

## MEASUREMENT OF EQUITY

S. DAVID PROMISLOW

### ABSTRACT

When an insurer classifies risks for the purpose of setting rates, one of the goals is to achieve equity. In general, however, inequities can be introduced when only certain factors are considered and others ignored. This paper, guided by ideas used by economists to measure income inequality, proposes a method of measuring this inequity. The proposed method is then used to investigate the effect on inequity of refining the classification of risks.

---

### INTRODUCTION

The concept of equity is one that is frequently encountered in actuarial terminology. Indeed, one possible definition of an actuary is an individual concerned with maintaining adequacy and equity in insurance schemes and similar financial programs. There is a problem, however, of deciding on a reasonable definition of equity. How do we justify our decisions when we say that one course of action is more equitable than another?

In an age where there is increasing concern for human rights, the insurance industry is often required to justify actions which appear to discriminate on the basis of such characteristics as age, sex, and race. The defense of equity is frequently invoked. Suppose we have two groups of individuals, group A and group B, such that the expected cost per individual of providing a certain insurance benefit is higher for those in group B. The standard and familiar argument is that it must be more equitable to charge a higher premium to those individuals in group B. If the groups consist of relatively homogeneous risks, this proposition seems quite valid. However it does not seem as obvious in the case where there is a great deal of variation in risk within the groups. A main goal of this paper is to try to provide a deeper analysis of this argument than it seems to have received.

There are many possible definitions and philosophical questions dealing with the idea of equity. In this paper we are going to focus on only a single aspect of this concept. We believe that it is easy in *theory* to give a perfectly

acceptable definition of equity in insurance. An individual wishes to insure against a loss. The amount of this loss is, of course, not strictly determined but is rather a random variable  $L$ . We postulate as a basic principle that equity simply involves *charging the expectation of this random variable as a net premium*, and we will take this as a departure point for our further investigations. The problem arises however since an insurer may be unwilling or unable to determine the random variable  $L$ . In assessing a risk one may look at certain obvious factors such as age or sex but ignore other characteristics. Our sole concern in this paper is the inequity that arises from making use of less than perfect information in determining the expectation of loss.

Consider again, the situation with the two groups of individuals. If one assesses the risk by looking at only certain factors and ignoring others, then it can indeed turn out that within each group there is a diverse collection of risks. Suppose an individual is in the group with the higher mean expectation but actually has an expectation lower than the average of the total collection of people. It is clear that such a person is treated less favorably than he or she deserves to be. This is an argument often put forth by the layman, but it tends to be dismissed by the expert as being somewhat naive. It would appear, however, that it merits more scrutiny. It is true, as the expert inevitably argues, that we must consider equity as a whole and not just that of particular individuals. However this brings us squarely face to face with the question of how to determine equity as a whole. The fact that *any* decision will necessarily result in *some* individuals being treated less fairly than they should be indicates that there is no obvious solution. We should not merely offer simple intuitive explanations without further analysis.

The first task then is to arrive at some method for the measurement of equity. To my knowledge this particular question has not been previously considered, but a similar problem has received a great deal of attention in the economic theory literature. This is the measurement of income inequality. In section 3 we use some results from this area as a guide and propose some formulas for measuring equity. Section 5 is devoted to giving some justification that our formulas really do achieve the main features that one would expect from such a measure.

The rest of the paper is mainly devoted to the investigation of whether or not these formulas lead to greater equity when the risk classification is refined. There are no definite answers as it depends on which one of many possible formulas in a given family is used, but an interesting fact is that it is *not* always 'the case. We show this in section 6 where we present an

example using a measurement formula based on the idea of entropy in information theory, a concept which has been used extensively for similar measurement problems. In section 7 this question is investigated in more detail, and at the same time a more general mathematical model is formulated. Some theorems are stated, giving conditions under which refining the risk classification will or will not achieve greater equity. These results depend heavily on a concept of comparing *riskiness* of random variables, which is discussed in section 4.

