\right)} \qquad \alpha > k \tag{4.5}$$

This form makes use of the incomplete beta function, defined below.

$$\beta(y; N, b) = \int_0^y \frac{t^{N-1}\cdot(1-t)^{b-1}}{B(N,b)}\, dt \qquad N, b > 0 \quad 0 < y < 1 \tag{4.6}$$

The incomplete beta function cannot be used directly when $\alpha \le k$. Klugman et al gives a recursive form that can be used for small values of α when N is large. This form will not work when k/α is exactly equal to an integer (e.g., the cases $\alpha = 1$ or $1/2$). A third alternative is needed for those cases.

Another form is written in terms of an infinite series. Formulas (4.7) and (4.8) provide two infinite series that converge to the expected dollar moments.

$$E_N\left(X_N^k\right) = \theta^k \cdot \sum_{j=0}^{\infty}\left(\frac{N}{N+j}\right)\cdot\left(\frac{\Gamma(j+k/\alpha)}{j!\cdot\Gamma(k/\alpha)}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.7}$$

$$E_N\left(X_N^k\right) = T^k - \theta^k \cdot \sum_{j=0}^{\infty}\left(\frac{j}{N+j}\right)\cdot\left(\frac{\Gamma(j+k/\alpha)}{j!\cdot\Gamma(k/\alpha)}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.8}$$

These series may be slow to converge when $\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)$ is close to 1.00, so this may not be an optimal formula for evaluating the moments. However, they do not have the numerical instability of formulas (4.4) or (4.5). Further, each term in the summation is a positive value, so the first series converges from below whereas the second series converges from above. The use of these two series together therefore lets us calculate moments to within any desired degree of accuracy.

We may note also that there are various recurrence relationships between moments of order statistics, for example as given by Khurana & Jha (1987), that can produce other methods for calculating the moments. However, these do not seem to offer more numerical stability than the formulas given above.

Just as we estimated $\hat{\alpha}$ and $\hat{T}$ by matching the mean and standard deviation of historical losses, we can alternatively estimate them by matching to the mean and largest value of the historical losses.[3]

$$\begin{aligned} E\left(X|\hat{\alpha},\hat{T}\right) &= \bar{x} \\ E_N\left(X_N|\hat{\alpha},\hat{T}\right) &= x_N \end{aligned} \tag{4.9}$$

However, we can make a better estimate by using the order statistics of the logarithms of the losses, instead of the losses themselves.

4.4 Evaluating Moments for the Largest Ln(Loss)

Where we had used $E_N\left(X_N|\hat{\alpha},\hat{T}\right)$ to represent the expected value of the largest loss in a sample of N, we now define $E_N\left(ln(X_N/\theta)|\hat{\alpha},\hat{T}\right)$ as the expected value of the logarithm of the largest loss, relative to the lower threshold.

The transformed variable $ln(X/\theta)$ follows an exponential distribution, and this allows for simpler calculations of the order statistic moments.

This form will turn out to have some advantages over working with the order statistics of the loss dollars themselves.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \int_{\theta}^{T} ln\left(\frac{x}{\theta}\right) \cdot \frac{N \cdot \left(1-\left(\frac{\theta}{x}\right)^{\alpha}\right)^{N-1} \cdot \alpha \cdot \theta^{\alpha} \cdot x^{-\alpha-1}}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}}\, dx \qquad \alpha \neq 0 \tag{4.10}$$

This integral can be evaluated as follows.

---

[3] This procedure is essentially the same as the recommendation in Cooke (1979).

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \frac{\left\{\frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{j}\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j}\right\} - ln\left(\frac{T}{\theta}\right)\cdot\left(1-\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}\right)}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{N}} \qquad \alpha \neq 0 \qquad (4.11)$$

We may also note that for the untruncated Pareto as $T \to \infty$ this expression simplifies to:

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)|T \to \infty\right) = \frac{1}{\alpha}\sum_{j=1}^{N}\frac{1}{j} \qquad \alpha \neq 0 \qquad (4.12)$$

These formulas can be re-written as a simple recurrence relationship between different sample sizes is given as follows.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = ln\left(\frac{T}{\theta}\right) + \frac{1}{\alpha\cdot N} - \frac{ln\left(\frac{T}{\theta}\right) - E_{N-1}\left(ln\left(\frac{X_{N-1}}{\theta}\right)\right)}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)} \qquad \alpha \neq 0 \qquad (4.13)$$

The sequence is starting by using the expected $E\big(ln(X/\theta)\big)$ for a single loss.

$$E_1\left(ln\left(\frac{X_1}{\theta}\right)\right) = E\left(ln\left(\frac{X}{\theta}\right)\right) = \frac{1}{\alpha} - \frac{ln\left(\frac{T}{\theta}\right)\cdot\left(\frac{\theta}{T}\right)^{\alpha}}{\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)} \qquad \alpha \neq 0 \qquad (4.14)$$

It can also be quickly recognized that if the expected value $E\big(ln(X)\big)$ is replaced by the mean of the logarithms of the sample of observed losses, then formula (4.14) is equivalent to the MLE formula (3.2). Matching the first moment of the logs is the same as performing a maximum likelihood estimate for the shape parameter α. This is a very useful result because it means that anyone currently using MLE to estimate the shape parameter will be able to use this moment matching strategy as an enhancement to their existing model.

Formulas (4.13) and (4.14) are not valid when $\alpha = 0$ but the moments for that special case are easily calculated.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \left(\frac{N}{N+1}\right)\cdot ln\left(\frac{T}{\theta}\right) \qquad \alpha = 0 \tag{4.15}$$

We can also evaluate the expected value of the largest log-loss using infinite series similar to those in formulas (4.7) and (4.8). As with those earlier expressions, we have a series that converges from below and a second that converges from above. These series are also somewhat faster to converge than those for the dollar moments.

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = \frac{1}{\alpha}\cdot\sum_{j=0}^{\infty}\left(\frac{N}{N+j}\right)\cdot\frac{1}{j}\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.16}$$

$$E_N\left(ln\left(\frac{X_N}{\theta}\right)\right) = ln\left(\frac{T}{\theta}\right) - \frac{1}{\alpha}\cdot\sum_{j=0}^{\infty}\left(\frac{1}{N+j}\right)\cdot\left(1-\left(\frac{\theta}{T}\right)^{\alpha}\right)^{j} \qquad \alpha > 0 \tag{4.17}$$

With these formulas, we are able to match the moments:

$$\begin{gathered} E\left(ln(X)|\hat{\alpha},\hat{T}\right) = \frac{1}{N}\cdot\sum_{j=1}^{N} ln(x_i) \\ E_N\left(ln(X_N)|\hat{\alpha},\hat{T}\right) = ln(x_N) \end{gathered} \tag{4.18}$$

These moment-matching equations make use of the same sufficient statistics identified in formula (3.4). The parameters $\hat{\alpha}$ and $\hat{T}$ are solved for numerically.

We may note a few advantages to the use of the estimators in formula (4.18):

1) The estimated $\hat{\alpha}$ is equivalent to the MLE estimate conditional upon $\hat{T}$
2) The estimates rely upon sufficient statistics, meaning they make use of all of the information about the truncated Pareto parameters contained in the sample

3) The recurrence formula is easily calculated

4) The estimate of $\hat{T}$ based on log-order statistics is slightly more conservative than the estimate based directly on the order statistics of the loss dollars. This is due to Jensen's Inequality:

$$E\big(ln(X)\big) \leq ln\big(E(X)\big)$$

We now go on to show how this procedure can be applied in real-world examples.

5. TWO ILLUSTRATIVE EXAMPLES

Having outlined a basic approach for estimating an upper truncation point, we will now look at two examples to illustrate the approach. The examples are not intended for use as actual pricing factors but just to show the thought process.

The numbers used in these examples are historical statistics related to natural disasters, and the samples are shown in Appendix A. The fact that these examples are from natural disasters does not mean that the same techniques could not be used for casualty events.

5.1 Earthquake Fatalities 1900-2011

The earthquake statistics are the estimated number of deaths for events from 1900 to 2011 as published by the U.S. Geological Survey (USGS). In many cases, these numbers are rough estimates. For this example, we look at the 21 earthquakes with 20,000 or more deaths. None of these figures has been adjusted for population changes or other factors.

The numbers can be summarized by the following statistics:

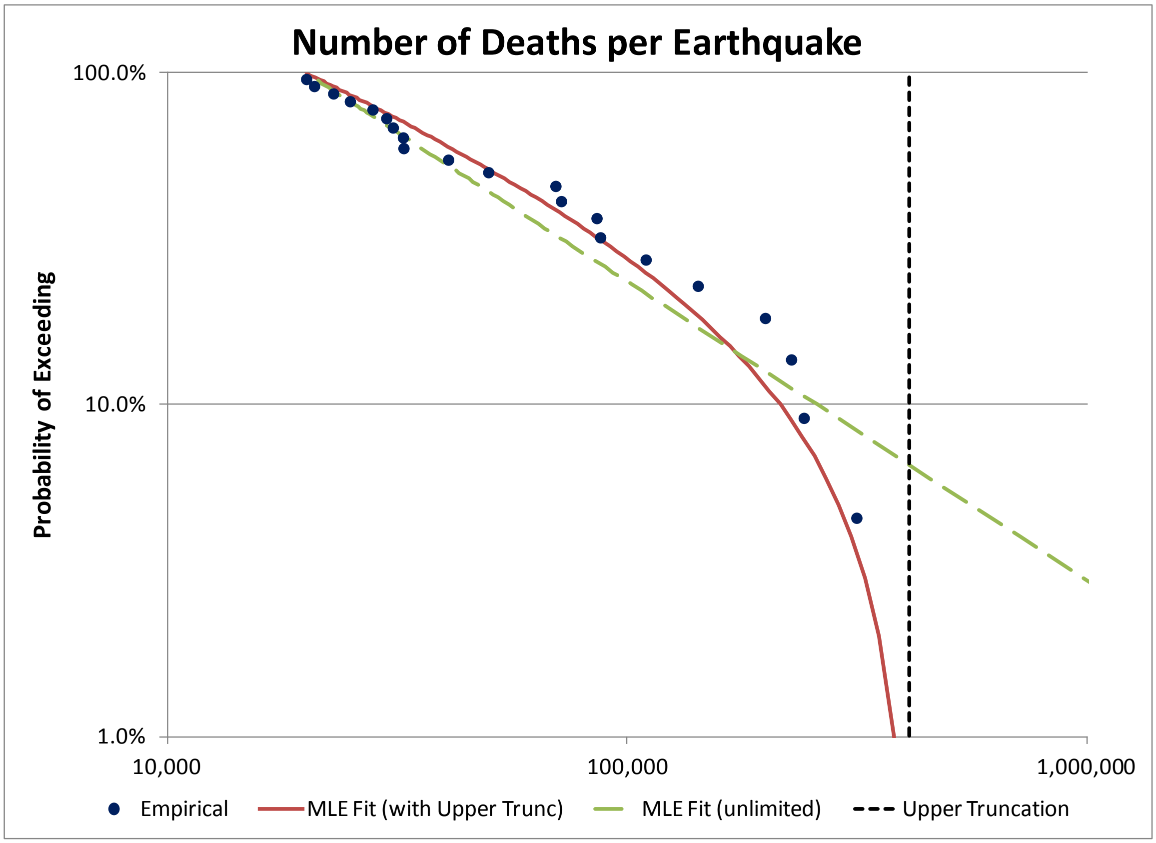
Number of Events	21
Lower Threshold θ	20,000
Average # Deaths	89,964
Standard Deviation of # Deaths	86,416
Largest # Deaths	316,000
Pareto Shape Parameter α (from MLE)	0.89993

The largest earthquake event, in terms of number of deaths, occurred in 2010 in Haiti, with 316,000 fatalities.

