INSURING RISKS WHEN PURE PREMIUM IS INFINITE?

Arthur CHARPENTIER
ENSAE-CREST
3 avenue Pierre Larousse, 92240 Malakoff cedex, France

ABSTRACT
Insurability is a major issue for risk managers in the insurance industry. ZAJDENWEBER (1996) mentioned that business interruption is hardly insurable, using extreme value results: the right tail of the distribution should be modeled using some Pareto distribution with parameter 1, which has none finite moment. Since the expected value in tails is infinite, on a theoretical point of view, it becomes impossible to assess the price of that risk, and to hedge it using standard insurance covers. As we shall see, the use of more advanced results in extreme value theory (a wide survey will be proposed) may let us think that the assumption of very fat tails may be not relevant. For instance, we will propose a test to see if a distribution has a finite mean. We shall also discuss at the end the use of the pure premium as a criteria to assess whether a risk is or not insurable.

Keywords: business interruption; distorted premium principle; extreme value theory; insurability; Pareto distribution; pure premium.

1. THE NOTION OF INSURABLE RISKS, A BASIC INTRODUCTION
Since risk managers are facing more and more new risks, they try to get a better understanding of their portfolio, seeking which risks can be insured, and which cannot. Therefore, the so-called notion of “insurability” become a major issue in the insurance industry. Inspired by BERLINER (1982), GODARD, HENRY, LAGADEC & MICHEL-KERJAN (2002), proposed several “axioms” that should satisfy a risk to be insurable. As in DENUIT & CHARPENTIER (2005), the following classification can be used,

1. judiciously, an insurance contract can be valid only if claim occurrence satisfy some randomness property,
2. the “game rule” (using Berliner’s expression, i.e. legal framework) should remain stable in time.
Those two notions yield the concept of “legal” insurability,
3. the possible maximum loss should not be huge, with respect to the insurer’s
solvency,
4.  the average cost should be identifiable and quantifiable,
5.  risks could be pooled so that the law of large numbers can be used
    (independent and identically distributed, i.e. the portfolio should be
    homogeneous).

These three notions define the concept of “actuarial” insurability, underlying the use
of the law of large numbers and the central limit theorem (see e.g. DENUIT & CHARPENTIER
(2004, 2005)), to assess premium levels and solvency margins.
6.  there should be no moral hazard, and no adverse selection,
7.  there should exist an insurance market, in the sense that offer and demand
    should meet, and a price (equilibrium price) should arise.

Those two last points define the concept of “economic” insurability, also called

Numerous studies of economic insurability present several ideas in order to avoid
those problems, and more specifically moral hazard and adverse selection (see e.g. WINTER
(1992) or DIONNE & DOHERTY (1992)). But actuarial insurability is perhaps more difficult
to escape from. The non-independent assumption (item 5) appears for instance in natural
hazard insurance (flood, hurricanes, earthquakes), where one event can hit many policies at
the same time: claims are not independent anymore. And as noticed in RUSSELL & JAFFE
(1997), BROWN & HOYT (1999) or FROOT & O’CONNELL (1999), those risks might be
hardly insurable.

RAJDENWEBER (1996) and RAJDENWEBER (2000) mention a particular case (business
interruption) where items 3 and 4 are not satisfied, but where insurance policies are sold
everyday. Section 2 of the present paper will focus on the example given in RAJDENWEBER
(1996), on almost the same dataset, showing that using advanced results in extreme value
theory, other conclusions are obtained. Several tools to test whether there is, or not, a finite
pure premium, will be presented. Section 3 will present some other way to look at the
insurance premium problem, when dealing with large risks. As we shall see, pure premium
might not be an appropriate benchmark. Actually, based on YAARI (1986) dual approach,
some distortion premium can be obtained, and they can be calculated simply using extreme
value results.
2. **TAIL BEHAVIOR AND EXISTENCE OF A PURE PREMIUM**

ZAJDENWEBER (1996) proposed to study only the tail of the distribution of claims, since non-existence of a pure premium can only come from fat tails.