## 2. THE RESTAURANT ANALOGY

The problem of determining equity arises in many situations other than insurance. We will attempt to motivate our theory by looking at a very simple situation. Suppose that three people enter a restaurant and order meals for one dollar, two dollars, and three dollars, respectively. The total bill of six dollars must then be divided up among the three. It seems likely that most people would agree that the most equitable way of doing this is to have each person pay his or her own share of the bill. (There of course may be other opinions. For example, some may believe that more equity is achieved by sharing the cost. We are not going to consider these here. As indicated in the introduction, it is in accordance with our basic principle to adopt as axiomatic the fact that each person paying his or her share is the most equitable procedure). Similarly most would agree that the least equitable method of division is to require the person with the minimum one dollar charge to pay the entire bill. It is not at all obvious however as how to rank the infinite number of possibilities between these two extremes (e.g., Is it more equitable for each to pay two dollars, or for the three dollar person to pay his or her share and let the other two each pay a dollar and a half?).

Consider a more elaborate situation. A large table of people is seated, and they order meals of various amounts. Nobody knows what the amount of his or her order is, and each will pay whatever bill he or she receives. The waiter decides that it is too much work to give separate bills and proposes to give only one, which will be divided equally among the entire table. Somebody objects that this is inequitable. The waiter then decides that it is possible to give two bills, one for the north side of the table, to be divided equally among those who sat on that side, and another for the south side. Is the second scheme more equitable than the first or not? It is this type of question that we wish to address.

Note that the waiter in the last example is in a similar position to the insurer mentioned in the introduction. Faced with the difficulty of making

an exact determination of actual charges, the tendency is to look at obvious, easily determinable factors such as "side of table" and ignore others.

Some will no doubt argue that the situation is different in that the factor "side of table" has no bearing on the person's bill. It is possible however to imagine situations where there could be some influence. Suppose for example that generally the large eaters tend to sit on the north side, where the view is not as good, since they are more interested in eating than looking. Consider a light eater who also sits on the north side since he does not like the particular view. He is in a similar position to the low cost individual in the high cost group which we referred to in the introduction.

### 3. THE DISCRETE MODEL

Using the restaurant situation as a guide, we will construct a mathematical model for measuring equity. In general let us consider for a group of  $n$  individuals, two nonnegative vectors.

$$\mathbf{a} = (a_1, a_2, \dots, a_n) \text{ and } \mathbf{b} = (b_1, b_2, \dots, b_n)$$

satisfying

$$\sum a_i = \sum b_i$$

and

$$(3.1)$$

$$a_i = 0 \text{ implies } b_i = 0.$$

In this discussion,  $a_i$  represents the fair cost incurred by the  $i$ th individual, and  $b_i$  represents the actual charge made to that individual. The first condition in (3.1) simply states that the total charges must equal the total costs, and the second means that we do not consider requiring payments from people who did not incur a cost from the beginning.

In certain cases we will want to postulate the further condition that

$$b_i = 0 \text{ implies } a_i = 0. \quad (3.2)$$

This says that anybody who incurs a cost must be charged some positive amount.

We want to define for all positive integers  $n$  and all such vectors  $\mathbf{a}$  and  $\mathbf{b}$  a real number  $U(\mathbf{a}, \mathbf{b})$  which measures the amount of inequity inherent when the fair costs are given by  $\mathbf{a}$  but the actual charges made are given by  $\mathbf{b}$ . (The letter  $U$  stands for *unfairness* as opposed to  $I$  for inequity which is an overworked letter in mathematics.)

