Implementation of PH-Transforms in Ratemaking

by Shaun Wang, Ph.D.

— Discussion Paper for the 1997 Casualty Actuarial Society Ratemaking Seminar

IMPLEMENTATION OF PH-TRANSFORMS IN RATEMAKING

SHAUN WANG, Ph.D.

ABSTRACT

In this article we introduce a relatively new method for deciding contingency provisions in insurance ratemaking by the use of proportional hazard(PH) transforms. This method is easy to understand, simple to use, and supported by theoretical properties as well as economic justification. After an introduction of the PH-transform method, we show through examples how it can be used in pricing ambiguous risks, excess-of-loss coverages, increased limits, and risk portfolios with dependency risk. We also show how a minimum rate-on-line can be achieved. As well, we propose a right-tail index for insurance risks.

1 INTRODUCTION

Recently, there has been considerable interest in and extensive discussion on risk loads by Fellows of the Casualty Actuarial Society. These discussions have focused on what measures a risk and methods to arrive at a 'reasonable' risk load. Although there are diverse opinions on the appropriate measurement of risk, there is general agreement on the distinction between process risks and parameter risks, and on the importance of parameter risks in ratemaking; see Finger (1976), Miccolis (1977), McClenahan (1990), Feldblum (1990), Philbrick (1991), Meyers (1991) and Robbin (1992).

Following Venter's (1991) advocacy of adjusted distribution methods, Wang (1995) proposes using proportional hazard (PH) transforms in the calculation of risk-adjusted premiums. Although extensive discussion on the economic justifications is valuable, this paper focuses on the practical aspects of implementation of PH-transforms in ratemaking. More specifically, we will show how it can be used to quantify process risks, parameter risks and dependency risks.

Consistent with previous papers, this paper will consider only pure premiums, excluding all expenses and commissions. To utilize the PH-transform in ratemaking, a probability distribution for the insurance claims is needed. With the advent of computerized technology, a probability distribution can often be estimated from industry claim data or by computer simulations. Even though a probability distribution can be obtained from past claim data, sound and knowledgeable judgements are always required to ensure that the estimated loss distribution is valid for ratemaking.

It is safe to say that no theoretical risk-load formula can claim to be the *right* one, since subjective elements always exist in any practical exercise of ratemaking. However, a good theoretical risk-load formula can assist actuaries and help maintain logical consistency in the ratemaking process. In this respect, it is hoped that the PH-transform method offers a useful tool to practicing actuaries in insurance ratemaking.

The remainder of this paper is divided into three sections. Section 2 introduces the PH-transform method and applies it to pricing of ambiguous risks, excess-of-loss layers, increased limits and risk portfolios. Section 3 discusses two simple mixtures of PH-transforms. The first mixture can yield a minimal rate-on-line, and the second mixture suggests a new index for the right tail risk. Section 4 briefly reviews the leading economic theories of risk and uncertainty, and their relations with insurance ratemaking.

2 PROPORTIONAL HAZARD TRANSFORM

An insurance risk X refers to a non-negative loss random variable, which can be described by the decumulative distribution function (ddf): $S_X(t) = \Pr\{X > t\}$. An advantage of using the ddf is the unifying treatment of discrete, continuous and mixed-type distributions. In general, for a risk X, the expected loss can be evaluated directly from its ddf:

$$\mathrm{E}(X) = \int_0^\infty S_X(t)dt.$$

Definition 1 *Given a best-estimate loss distribution $S_X(t) = Pr\{X > t\}$, for some exogenous* **index** r *($0 \leq r \leq 1$), the* **proportional hazard (PH)** *transform refers to a mapping $S_Y(t) := [S_X(t)]^r$, and the* **PH-mean** *refers to the expected value under the transformed distribution:*

$$\mathrm{H}_r(X) = \int_0^\infty [S_X(t)]^r dt, \qquad (0 \leq r \leq 1).$$

The PH-mean was introduced by Wang (1995) to represent risk-adjusted premiums.

Example 1: The following three loss distributions

$$\begin{aligned} S_U(t) &= 1 - \tfrac{1}{2b}t, \quad 0 \leq t \leq 2b && \text{(uniform)} \\ S_V(t) &= e^{-\frac{t}{b}} && \text{(exponential)} \\ S_W(t) &= (\tfrac{b}{b+t})^2 && \text{(Pareto)}, \end{aligned}$$

have the same expected loss, b. One can easily verify that

$$\mathrm{H}_r(U) = \frac{2b}{1+r}, \qquad \mathrm{H}_r(V) = \frac{b}{r}, \qquad \mathrm{H}_r(W) = \begin{cases} \frac{b}{2r-1}, & r > 0.5; \\ \infty, & r \leq 0.5. \end{cases}$$

Table 1: Some values of PH-mean $\mathrm{H}_r(.)$

	U	V	W
$r_1 = \frac{5}{6}$	$1.09\,b$	$1.2\,b$	$1.5\,b$
$r_2 = \frac{2}{3}$	$1.2\,b$	$1.5\,b$	$3.0\,b$

The PH-mean, interpreted as risk-adjusted premium, preserves the ordering of relative riskiness among those three distributions (see Table 1).

Example 2: When X has a Pareto distribution with parameters (α, λ):

$$S_X(t) = (\frac{\lambda}{\lambda + t})^{\alpha},$$

the PH-transform $S_Y(t)$ also has a Pareto distribution with parameters $(r\alpha, \lambda)$.