2.1 **Basic on extreme values**

There are three equivalent techniques used to study extremes, in all yield the same results. Consider some i.i.d. random variables (e.g. insurance costs) $X, X_1, X_2, \ldots, X_n$ with distribution function $F(x) = P(X \leq x)$ and density $f$, and set $x_F = \sup \{x, F(x) < 1\} \leq \infty$.

The first idea was introduced by Fisher and Tippet in 1928, based on the study of the maximum of an i.i.d. sample, and has be proved by Gnedenko in 1943. Let $X_{n,n} = \sup \{X_1, \ldots, X_n\}$, then $X_{n,n} \mathcal{P} \to x_F$. Hence, there is no chance to derive some non-degenerated limiting distribution for the maxima. One idea can be to use the same approach as the empirical mean. From the law of large numbers, under the assumption that the variance is finite, $\frac{\bar{X}_n}{n} \mathcal{P} \to E(X)$. In order to obtain a limiting distribution for $\frac{\bar{X}_n}{n}$, a normalized version of the empirical mean is considered and

$$\frac{\bar{X}_n}{\frac{\sqrt{\text{Var}(X)}}{n}} \mathcal{L} \to N(0,1),$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Recall that in the case where neither the expected value, nor the variance is finite, other limiting distributions can be obtained (the so-called stable laws). For the maxima, the idea is also to consider a normalized version of the maxima, i.e. some $a_n$ and $b_n > 0$ such that

$$\frac{X_{n,n} - a_n}{b_n} \mathcal{L} \to Z,$$

where $Z$ has a non-degenerate distribution. A proved by Gnedenko (see e.g. EMBRECHTS, KLÜPPELBERG & MIKOSH (1997)), if the limiting distribution is non-degenerated, it is either Fréchet, Gumbel or Wiebull, i.e. there is $\xi \in R$ such that, up to a scaling and a location parameter,

$$P(Z \leq x) = G_{\mu,\sigma,\xi}(x) = \begin{cases} \exp \left( -\left[1 - \xi \frac{x - \mu}{\sigma} \right]^\frac{1}{\xi} \right) & \text{if } \xi \neq 0 \\ \exp \left( -\exp \left[ -\frac{x - \mu}{\sigma} \right] \right) & \text{if } \xi = 0 \end{cases}$$
Note that normalizing sequences \((a_n, b_n)\) influence only \(\mu\) and \(\sigma\), i.e. \(\xi\) will always be the same, given the distribution of the \(X_i\)'s. Hence, if there are \(\mu\) and \(\sigma\) such that \(G_{\mu,\sigma,\xi}\) is a limiting distribution for a normalized version of the maximum of some i.i.d. sample \(X_i\) with distribution function \(F\), \(F\) is said to be in the max-domain of attraction of \(G_{\cdot,\cdot,\xi}\), denoted \(\text{GEV}(\xi)\).

The second idea is based on old properties obtained by Pareto, but proved by Balkema, de Haan and Pickands in 1974. The limiting exceedance distribution, i.e. of \(X - u \mid X > u\) when \(u \to \infty\) is closely related to Pareto distribution, in the sense that for all \(u\) large enough, there exists \(\sigma\) such that

\[
P(X - u \leq x \mid X > u) \approx H_{\xi,\sigma}(x) = 1 - \left(1 + \frac{x}{\xi \sigma}ight)^{-1/\xi}.
\]

\(H_{\xi,\sigma}\) is called Generalized Pareto distribution (GPD(\(\xi\))). More precisely, \(F\) is in the max-domain of attraction of \(\text{GEV}(\xi)\) if and only if there exist \(\sigma(\cdot)\)

\[
\lim_{u \to \infty} \sup_{0 < x < \sigma(u)} \{P(X - u \leq u \mid X > u) - H_{\xi,\sigma(u)}(x)\} = 0.
\]

Those two approaches are closely related since the same tail index parameter \(\xi\) appears.