As mentioned in the introduction, a guide to defining  $U(\mathbf{a}, \mathbf{b})$  can be found by looking at methods of measuring inequality of incomes. Suppose we have a population of  $n$  individuals, and the  $i$ th individual receives a proportion  $p_i$  of the total income allocated to the group. Economists wish to measure the inequality resulting from such a division. This is done for a variety of purposes such as comparing inequalities between certain countries or comparing the equality-producing effects of different tax measures. Although the emphasis may be different, from a mathematical point of view we really have the situation described at the beginning of the section where for all  $i$ ,  $a_i = 1/n$  and  $b_i = p_i$ . In other words, the income inequality case deals with the situation where all fair costs are equal (making the usual assumption of zero-order homogeneity).

There is a great deal of literature on the question of measuring income inequality, and many methods have been proposed. See Marshall and Olkin [13, section 13F] for a brief survey. We do not want to give an exhaustive treatment here, but simply want to note that much of the recent literature, Cowell and Kuga [4], Shorrocks [16], and Theil [17], suggests that an appropriate form of measurement in our model is as follows. We first define for each  $i$ , an *inequity ratio*

$$r_i = b_i/a_i$$

which is simply the ratio of actual charge to fair cost for the  $i$ th individual. We will always take  $r_i = 1$  if  $a_i = b_i = 0$ . We then define

$$U(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n a_i g(r_i) \quad (3.3)$$

where  $g$  is a *convex* function with  $g(1) = 0$ . (Some basic facts about convexity are reviewed in section 4.)

Note that we do not have a single formula for measuring inequity but rather a family of formulas depending on the convex function  $g$ .

In section 5 we will present evidence that formula (3.3) does indeed achieve many of the features that one would expect from a measure of inequity. We would first like to recast the definition in a more probabilistic form.

For any vector  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  we let  $\sum_i |a_i|$  be denoted by  $\|\mathbf{a}\|$ . (The usual mathematical notation is  $\|\mathbf{a}\|_1$  indicating the  $l_1$  norm. We will suppress the subscript here as no other norms are involved.)

Suppose we choose *one unit* of fair cost at random, where we consider all such units to be equally likely. Let  $R$  be the random variable giving the

equity ratio for that unit. For example, in the case where  $\mathbf{a} = (1,2,3)$  and  $\mathbf{b} = (2,2,2)$ , we would have  $R$  taking the values 2, 1, and  $2/3$  with probabilities  $1/6$ ,  $2/6$ , and  $3/6$ , respectively.

In general,  $R$  will take the value  $b_i/a_i$  with probability  $a_i/\|\mathbf{a}\|$ , and it follows easily that

$$E(R) = 1. \quad (3.4)$$

We can now write (3.3) as

$$U(\mathbf{a}, \mathbf{b}) = \|\mathbf{a}\| E[g(R)]. \quad (3.5)$$

We can already see that our formula has some intuitive appeal. Inequity is measured by simply multiplying the total costs by the expected value of a certain function of the inequity ratio.

In many cases we will want to consider the amount of inequity per unit of fair cost, and so we define what we will call the *inequity index*

$$U_o(\mathbf{a}, \mathbf{b}) = \|\mathbf{a}\|^{-1} U(\mathbf{a}, \mathbf{b}) = E[g(R)].$$

One point to note is that, in general, we will view  $U$  as an ordinal rather than a cardinal measure. That is, we will not be interested in the actual values of the function but rather in the order relation which it induces on the possible actual charges for given fair costs. We will consider therefore two such functions  $U$  and  $U'$  to be equivalent if they result in the same such relation. To state this precisely,  $U$  and  $U'$  are *equivalent* if for all pairs of vectors with the same first entry,  $(\mathbf{a}, \mathbf{b})$  and  $(\mathbf{a}, \mathbf{c})$  we always have

$$U(\mathbf{a}, \mathbf{b}) \leq U(\mathbf{a}, \mathbf{c}) \text{ if and only if } U'(\mathbf{a}, \mathbf{b}) \leq U'(\mathbf{a}, \mathbf{c}).$$

If  $U$  is given by (3.3), multiplying  $g$  by any positive constant gives an equivalent formula. Note also adding a multiple of  $(x-1)$  to  $g$  gives in fact exactly the same formula in virtue of (3.4).