When X has a Burr distribution with parameters (α, λ, τ):

$$S_X(t) = (\frac{\lambda}{\lambda + t^{\tau}})^{\alpha},$$

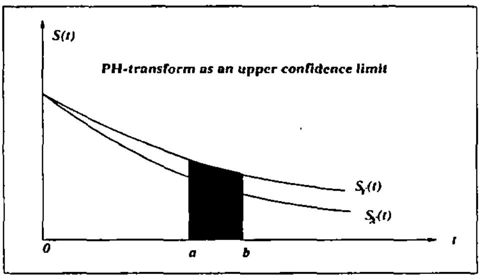
the PH transform $S_Y(t)$ also has a Burr distribution with parameters $(r\alpha, \lambda, \tau)$.

When X has a gamma (or log-normal) distribution, the PH transform $S_Y(t)$ is no longer a gamma (or log-normal). In such cases, numerical integration may be required to evaluate the PH-mean.

2.1 Pricing of Ambiguous Risks

In practice, the underlying loss distribution is seldom known with precision. There are always uncertainties regarding the best-estimate loss distribution. Insufficient data or poor-quality data often results in sampling errors. Even if a large amount of high-quality data is available, due to changes in the claim generating mechanisms, past data may not fully predict the the future claim distribution.

Figure 1: Margins for parameter uncertainty by PH-transforms



As illustrated in Figure 1, the PH-transform, $S_Y(t) = [S_X(t)]^r$, can be viewed as an upper confidence limit for the best-estimate loss distribution $S_X(t)$. A smaller index r yields a wider range between the curves S_Y and S_X. This upper confidence limit interpretation has support in a statistical estimation theory (see Appendix). The index r can be assigned accordingly with respect to the level of confidence in the estimated loss distribution. The more ambiguous the situation is, the lower the value of r should be used.

Example 3: Consider the following experiment conducted by Hogarth and Kunreuther (1992). An actuary is asked to price warranties on the performance of 10,000 units of a new line of microcomputers. Suppose that the cost of repair is \$100 per unit, and there can be at most one breakdown per period. Also, suppose that the risks of breakdown associated with any two units are independent. The best-estimate of the probability of breakdown has three scenarios:

$$\theta = 0.001, \qquad \theta = 0.01, \qquad \theta = 0.1.$$

The level of confidence regarding the best estimate has two scenarios:

Non-ambiguous: There is little ambiguity regarding the best-estimate loss distribution. Experts all agree with confidence on the chances of a breakdown.

Ambiguous: There is considerable ambiguity regarding the best-estimate loss distribution. Experts disagree and have little confidence in the estimate of the probabilities of a breakdown.

Note that the loss associated with a computer component can only assume two possible values, either zero or \$100. For any fixed $t < 100$, the probability that the loss exceeds t is the same as the probability of being exactly \$100, θ. For a fixed $t \geq 100$, it is impossible that the loss exceeds t. Thus, the best-estimate ddf of the insurance loss cost is

$$S_X(t) = \begin{cases} \theta, & 0 < t < 100; \\ 0, & 100 \leq t. \end{cases}$$

A PH-transform with index r yields a risk-adjusted premium at $100\theta^r$. If we choose $r = 0.97$ for the non-ambiguous case, and $r = 0.87$ for the ambiguous case, we get the following premium structures as in Table 2:

Table 2: The ratio of the risk-adjusted premium to the expected loss

	$\theta = 0.001$	$\theta = 0.01$	$\theta = 0.1$
Non-ambiguous ($r = 0.97$)	1.23	1.15	1.07
Ambiguous ($r = 0.87$)	2.45	1.82	1.35

In summary, the PH-transform can be used as a means of provision for estimation errors. The actuary can subsequently set up a table for the index r according to different levels of ambiguity, such as the following:

Ambiguity Level	Index r
Non-ambiguous	0.96-1.00
Slightly ambiguous	0.90-0.95
Moderately ambiguous	0.80-0.89
Highly ambiguous	0.50-0.79
Extremely ambiguous	0.00-0.49

2.2 Pricing of Excess-of-Loss Layers

Since most practical contracts contain clauses such as a deductible and a maximum limit, it is convenient to use the general language of excess-of-loss layers. A layer $(a, a+h]$ of a risk X is defined by the loss function:

$$I_{(a,a+h]} = \begin{cases} 0, & 0 \le X < a; \\ (X-a), & a \le X < a+h; \\ h, & a+h \le X, \end{cases}$$

where a is the **attachment point (retention)**, and h is the **limit**.

One can verify that the loss variable $I_{(a,a+h]}$ has a ddf:

$$S_{I_{(a,a+h]}}(t) = \begin{cases} S_X(a+t), & 0 \le t < h \\ 0, & h \le t, \end{cases}$$

and that the average loss cost for the layer $(a, a+h]$ is

$$\mathrm{E}[I_{(a,a+h]}] = \int_0^h S_X(a+t)dt = \int_a^{a+h} S_X(t)dt.$$

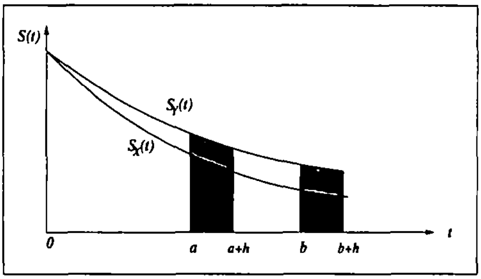
Note that $S_X(t)dt$ represents the net premium for an infinitesimal layer at $(t, t+dt]$. Thus, the ddf $S_X(t)$ plays an important role of *layer net premium density*.

Under the PH-transform $S_Y(t) = [S_X(t)]^r$, the PH-mean for the layer $(a, a+h]$ is

$$\mathrm{H}_r(I_{(a,a+h]}) = \int_0^{\infty} [S_{I_{(a,a+h]}}(t)]^r dt = \int_0^h [S_X(a+t)]^r dt = \int_a^{a+h} [S_X(t)]^r dt.$$

In other words, the net premium and the risk-adjusted premium for the layer $(a, a+h]$ are represented by the areas over the interval $(a, a+h]$ under the curves $S_X(t)$ and $S_Y(t)$, respectively (see Figure 2).