The third idea is linked with the concept of regular variation. Recall that function \(h\) is said to be regularly varying if there is \(g\) such that

\[
\lim_{t \to \infty} \frac{h(tx)}{h(t)} = g(x) \text{ for all } x > 0.
\]

Using some stability property, \(g\) is necessarily the solution of Cauchy functional equation \(g(x \cdot y) = g(x) \cdot g(y)\), and therefore, there if \(\alpha \in \mathbb{R}\) such that \(g(x) = x^\alpha\). Define also the notion of slowly varying function, when \(\alpha = 0\). From this notion, recall the so-called Tauberian theorem which allows to define a notion of regular variation for random variables. Random variable \(X\), with distribution function \(F\) and Laplace transform \(L_F\), is said to be regularly varying with index \(-\alpha\), where \(\alpha \leq 0\), if one of the following equivalent condition is fulfilled,

the survival distribution \(\overline{F}(\cdot) = P(X > \cdot)\) is regularly varying with index \(-\alpha\),

\[
\overline{F}(x) = x^{-\alpha} L_F(x),
\]

the quantile function (also called Value-at-Risk) is regularly varying
the Laplace transform satisfies \( 1 - L_F(t) \sim t^\alpha L_L(1/t) \),

if \( X \) has a density \( f \) which satisfies \( xf(x)/F(x) \to \alpha \) as \( x \to \infty \), the the density is regularly varying with index \( -(1+\alpha) \),

\[
f(x) = x^{-(1+\alpha)} L_f(x),
\]

where \( L_F \), \( L_{P^{-1}} \), \( L_L \) and \( L_f \) are slowly varying function. This concept is useful when studying tails of random variables since \( F \) is in the max-domain of attraction of GEV(\( \xi \)) if and only if \( X \) is regularly varying with index \( \alpha = -1/\xi \).

Hence, those three definitions of extremes (based on the limiting distribution of the maxima, of the exceedance distribution, and the fatness of the cdf or the density function) are all based on the tail index \( \xi \).

From the second idea, note furthermore that if \( X \) has a generalized Pareto distribution with tail index \( \xi > 0 \), then

\[
E[X^r] = \frac{\sigma^r}{\xi^{r+1}} \frac{\Gamma(1/\xi - r)}{1/\xi + 1} r!, \quad \text{when } r \in \mathbb{N} \text{ and } r < 1/\xi.
\]

Hence, if \( \xi \geq 1 \), \( E(X) \) is infinite. ZAJDENWEBER (1996) used this idea to assess the insurability of a risk: if \( \xi \geq 1 \), \( E(X - u | X > u) \) is infinite and therefore risk \( X \) can not have a finite pure premium.

### 2.2 Estimation of the tail index

ZAJDENWEBER (1996) considered some log-log scatterplot to estimate the tail index, as shown here in Figure 1. Using the exceedances idea, if \( u \) is large enough,

\[
P(X \leq x | X > u) \approx 1 - \left(1 + \frac{x + u}{\sigma} \right)^{-1/\xi},
\]

for some \( \sigma > 0 \), or equivalently

\[
\log \left(1 - \frac{F(x)}{1 - F(u)} \right) \approx - \frac{1}{\xi} \log \left(1 + \frac{x + u}{\sigma} \right) \approx - \frac{1}{\xi} \log x + \text{constant}.
\]

Hence, if the \( X_1, \ldots, X_n \) where i.i.d., with distribution function \( F \) in the max-domain of GEV(\( \xi \)) then the \( \left( \log X_i, \log(1 - \hat{F}(X_i)) \right) \)'s should be sensibly on a line of slope \(-1/\xi \) (\( \hat{F} \) denotes here the usual empirical distribution function). A natural estimator is then the least square estimator of the slope the straight line fitted to points on a
scatterplot (log \(X_i\), log(1 - \(\hat{F}(X_i)\)))’s, where that line that is in some sense (\(L^2\) distance) closest to all of the data points simultaneously.