The p-value from this data is .173, meaning that there is a 17.3% chance that the largest event in a sample of 21 would be 316,000 or less from an untruncated Pareto. This is not strong evidence for rejecting the untruncated Pareto but does not rule out the possibility of including an upper truncation point.

A Pareto fit with no upper truncation indicates a shape parameter of 0.89993. Because this is less than 1.00000, the expected value would be undefined (infinite). This would create a serious problem in modeling the events, as simulation results could be chaotic. It is desirable to include an upper truncation.

We can select an indicated upper truncation point by matching the expected values to the average and largest amounts in our sample. As the graph below shows, this is an improved fit to the data also. The empirical points on the log-log graph show a downward curving shape, rather than a pure linear relationship that would indicate an untruncated curve.



The values from this moment-matching calculation are listed below:

Lower Threshold θ	20,000
Estimated Upper Truncation T	437,171
Pareto Shape Parameter α conditional on T	0.57122
Expected Value of # Fatalities	88,563
Expected Standard Deviation	88,334
Expected Largest of 21 Events $E_N(X_N)$	326,681

The key output from this analysis is the estimated upper truncation point as 437,171. This implies that the maximum possible number of deaths from an earthquake is 437,171 or about 38% higher than the worst event seen in the history.

The standard deviation and actual observed largest loss the actual data are slightly lower than would have been predicted by the model. This means our estimate of the upper truncation point is slightly higher than what would be needed to exactly match the sample; this conservatism is desirable since our goal is to select an upper truncation point that represents the largest possible loss.

We can also re-fit the model with different lower thresholds to include more or fewer losses to evaluate the sensitivity of the calculation.

Most importantly, we want to compare the moment-matching indication to what is known about the physical world that might create an upper bound on the possible number of deaths. Factors such as population density, construction of buildings and the possible intensities of earthquakes should be considered. Catastrophe models attempt to estimate the probability distribution based on these factors, and output from these models should be compared.

5.2 Large U.S. Weather Losses 1980-2011

The weather statistics come from the National Climatic Data Center (NCDC), a part of the National Oceanic and Atmospheric Administration (NOAA). The dollars are listed in thousands, and have been adjusted (by NCDC) to 2012 cost levels using the CPI. The losses represent estimates of total damages, not limited to just the insured portion. The sample in Appendix A are those events that caused $5 billion or more in 2012 dollars.

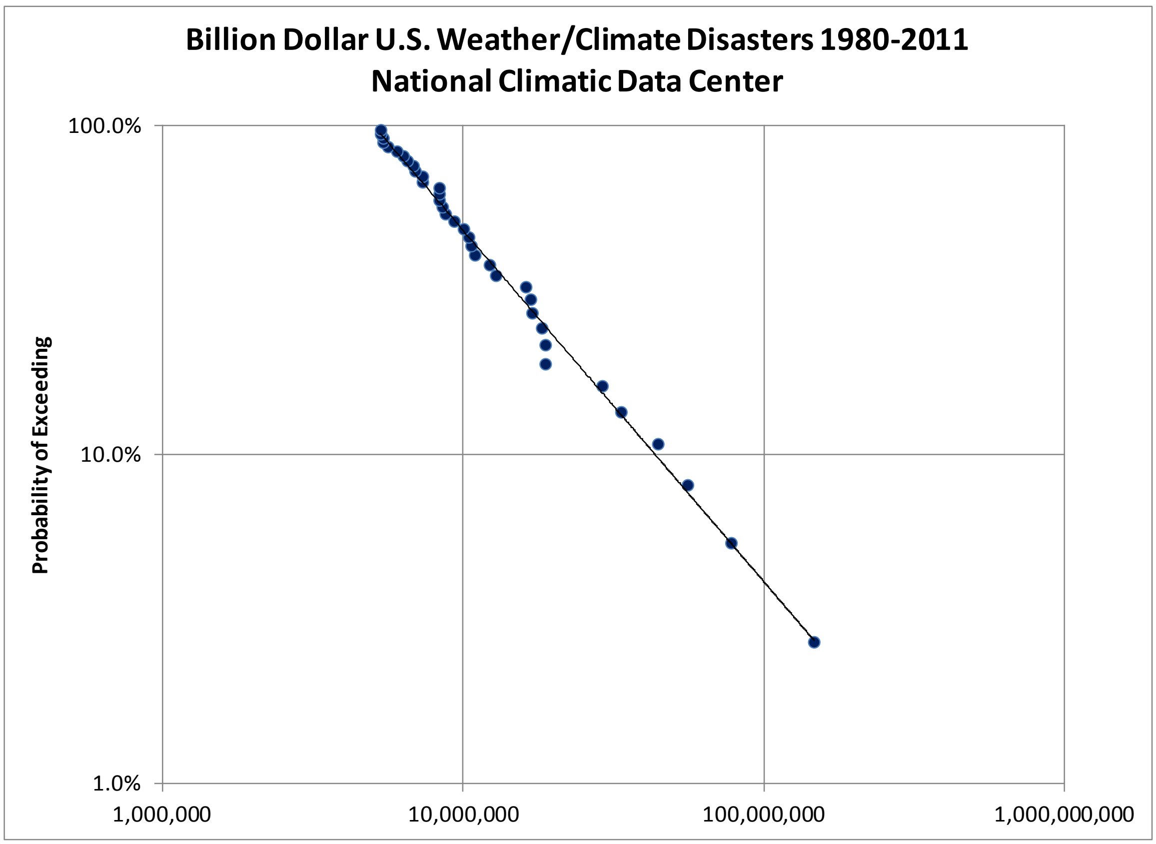
The numbers can be summarized by the following statistics:

Number of Events	36
Lower Threshold θ	5,000,000
Average Loss > Threshold	18,994,444
Standard Deviation of Losses	26,701,171
Largest Loss Damage	146,300,000
Pareto Shape Parameter α (from MLE)	1.11299

The largest weather event in this sample was Hurricane Katrina in 2005, estimated to be $146 billion in 2012 dollars.

The p-value from this data is .432, meaning we fail to reject the null hypothesis that the losses came from an untruncated Pareto. In practice, this implies that if we want to include an upper truncation point, it should be well above the largest order statistic.

The graph below shows the log-log graph of damage amount (in thousands) compared to the empirical survival probabilities (probability of exceeding the dollar amount). The historical amounts line up pretty closely along a straight line indicating, again, that if there is an upper truncation point then it must be much larger than the largest historical point.



If we calculate an upper truncation point so as to match the average and largest of the historical events, we find the following:

Lower Threshold θ	5,000,000
Estimated Upper Truncation T	480,073,321
Pareto Shape Parameter α conditional on T	1.07182
Expected Amount of Damage	21,014,276
Expected Standard Deviation	39,261,964
Expected Largest of 36 Events $E_N(X_N)$	178,675,516

The estimated upper truncation point of 480,073,321 is more than three times the largest observed historical event. The conclusion is that the largest possible hurricane damage is significantly larger than 2005's Hurricane Katrina. This indication is itself subject to estimation uncertainty but it does provide one more piece of information for use in modeling loss exposure.

5.3 Discussion of the Examples

These two numerical examples illustrate some of the assumptions and limitations of this estimation process.

First, we may note that the estimation is dependent upon the truncated Pareto being the "true" distribution for the phenomenon. Our estimate does not reflect the possibility that some other distributional model might be correct. If a different model would have been better, then it is possible that a higher upper truncation point would have been estimated.

Second, we are assuming that the sample we have observed is representative, and that future events will be of the same kind as those that have taken place historically. Events that are qualitatively different (not just bigger events of the same kind) need to be modeled separately. It is common to talk of events that have never been observed as "black swans" and we should recognize that a model that is parameterized based on past observations cannot account for these.

Third, we note that in both of the examples above the amounts observed were only estimates of the actual values, and include estimating error in themselves. An exact count of deaths from the Haiti earthquake was not made, so the upper truncation point is also less exact. This estimation error is compounded by uncertainty in inflation or demographic trends. The number of earthquake deaths or losses from weather events was gathered from a variety of sources, including newspaper reports. This type of uncertainty is also an issue in insurance losses where claim values may include case reserves rather than actual ultimate payments.

These factors are common to many statistical estimation problems. In the case of estimating an upper truncation point, we have the further difficulty that we are necessarily extrapolating beyond the range represented in our sample data. Given this level of uncertainty, our final reality check needs to be to ask if the upper truncation point corresponds to some true physical limit on the size of the loss; and if not to consider it a lower bound on the MPL.

6. CONCLUSIONS

The selection of an upper truncation point for the Pareto can be difficult in insurance applications. It represents, in theory, the Maximum Possible Loss (MPL) that could occur on the exposures written by the insurance company. This amount is generally selected based upon management's judgment about possible loss events. The use of order statistics allows us to squeeze some additional information out of the observed historical losses.

At the very least, we are able to calculate statistics such as standard deviation and expected largest loss for the upper-truncated Pareto, and compare these to the historical loss data. This provides more information to inform the judgment being made.

Appendix A - Data Sets for Examples

The data sets below are used as examples in Section 5. The earthquake statistics come from the U.S. National Geological Survey and represent estimated fatalities for international earthquakes since 1900. The U.S. Weather/Climate Disasters come from the National Climatic Data Center and represent total economic damages from weather events in the United States for 1980-2011, adjusted to 2012 dollars.