Figure 2: Risk load by layers: an illustration



In Wang (1995), it is shown that, for $0 < r < 1$, the PH-mean has the following properties:

- Positive loading:

$$\mathrm{H}_r(I_{(a,a+h]}) > \mathrm{E}(I_{(a,a+h]}).$$

- Decreasing risk-adjusted premiums:

$$\text{For } a < b, \qquad \mathrm{H}_r(I_{(a,a+h]}) > \mathrm{H}_r(I_{(b,b+h]}).$$

- Increasing relative loading:

$$\text{For } a < b, \qquad \frac{\mathrm{H}_r(I_{(a,a+h]})}{\mathrm{E}(I_{(a,a+h]})} < \frac{\mathrm{H}_r(I_{(b,b+h]})}{\mathrm{E}(I_{(b,b+h]})}.$$

These properties are consistent with market premium structures (Patrick, 1990; Venter, 1991).

Example 4 : A risk has a 10% chance of incurring a claim, and if a claim occurs the claim size has a Pareto distribution ($\lambda = 2,000$, $\alpha = 1.2$). Putting frequency and severity together, we have

$$\begin{aligned} S_X(t) &= \Pr\{X > t\} \\ &= \text{Probability of occurrence} \times \Pr\{\text{Loss Size } > t\} \\ &= 0.1 \times (\tfrac{2000}{2000+t})^{1.2}. \end{aligned}$$

Suppose that, the actuary infers an index, say $r = 0.833$, from individual risk analysis and market conditions. The actuary may need to compare with the risk loads for other contracts with similar characteristics in the market. The PH-transform with $r = 0.833$ yields a ddf:

$$S_Y(t) = 0.1^{0.833} \times (\frac{2000}{2000+t})^{1.2\times 0.833},$$

which produces risk-adjusted layer premiums as shown in Table 3.

Table 3: Layer premiums using PH-transforms

Layer	Net Premium	Risk-adjusted Premium	Percentage Loading
(0, 1000]	77.892	119.129	53%
(5000, 6000]	20.512	39.250	91%
(10000, 11000]	11.098	23.533	112%
(50000, 51000]	1.982	5.603	183%
(100000, 101000]	0.888	2.870	223%
(500000, 501000]	0.132	0.587	345%
(1000000, 1001000]	0.058	0.294	412%

2.3 Increased Limits Ratemaking

In commercial liability insurance, a policy generally covers a loss up to a specified maximum dollar amount that will be paid to any individual loss.

It is general practice to publish rates for some standard limit called the basic limit (used to be \$25,000 and nowadays \$100,000). Increased limit rates are calculated using a multiple factor, called the increased limit factors (ILFs). Without risk load, the increased limit factor is the expected loss at the increased limit divided by the expected loss at the basic limit. The increased limit factor with risk load is the sum of the expected loss and the risk load at the increased limit divided by the sum of the expected loss and the risk load at the basic limit:

$$\mathrm{ILF}(\omega) = \frac{\mathrm{E}[X;\omega] + \mathrm{RL}_{(0,\omega]}}{\mathrm{E}[X;\ 100,000] + \mathrm{RL}_{(0,\,100,000]}}.$$

It is widely felt that ILFs should satisfy the following conditions (Rosenberg, 1977; Meyers, 1991; Robbin, 1992):

1. The relative loading with respect to the expected loss is higher for increased limits.

2. ILFs should produce the same price under any arbitrary division of layers.

3. The ILFs should exhibit a pattern of declining marginal increases as the limit of coverages is raised. In other words, when $x < y$,

$$\mathrm{ILF}(x+h) - \mathrm{ILF}(x) \geq \mathrm{ILF}(y+h) - \mathrm{ILF}(y),$$

In the U.S., most companies use the Insurance Service Office (ISO) published ILFs. Traditionally, only the severity distribution is used (ISO assumes a Pareto loss severity distribution) when producing ILFs. Until the mid-1980's, ISO used the variance of the loss distribution to calculate risk loads, a method proposed by Robert S. Miccolis (1977). From mid-1980's to early 1990's, ISO used the standard deviation of the loss distribution to calculate risk loads (e.g. Feldblum, 1990). Meyers (1991) presents a Competitive Market Equilibrium approach, which yields a variance-based risk load method; however, some authors have questioned the appropriateness of the variance-based risk load method for the calculation of ILFs (e.g. Robbin, 1992).

The following is an illustrative example to show how the PH-transform method can be used in increased limits ratemaking.

Example 5: Assume that the claim severity distribution has a Pareto distribution with ddf:

$$S_X(t) = \left(\frac{\lambda}{\lambda + t}\right)^{\alpha},$$

with $\lambda = 5,000$ and $\alpha = 1.1$. This is the same distribution used by Meyers (1991), although he also considered parameter uncertainty.

Assume that. based on the market premium structure, the actuary feels that (for illustration only) an index $r = 0.8$ provides an appropriate provision for parameter uncertainty. When using a Pareto severity distribution, there is a simple analytical formula for the ILFs:

$$\mathrm{ILF}(\omega) = \frac{1 - (\frac{\lambda}{\lambda+\omega})^{r\alpha-1}}{1 - (\frac{\lambda}{\lambda+100,000})^{r\alpha-1}}.$$

One can then easily calculate the increased limit factors at any limit (see Table 4).

Table 4: Increased limit Factors using PH-transforms

Policy Limit ω	Expected Loss $E[X;\omega]$	ILF Without RL	Risk Load	ILF With RL
100000	13124.	1.00	5251.	1.00
250000	16255.	1.24	8866.	1.37
500000	18484.	1.41	12344.	1.68
750000	19726.	1.50	14687.	1.87
1000000	20579.	1.57	16490.	2.02
2000000	22543.	1.71	21330.	2.39

2.4 Pricing of Risk Portfolios and Dependency Risk

For ratemaking based on the aggregate claims from a risk portfolio, the actuary often considers claim frequency and claim severity separately, due to the type of information available.