**Example 1.** Figure 1 is based on some dataset, kindly provided by the French Federation of Insurers (as in ZAJDENWEBER (1996)), on the period 1992-2000- in order to fulfill the i.i.d. assumption (at least no sensible inflation). The slope estimator is –1.47, or \(\hat{\xi} = 0.678\).

![Figure 1. Empirical cumulative distribution function \(\hat{F}(x)\) for business interruption (from 1992 to 2000), and the log-transformation (log(1 - \(\hat{F}(x)\)) as a function of log(x))](image)

In order to use the maxima approach to get an estimator for \(\xi\), the natural idea is to consider some “bloc-maxima techniques”, as described in EMBRECHTS, KLUPPELBERG & MIKOSH (1997). Consider some i.i.d. sample \(X_1, \ldots, X_n\) where \(n = m \cdot k\), and consider \(m\) subsamples with respective sizes \(k\), i.e. \(\{X_1, \ldots, X_k\}\), \(\{X_{k+1}, \ldots, X_{2k}\}\), ..., \(\{X_{(m-1)k+1}, \ldots, X_{km}\}\). Let \(Y_i\) denote the maximum of the \(i\) th subsample. If \(k\) and \(m\) are both large enough, the distribution of the \(Y_i\)’s should be GEV\(\xi\). A natural estimator for \(\xi\) is then the maximum likelihood estimator of the GEV distribution based on bloc maxima \(Y_1, \ldots, Y_m\). Using some asymptotic normality property, some confidence interval can be derived using appropriate likelihood profile.

An other technique can be to fit some generalized Pareto. Since the distribution over
a given threshold should be close to the generalized Pareto, for a threshold large enough. Given \( u \), consider the subsample of observations that exceed \( u \), i.e. \( Z_1, \ldots, Z_m \) (\( m \) being here the number of observations exceeding threshold \( u \)). A natural estimator for \( \xi \) is then the maximum likelihood estimator of the GPD distribution based on exceeding observations \( Z_1, \ldots, Z_m \). Using some asymptotic normality property, some confidence interval can be derived using appropriate likelihood profile.

HILL (1981) considers the following estimator of \( \xi \), when \( \xi > 0 \), using the regular variation concept. \( \widetilde{F}(x) = x^{\xi/\xi} L(x) \), where \( L \) is a slowly varying function, or, expressed through the quantile function \( F^{-1}(1 - p) = p^{-\xi} L^{-1}(1/p) \).

Hence \( \log F^{-1}(1 - p) = -\xi \log p + \log L^{-1}(1/p) \).

A natural estimator of \( F^{-1}(1 - k/(n+1)) \) is \( X_{n-k+1} \). And therefore, points
\[
(\log(X_{n-j+1}), \log\left(\frac{j}{n+1}\right)), j = 1, \ldots, k,
\]
should be on a line with slope \( \xi \). The natural estimator of the slope being
\[
\hat{\xi} = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1} - \log X_{n-k+n} - \frac{1}{k} \sum_{j=1}^{k} \log \frac{j}{n+1} - \log \frac{k}{n+1},
\]
and since for \( k \) large enough, the denominator is almost 1, HILL (1975) considered,
\[
\hat{\xi} = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1} - \log X_{n-k+n}.
\]

It is then possible to plot \( \hat{\xi} \) as a function of \( k \). Furthermore, since
\[
\sqrt{k} \left( \hat{\xi} - \xi \right) L \to N\left(0, \xi^2\right) \quad \text{for} \quad \xi > 0,
\]
some confidence interval can be derived.

**Example 2.** On the same dataset, Figure 2 shows Hill plot (evolution of \( \hat{\xi} \) as a function of the number of exceedances \( k \)) on the left, while on the right is presented an estimation of \( \xi \) using maximum likelihood techniques on exceedances \( u \). Note that \( \hat{\xi} \) obtained using maximum likelihood techniques on exceedance distributions is significatively lower than 1 with probability 97.5 % (using more than 50 exceeding observations, i.e. 5 %), and similarly for Hill plot, where \( \hat{\xi} \) which is significatively lower than 1 with probability 97.5 % (using more than 50 exceeding observations). Table 1
summarizes most of the results, with several estimators, using Hill plot, GPD approximation and bloc maxima techniques.