Earthqake Deaths Since 1900

Rank	# Deaths
1	316,000
2	242,769
3	227,898
4	200,000
5	142,800
6	110,000
7	87,587
8	86,000
9	72,000
10	70,000
11	50,000
12	40,900
13	32,700
14	32,610
15	31,000
16	30,000
17	28,000
18	25,000
19	23,000
20	20,896
21	20,085

NOAA Weather Losses 1980-2011

Rank	$ Damage
1	146,300,000
2	77,600,000
3	55,600,000
4	44,300,000
5	33,400,000
6	28,900,000
7	18,700,000
8	18,700,000
9	18,200,000
10	16,900,000
11	16,700,000
12	16,100,000
13	12,800,000
14	12,200,000
15	10,900,000
16	10,600,000
17	10,400,000
18	10,000,000
19	9,300,000
20	8,700,000
21	8,500,000
22	8,300,000
23	8,300,000
24	8,300,000
25	7,300,000
26	7,300,000
27	6,900,000
28	6,800,000
29	6,500,000
30	6,300,000
31	6,000,000
32	5,600,000
33	5,400,000
34	5,400,000
35	5,300,000
36	5,300,000

7. REFERENCES

[1] I. Aban, M. Meerschaert, A. Panorska, "Parameter Estimation for the Truncated Pareto Distribution," *Journal of the American Statistical Association* **2006**, Vol. 101, 270-277.
[2] Arnold, Barry C., N. Balakrishnan and H.N. Nagaraja, "A First Course in Order Statistics," Society for Industrial and Applied Mathematics (SIAM), 2008.
[3] Cooke, Peter, "Statistical inference for bounds of random variables," *Biometrika* **1979**, 66, 2, pp. 367-74.
[4] Johnson, Norman L., Samuel Kotz and Adrienne W. Kemp, "Univariate Discrete Distributions," Second Edition, John Wiley & Sons, 1993.
[5] Johnson, Norman L., Samuel Kotz and N. Balakrishnan, "Continuous Univariate Distributions," Second Edition, John Wiley & Sons, 1994
[6] A.P. Khurana and V.D. Jha, "Recurrence Relations Between Moments of Order Statistics from a Doubly Truncated Pareto Distribution," *Sankhya: The Indian Journal of Statistics* **1987**, Volume 53, Series B, Pt.1, 11-16.
[7] Klugman, Stuart A., Harry H. Panjer and Gordon E. Willmot, "Loss Models: From Data to Decisions," John Wiley & Sons, 1998.
[8] Philbrick, Stephen, "A Practical Guide to the Single Parameter Pareto," *PCAS* **1985**, *Vol. LXXII, 44-84*.
[9] Robson, D.S. and J.H. Whitlock, "Estimation of a Truncation Point," *Biometrika* **1964**, 51, 1 and 2, p. 33.
[10] Rytgaard, Mette, "Estimation in the Pareto Distribution," *ASTIN Bulletin* **1990**, Vol. 20, No. 2, 201-216.
[11] Wilkinson, Margret E., "Estimating Probable Maximum Loss with Order Statistics," *PCAS* **1982,** Vol. LXIX, 195-209.

Abbreviations and notations

N, number of large losses observed in a sample
$E_N(X_N)$, expected value of largest loss in a sample size of N
X, random variable representing a single loss amount; $\theta \leq X \leq T$
x_N, largest observed loss in a sample size of N
θ, Theta, representing the lower threshold of losses
T, Upper truncation point – loss amount above this are not considered possible
α, Alpha, representing the "shape parameter" of the Pareto distribution

Biography of the Author

David R. Clark is a Senior Actuary with Munich Reinsurance America, and a Fellow of the Casualty Actuarial Society (FCAS). His prior papers include "LDF Curve-Fitting and Stochastic Reserving: A Maximum Likelihood Approach" which received the 2003 Best Reserve Call Paper prize, and "Insurance Applications of Bivariate Distributions" co-written with David Homer which received the 2004 Dorweiler Prize.

Acknowledgments

The author gratefully acknowledges the valuable review given by Michael Fackler, David Homer, Hou-Wen Jeng and James Sandor.

Calibration Of A Jump Diffusion

Rasa Varanka McKean, ACAS, MAAA

1 Introduction

This paper outlines an application of a weighted Monte Carlo method to a jump diffusion model in the presence of clustering and runs suggestive of contagion. The paper was originally submitted as a master's thesis in the Mathematics in Finance program at the Courant Institute of Mathematical Sciences, New York University, on March 15, 2003. The author wishes to make the material available to a wider audience. Explanatory material has been added to make the paper easier to read. The mathematics is unchanged.

Although the motivation for this application is actuarial in nature, the method is not limited to insurance accidents. In fact, the method has broad application to financial analysis. The first equation below may be seen as a theoretical bridge between the two fields of Actuarial Science and Finance; the underlying processes have a common construct.

In its most general form, a sample path corresponding to a stochastic process which is *differential*, *homogeneous*, and *increasing* may be decomposed as a linear part plus a pure jump function [12][10]. The process is referred to as a homogenous differential process with increasing paths. The process will be defined in Section 2. For now, we state only the equation:

$$p_t = p(t) = mt + \int_0^\infty \ell \wp([0,t] \times d\ell); t \geq 0 \tag{1}$$

The term ℓ corresponds to the size of a jump. The function $\wp$ counts the number of jumps. In this paper, the counts will be Poisson distributed. The term $[0,t] \times d\ell$ corresponds to a set over which the counts are taken. The expression $\wp([0,t] \times d\ell)$ is the number of jumps occuring up to time t of severity between ℓ and $\ell + d\ell$. In other words, the integrand corresponds to the well-known actuarial phrase "frequency times severity".

The linear part of the decomposition is mt. If the linear term is dropped, i.e. $mt = 0$, then the resulting pure jump function resembles an insurance aggregate loss. Frequency of accidents times severity of those accidents are summed over a given population to obtain the total loss amount [8]. If the linear part is replaced by a Brownian motion with or without drift, a financial model results. It's important to note here that in the financial model, there is a stochastic differential equation where the Brownian motion is in the exponential. There is no exponential term in the actuarial model; a jump is a jump.

Avellaneda has calibrated a variety of financial instruments [1]. Throughout this reference, a pricing model refers to a model for pricing less liquid instruments relatively to more liquid instruments (the benchmarks). Calibration of the Monte Carlo model is performed by assigning probability weights to the simulated paths. The weights are derived by minimizing the Kullback-Leibler relative entropy of the posterior measure to the prior (empirical) measure.

Recall the definition of entropy and don't feel bad if you look it up in Wikipedia. Entropy is a measure of how evenly energy is distributed in a system. Entropy is a measure of order versus disorder or randomness. Relative entropy measures the entropy of one state as compared to the entropy of a second state. A very rough analogy can be found in the measurement of temperature. Temperature is measured with thermometers, which may be calibrated to a variety of temperature scales such as degrees Fahrenheit, Celsius, or Kelvin. Relative entropy would be very roughly similar to a comparison of any two of these three temperature scales.

2 Modelling A Jump Diffusion

Applications of jump diffusions include option pricing, credit risk, and actuarial science. Applications in option pricing and actuarial science are outlined in the history of jump diffusion models below. For a financial model of credit risk, a suggested reference is [15].

A stochastic process with sample paths $p(t)$, $p(0) = 0$ is said to be *differential* if its increments $p[t_1, t_2) = p(t_2) - p(t_1)$ over disjoint intervals $[t_1, t_2)$ are independent, *homogenous* if the law of $p[t_1 + s, t_2 + s)$is independent of $s(\geq 0)$, and *increasing* if $p(t_1) \leq p(t_2)$ for $t_1 \leq t_2$.

A sample path may be decomposed into a linear part plus an integral of Poisson processes:

$$p_t = p(t) = mt + \int_0^\infty \ell \wp([0,t] \times d\ell); t \geq 0 \tag{2}$$

$\wp(dt \times d\ell)$ being Poisson distributed with mean $dt \times \nu' d\ell$ where $d\nu = \nu' d\ell$, ν' being the density function of the measure $d\nu$. The measure ν shouldn't be too large in the sense that the integral is finite:

$$\int_0^1 \ell d\nu + \int_1^\infty d\nu < \infty \tag{3}$$

Then:

$$\mathcal{P}[\wp(B) = n] = \frac{\beta^n}{n!} e^{-\beta}; for\ n \geq 0, B \subset ([0, +\infty) \times (-\infty, +\infty)), \tag{4}$$

$$\beta = \int_B dt \nu' d\ell \tag{5}$$

The process p_t is *differential* because the counts $\wp(B)$ attached to disjoint $B \subset [0, +\infty) \times (0, +\infty)$ are independent, and *additive* in the sense that $\wp(\bigcup_{n \geq 1}) = \sum_{n \geq 1} \wp(B_n)$ for disjoint B_1, B_2, etc. $\subset ([0, +\infty) \times (0, +\infty))$. $\wp([t_1, t_2) \times [\ell_1, \ell_2))$ is just *the number of jumps of* $p(t); t_1 \leq t < t_2$ of magnitude $\ell_1 \leq \ell < \ell_2$.

Natural extensions of the basic model would include stochastic volatility, mean reversion, and multiple jump processes.

Stochastic volatility accommodates volatility clustering, an important feature of the data. Mean reversion may account for perturbations induced by diffusion vs. perturbations created by jumps. Multiple jump processes may be used to distinguish between types of jumps or sizes of jumps. Additionally, the jump sizes may be modelled by various distributions.

Each of these scenarios will be described in turn.

3 Jump Diffusion Processes In The Literature

3.1 The Initial Pricing Model

The first pure jump model in the financial literature is attributed to Cox, Ingersoll and Ross [4].The model is illustrated as follows in the usual discrete pricing diagram.

$$S \begin{smallmatrix} \nearrow Su \\ \searrow S\exp^{(-w\Delta t)} \end{smallmatrix}$$

The upward and downward probabilities for stock price movement in time Δt are $\lambda\Delta t$ and $1-\lambda\Delta t$ respectively. The asset price declines at rate w except for occasional jumps occurring as a Poisson process with rate λ. The size of the jumps are modelled as u times the current asset price S.

Criticism of the model notes that jumps here can only be positive which is unrealistic in the financial markets except for the probability of ruin. Note that the exponential term in the diagram above cannot yield a negative value. However, this criticism would not hold for insurance losses. For instance, the value of a property lost in a fire cannot be negative, one does not lose negative time on the job due to an injury, and so forth.

Arguably, a reserve for future indemnity benefits, as an example, may be posted and subsequently netted down due to the death of the claimant. In such a case, however, the life expectancy of the claimant would have been quantified and posted as the initial reserve. The resulting downward movement could be seen as parameter risk. In other words, if the reserve had been estimated with greater accuracy, the downward movement would not have occurred. Further such arguments could be made to show that a pure jump process is useful in modelling insurance losses.

Another criticism of the financial model is that the process leads to a distribution of stock price values with a fat right tail and a thin left tail, the opposite to that observed for equities. Such a distribution, however, is common in insurance and especially in reinsurance. The time lags in discovering and reporting losses such as medical malpractice or products liability create a fat right tail. inflationary and social trends in jury awards may be very different ten years hence, leading to unexpected increases in the size of awards.

In Actuarial Science, pricing models for aggregate distributions of claim data occur in the cohort approach to collective risk theory. An analysis of collective risk theory and insurance models is beyond the scope of this paper. Aggregate loss distributions have been widely discussed in the actuarial literature. The interested reader is referred to basic, comprehensive treatments [2] [8] [6].

3.2 An Application To Option Pricing

Merton first suggested a modification to the standard option pricing model, a jump function added to the Brownian motion term [13]. The jump component represents the occasional discontinuous breaks observed in the financial markets.

Define:

μ_B the expected return from an asset associated with the Brownian motion
σ_B the volatility of the Brownian motion
λ the rate of occurrence of a jump
κ the average jump size (amplitude) as the change in asset price.

Then the model is written in the following form:

$$\frac{dS}{S} = (\mu_B - \lambda\kappa)dt + \sigma_B dW + \kappa dq \tag{6}$$

where κ is drawn from a normal distribution $\kappa \sim N(\mu_J, \sigma_J^2)$ for μ_J and σ_J the mean and standard deviation of the jump respectively, W is a Wiener process, and q is a Poisson process generating the jumps. The processes W and q are assumed to be independent.

If $\lambda\kappa$ is the contribution from the jumps then the remainder $\mu_B - \lambda\kappa$ is the expected growth rate provided by the geometric Brownian motion.