Let N denote the claim frequency with probability function $f_N(k) = \Pr\{N = k\}$ and ddf: $S_N(k) = f(k+1) + f(k+2) + \cdots$, $(k = 0, 1, 2, \cdots)$.

Let X denote the claim severity and let

$$Z = X_1 + X_2 + \cdots X_N = \sum_{i=1}^{N} X_i$$

represent the aggregate claims from the risk portfolio.

Depending on the available information, the actuary may have different levels of confidence in the estimates for the frequency and severity distributions. According to the level of confidence in the estimated frequency and severity distributions, the actuary can choose an index r_1 for the frequency and an index r_2 for the severity. As a result, the actuary can calculate the risk-adjusted premium for the risk portfolio as:

$$\mathrm{H}(Z) = \mathrm{H}_{r_1}(N) \times \mathrm{H}_{r_2}(X).$$

Example 6: Consider a group coverage of liability insurance. The actuary has estimated the following loss distributions: (i) the claim frequency has a Poisson distribution with $\lambda = 2.0$, and (ii) the claim severity is modeled by a log-normal distribution with a mean of \$50,000 and coefficient of variation of 3, which was used by Finger (1976) for liability claim severity distribution. Suppose that the actuary has low confidence in the estimate of claim frequency, but higher confidence in the estimate of the claim severity distribution, thus chooses $r_1 = 0.7$ for the claim frequency and $r_2 = 0.8$ for the claim severity. The premium can be calculated using numerical integrations:

$$\mathrm{H}_{0.7}(N) = 2.527, \qquad \text{and} \qquad \mathrm{H}_{0.8}(X) = 82,960.$$

Thus, the required total premium is

$$\mathrm{H}_{0.7}(N) \times \mathrm{H}_{0.8}(X) = 209,640.$$

Kunreuther et al (1993) discussed the ambiguities associated with the estimates for claim frequencies and severities. They mention that for some risks such as playground accidents, there are considerable data on the chances of occurrence but much uncertainty about the potential size of the loss due to arbitrary court awards. On the other hand, for some risks such as satellite losses or new product defects, the chance of a loss occurring is highly ambiguous due to limited past claim data, however, the magnitude of such a loss is reasonably predictable.

For some risk events such as earthquake insurance, it is more plausible to consider the dependency between claim frequency and claim severity. For instance, the Richter scale value of an earthquake may affect both the frequency and severity simultaneously; and for hurricane losses the wind velocity would affect both the frequency and severity simultaneously.

Regardless of the dependency structure, computerized simulation methods can always be used to model the total claims based on given geographic concentration.

For instance, in simulating earthquake losses, one can use the following procedures: (i) simulate some numerical values of the Richter scale; (ii) conditional on the simulated Richter scale values, run a secondary generator for the claim frequency and the claim severity (of course both the frequency and the severity depend on the Richter scale values). Once the actuary has obtained sample distributions for the claim frequencies and severities, or a sample distribution for the total claims, he or she can apply a PH-transform directly to the simulated sample distributions.

2.5 Some Properties of the PH-Mean

In general, for $0 \leq r \leq 1$, the PH-mean has the following properties:

- $\mathrm{E}(X) \leq \mathrm{H}_r(X) \leq \max(X)$. When r declines from one to zero, $\mathrm{H}_r(X)$ increases from the expected loss, $\mathrm{E}(X)$, to the maximum possible loss, $\max(X)$.
- Scale and translation invariant: $\mathrm{H}_r(aX+b) = a\mathrm{H}_r(X) + b$, for $a,\ b \geq 0$.
- Sub-additivity: $\mathrm{H}_r(X_1 + X_2) \leq \mathrm{H}_r(X_1) + \mathrm{H}_r(X_2)$.
- Layer additivity: when a risk X is split into a number of layers

$$\{(x_0, x_1],\ (x_1, x_2], \cdots\},$$

the layer premiums are additive (the whole is the summation of the parts):

$$\mathrm{H}_r(X) = \mathrm{H}_r(I_{(x_0,x_1]}) + \mathrm{H}_r(I_{(x_1,x_2]}) + \cdots.$$

Pricing often assumes that a certain degree of diversification will be reached through the market efforts. In real life examples, risk-pooling is a common phenomena. It is assumed that, in a competitive market, the benefit of risk-pooling is transferred back to the policy-holders (in the form of premium reduction). In the PH-model, the layer-additivity property has already taken into account of the effect of risk-pooling.

Theoretically, in an efficient market (no transaction expenses in risk-sharing schemes) with complete information, the optimal cooperation among insurers is to form a market insurance portfolio (like the Dow Jones index), and each insurer takes a layer or quota-share of the market insurance portfolio.

In real life, however, the insurance market is *not* efficient. This is mainly because of incomplete information (ambiguity) and extra expenses associated with the risk-sharing transactions. There exist distinctly different local market climates in different geographic areas and in different lines of insurance. For instance, one can compare the automobile insurance market with the market for earthquake damage coverages in both California and Ontario. As a result, the value of the index r may vary with respect to the local market climate, which is characterized by the levels of ambiguity, risk concentration, and competition.