Figure 2. Estimations of the tail index $\xi$, Hill’s estimator on the left, and maximum likelihood estimator of the Pareto parameter when considering exceeding observations on the right.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\xi}$</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hill</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.5580</td>
<td>0.7010</td>
<td>0.7415</td>
</tr>
<tr>
<td>50</td>
<td>0.5919</td>
<td>0.6991</td>
<td>0.7296</td>
</tr>
<tr>
<td>100</td>
<td>0.7046</td>
<td>0.7949</td>
<td>0.8205</td>
</tr>
<tr>
<td>GPD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.7400</td>
<td>1.162</td>
<td>1.282</td>
</tr>
<tr>
<td>50</td>
<td>0.4831</td>
<td>0.7307</td>
<td>0.8009</td>
</tr>
<tr>
<td>100</td>
<td>0.5458</td>
<td>0.7363</td>
<td>0.7903</td>
</tr>
<tr>
<td>GEV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.6744</td>
<td>0.8129</td>
<td>0.8521</td>
</tr>
<tr>
<td>20</td>
<td>0.6991</td>
<td>0.9196</td>
<td>0.9821</td>
</tr>
<tr>
<td>50</td>
<td>0.5802</td>
<td>0.8341</td>
<td>0.9061</td>
</tr>
</tbody>
</table>

Table 1. Estimation of $\xi$ using three techniques, and different thresholds (here is given the number of exceeding observations) or bloc size, including the upper bound of the confidence interval.
2.3 Testing whether $\xi \geq 1$ (non finite pure premium) or $\xi < 1$

Looking at Table 1, we should be suspicious about the non existence of a finite premium. It is also possible to test whether $\xi \geq 1$ ($H_0$ hypothesis) or $\xi < 1$ (the alternative hypothesis $H_1$). REISS & THOMAS (2001) or COLES (2001) suggest some likelihood ratio tests, based either on the maxima approach, or the exceedance, which can be performed when changing $H_0$. The likelihood ratio test for testing $H_0^* : \xi = 1$ against $H_1 : \xi < 1$,

with unknown $(\mu, \sigma)$ parameters for GEV distributions, or unknown $\sigma$ parameter for GPD distributions, is

$$LR(Y_1, ..., Y_m) = 2 \log \frac{\sup \{ L(\xi | Y_1, ..., Y_m, \xi = 1) \}}{\sup \{ L(\xi | Y_1, ..., Y_m, \xi < 1) \}},$$

where the $Y_1, ..., Y_m$ are either block maxima, or exceedance observations, and $L$ denotes the respective likelihood functions (product of GEV or GPD densities). Because of the dimension of the parameters, the $LR$-statistic is asymptotically distributed according to a chi-square distribution with $k = 1$ degree of freedom (under the null hypothesis, where $k$ is the difference in the dimensionality of the two hypothesis). Consequently, the associated $p$-value is $p = 1 - \chi^2_1 (LR)$. As mentioned in HOSKING (1984) (when testing if $\xi = 0$ but the result still holds here), the significance level is attained with a higher accuracy by employing Bartlett correction, when the $LR$ statistic is replaced by $LR/(1 + b/m)$ (where $b = 4$ in the GPD model, and $b = 2.8$ in the GEV model), and the associated $p$-value of the test will be denoted $\tilde{p}$.

**Example 3.** Table 2 shows some Likelihood Ratio tests of $H_0^* : \xi = 1$ against $H_1 : \xi < 1$. Since $p$-values almost always exceed 5% we should confidently reject the assumption of $\xi = 1$. The only cases where it might be accepted is when 100 exceeding observations are considered, but this might yield us far away from the asymptotic assumption in de Haan-Balkema theorem, and when 10 blocks are considered (only 10 observations) to fit 3 parameters (or 2 when model is constrained).
Table 2. Estimation of non-constrained and constrained (\(\xi = 1\)) GEV and GPD distribution, including likelihood ratio statistics, and the associated p-value of the test.