A concept of duality for measurement formulas was suggested by Bourguignon in [3]. Similar ideas appear in [17]. We would like to develop this idea in more generality here for application to later sections of the paper.

Suppose now that (3.2) holds, which means that the vectors  $\mathbf{a}$  and  $\mathbf{b}$  can be treated symmetrically. Given any function  $U$  for measuring inequity, we can naturally associate another function with it by simply reversing the roles of the fair costs and actual charges. To be precise, we have a dual function  $U^*$  given by

$$U^*(\mathbf{a}, \mathbf{b}) = U(\mathbf{b}, \mathbf{a}) \quad \text{for all } \mathbf{a} \text{ and } \mathbf{b}.$$

In the case that  $U$  is given by (3.3) for the convex function  $g$ , it is not difficult to see that  $U^*$  is also of this form, given by the function

$$g^*(x) = x g(1/x)$$

which is convex on the interval  $(0, \infty)$ .

4. RISKINESS

At this point, we will digress somewhat and discuss a result from risk theory which we will eventually apply to our random variable  $R$ . Consider two random variables  $X$  and  $Y$  with the same expectation. There are many situations (insurance, gambling, investment strategy) which involve an exchange of one such random variable for another and for which we wish to compare the random variables as to the degree of risk. One of the most appropriate and useful methods is as follows. For  $X$  and  $Y$  as above, let us say that  $X$  is less risky than  $Y$  if

$$\int_t^\infty 1 - F(x) dx \leq \int_t^\infty 1 - G(x) dx \text{ for all } t \tag{4.1}$$

where  $F$  and  $G$  represent the distribution functions of  $X$  and  $Y$ , respectively. In view of the fact that  $E(X) = E(Y)$ , formula 4.1 is equivalent to

$$\int_{-\infty}^t F(x) dx \leq \int_{-\infty}^t G(x) dx \text{ for all } t. \tag{4.2}$$

In the literature, (4.2) is usually referred to as *second order dominance* of  $X$  over  $Y$ , while (4.1) is referred to as *second order stop-loss dominance* of  $Y$  over  $X$ . See Goovaerts, De Vylder, and Haezendonck [10] for the general  $n$ th order definitions. Also see Rothschild and Stiglitz [15] for an extensive discussion of this concept together with other equivalent formulations. Before stating our main result, we will review some facts about convex functions. Recall that a real valued function  $g$  defined on an interval  $I$  of the real line is said to be *convex* if for all  $s$  and  $t$  in  $I$  and all  $\alpha$  in the interval  $(0,1)$

$$g[\alpha s + (1 - \alpha)t] \leq \alpha g(s) + (1 - \alpha)g(t) \tag{4.3}$$

One of the most useful consequences of this is the *increasing slope condition*. Given any points  $x < y < z$  in  $I$ , we apply (4.3) with  $s = x$ ,  $t = z$  and  $\alpha = (z - y)/(z - x)$  to obtain that

$$[g(y) - g(x)]/(y - x) \leq [g(z) - g(y)]/(z - y). \tag{4.4}$$

From (4.4) it is not hard to deduce that  $g$  is continuous at all interior points of the interval  $I$ , although it is not necessarily continuous at the end points of  $I$ , if any.

The function  $g$  is said to be *strictly convex* if the inequality sign is strict in (4.3) and hence also in (4.4). Geometrically, this means there is no subinterval of  $I$  over which the graph of  $g$  is a straight line.

In the case that  $g$  is twice differentiable, convexity of  $g$  is equivalent to the condition that  $g'' \geq 0$ .