3 MIXTURE OF PH-TRANSFORMS

While a single index PH-transform has one parameter r to control the relative premium structure, one can obtain more flexible premium structures by using a mixture of PH-transforms:

$$p_1 \mathrm{H}_{r_1} + p_2 \mathrm{H}_{r_2} + \cdots + p_n \mathrm{H}_{r_n}, \qquad \sum_{j=1}^{n} p_j = 1, \qquad 0 \le r_j \le 1 \; (j = 1, \cdots, n).$$

Let $\bar{r} = \sum_{j=1}^{n} p_j r_j$ be the weighted average index. It can be verified that

- For any risk X, $p_1 \mathrm{H}_{r_1}(X) + p_2 \mathrm{H}_{r_2}(X) + \cdots + p_n \mathrm{H}_{r_n}(X) \ \ge \ \mathrm{H}_{\bar{r}}(X)$.
- For a layer $I_x = (x, x+h)$, the ratio

$$\frac{p_1 \mathrm{H}_{r_1}(I_x) + p_2 \mathrm{H}_{r_2}(I_x) + \cdots + p_n \mathrm{H}_{r_n}(I_x)}{\mathrm{H}_{\bar{r}}(I_x)}$$

  is an increasing function of x.

The PH-measure mixture can be interpreted as a collective decision-making process. Each member of the decision-making 'committee' chooses a value of r, and the index mixture represents different r's chosen by different members. It also has interpretations as (i) an index mixture chosen by a rating agency according to the indices for all insurance companies in the market; (ii) an index mixture which combines an individual company's index with the rating agency's index mixture.

For ratemaking purposes, mixtures of PH-transforms add more flexibility than a single index. In the remaining sections of this article, we shall discuss some special two-point mixtures of PH-transforms:

$$(1-\alpha)\mathrm{H}_{r_1}(X) + \alpha \mathrm{H}_{r_2}(X), \qquad 0 \le \alpha \le 1, \qquad r_1, r_2 \le 1.$$

3.1 Minimum Rate-on-Line

In most practical circumstances, very limited information is available for claims at extremely high layers. In such highly ambiguous circumstances, most (re)insurers adopt a survival rule of minimum rate-on-line. The rate-on-line is the premium divided by the coverage limit, and most (re)insurers establish a minimum they will accept for this ratio (see Venter, 1991).

By using a two-point mixture of PH-transforms with $r_1 \leq 1$ and $r_2 = 0$, the premium functional

$$(1-\alpha)\mathrm{H}_{r_1}(X) + \alpha\mathrm{H}_0(X) = (1-\alpha)\mathrm{H}_{r_1}(X) + \alpha \max(X)$$

can yield a minimum rate-on-line at α.

Example 7: Reconsider Example 4, the best-estimate loss distribution (ddf) is

$$S_X(t) = 0.1 \times (\frac{2000}{2000+t})^{1.2}.$$

By choosing a two-point mixture with $r_1 = 0.85$, $r_2 = 0$, and $\alpha = 0.02$, we get an adjusted distribution:

$$S_Y(t) = 0.98 \times 0.1 \times (\frac{2000}{2000+t})^{1.2\times 0.85} + 0.02.$$

As shown in the table below, this two-point mixture guarantees a minimum-rate-on-line at 0.02 (1 full payment out of 50 years). By comparing Table 5 with Table 3 one can see that, at higher layers, this method yield distinctly different premiums from those in Example 4.

3.2 The Right-Tail Deviation

Consider a two-point mixture of PH-transforms with $r_1 = 1$ and $r_2 = \frac{1}{2}$:

$$(1-\alpha)\mathrm{H}_1(X) + \alpha\mathrm{H}_{\frac{1}{2}}(X), \qquad 0 < \alpha < 1,$$

which can be rewritten as (noting that $\mathrm{H}_1(X) = \mathrm{E}(X)$):

$$\mathrm{E}(X) + \alpha\,[\mathrm{H}_{\frac{1}{2}}(X) - \mathrm{E}(X)],$$

which is analogous to the standard deviation method: $\mathrm{E}(X) + \alpha\,\sigma(X)$.

Now we introduce a new risk-measure analogous to the standard deviation.

Table 5: Layer premiums under an index mixture

Layer	Net Premium	Risk-adjusted Premium
(0, 1000]	77.892	131.802
(5000, 6000]	20.512	56.006
(10000, 11000]	11.098	41.363
(50000, 51000]	1.982	24.940
(100000, 101000]	0.888	22.497
(500000, 501000]	0.132	20.493
(1000000, 1001000]	0.058	20.244

Definition 2 *The* **right-tail deviation** *is defined as*

$$\mathrm{D}(X) = \mathrm{H}_{\frac{1}{2}}(X) - \mathrm{E}(X) = \int_0^\infty \sqrt{S_X(t)}dt - \int_0^\infty S_X(t)dt.$$

and the **right-tail index** *is defined as*

$$d(X) = \frac{\mathrm{H}(X)}{\mathrm{E}(X)}.$$

Analogous to the standard deviation, the right-tail deviation $\mathrm{D}(X)$ satisfies:

- If $\Pr\{X = b\} = 1$, then $\mathrm{D}(X) = 0$.
- Scale-invariant: $\mathrm{D}(cX) = c\mathrm{D}(X)$ for $c > 0$.
- Sub-additivity: $\mathrm{D}(X + Y) \leq \mathrm{D}(X) + \mathrm{D}(Y)$.
- If X and Y are perfectly correlated, then $\mathrm{D}(X + Y) = \mathrm{D}(X) + \mathrm{D}(Y)$.

At very high layers, the standard deviation and the right-tail deviation converge to each other, as demonstrated in the following example.

Example 8 : Re-consider the claim distribution in Example 4 with a ddf:

$$S_X(t) = 0.1 \times (\frac{2000}{2000 + t})^{1.2}.$$

For different layers with fixed limit at 1000, we compare the standard deviation and the right-tail deviation in the following table.