<table>
<thead>
<tr>
<th>(\xi)</th>
<th>(\mu)</th>
<th>(\sigma)</th>
<th>Deviance</th>
<th>(LR)</th>
<th>(p)</th>
<th>(\hat{p})</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.6743</td>
<td>2.1926</td>
<td>1.7189</td>
<td>600.0943</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2.100</td>
<td>1.904</td>
<td>607.3329</td>
<td>7.238</td>
<td>0.007</td>
</tr>
<tr>
<td>20</td>
<td>0.6985</td>
<td>3.6262</td>
<td>2.4885</td>
<td>345.7140</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3.440</td>
<td>2.587</td>
<td>348.0647</td>
<td>2.351</td>
<td>0.125</td>
</tr>
<tr>
<td>50</td>
<td>0.5801</td>
<td>7.7519</td>
<td>4.6499</td>
<td>166.0185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>7.496</td>
<td>5.584</td>
<td>168.9390</td>
<td>2.921</td>
<td>0.087</td>
</tr>
<tr>
<td>GPD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.7398</td>
<td>–</td>
<td>3.6227</td>
<td>151.3493</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>–</td>
<td>3.224</td>
<td>151.8389</td>
<td>0.489</td>
<td>0.484</td>
</tr>
<tr>
<td>50</td>
<td>0.4831</td>
<td>–</td>
<td>4.1212</td>
<td>237.7393</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>–</td>
<td>3.428</td>
<td>241.6838</td>
<td>3.944</td>
<td>0.047</td>
</tr>
<tr>
<td>100</td>
<td>0.566</td>
<td>–</td>
<td>2.057</td>
<td>461.9841</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>–</td>
<td>1.654</td>
<td>467.1144</td>
<td>5.130</td>
<td>0.023</td>
</tr>
</tbody>
</table>

In the general case (see Lehman (1986)), the likelihood ratio test can not be performed when testing \(H_0\) against \(H_1\).

### 2.4 Application to reinsurance premium calculation

Assuming that the generalized Pareto can be assumed to model a tail, it becomes possible to calculate easily some reinsurance XL covers (with even infinite limit). Hence, if \(Y\) has a generalized Pareto distribution, the survival distribution function of \(Y - d\) given \(Y > d\) is

\[
P(Y - d > x | Y > d) = \frac{P(Y - d > x)}{P(Y > d)} = \left(\frac{\sigma + \xi (d + x)}{\sigma + \xi d}\right)^{-\frac{1}{\xi}}.
\]

Therefore, the associate pure premium can be derived, since

\[
E(Y - d | Y > d) = \int_0^\infty P(Y - d > x | Y > d)dx = \frac{\sigma + \xi d}{1 - \xi}, \text{ and finally, one gets}
\]

\[
E(Y | Y > d) = d + \frac{\sigma + \xi d}{1 - \xi}.
\]

The first step is to chose some threshold \(u\) so that \(YL = (X - u | X > u)\) has a
generalized Pareto distribution.

**Example 4.** On the same dataset of business interruption claims, four reinsurance contracts have been priced, using different thresholds over which the claims should have a generalized Pareto distribution. Even without finite limit, premiums can be obtained easily, and the results are relatively stable, as shown in Table 3.

<table>
<thead>
<tr>
<th>( u )</th>
<th>( n )</th>
<th>( \xi )</th>
<th>( \sigma )</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>478</td>
<td>0.6030423</td>
<td>0.7811347</td>
<td>4.4869</td>
<td>14.5636</td>
<td>27.1594</td>
<td>52.3510</td>
</tr>
<tr>
<td>2</td>
<td>129</td>
<td>0.505470</td>
<td>2.003365</td>
<td>-</td>
<td>14.1616</td>
<td>24.2722</td>
<td>44.4934</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>0.5575541</td>
<td>3.5601549</td>
<td>-</td>
<td>-</td>
<td>30.6481</td>
<td>53.2497</td>
</tr>
</tbody>
</table>

Table 3. Prices for some Excess of Loss treaties, with priority \( d \) and no finite limit.