In a special case of Merton's model, the logarithm of the jump amplitude is normally distributed. The European call option price is then written as:

$$C = \Sigma_{n=0}^{\infty} \frac{e^{(-\lambda' T)}(\lambda'\tau)^n}{n!} f_n \tag{7}$$

where $\tau = T - t$, $\lambda' = \lambda(1+\kappa)$ and f_n is the Black-Scholes option price with parameters

$$\sigma_n^2 = \sigma^2 + \frac{n\sigma^2}{\tau} \tag{8}$$

for σ, the standard deviation of the normal distribution, and

$$r_n = r - \lambda\kappa + \frac{n(ln(1+\kappa))}{\tau} \tag{9}$$

for r, the interest rate. Terminating the infinite sum is not problematic since the factorial function grows rapidly.

Note that the model gives rise to fatter left and right tails than Black-Scholes and is consistent with implied volatilities in currency options but not in insurance losses.

The key assumption in Merton's model is that the jump component of the asset return models non-systematic risk. This assumption would be difficult to make in an actuarial model where the jumps may represent a mixture of both systematic risk and non-systematic risk. Insurance claim sizes tend to cluster due to insurance policy limits, trends in jury awards, and claims adjusters' case reserving practices. The jump component of risk may involve contract law, procedural law, and insurance goodwill. A more sophisticated model incorporating multiple jump processes may be required.

3.3 Stochastic Volatility

A stochastic volatility model may be of the following form.

$$\frac{dS}{S} = \mu_B dt + \sigma dW + kdq \tag{10}$$

where $k \sim N(\mu_J, \sigma_J^2)$ and

$$d(ln\sigma^2) = b(\mu_h - ln(h^2))dt + cdZ \tag{11}$$

The logarithm of the variance σ^2 follows a mean-reverting process with the Wiener error term dZ. This model is termed a stochastic volatility jump diffusion process (SVJD)[5]. The model has constant jump amplitude and a mean-reverting process for the volatility. In other words, the path of the volatility parameter is a mean-reverting process. Note the drift is also a mean-reverting process.

3.4 Multiple Jump Processes

In the general form of the sample path, denote the pure jump process by $\mathcal{J}_t = \int_0^\infty \ell\wp([0,t] \times d\ell)$ and replace mt by $d\mathcal{W}$ where W is a Wiener process with drift $d\mathcal{W} = \sigma d\mathcal{Z} + \mu dt$.

$d\mathcal{Z}$ is an independent Gaussian shock
σ is the variance
μ is the drift

$\mathcal{J}_t$ may be further decomposed as a multiple jump process $\mathcal{J}_t = \mathcal{J}_t^1 + \mathcal{J}_t^2$ where $\mathcal{J}_t^1$ has jump-amplitudes ≤ 1 and $\mathcal{J}_t^2$ has jump-amplitudes >1. The $\mathcal{J}_t^1$ term may be comprised of an mininite number of small jumps. In financial and actuarial applications, the $\mathcal{J}_t^1$ term would assume a finite number of jumps or insurance losses in a given period of time.

Then, referring back to equation (2):

$$p_t = mt + \int_0^\infty \ell\wp([0,t] \times d\ell) = d\mathcal{W} + \mathcal{J}_t^1 + \mathcal{J}_t^2 = \sigma d\mathcal{Z} + \mu dt + \mathcal{J}_t^1 + \mathcal{J}_t^2 \tag{12}$$

3.5 Jump Amplitudes

The jump amplitudes can give rise to a variety of models. The large jumps may be seen as rare events relative to the background noise of the diffusion. The jump amplitude may be time dependent.

Denote the positive measure by ν and the associated density function as ν', as before in the general form of the model. The measure ν is the product of the Poisson rate λ and the size of the jump.

The measure ν may not be a probability measure. In such a case, the jump diffusion model is not of the compound Poisson type. Further, $\int \nu'(dx) = \int d\nu$ may not be finite.

Processes with an infinite number of jumps may be modelled by jump amplitudes with densities given by:

1. $\nu'(x) = A \mid x \mid^{-1} e^{(-\eta_{\pm}|x|)}$ a variance gamma function
2. $\nu'(x) = A_{\pm} \mid x \mid^{-(1+\alpha)} e^{(-\eta_{\pm}|x|)}$ a tempered ("truncated") stable process
3. $\nu'(x) = \frac{Ae^{(-\lambda x)}}{\sinh(x)}$ a Meixner process

Note that in these cases, the singularity occurs near the origin as the denominator approaches zero. The small jumps may be truncated or the singularities may be omitted by dropping the $\mathcal{J}_t^1$ term.

In cases where $\lambda = \int \nu'(x)dx < +\infty$, the measure is finite and the measure ν can be normalized to define a probability measure μ which can be interpreted as the distribution of jump sizes:

$$\mu(dx) = \frac{\nu(dx)}{\lambda}$$

In these cases, it may be shown that one necessarily obtains a compound Poisson process as in formula (1). Processes constituted by stochastic variation in both the number of jumps and amplitude of the jumps are termed compound processes. The independence of the variables denoting the number of jumps and the jump amplitudes follows from the assumption of independent increments for the sample paths. The jump amplitudes x_i and x_j are independent of each other $\forall\ i \neq j$ by independence of increments. Each x_i has the same distribution by homogeneity.

Consider sample paths. Let $X(t) = \sum_{n=0}^{N(t)} x_n$ where x_n are independent identically distributed random variables, $N(t)$ is a Poisson process with rate λ. The sum of the jumps is compound Poisson. Without loss of generality, we may assume $X(0) = 0$. Let $d\mu = \frac{d\nu}{\lambda}$ be the normalization of the measure to a probability measure. Compute the characteristic function of the sample path:

$$\mathbb{E}(e^{ikX(t)}) = \mathbb{E}(e^{ik\sum_{n=0}^{N(t)} x_n}) = \mathbb{E}(e^{ikx_0} ... e^{ikx_{N(t)}}) \tag{13}$$

By independence of N and x:

$$= \mathbb{E}((\int e^{ikx} d\mu(x))^{N(t)}) = \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} e^{-\lambda t} (\int e^{ikx} d\mu(x))^j \tag{14}$$

$$= e^{-\lambda t} \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} (\int e^{ikx} d\mu(x))^j = e^{-\lambda t} (e^{\lambda t \int e^{ikx} d\mu(x)})$$

$$= e^{\lambda t (\int e^{ikx} d\mu(x) - 1)} = e^{\lambda t \int e^{ikx} d\mu(x) - \lambda t \int d\mu(x)}$$

Substituting the normalized probability measure, one obtains:

$$e^{t \int (e^{ikx(t)} - 1) d\nu(x)} \tag{15}$$

In a financial or actuarial application, the number of jumps per unit of time is finite so the application may be described by a jump process of compound Poisson type.

4 Contagion

The next step in this exposition is a description of contagion as it affects the statistical properties of the number of jumps seen as a random process. We introduce contagion into the jump diffusion process by considering a "mixed" compound Poisson process. This type of process is often used in actuarial work when one accident effectively increases the probability of future accidents through a conditional probability.

We begin at the beginning with Polya's urn scheme and Polya's scheme of contagion [7]. Suppose an urn contains b black balls and r red balls. A ball is drawn at random. The ball drawn is always replaced and in addition, c balls of the same color are added to the urn. The absolute probability of the sequence *black, black* is by Bayes' theorem below. Let H denote the first drawing of *black* and let A denote the second drawing of *black*. The sequence *black, black* is denoted by AH. If the first ball drawn is black, the conditional probability of a black ball at the second drawing is $\frac{(b+c)}{(b+c+r)}$. The probability of the sequence AH is, by Baye's theorem:

$$P[AH] = P[A \mid H]P[H] = \frac{b}{(b+r)} \times \frac{(b+c)}{(b+c+r)} \tag{16}$$

If the first two drawings result in black, the urn contains $b + 2c$ black balls and $b + r + 2c$ balls in total. The conditional probability of a black ball at the third trial becomes $\frac{(b+2c)}{(b+2c+r)}$. The probability of any sequence can be calculated in this way. The ordering of the sequence is immaterial. Any sequence of n drawings resulting in n_1 black and n_2 red balls for $n_1+n_2=n$ has the same probability as the sequence first n_1 black balls and then n_2 red balls given by:

$$P_{n_1,n_2} = \frac{b(b+c)(b+2c)...(b+n_1c-c)r(r+c)...(r+n_2c-c)}{(b+r)(b+r+c)(b+r+2c)...(b+r+nc-c)} \tag{17}$$

The Polya process describes a model for contagion, where every accident increases the probability of future accidents. The applications of the process include contagious diseases, meteorology, lattices in crystal structure, industrial quality control, and insurance, where long runs are suggestive of contagion or accumulated chance effect. It can be shown that the limiting form of Polya's distribution of probabilities is the negative binomial distribution. The limiting form may be used in a mixed compound Poisson process as the distribution for the Poisson rate variable λ [7].

5 Calibration Of The Pricing Model

The calibration methods are taken from a pricing model developed for the financial markets due to Avellaneda [1]. One purpose of the model is to price less liquid instruments relative to more liquid instruments. Avellaneda has calibrated a variety of financial instruments. Further work was done by Cont [3]. Avellaneda's model for bid-ask spreads admits a jump diffusion, an enhancement proposed by this paper. The enhancement will be shown in the next section.

5.1 Theory Of The Model

Consider a simulation with sample paths denoted by $\omega_1, ..., \omega_\nu$. Define a uniformly weighted simulation to be one where each path has equal probability of occurrence. In a non-uniformly weighted simulation, we assign probabilities $p_1, ..., p_\nu$ to each path where the probabilities are not necessarily equal.

For a contingent claim that pays the holder h_i dollars if the path ω_i occurs, the value of the contingent claim in the non-uniformly and uniformly weighted scenario where $p_i = \frac{1}{\nu}\ \forall i$ is:

$$\Pi_h = \sum_{i=1}^{\nu} h_i p_i \tag{18}$$

and

$$\Pi_h = \frac{1}{\nu}\sum_{i=1}^{\nu} h_i \tag{19}$$

respectively.

A prior distribution is generated by simulating the paths of a stochastic process which are uniformly weighted. Probabilities $p_1, ..., p_\nu$ are then determined to simulate a posterior distribution comprised of non-uniformly weighted sample paths.

For two probability vectors $p_1, ..., p_\nu$ and $q_1, ..., q_\nu$, the relative entropy of p with respect to q is defined as:

$$D(p \mid q) = \sum_{i=1}^{\nu} p_i \log(\frac{p_i}{q_i}) \tag{20}$$

For the Monte Carlo simulation with equal weights, denote the uniform probability vector by $u = (1/\nu, \ldots, 1/\nu)$. Then substitute $q_i = 1/\nu \equiv u_i$ into the above equation to derive the relative entropy distance immediately below.

The calibrated posterior probability measure is found by minimizing the Kullback-Leibler relative entropy of the prior and posterior measures. The relative entropy distance

$$D(p \mid u) = \log \nu + \sum_{i=1}^{\nu} p_i \log p_i \tag{21}$$

measures the deviation of the calibrated model from the prior data. Note that $D \geq 0$ with equality holding only if $p_i = \frac{1}{\nu}$.

For $p_i = \frac{1}{\nu}$:

$$D(p \mid u) = \log \nu + \frac{1}{\nu} \log \prod_{i=1}^{\nu} p_i = \log \nu + \frac{1}{\nu} \log(\frac{1}{\nu})^{\nu} = 0 \tag{22}$$

The relative entropy is directly related to the support of the measure. Suppose $p_i = \frac{1}{\nu^{\alpha_i}}$ for $i = 1, 2, ..., \nu$. Let N_α represent the number of paths with $\alpha_i = \alpha$.