Layer	Expected loss	Std-deviation of the loss	Right-tail deviation	Percentage difference
I	$\mathrm{E}(I)$	$\sigma(I)$	$\mathrm{D}(I)$	$\frac{\sigma(I)}{\mathrm{D}(I)} - 1$
(0, 1000]	77.89	256.0	200.5	27.7%
(1000, 2000]	51.56	214.3	175.2	22.3%
(10000, 11000]	11.10	103.9	94.24	10.3%
(100000, 101000]	.8879	29.76	28.91	2.93%
(1000000, 1001000]	.05754	7.584	7.528	.75%
(10000000, 10001000]	.003640	1.908	1.904	.19%
(100000000, 100001000]	.0002297	.4793	.4791	.05%
(1000000000, 1000001000]	.00001450	.1204	.1204	.01%

It can be shown that, for any small layer $[a, a+h)$, $\mathrm{D}(I_{(a,a+h]}) \leq \sigma(I_{(a,a+h]})$, $\mathrm{D}(I_{(a,a+h]})$ converges to $\sigma(I_{(a,a+h]})$ at upper layers (i.e. the relative error goes to zero when a becomes large). As a result, for any non-negative random variable X, the right-tail deviation $\mathrm{D}(X)$ is finite, if and only if, the standard deviation $\sigma(X)$ is finite.

Having stated a number of similarities, here we point out some crucial differences between the right-tail deviation $\mathrm{D}(X)$ and the standard deviation $\sigma(X)$:

- $\mathrm{D}(X)$ is layer-additive, but $\sigma(X)$ is *not* additive.
- $\mathrm{D}(X)$ preserves some natural ordering of risks such as first stochastic dominance[1], but $\sigma(X)$ does not.

3.3 Links to the Gini Index in Welfare Studies

Historically, some long-tailed distributions have an origin in income distributions (e.g. Pareto, log-normal distributions, see Arnold, 1983). In social welfare studies, a celebrated measure for income inequality[2] is the Gini index. Assume that individual's wealth level in a country (community) can be summarized by a distribution: $S_X(u) = \mathsf{Proportion}\{X > u\}$. As a measure of income inequality of a society, the Gini index is

[1] Risk X is small than risk Y in first stochastic dominance if $S_X(t) \leq S_Y(t)$ for all $t \geq 0$; or equivalently, Y has the same distribution as $X + Z$ where Z is another non-negative random variable.

[2] Here 'income inequality' refers to the polarization of the wealth distribution.

defined as

$$G(X) = \frac{2E(|X - Y|)}{E(X)},$$

where X and Y are independent and identically distributed.

An equivalent definition of the Gini index is

$$G(X) = 1 - \frac{\int_0^\infty [S_X(u)]^2 du}{\int_0^\infty S_X(u) du}.$$

The higher the Gini index is, the more polarized a society is. As a measure of welfare inequality, the Gini index has the following properties:

- Each dollar transferred from the rich to the poor will lower the Gini index.
- Adding an equal amount to all persons' wealth will decrease the Gini index.

It is noted that $d(X)$ and $G(X)$ are similar in their definition formulae. This similarity may suggest that the role of the right-tail index $d(X)$ in measuring the right-tail risk is parallel to the role of the Gini index $G(X)$ in measuring income inequalities.

Consider the following loss distributions each with the same mean(=1) and variance(=3). Without referring to higher moments, we can order them by the right-tail index $d(X)$.

Risk X_i	Distribution	$E(X_i)$	$\sigma(X_i)$	$d(X_i)$	Gini index
Pareto	$S(t) = (\frac{2}{2+x})^3$	1	$\sqrt{3}$	**3.00**	0.600
Log-normal	$\mu = -\log(2),\ \sigma = \log(4)$	1	$\sqrt{3}$	**2.46**	0.595
Inverse-Gaussian	$f(x) = \frac{\exp\{-\frac{(x-1)^2}{6x}\}}{\sqrt{6\pi x^3}}$	1	$\sqrt{3}$	**2.17**	0.632
Gamma	$\alpha = \beta = \frac{1}{3}$	1	$\sqrt{3}$	**1.96**	0.713
Bernoulli	$f(0) = \frac{3}{4},\ f(4) = \frac{1}{4}$	1	$\sqrt{3}$	**1.00**	0.750

As its name may suggest, the right-tail deviation measures the right-tail risk, as opposed to the standard deviation which measures the deviation about the mean, and as opposed to the Gini index which measures the polarization of the wealth distribution.

4 ECONOMIC THEORIES

4.1 Expected Utility Theory

Traditionally, expected utility (EU) theory has played a dominant role in modeling decisions under risk and uncertainty. To a large extent, the popularity of EU was attributed to the axiomatization of von Neumann and Morgenstern (1947). They proposed five axioms (somewhat self-evident) and showed that any decision-making which is consistent with these axioms can be modeled by using a utility function of wealth. However, due to difficulties associated with implementation, EU remains as an academic pursuit and has had little direct impact in practice.

When EU is applied to produce an insurance premium for a risk X, the minimum premium P that an insurance company will accept for full insurance is defined by $u(w) = \mathrm{E}[u(w + P - X)]$, in which u and w refer to the insurer's utility and wealth (see Bowers et al, 1986). As pointed out by Meyers (1995), EU gives lower and upper bounds of an insurance premium, without due consideration of the market setting.

The EU does have an indirect application in actuarial work via the mean-variance analysis, which is viewed by some authors as a rough approximation of utility theory (Meyers, 1995). A commonly used actuarial method for deciding risk loads is based on the first two moments. Since loss distributions are often highly skewed, the first two moments cannot accurately reflect the level of insurance risk. In fact, actuaries often find that long tailed claim distributions, such as Pareto distributions, are more appropriate to describe the potential losses for some insurance contracts (e.g. liability insurance). Even for a large risk portfolio, the total claim distribution can be highly non-normal due to correlations or ambiguities in the initial estimates of individual risks.