3. **AN ALTERNATIVE TO PURE PREMIUM**

Even if a pure premium can be calculated, since we should reject the assumption of infinite mean for business interruption claims, we can still wonder if it is still an interesting benchmark.

3.1 **Why is the pure premium the usual benchmark?**

The pure premium \( \pi \), or the expected value, is a key notion in insurance business. As mentioned in Denis & Charpentier (2004), it is the constant value that is the closed from the random claims, in the \( L^2 \) distance,

\[ \pi = E(X) = \arg\min\{E(X - c)^2, c \in \mathbb{R}\}, \]

where \( X \) is the random cost of claims,

because of the law of large numbers, the average cost converges to the pure premium,

\[ \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i P \rightarrow \pi = E(X), \]

where the convergence in probability means that the probability that \( \overline{X}_n \) becomes sensibly different from \( \pi \) tends to 0 as \( n \to \infty \) : for all \( \varepsilon > 0 \),

\[ P\left( \left| \overline{X}_n - \pi \right| > \varepsilon \right) \to 0 \text{ when } n \to \infty \]
if the premium is lower than the pure premium the insurance company will default
with probability one, i.e. for all \( p < \pi \)

\[
\lim_{n \to \infty} P(X_n > p) = 1
\]

based on von Neuman & Morgenstern approach, given a utility function \( u \)
increasing and concave (risk aversion), a certainty equivalent is \( p \) such that

\[
E(u(X)) = u(p).
\]

Since \( u \) is concave, using Jensen inequality, note that

\[
p = u^{-1}(E(u(X))) \geq u^{-1}(u(E(X))) = E(X) = \pi.
\]

Hence, any premium for risk adverse insured (\( u \) concave) should always be higher
than the pure premium.

Because of those four results, the pure premium became the actuarial basis to
calculate a premium. Note that the explanations based on the distance minimization and the
law of large number both rely on the assumption of finite variance: they can only be used
when dealing with “light” tails (excluding large claims risks, for instance business
interruption). One can also wonder why considering the \( L^2 \) distance, and not another one
(which yields different results). And finally, recall that the law of large numbers, and the
ruin probability property are both asymptotic: they can be used only when \( n \) is large
enough, i.e. for an infinite portfolio, or an infinite time horizon. Here, in business
interruption line of business, portfolio are relatively small, and in practice, insurance
companies assess solvency rules using finite horizon ruin probabilities. Further, most of the
insurance company can afford to loose money on that line of business (i.e. ruin occurs)
since large insurance groups sell those policies, and using diversification of their portfolio,
they can afford non-null ruin probabilities. For all those reasons, the pure premium might
not be an appropriate and relevant benchmark to assess insurability of those large risks.

3.2 Distorted premium principle

Hence, other premium principles can be considered. A large survey on premium
principles in insurance can be found in GOOVAERTS, DE VYLDER & HAEZENDONCK (1984).
For instance, distorted premiums can be considered: instead of considering, for positive
losses \( \pi = E(X) = \int_{\theta}^{\infty} [1 - F(x)] dx \),

one idea (see, or DENUIT & CHARPENTIER (2004, 2005) for a general survey) can be
to distort the survival function,
\[ \pi_g = \int_0^\infty g([1 - F(x)])dx, \]

where \( g : [0,1] \to [0,1] \) is some distortion function, increasing and continuous, with \( g(0) = 0 \) and \( g(1) = 1 \).

An economic justification of this concept can be Yaari (1987) dual’s approach. As pointed out by Allais (1953) or Ellsberg (1961), most of the axioms on which is based von Neuman & Morgenstern approach are not relevant (and therefore, there is no reason for item 4 to be fulfilled, and the premium might be lower than the pure premium). The premium is obtained here using a distortion of cumulated probabilities, rather than a utility function. Let \( g \) denote a distortion function then set

\[ \pi_g (X) = \int_0^\infty g(1 - F(x))dx \]

where \( F(x) = P(X \leq x) \).