Then, $\sum_\alpha N_\alpha = \nu$, $\sum_\alpha \frac{N_\alpha}{\nu^\alpha} = 1$, and:

$$D(p \mid u) = \log \nu + \sum_{\alpha} \frac{N_\alpha}{\nu^\alpha} \log \frac{1}{\nu^\alpha} \tag{23}$$

which reduces to:

$$\log \nu - \sum_{\alpha} \frac{\alpha N_\alpha}{\nu^\alpha} \log \nu = \log \nu (1 - \sum_{\alpha} \frac{N_\alpha}{\nu^\alpha} \alpha) = \log \nu (1 - \mathbb{E}^p(\alpha)) \tag{24}$$

A small relative entropy corresponds to a large expected value of α. A small α corresponds to a thin support, which implies that a large number of paths are discarded by the algorithm. A small α may also be seen as a mismatch of probabilities between the prior and posterior measures since the measure will be concentrated on a small number of paths in the posterior measure. One sees that it all depends on the measure. Therein lies the difficulty in calibrating a jump diffusion when the frequency of jumps is small.

To elucidate the theory, denote the set of sample paths as: $\omega^{(i)} = (x_1(\omega^{(i)}), ..., x_N(\omega^{(i)})$ for

$i = 1, 2, ..., \nu$

and the associated stochastic differential equation with Wiener process W

$$dX = \sigma(X, t)dW + \mu(X, t)dt \tag{25}$$

Denote the market prices of N benchmark instruments by $C_1, ..., C_N$ and the present value of the jth cashflow as $g_{1j}, g_{2j}, ..., g_{\nu j}$ for $j = 1, ..., N$. The price relations for the benchmark instruments, for $j = 1, ..., N$ are:

$$\sum_{i=1}^{\nu} p_i g_{ij} = C_j \tag{26}$$

5.2 The Calibration Algorithm

As before, denote the uniform probability vector by

$$u = (\frac{1}{\nu}, ..., \frac{1}{\nu})$$

In the case of the prior measure, we consider the following minimization problem.

Minimize:

$$D(p \mid u) = \log \nu + \sum_{i=1}^{\nu} p_i \log p_i \tag{27}$$

under linear constraints $C_j = \sum_{i=1}^{\nu} p_i g_{ij}$ for Lagrange multipliers $\lambda_1, ..., \lambda_N$:

$$\min_{\lambda}[\max_{p}\{-\log \nu - \sum_{i=1}^{\nu} p_i \log p_i + \sum_{j=1}^{N} \lambda_j (\sum_{i=1}^{\nu} p_i g_{ij} - C_j)\}] \tag{28}$$

Consider the max first. Differentiate with respect to p_i, for fixed i and equate the derivative to the Lagrange multiplier ϕ for the additional constraint $\sum p_i = 1$:

$$-\log p_i - 1 + \sum_{j=1}^{N} \lambda_j g_{ij} = \phi$$

Let $\phi = -\mu - 1$

Then

$$-\log p_i - 1 + \sum_{j=1}^{N} \lambda_j g_{ij} = -\mu - 1$$

$$\log p_i = \mu + \sum_{j=1}^{N} \lambda_j g_{ij}$$

Let $e^{\mu} = \frac{1}{Z}$. Then the maximum occurs at the value p_i^* $\forall i$,

$$p_i^* = \frac{e^{(\sum_{j=1}^{N} \lambda_j g_{ij})}}{Z} \tag{29}$$

The constraint

$$\sum_{i=1}^{\nu} p_i = 1 = \sum_{i=1}^{\nu} \frac{e^{(\sum_{j=1}^{N} \lambda_j g_{ij})}}{Z}$$

shows that Z is a normalizing constant.

Note:

$$\log p_i = \sum_{j=1}^{N} \lambda_j g_{ij} - \log Z \tag{30}$$

at the max p.

Thus, at the maximum,

$$\max_p \{-\log \nu - \sum_{i=1}^{\nu} p_i \log p_i + \sum_{j=1}^{N} \lambda_j (\sum_{i=1}^{\nu} p_i g_{ij} - C_j)\}$$

$$= -\log \nu - \sum_{i=1}^{\nu} p_i (\sum_{j=1}^{N} \lambda_j g_{ij} - \log Z) + \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{\nu} p_i g_{ij} - \sum_{j=1}^{N} \lambda_j C_j$$

$$= -\log \nu + \sum_{i=1}^{\nu} p_i \log Z - \sum_{j=1}^{N} \lambda_j C_j = -\log \nu + \log Z - \sum_{j=1}^{N} \lambda_j C_j$$

Now consider the minimum:

$$\min_{\lambda}[-\log \nu + \log Z - \sum_{j=1}^{N} \lambda_j C_j]$$

Differentiate with respect to λ_k and equate to zero:

$$\frac{1}{Z} \frac{\partial}{\partial \lambda_k} \sum_{i=1}^{\nu} e^{(\sum_{j=1}^{N} \lambda_j g_{ij})} - C_k = \frac{1}{Z} \sum_{i=1}^{\nu} g_{ik} e^{(\sum_{j=1}^{N} \lambda_j g_{ij})} - C_k = 0$$

Let

$$V(\lambda) = -\log \nu + \log Z(\lambda) - \sum_{j=1}^{N} \lambda_j C_j \tag{31}$$

By substituting (28) into (27), one sees that the optimization of (27) is equivalent to minimizing $V(\lambda)$. For the minimizing λ_k so determined, define the calibrated instrument:

$$\frac{\partial V(\lambda)}{\partial \lambda_k} = \sum_{i=1}^{\nu} p_i g_{ik} - C_k = \mathbb{E}^p(g_k(\omega)) - C_k \tag{32}$$

where $g_{ik} = g_k(\omega_i) = g_k(\omega)$ and $\mathbb{E}^p(g_k(\omega)) = \sum_{i=1}^{\nu} p_i g_k(\omega_i)$

6 Calibration Of The Jump Diffusion Model

Avellaneda models a bid-ask spread by minimizing the relative entropy and the sum of the weighted least-squares residuals:

$$\chi_w^2 = \frac{1}{2} \sum_{j=1}^{N} \frac{1}{w_j} (\mathbb{E}^p(g_j(w)) - C_j)^2 \tag{33}$$

where $w = (w_1, ..., w_N)$ is a vector of positive weights.

6.1 The Minimization

Minimize:

$$D(p \mid u) + \chi_w^2 \tag{34}$$

Preliminaries:

Denote $\mathbb{E}^p\{g_i(w))\}$ by E_i. Then $\chi_w^2 = \sum_{i=1}^N \frac{1}{w_i}(E_i - C_i)^2$.

Let

$$a_i = \frac{1}{\sqrt{w_i}}(E_i - C_i)$$

and

$$-b_i = \lambda_i \sqrt{w_i}$$

Utilizing the inequality $\frac{1}{2}a^2 + \frac{1}{2}b^2 \geq ab$ (right hand side is the inner product) and summing over i:

$$\frac{1}{2}\sum_{i=1}^N \frac{1}{w_i}(E_i - C_i)^2 + \frac{1}{2}\sum_{i=1}^N w_i \lambda_i^2 \geq -\sum_{i=1}^N \lambda_i (E_i - C_i)$$

i.e.

$$\chi_w^2 \geq -\sum_{i=1}^N \lambda_i (\mathbb{E}\{g_i(w)\} - C_i) - \frac{1}{2}\sum_{i=1}^N w_i \lambda_i^2$$

It follows that

$$\min_p [D(p \mid u) + \chi_w^2] \geq \max_\lambda \{\min_p [D(p \mid u) - \sum_{j=1}^N \lambda_j (\mathbb{E}^p\{g_j(w)\} - C_j)] - \frac{1}{2}\sum_{j=1}^N w_j \lambda_j^2\} \tag{35}$$

The next equality holds by the following logic: $max(x) = -min(-x)$.

$$= -\min_\lambda [-\min_p [D(p \mid u) - \sum_{j=1}^N \lambda_j (\mathbb{E}^{\text{'}}\{g_j(w)\} - C_j)] - \frac{1}{2}\sum_{j=1}^N w_j \lambda_j^2]$$

and since the last two terms are independent of p

$$= -\min_\lambda [\max_p [-D(p \mid u) + \sum_{j=1}^N \lambda_j \mathbb{E}^p\{g_j(w)\}] - \sum_{j=1}^N \lambda_j C_j + \frac{1}{2}\sum_{j=1}^N w_j \lambda_j^2]$$

It can be shown that the inequality is in fact an equality.

Now let D be the entropy rather than relative entropy $D = \sum_{i=1}^{\nu} p_i \log p_i$, where the term $\log \nu$ is dropped without loss of generality.

Then

$$\max_p[-D + \sum_{j=1}^{N} \lambda_j \mathbb{E}^p(g_j)] = \max_p[-D + \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{\nu} p_i g_{ij}]$$

$$= \max_p[-\sum_{i=1}^{\nu} p_i \log p_i + \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{\nu} p_i g_{ij}]$$

$$= -\sum_{i=1}^{\nu} p_i \sum_{j=1}^{N} \lambda_j g_{ij} + \sum_{i=1}^{\nu} p_i \log Z + \sum_{j=1}^{N} \lambda_j \sum_{i=1}^{\nu} p_i g_{ij}$$

where the last equality holds at the maximum $p = p^*$. This line now reduces to:

$$\sum_{i=1}^{\nu} p_i \log Z = \log Z = \log Z$$

since $\sum_{i=1}^{\nu} p_i = 1$) and Z does not depend on p.

Therefore

$$\min_p[D(p \mid u) + \chi_w^2] = -\min_\lambda[\log Z - \sum_{j=1}^{N} \lambda_j C_j + \frac{1}{2} \sum_{j=1}^{N} w_j \lambda_j^2] \tag{36}$$

$$= -\min_\lambda[V(\lambda) + \frac{1}{2} \sum_{j=1}^{N} w_j \lambda_j^2] \tag{37}$$

Here $V(\lambda) = \log Z(\lambda) - \sum_j \lambda_j C_j$ is the function used in the case of exact fitting.

Differentiating with respect to λ_k,

$$\frac{\partial V(\lambda)}{\partial \lambda_k} + w_k \lambda_k = \sum_{i=1}^{\nu} p_i g_{ik} - C_k + w_k \lambda_k$$

$$= \mathbb{E}^p\{g_k(w)\} - C_k + w_k\lambda_k = 0$$

and we have the optimal λ_k:

$$\lambda_k^* = -\frac{1}{w_k}[\mathbb{E}^{p^*}\{g_k(w)\} - C_k] \tag{38}$$

Note, the minimization over p is the same as in the case of exact fitting since χ_w^2 does not depend on p, and so leads to the same values of p_i^*:

$$p_i^* = \frac{1}{Z(\lambda_i^*)} e^{(\sum_{j=1}^N \lambda_j^* g_{ij})} \tag{39}$$

6.2 The Calibration Algorithm

The minimizing function in the case of least-squares fitting is

$$\log(Z(\lambda) - \sum_{j=1}^N \lambda_j(\mathbb{E}^{p^*}\{g_j(w)\} - C_j) + \frac{1}{2}\sum_{j=1}^N w_j\lambda_j^2 \tag{40}$$

This may be seen where, in the minimization of $W(\lambda)$, the term $\sum_{j=1}^N \lambda_j(\mathbb{E}^{p^*}\{g_j(w)\} - C_j)$ is substituted for $-\sum_{j=1}^N \lambda_j C_j$. The substitution occurs since our assumption $C_k = \mathbb{E}^p\{g_k(w)\}$ no longer holds.