The inconsistency of moment-based methods in calculating layer premiums are discussed by a number of authors (e.g. Venter, 1991; Robbin, 1992).

4.2 The Dual Theory of Yaari

A new theory of decision under uncertainty has been developed in the last decade by a group of economists (e.g. Quiggin, 1982; Yaari 1987). Analogous to the development of non-Euclidean geometry, Yaari (1987) formalized an alternative set of axioms and developed a *dual* theory of decision under uncertainty. In Yaari's dual theory, risk-aversion is described by a distortion function (increasing and convex) $g : [0, 1] \mapsto [0, 1]$

which is applied to probability distributions. The certainty equivalent to a bounded random economic prospect V $(0 \leq V \leq m)$ is

$$\int_0^m g[S_V(t)]dt, \qquad \text{where } S_V(t) = \Pr\{V > t\}.$$

In other words, the certainty equivalent to a random economic prospect, V, is just the expected value under the distorted probability distribution, $g[S_V(t)]$.

4.3 Schemeidler's Ambiguity-Aversion

As early as 1921, John Keynes identified a distinction between the *implication* of evidence (the implied likelihood) and *weight* of evidence (confidence in the implied likelihood). Frank Knight (1921) also drew a distinction between *risk* (with known probabilities) and *uncertainty* (ambiguity about the probabilities). A famous example on ambiguity-aversion is Ellsberg's (1961) paradox which can be briefly described as follows: There are two urns each containing 100 balls. One is a non-ambiguous urn which has 50 red and 50 black balls; the other is an ambiguous urn which also contains red and black balls but with unknown proportions. When subjects are offered \$100 for betting on a red draw, most subjects choose the non-ambiguous urn (and the same for the black draw). Such a pattern of preference *cannot* be explained by EU (Quiggin, 1993, p.42).

Ellsberg's work has spurred much interest in dealing with ambiguity factors in risk analysis. Schmeidler (1989) brought to economists *non-additive probabilities* in his axiomization of preferences under uncertainty. For instance, in Ellsberg's experiment, the non-ambiguous urn, with 50 red and 50 black balls, is preferred to the ambiguous urn with unknown proportions of red or black balls. This phenomenon can be explained if we assume that one assigns a subjective probability $\frac{3}{7}$ to the chance of getting a red draw (or black draw). Since $\frac{3}{7} + \frac{3}{7} = \frac{6}{7}$ which is less than one, the difference $1 - \frac{6}{7} = \frac{1}{7}$ may represent the magnitude of ambiguity aversion.

Built on its own axiomatic system, Schmeidler's theory leads to the same mathematical formulation as that of Yaari; that is, a certainty equivalent to a random economic prospect V $(0 \leq V \leq m)$ can be evaluated as

$$\mathrm{H}(V) = \int_0^m g[S_V(t)]dt,$$

where $g : [0, 1] \mapsto [0, 1]$ is a distortion function and $g[S_X(t)]$ represents the subjective probabilities.

The method of using adjusted distributions is widely known by actuaries. However, actuaries often use a transformed random variable, $Y = g(X)$, which yields $S_Y(t) = S_X(g^{-1}(t))$, a different formulation from Yaari's and Schmeidler's. A key point in the theories of Yaari and Schmeidler is that one needs to transform *directly* the distribution function $S_X(t)$.

Using a market argument, Venter (1991) discussed the no-arbitrage implications of insurance pricing. He observed that in order to ensure additivity when layering a risk, it is necessary to adjust the loss distribution so that layer premiums are expected losses under the adjusted loss distribution. Inspired by Venter's insightful observation, Wang (1995, 1996a) proposed the PH-transform method, which is in agreement with the formulation in Yaari and Schmeidler, thus is supported by their economic theories.

5 SUMMARY

In this paper we have introduced the basic methodologies of the PH-transform method and have shown by example how it can be used in insurance ratemaking. We did not discuss how to decide the overall level of contingency margin, which depends greatly on market conditions. An important avenue for future research is to link the overall level of risk load with the required surplus for supporting the written contract. Some pioneer work in this direction can be found in Kreps (1990) and Philbrick (1994).

The use of adjusted/conservative life tables has long been practiced by life actuaries (see Venter, 1991). To casualty actuaries, the PH-transform method contributes a theoretically sound and practically plausible way to adjust the loss distributions. For economic interpretations and empirical tests of the PH-transform method, see Wang (1996b). For updating risk-adjusted premiums in the light of new information, see Wang and Young (1996).

Acknowledgments: The author thanks the CAS Ratemaking Committee, especially Israel Krakowski, for numerous comments and suggestions.

REFERENCES

Arnold, Barry C. (1983). *Pareto Distributions*, International Co-operative Publishing House, Burtonsville, Maryland.

Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt. (1986). *Actuarial Mathematics*, Itasca, Ill.: Society of Actuaries, 1986.

Ellsberg, Daniel (1961) "Risk, Ambiguity, and the Savage Axioms", *Quarterly Journal of Economics*, **75**, 643-669.

Feldblum, Sholom (1990). "Risk loads for insurers", *PCAS*, LXXVII, 160-195.

Finger, Robert J. (1976). "Estimating pure premiums by layer – an approach", *PCAS*, LXIII, 34-52.

Hogarth, Robin M. and Howard Kunreuther (1992). "Price insurance and warranties: ambiguity and correlated risks", *The Geneva Papers on Risk and Insurance Theory*, **17**(1), 35-60.

Hogg, Robert and Stuart Klugman (1984). *Loss Distributions*, John Wiley, New York.

Keynes, John M. (1921). *A Treatise on Probability*, London: Macmillian.

Knight, Frank H. (1921). *Risk, Uncertainty and Profit*, Boston: Houghton Mifflin.

Kreps, Rodney. (1990). "Reinsurer risk loads from marginal surplus requirements", *PCAS*, Vol. LXXVII, 196-203.