Recall that a premium principle is interesting in the case it conserves stochastic inequalities: if \( \preceq \) is a stochastic ordering, one can expect that if \( X \preceq Y \), \( \pi(X) \leq \pi(Y) \). Here the premium \( p \) should then be a solution of \( \pi_g (X) = \pi_g (p) = p \), hence, a distortion risk measure can be seen as a premium principle: \( X \preceq Y \) where \( \preceq \) is a stochastic order satisfying Yaari’s axioms if an only if the premium of \( X \) is smaller than the premium associated to \( Y \) (see also Denneberg (1990, 1994), Wang (1995, 1996) or Wang, Young & Panjer (1997)).

**Remark 5.** Comparing with the pure premium can also be done in the dual theory: distortion functions \( g \) such that \( \pi_g (X) \geq \pi = E(X) \) for all \( X \) (as the convexity property in the expected value context) is obtained when the distortion function is below the first diagonal. If such a result can be obtained for several distortion functions (see e.g. Wang (1996)), it is usually not the “empirical” shape, also called “inverted-S shape of distortion functions (from Preston & Baratta (1948) to Ryan & Vaithianathan (2003)).

Distortion premiums satisfies several properties, such as positive homogeneity, translation invariance, monotonicity, or additivity for comonotonic risks. It is also subadditive in the case where \( g \) is concave (see e.g. Hürliman (1998)).

### 3.3 Some examples of distorted premiums

Most of the standard risk measures can be seen as distorted premiums.

**Example 6.** Value-at-Risk (VaR) and Tail-Value-at-Risk (TVaR) are two well-known distorted premium, where the distortion functions are respectively
\( g(x) = I(x \in (p, 1]) \) and \( g(x) = \min\{x/p, 1\} \). Proportional hazard transform principle is also obtain when \( g(x) = x^\rho \). In the case where the claims have a generalized Pareto distribution in tails (when threshold \( u \) is exceeded), then

\[
VaR(X, p) = u + \frac{\sigma}{\xi \left( \frac{1 - p}{P(X > u)} \right)^{-\xi} - 1}
\]

when \( p \) is large enough,

and

\[
TVaR(X, p) = VaR(X, p) \left[ \frac{1}{1 - \xi} + \frac{\sigma - \xi u}{1 - \xi VaR(X, p)} \right]
\]

when \( p \) is large enough

Hence, using estimation techniques mentioned in Section 2, those premiums can be estimated easily. In the case of business interruption claims, those two premiums can be obtained, for large \( p \), including some confidence interval (see Figure 3 and Table 4).

**Figure 3.** Estimation of the VaR and the TVaR with levels 99.9%, using GPD maximum likelihood techniques, where \( u = 1 \) on the left, \( u = 5 \) on the right.

<table>
<thead>
<tr>
<th>( u )</th>
<th>( \text{Var}(99.9%) )</th>
<th>( \text{TVar}(99.9%) )</th>
<th>( \text{Var}(99.9%) )</th>
<th>( \text{TVar}(99.9%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>27.50320</td>
<td>46.57878</td>
<td>99.60774</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>48.80638</td>
<td>115.46481</td>
<td>205.12015</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>29.09757</td>
<td>47.26090</td>
<td>89.10651</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>56.36115</td>
<td>120.26600</td>
<td>205.12015</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25.73079</td>
<td>43.85056</td>
<td>144.09078</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>42.10588</td>
<td>238.93939</td>
<td>205.12015</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Distorted premiums for business interruption claims, using GPD approximation (with different thresholds \( u \)).
Acknowledgment: The author would like to thank Johan Segers for interesting discussions on testing whether a distribution has finite moments or not, and the French Federation of Insurers for kindly providing a dataset.

REFERENCES