The term $\mathbb{E}^{p^*}\{g_j(w)\} - C_j$ is precisely the modelled bid-ask spread, the bid-ask spread being a small constant-valued jump. This term, however, may be any constant value. C_k is a constant, the instantaneous price observed in the market. As an expected value, $\mathbb{E}^{p^*}\{g_j(w)\}$ is not a stochastic term. In fact, an expected value is a constant.

If we replace the bid-ask spread with a larger jump term, the minimization is essentially unchanged. The mispriced asset value represented by the bid-ask spread is replaced by a larger mispriced value representing a shock. This paper proposes that if the shock occurs as a compound Poisson process, one may replace the bid-ask spread by the expected value of the compound Poisson jump.

The exhibits of the next section illustrate the concepts. See [9], [11], and [14] for background material, basic concepts, and formulas.

7 An Example: Exhibits

U.S. Treasury Yield Curve Rates as of 12/16/02
Term Structure of Interest Rates

Maturity	Zero Price	Zero Yields (continuous)
1	**0.985**	0.015
2	**0.962**	0.0194
3	**0.932**	0.0235
5	**0.855**	0.0314
7	**0.769**	0.0375
10	**0.660**	0.0415

`

P = exp (-r * t)

Zero Yields for various maturities obtained from the U.S. Treasury website.

Zero Price calculated using continuous compounding and fit to a parabola with R squared = 99.75% to obtain Zero Price via function.

Fiting Parameters:	
y = c + b*(x-d)^2	
1.483766	c
-0.00053	b
-29.3765	d

Valuing Coupon Bonds	
0 (nowyear)	0
s (zeroyear)	10
Bond Face Value (L)	1.00
Bond Coupon (cL)	0.05

Bond value (using ytm)	Bond yield to maturity
1.072	4.03%

ln (1+IRR (FV cash flows))

Check On Bond value
1.072

time in years

Maturity	Zero Price (via function)	PV Bond Cash Flows	FV Bond Cash Flows	Zero Yields (via function)	Forward Rates (via function)	PV Bond Cash Flows (continuous)
0			-1.072			
1	0.992	0.050	0.050	0.84%	0.84%	0.048
2	0.959	0.048	0.050	2.11%	3.38%	0.046
3	0.925	0.046	0.050	2.61%	3.61%	0.044
4	0.890	0.044	0.050	2.93%	3.87%	0.043
5	0.853	0.043	0.050	3.17%	4.15%	0.041
6	0.816	0.041	0.050	3.38%	4.46%	0.039
7	0.778	0.039	0.050	3.59%	4.80%	0.038
8	0.739	0.037	0.050	3.79%	5.19%	0.036
9	0.698	0.035	0.050	3.99%	5.63%	0.035
10	0.657	0.690	1.050	4.20%	6.12%	0.702

fitted bond prices
these are PV

price * FV cashflow

coupons + face value
FV (full value) cash
flows are undiscounted

- ln (p_i) / t_i

- ln (p_i+1 / p_i)

CF * exp (-t_i * IRR)

Zero Yields and Forward Prices both based on Zero Price.

Exhibit 7.1

---- Descriptive Statistics for Variables ----

Variable	Minimum value	Maximum value	Mean value	Standard dev.
x	1	10	4.666667	3.386247
y	0.66	0.985	0.8605	0.1261091

---- Calculated Parameter Values ----

Parameter	Initial guess	Final estimate	Standard error	t	Prob(t)
c	1	1.48376593	0.5056422	2.93	0.06078
b	1	-0.000533394524	0.0004189297	-1.27	0.29264
d	1	-29.3764531	27.38353	-1.07	0.36199

---- Analysis of Variance ----

Source	DF	Sum of Squares	Mean Square	F value	Prob(F)
Regression	2	0.07931547	0.03965773	588.88	0.00013
Error	3	0.0002020314	6.73438E-005		
Total	5	0.0795175			

Fitting Program Output:

```
1: Title "Parabola";
2: Variables x,y;
3: Parameters c,b,d;
4: Function y = c + b*(x-d)^2;
5: plot;
6: data;
```

Beginning computation...

---- Final Results ----

NLREG version 5.3
Copyright (c) 1992-2002 Phillip H. Sherrod.

Parabola
Number of observations = 6
Maximum allowed number of iterations = 500
Convergence tolerance factor = 1.000000E-010
Stopped due to: Relative function convergence.
Number of iterations performed = 37
Final sum of squared deviations = 2.0203139E-004
Final sum of deviations = -3.3234526E-012
Standard error of estimate = 0.00820633
Average deviation = 0.00518032
Maximum deviation for any observation = 0.00895358
Proportion of variance explained (R^2) = 0.9975 (99.75%)
Adjusted coefficient of multiple determination (Ra^2) = 0.9958 (99.58%)
Durbin-Watson test for autocorrelation = 2.021
Analysis completed 29-Dec-2002 16:38. Runtime = 0.04 seconds.

Exhibit 7.2

Vasicek Model : see Hull (4th Edition) pp 567-9

RN model dr = a(b-r) dt + sr dz

a	**0.1779**	**Zero-coupon bond price**	
b	**0.0866**	P(0,s)	0.6006
r	**1.50%**	via fn	0.6006
0 (nowyr)	0.00		
s (zeroyr)	**10.00**	**Zero yield**	
zero life	10.00	R(0,s)	5.10%
σ_r	**2.00%**	via fn	5.10%
B(0,s)	4.6722	**Zero yield (infinite maturity)**	
A(0,s)	0.6442	R(∞)	8.02%
		Volatility of zero yield	
		σ_R(0,s)	0.93%

Cox, Ingersoll and Ross Model : see Hull (4th edition) pg 570

RN model dr = a(b-r) dt + s sqrt(r) dz

a	**0.2339**	**Zero-coupon bond price**	
b	**0.0808**	P(0,s)	0.5752
r	**1.50%**	via fn	0.5752
0 (nowyr)	0.00		
s (zeroyr)	**10.00**	**Zero yield**	
zero life	10.00	R(0,s)	5.53%
σ	**0.0200**	via fn	5.53%
γ	0.2356	**Zero yield (infinite maturity)**	
exp(γ(s-0))-1	9.5491	R(¥)	8.05%
B(0,s)	3.8547		
A(0,s)	0.6094	**Volatility of zero yield**	
		σP(0,σ)	0.09%

Vasicek Term Structure

Maturity	Vasicek Zero Yield	Zero Yield Volatility	Zero Price	Vasicek Forward Rate
	5.10%	0.93%		
0	1.50%	2.00%		
1	2.09%	1.83%	0.979	2.09%
2	2.61%	1.68%	0.949	3.13%
3	3.07%	1.55%	0.912	3.98%
4	3.47%	1.43%	0.870	4.68%
5	3.83%	1.32%	0.826	5.25%
6	4.14%	1.23%	0.780	5.72%
7	4.42%	1.14%	0.734	6.11%
8	4.67%	1.07%	0.688	6.43%
9	4.90%	1.00%	0.644	6.69%
10	5.10%	0.93%	0.601	6.91%

CIR Term Structure

Maturity	CIR Zero Yield	Zero Yield Volatility	Zero Price	CIR Forward Rate
	5.53%	0.09%		
0	1.50%	0.24%		
1	2.21%	0.22%	0.978	2.21%
2	2.82%	0.20%	0.945	3.44%
3	3.35%	0.18%	0.904	4.40%
4	3.80%	0.16%	0.859	5.17%
5	4.20%	0.14%	0.811	5.77%
6	4.54%	0.13%	0.762	6.25%
7	4.84%	0.12%	0.713	6.63%
8	5.10%	0.11%	0.665	6.93%
9	5.33%	0.10%	0.619	7.16%
10	5.53%	0.09%	0.575	7.35%

Exhibit 7.3

Using Monte Carlo Simulation to Value Zero Yields

Random Numbers (Excel)

Vasicek stochastic DE	
1st term	0.00127
2nd term	0.02000
CIR stochastic DE	
1st term	0.00154
2nd term	0.00245

r + dr	Vasicek	CIR
MC value	**0.0120**	**0.0160**
MC stdev	**0.0214**	**0.0026**

Vasicek Model : see Hull (4th Edition) p567-9

RN model dr = a(b-r) dt + sr dz

a	0.1779	**Zero-coupon bond price**	
b	0.0866	P(0,s)	0.6006
r	1.50%	MC Value	0.6089
0 (nowyr)	0.00	**Zero yield**	
s (zeroyr)	10.00	R(0,s)	5.10%
zero life	10.00	MC value	4.96%
σ_r	2.00%		
dt	0.10	**Volatility of zero yield**	
B(0,s)	4.6722	$\sigma_R(0,s)$	0.93%
A(0,s)	0.6442	MC value	1.00%

Cox, Ingersoll and Ross Model : see Hull (4th edition) p570

RN model dr = a(b-r) dt + s sqrt(r) dz

a	0.2339	**Zero-coupon bond price**	
b	0.0808	P(0,s)	0.5752
r	1.50%	MC Value	0.5729
0 (nowyr)	0.00	**Zero yield**	
s (zeroyr)	10.00	R(0,s)	5.53%
zero life	10.00	MC Value	5.57%
σ_r	2.00%		
dt	0.10		
γ	0.2356	**Volatility of zero yield**	
exp(γ(s-0))-	9.5491	$\sigma_R(0,s)$	0.09%
B(0,s)	3.8547	MC Value	0.01%
A(0,s)	0.6094		

Exhibit 7.4

Using Monte Carlo Simulation to Value Zero Yields

Quasi Random Numbers

Vasicek stochastic DE	
1st term	0.00127
2nd term	0.02000
CIR stochastic DE	
1st term	0.00154
2nd term	0.00245

r + dr	**Vasicek**	**CIR**
MC value	0.0154	0.0164
MC stdev	0.0194	0.0024

Vasicek Model : see Hull (4th Edition) pp 567-9
RN model dr = a(b-r) dt + sr dz

a	0.1779	**Zero-coupon bond price**	
b	0.0866	P(0,s)	0.6006
r	1.50%	MC Value	0.5993
0 (nowyr)	0.00	**Zero yield**	
s (zeroyr)	10.00	R(0,s)	5.10%
zero life	10.00	MC value	5.12%
σ_r	2.00%		
dt	0.10	**Volatility of zero yield**	
B(0,s)	4.6722	$\sigma_R(0,s)$	0.93%
A(0,s)	0.6442	MC value	0.91%