The main result about riskiness that we want is:

**THEOREM 4.1.** *Suppose that  $X$  is less risky than  $Y$ . Then for any convex function  $g: R \rightarrow R$  (or  $R^+ \rightarrow R$  in the case that  $X$  and  $Y$  are nonnegative),*

$$E[g(X)] \leq E[g(Y)]$$

*provided that both expectations exist.*

There are several proofs given in the literature. See, e.g., [10] or [15]. Most proofs seem to involve some restriction, such as requiring boundedness of the random variables or differentiability of  $g$ , but the result is true in the full generality as stated in Theorem 4.1.

We can easily show that the conclusion of Theorem 4.1. is a *sufficient* condition for  $X$  to be less risky than  $Y$ . For any real number  $t$ , define the functions

$$\alpha_t(x) = \begin{cases} 0, & \text{if } x \leq t \\ x-t, & \text{if } x \geq t \end{cases} \quad \beta_t(x) = \begin{cases} t-x, & \text{if } x \leq t \\ 0, & \text{if } x \geq t. \end{cases}$$

These are convex functions and (4.1) says precisely that the conclusion of the theorem holds for all  $\alpha_t$ , while (4.2) says that it holds for all  $\beta_t$ . The proof of Theorem 4.1 can be achieved by showing that any convex function can be suitably approximated by a positive linear combination of the  $\alpha_t$ 's and  $\beta_t$ 's. We will not go into the details here. We will return however to these functions in section 7.

For a typical interpretation of Theorem 4.1, as given in [15], note that a risk adverse individual will have a *concave* utility function. The inequality in the conclusion reverses and, as terminology suggests, the less risky choice is preferred.

For one application to our problem at hand, observed by Atkinson [1] in the income inequality case, note that if we measure equity according to formula (3.3), then the ranking of possibilities may depend on the



choice of the function  $g$ . However suppose we have two choices of distributing the actual charges leading to the random variables  $R_1$  and  $R_2$  such that  $R_1$  is less risky than  $R_2$ . Then it does not matter what  $g$  is. We always get more equity by choosing the distribution leading to  $R_1$ .

We will give more applications of the riskiness concept in section 7.

#### 5. PROPERTIES OF $U(a,b)$

We now look at certain features which result from our definition of inequity. Some of these ideas have been extensively discussed in the income inequality case, and we want to adapt them to the case of inequity measurement.

*The Transfer Principle.* In the income inequality case, this was formulated by some of the earliest writers on the subject, see [6, 14], and accordingly is often referred to as the Pigou-Dalton property. Taken in our present context of inequity measurement, it would say that, if a person who is treated more unfairly than another has a sufficiently small portion of his charge transferred to that other person, then the equity, as a whole, should increase. To state this precisely, suppose that for given vectors  $a$  and  $b$  we have indices  $i$  and  $j$  and a positive constant  $s$  such that

$$(b_i - s)/a_i \geq (b_j + s)/a_j.$$

Let  $b'$  be defined as  $b$  except with  $(b_i - s)$  replacing  $b_i$  and  $(b_j + s)$  replacing  $b_j$ . Using the increasing slope property of convexity, we can see that

$$\frac{g[(b_j + s)/a_j] - g(b_j/a_j)}{s/a_j} \leq \frac{g(b_i/a_i) - g[(b_i - s)/a_i]}{s/a_i},$$

and after some algebraic manipulation, we get

$$a_j g[(b_j + s)/a_j] + a_i g[(b_i - s)/a_i] \leq a_j g(b_j/a_j) + a_i g(b_i/a_i),$$

and we arrive at the transfer principle

$$U(a, b') \leq U(a, b). \quad (5.1)$$

Another argument, essentially used in [1] for the income inequality case, is to note that a transfer as shown leads to a less risky distribution for  $R$  and then to invoke Theorem 4.1.

Many writers have advocated a stronger version of (5.1) in which the inequality sign is strict. In our context, this would say that a transfer should result in a definite decrease in inequity and not merely leave it

unchanged. This, of course, will occur if we require that  $g$  be strictly convex, since then the inequality sign will remain strict throughout the derivation.