Kunreuther, Howard, Robin Hogarth and Jacqueline Meszaros (1993). "Insurer ambiguity and market failure", *Journal of Risk and Uncertainty*, **7**, 71-87.

Lawless, Jerry F. (1982). *Statistical Models and Methods For Lifetime Data*, John Wiley & Sons, New York, 1982.

McClenahan, Charles L. (1990). Chapter 2. Ratemaking in *Foundations of Casualty Actuarial Science*, published by the Casualty Actuarial Society, New York.

Meyers, Glenn G. (1991). "The competitive market equilibrium risk load formula for increased limits ratemaking", *PCAS*, Vol. LXXVIII, 163-200.

Meyers, Glenn G. (1995) "Managing the catastrophe risk", Casualty Actuarial Society Discussion Paper Program, May 14-17, 1995.

Miccolis, Robert S. (1977). "On the theory of increased limits and excess-of-loss pricing", *PCAS*, Vol. LXIV, 1977, 27-59.

Patrick, S. Gary (1990). "Reinsurance", *Foundations of Casualty Actuarial Science*, Chapter 6, published by the Casualty Actuarial Society, New York.

Philbrick, Stephen (1991). Discussion of Feldblum: "Risk loads for insurers", *PCAS*, LXXVIII, 56-63.

Philbrick, Stephen W. (1994) "Accounting For Risk Margins", *Casualty Actuarial Society Forum*, Spring 1994, Vol. 1, 4-87.

Quiggin, John (1982). "A theory of anticipated utility", *Journal of Economic Behavior and Organization* **3**(4), 323-43.

Quiggin, John (1993). *Generalized Expected Utility Theory: the rank-dependent model*, Kluwer Academic Publishers, Boston.

Robbin, Ira (1992). Discussion of Meyers: "The competitive market equilibrium risk load formula for increased limits ratemaking", *PCAS*, LXXIX, 367-384.

Rosenberg, Sheldon (1977). Discussion of Robert S. Miccolis' paper — "On the theory of increased limits and excess-of-loss pricing" , *PCAS*, Vol. LXIV, 1977, 60-73.

Schmeidler, David (1989). "Subjective probability and expected utility without additivity", *Econometrica*, **57**, 571-587.

Venter, Gary G. (1991). "Premium calculation implications of reinsurance without arbitrage", *ASTIN Bulletin*, **21**, 223-230.

von Neumann, John and Oskar Morgenstern (1947). *Theory of Games and Economic Behavior*, (2nd edition), Princeton, NJ: Princeton University Press.

Wang, Shaun (1995). "Insurance pricing and increased limits ratemaking by proportional hazards transforms", *Insurance: Mathematics and Economics*, **17**, 43-54.

Wang, Shaun (1996a). "Premium calculation by transforming the layer premium density", *ASTIN Bulletin*, **26**(1), 71-92.

Wang, Shaun (1996b). "Ambiguity-aversion & economics of insurance", *Proceedings of the 1996 Risk Theory Seminar*, Wisconsin-Madison.

Wang, Shaun and Virginia R. Young. (1996). "Risk-adjusted credibility premiums using distorted probabilities", submitted to *Scandinavian Actuarial Journal.*

Yaari, Menahem E. (1987). The dual theory of choice under risk, *Econometrica*, **55**, 95-115.

APPENDIX: Ambiguity and Parameter Risk

Most insurance risks are characterized by the uncertainty about the estimate of the tail probabilities. This is often due to data sparsity for rare events (small tail probabilities), which in turn causes the estimates for tail probabilities to be unreliable.

To illustrate, assume that we have a finite sample of n observations from a class of identical insurance policies. The empirical estimate for the loss distribution is

$$\hat{S}(t) = \frac{\#\text{ of observations } > t}{n}, \quad t \geq 0.$$

Let $S(t)$ represent the *true* underlying loss distribution, which is generally unknown and different from the empirical estimation $\hat{S}(t)$. From statistical estimation theory (e.g., Lawless, 1982, pp. 402; Hogg and Klugman, 1984), for some specified value of t, we can treat the quantity

$$\frac{\hat{S}(t) - S(t)}{\sigma(\hat{S}(t))},$$

as having a standard normal distribution for large values of n, where

$$\sigma(\hat{S}(t)) \approx \frac{\sqrt{\hat{S}(t)[1 - \hat{S}(t)]}}{\sqrt{n}}.$$

The η% **upper confidence limit (UCL)** for the true underlying distribution $S(t)$ can be approximated by

$$\text{UCL}(t) = \hat{S}(t) + \frac{q_\eta}{\sqrt{n}}\sqrt{\hat{S}(t)[1 - \hat{S}(t)]},$$

where q_η is a quantile of the standard normal distribution: $\Pr\{N(0,1) \leq q_\eta\} = \eta$. Keeping n fixed and letting $t \to \infty$, the ratio of the UCL to the best-estimate $\hat{S}(t)$ is

$$\frac{\text{UCL}(t)}{\hat{S}(t)} = 1 + \frac{q_\eta}{\sqrt{n}}\sqrt{\frac{1 - \hat{S}(t)}{\hat{S}(t)}} \to \infty,$$

which grows without bounds as t increases.

As a means of dealing with ambiguity regarding the best-estimate, the PH-transform:

$$\hat{S}_Y(t) = [\hat{S}_X(t)]^r, \quad r \leq 1,$$

can be viewed as an upper confidence limit (UCL) for the best-estimate $\hat{S}_X(t)$. It automatically gives higher relative safety margins for the tail probabilities, and the ratio

$$\frac{[\hat{S}_X(t)]^r}{\hat{S}_X(t)} = [\hat{S}_X(t)]^{r-1} \to \infty, \quad \text{as } t \to \infty,$$

increases without bound to infinity.