Cox, Ingersoll and Ross Model : see Hull (4th edition) pg 570
RN model dr = a(b-r) dt + s sqrt(r) dz

a	0.2339	**Zero-coupon bond price**	
b	0.0808	P(0,s)	0.5752
r	1.50%	MC Value	0.5720
0 (nowyr)	0.00	**Zero yield**	
s (zeroyr)	10.00	R(0,s)	5.53%
zero life	10.00	MC Value	5.59%
σ_r	2.00%		
dt	0.10		
γ	0.2356	**Volatility of zero yield**	
exp(γ(s-0))-	9.5491	$\sigma_R(0,s)$	0.09%
B(0,s)	3.8547	MC Value	0.01%
A(0,s)	0.6094		

Exhibit 7.5

Using Monte Carlo Simulation to Value Bond Prices
Comparison of Results
MRE (Minimum Relative Entropy) Prior Distribution

Zero Yield

time in years	Vasicek equation	Vasicek MC1	Vasicek MC2	CIR equation	CIR MC1	CIR MC2
0	1.50%	1.64%	1.42%	1.50%	1.45%	1.49%
1	2.09%	1.93%	2.03%	2.21%	2.23%	2.22%
2	2.61%	2.39%	2.57%	2.82%	2.84%	2.84%
3	3.07%	3.12%	3.03%	3.35%	3.41%	3.38%
4	3.47%	3.61%	3.45%	3.80%	3.83%	3.84%
5	3.83%	3.93%	3.81%	4.20%	4.26%	4.24%
6	4.14%	4.17%	4.14%	4.54%	4.57%	4.58%
7	4.42%	4.47%	4.43%	4.84%	4.89%	4.89%
8	4.67%	4.79%	4.68%	5.10%	5.16%	5.15%
9	4.90%	5.03%	4.91%	5.33%	5.39%	5.38%
10	5.10%	5.27%	5.12%	5.53%	5.59%	5.59%
Bond Price	**0.54706**	**0.54012**	**0.54478**	**0.52858**	**0.52478**	**0.52521**

Using Monte Carlo Simulation to Value Bond Prices
Comparison of Results
MRE (Minimum Relative Entropy) Prior Distribution

Forward Rate

time in years	Vasicek equation	Vasicek MC1	Vasicek MC2	CIR equation	CIR MC1	CIR MC2
0	2.09%	1.93%	2.03%	2.21%	2.23%	2.22%
1	2.09%	1.93%	2.03%	2.21%	2.23%	2.22%
2	3.13%	2.85%	3.10%	3.44%	3.45%	3.46%
3	3.98%	4.59%	3.97%	4.40%	4.57%	4.45%
4	4.68%	5.08%	4.69%	5.17%	5.08%	5.22%
5	5.25%	5.22%	5.27%	5.77%	5.97%	5.84%
6	5.72%	5.37%	5.76%	6.25%	6.12%	6.32%
7	6.11%	6.27%	6.15%	6.63%	6.82%	6.70%
8	6.43%	7.05%	6.48%	6.93%	7.05%	7.00%
9	6.69%	6.95%	6.75%	7.16%	7.23%	7.24%
10	6.91%	7.35%	6.98%	7.35%	7.39%	7.42%

Exhibit 7.6

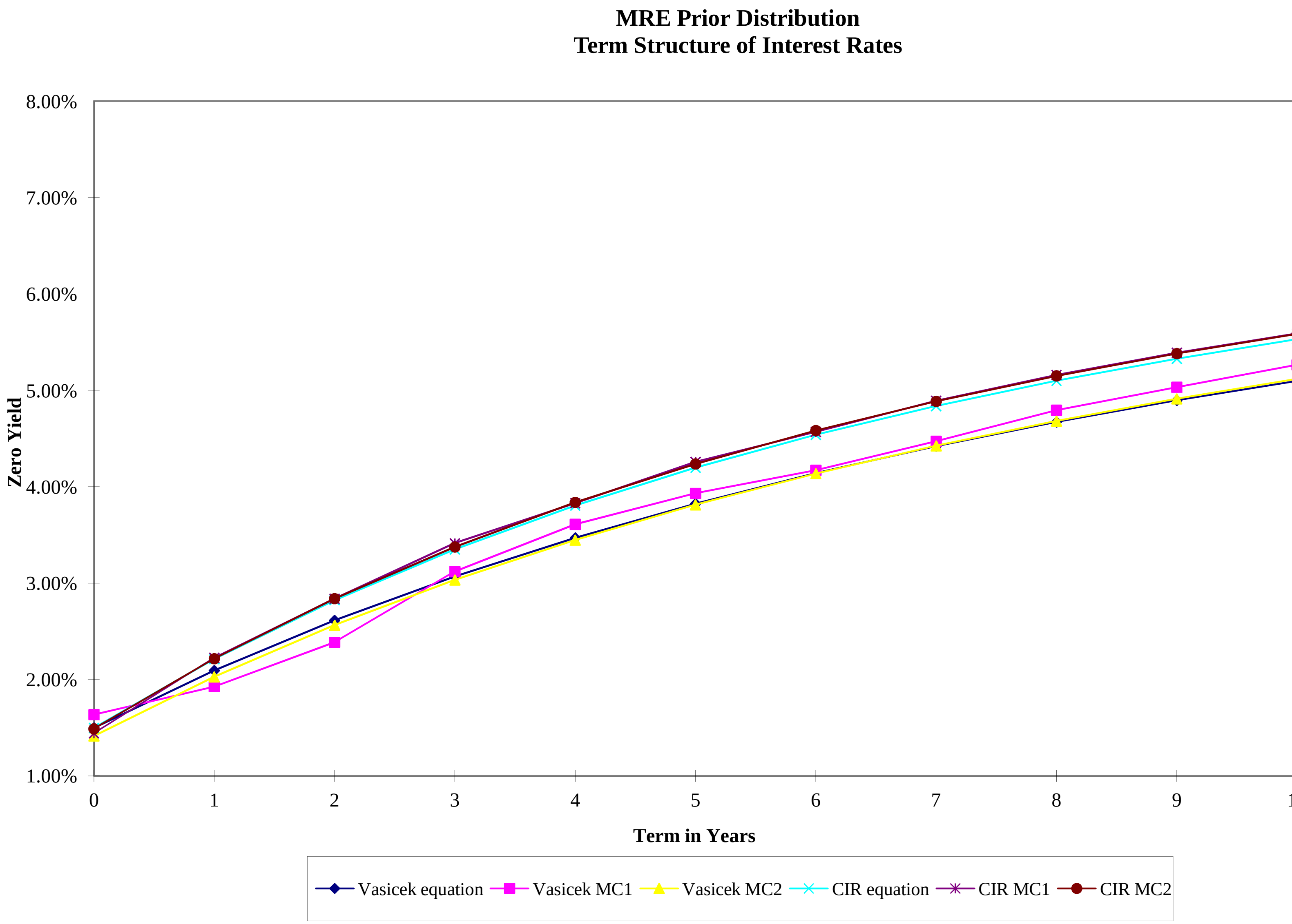

MRE Prior Distribution
Term Structure of Interest Rates
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
Zero Yield
0
1
2
3
4
5
6
7
8
9
10
Term in Years
Vasicek equation
Vasicek MC1
Vasicek MC2
CIR equation
CIR MC1
CIR MC2

Exhibit 7.7

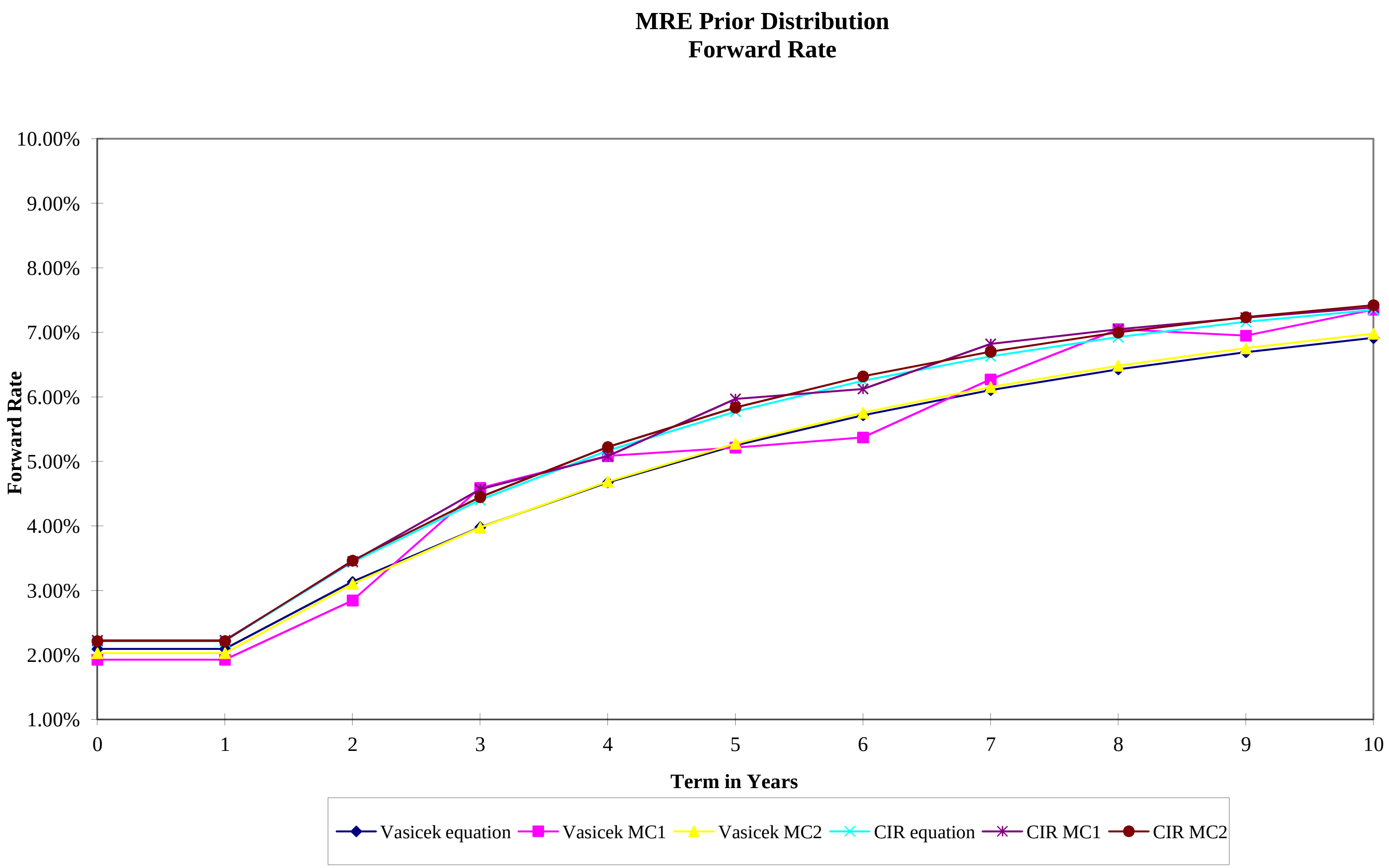

MRE Prior Distribution
Forward Rate
10.00%
9.00%
8.00%
7.00%
6.00%
5.00%
4.00%
3.00%
2.00%
1.00%
Forward Rate
0
1
2
3
4
5
6
7
8
9
10
Term in Years
Vasicek equation
Vasicek MC1
Vasicek MC2
CIR equation
CIR MC1
CIR MC2

Exhibit 7.8

Initial Bond Portfolio Value	0.522890

Final Bond Portfolio Value	0.506413

Benchmark instruments N=5, nu=11

price C_1	PV cash flow g_1j	price C_2	PV cash flow g_2j	price C_3	PV cash flow g_3j	price C_4	PV cash flow g_4j	price C_5	PV cash flow g_5j
0.693	0.059	0.608	0.050	0.523	0.040	0.438	0.030	0.353	0.020
	0.058		0.048		0.038		0.029		0.019
	0.055		0.046		0.037		0.028		0.018
	0.053		0.044		0.036		0.027		0.018
	0.051		0.043		0.034		0.026		0.017
	0.049		0.041		0.033		0.024		0.016
	0.047		0.039		0.031		0.023		0.016
	0.044		0.037		0.030		0.022		0.015
	0.042		0.035		0.028		0.021		0.014
	0.039		0.033		0.026		0.020		0.013
	0.194		0.192		0.191		0.189		0.187

The Vector Of Benchmark Prices C_j

0.692669	0.607779	0.522890	0.438000	0.353111

The Vector Of Final Values p_i

0.173543	0.171769	0.193366	0.209694	0.251628

The Matrix Of Present Valued Cash Flows g_ij

0.059495	0.049579	0.039663	0.029748	0.019832
0.057519	0.047932	0.038346	0.028759	0.019173
0.055479	0.046232	0.036986	0.027739	0.018493
0.053374	0.044479	0.035583	0.026687	0.017791
0.051206	0.042672	0.034137	0.025603	0.017069
0.048974	0.040811	0.032649	0.024487	0.016325
0.046677	0.038898	0.031118	0.023339	0.015559
0.044317	0.036931	0.029545	0.022158	0.014772
0.041892	0.034910	0.027928	0.020946	0.013964
0.039404	0.032837	0.026269	0.019702	0.013135
0.194332	0.192498	0.190665	0.188832	0.186999

The Matrix Of Lagrange Multipliers lambda*_j

-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000
-0.026586	-0.047202	0.171634	0.389973	1.000000

Exhibit 7.9

MRE: Find LaGrange Multipliers
Maximum allowed number of iterations = 500
Convergence tolerance factor = 1.000000E-010

Number of iterations performed = 5
Final function value = 4.9005938E-017
Analysis completed 2-Jan-2003 07:54. Runtime = 0.02 seconds.

---- Calculated Parameter Values ----

Parameter	Initial guess	Final estimate
L1	1	-0.26586057
L2	1	-0.0472018362
L3	1	0.171634464
L4	1	0.389972719
L5	1	1

Exhibit 7.10

Expected Value Of The Jump	0.255222

Severity	***Frequency***
0.10	6
0.15	263
0.20	741
0.25	765
0.30	503
0.35	265
0.40	111
0.45	31
0.50	13
0.55	2
0.60	0

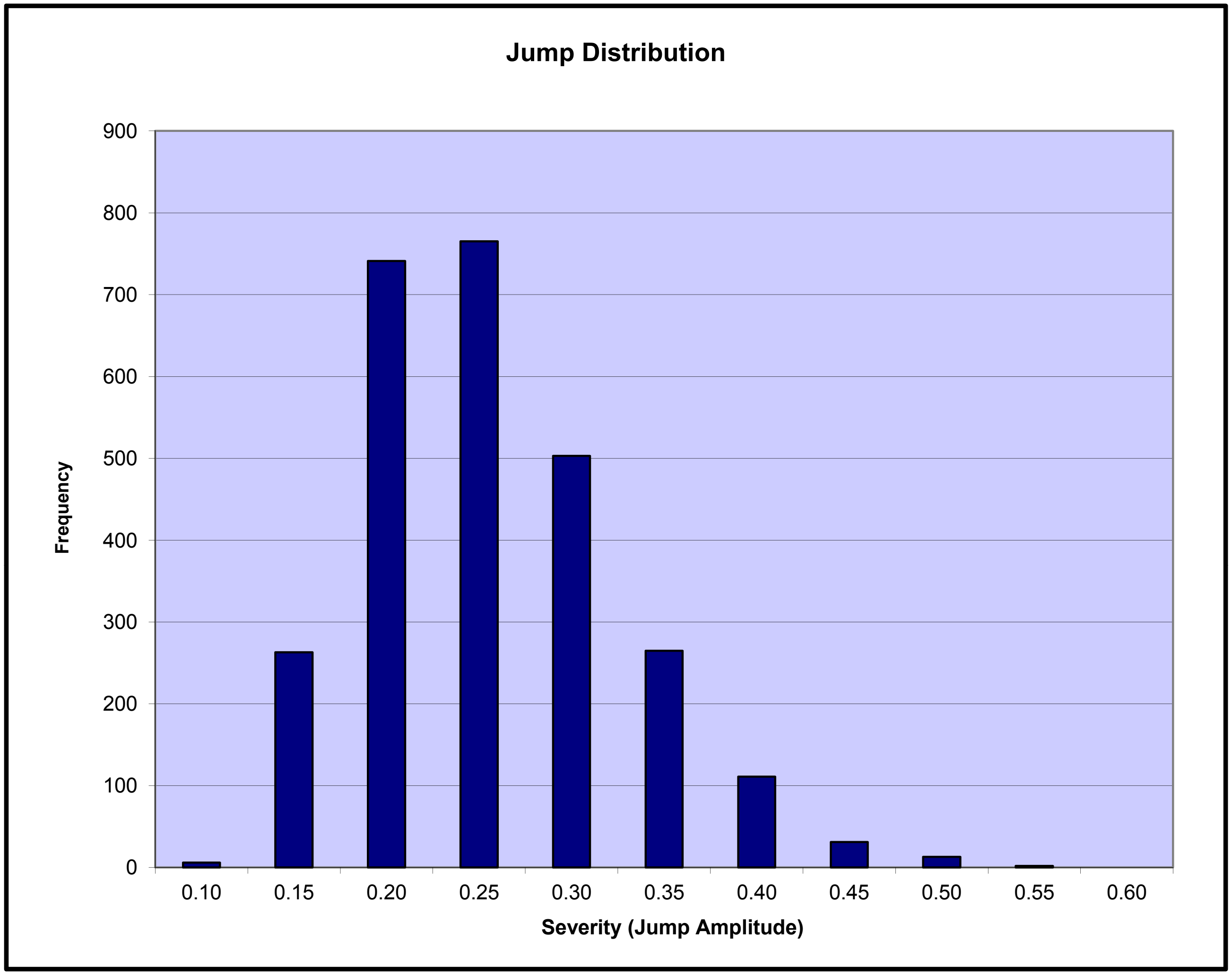


Exhibit 7.11

Initial Bond Portfolio Value	0.522890

Final Bond Portfolio Value	0.504591

Benchmark instruments N=5, nu=11

price C_1	PV cash flow g_1j	price C_2	PV cash flow g_2j	price C_3	PV cash flow g_3j	price C_4	PV cash flow g_4j	price C_5	PV cash flow g_5j
0.693	0.059	0.608	0.050	0.523	0.040	0.438	0.030	0.353	0.020
	0.058		0.048		0.038		0.029		0.019
	0.055		0.046		0.037		0.028		0.018
	0.053		0.044		0.036		0.027		0.018
	0.051		0.043		0.034		0.026		0.017
	0.049		0.041		0.033		0.024		0.016
	0.047		0.039		0.031		0.023		0.016
	0.044		0.037		0.030		0.022		0.015
	0.042		0.035		0.028		0.021		0.014
	0.039		0.033		0.026		0.020		0.013
	0.194		0.192		0.191		0.189		0.187

The Vector Of Benchmark Prices C_j

0.692669	0.607779	0.522890	0.438000	0.353111

The Vector Of Final Values p_i

0.159167	0.177378	0.197673	0.220289	0.245493

The Matrix Of Present Valued Cash Flows g_ij

0.059495	0.049579	0.039663	0.029748	0.019832
0.057519	0.047932	0.038346	0.028759	0.019173
0.055479	0.046232	0.036986	0.027739	0.018493
0.053374	0.044479	0.035583	0.026687	0.017791
0.051206	0.042672	0.034137	0.025603	0.017069
0.048974	0.040811	0.032649	0.024487	0.016325
0.046677	0.038898	0.031118	0.023339	0.015559
0.044317	0.036931	0.029545	0.022158	0.014772
0.041892	0.034910	0.027928	0.020946	0.013964
0.039404	0.032837	0.026269	0.019702	0.013135
0.194332	0.192498	0.190665	0.188832	0.186999

The Matrix Of Lagrange Multipliers lambda*_j

-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111
-1.276111	-1.276111	-1.276111	-1.276111	-1.276111

Exhibit 7.12

References

[1] M. Avellaneda, R. Buff, C. Friedman, N. Grandchamp, L. Kruk, J. Newman, (2001) Weighted Monte Carlo: a new technique for calibrating asset pricing models, International Journal of Theoretical and Applied Finance Bull. London Math. Soc., Vol. 4, No. 1, pp 91-119, Available at SSRN: http://ssrn.com/abstract=1759452.

[2] R.E. Beard, T. Pentikainen, E. Pesonen, (1984), Risk Theory: The Stochastic Basis of Insurance, Third Edition, Monographs on Statistics and Applied Probability, Chapman and Hall, Chapter 8.

[3] R. Cont, P. Tankov, (2004), Non-parametric calibration of jump-diffusion processes, Journal of Computational Finance, Vol .7, No 3, pp 1-49, Working paper available at SSRN: http://ssrn.com/abstract=332400.

[4] J. Cox, J. Ingersoll, Jr., S. Ross, (1985), A Theory of the Term Structure of Interest Rates, Econometrica, Vol. 53, No. 2, pp 385-408.

[5] R. Craine, L. A. Lochstoer, K. Syrtveit, (2000), Estimation of a Stochastic-Volatility Jump-Diffusion Model, Revista de Analisis Economico, 2000, (15), pp 61-87 Available at http://www.columbia.edu/ ll2609/.

[6] H. Cramer, (1930), Collective Risk Theory, Skandia Jubilee Volume, Stockholm, pp 46-83.

[7] William Feller, (1950), An Introduction to Probability Theory and Its Applications, Volume I, John Wiley and Sons, Inc., pp 82-83.

[8] P. Heckman, G. Meyers, (1983), The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions, Proceedings of the Casualty Actuarial Society, PCAS LXX, pp 22-61.

[9] J.C. Hull, (2000), Options, Futures, and Other Derivatives, Fourth Edition, Prentice Hall

[10] K. Itō, H. P. McKean Jr., (1974), Diffusion Processes and Their Sample Paths, Springer-Verlag, pp 31-33.

[11] F. Jamshidian, (1989), An Exact Bond Option Formula, Journal of Finance, Volume 44, No. 1, pp 205-209.

[12] P. Lévy, (1954), Théorie de l'addition des variables aléatoires, Paris, 2nd edition, pp 173-180.

[13] R. C. Merton, (1976), Option Pricing When Underlying Stock Returns are Discontinuous, Journal of Financial Economics, Vol. 3, pp 125-144.

[14] (1977), O. Vašíček, An equilibrium characterization of term-structure, Journal of Financial Economics, Volume 5, pp 177-188.

[15] C. Zhou, (1997) A Jump-Diffusion Approach to Modeling Credit Risk and Valuing Defaultable Securities, Available at SSRN: http://ssrn.com/abstract=39800.