*Minimum and Maximum Inequity.* We now show that the transfer principle does indeed give us the expected results regarding the points of minimum and maximum inequity. The postulate that  $g(1) = 0$  tells us that  $U(a,a) = 0$  for all  $a$ , obviously a required feature since we should have no inequity if everybody pays the fair cost. Moreover we see that this is the minimum value of  $U$ , since if  $b \neq a$ , we could reduce inequity by making a transfer. A consequence of this is that  $U(a,b)$  is always nonnegative, which is not completely obvious from the original definition since  $g$  can take negative values. Similarly, the maximum inequity occurs when the individual with a minimal fair charge must pay the total. In any other case, we could increase inequity by a transfer reverse to the type described. These same principles hold in more general situations. Suppose we fix the actual charges for certain individuals and then ask what is the fairest way to divide the remaining charges among those remaining. The transfer principle again tells us that we must divide proportionally so that all equity ratios of the remaining individuals are equal. Similarly the maximum inequity results when an individual with a minimum fair cost is required to pay all the remaining charges.

So, for example, given  $a = (1,2,3,4,5)$ , suppose we impose the constraint that  $b_4 = 1$  and  $b_5 = 2$ . Then  $U(a,b)$  is minimized for  $b = (2,4,6,1,2)$  and maximized for  $b = (12,0,0,1,2)$

*Subadditivity.* A direct consequence of the convexity of  $g$  is that, given two pairs of vectors  $(a,b)$  and  $(a',b')$ , we have that

$$U(a+a',b+b') \leq U(a,b) + U(a',b').$$

This property does not seem to have been considered in the income inequality case. In our context, we can give an interpretation using the restaurant analogy. Suppose that the same group of people visit two restaurants, each of which uses a different scheme for dividing the total bill. The preceding says that total inequity resulting from both visits is less than the sum of the inequities from each of the visits. This seems reasonable in view of the fact that individuals who were treated very unfairly in one case may receive very favorable treatment in the other, allowing some cancellation of inequity.

*Equity within and between groups.* We now discuss another important principle which has been extensively considered in measurement questions of the type

we are investigating. It gives some insight into a possible choice for the function  $g(x)$  appearing in (3.3).

It will be more convenient here to work with the inequity index  $U_0$  rather than  $U$ .

Fix  $\mathbf{a}$  and  $\mathbf{b}$  and select a certain subset  $S$  of our individuals. It is natural to try to compute that portion of the inequity, which results solely from the treatment of individuals within that subset. For simplicity in notation, suppose that  $S = \{1, 2, \dots, k\}$ , and let

$$\mathbf{a}_S = (a_1, a_2, \dots, a_k), \quad \mathbf{b}_S = (b_1, b_2, \dots, b_k).$$

If  $\|\mathbf{a}_S\| = \|\mathbf{b}_S\|$ , then it is clear that  $U_0(\|\mathbf{a}_S\|, \|\mathbf{b}_S\|)$  is the quantity we want. In general, however, these two subvectors will have different norms. Let

$$t = \|\mathbf{b}_S\| / \|\mathbf{a}_S\|$$

and define the inequity index pertaining to subset  $S$  by

$$U_0(S) = U_0(t \mathbf{a}_S, \mathbf{b}_S)$$

which is also equal to  $U_0(\mathbf{a}_S, t^{-1} \mathbf{b}_S)$  provided that  $\mathbf{b}_S$  is not the zero vector. In the case that  $U$  is given by (3.3), we see that

$$U_0(S) = E(R'),$$

where  $R'$  is the random variable equal to  $tR$ , conditioned on the fact that the unit of fair cost comes from subset  $S$ .

Note that if  $t$  is not equal to one, then the group  $S$ , as a whole, is treated more or less favorably than it should be. The quantity  $U_0(S)$ , however, ignores this source of the inequity and concentrates solely on the relative treatment of individuals within this group.

Suppose now that we have a partition of  $\{1, 2, 3, \dots, n\}$  into pairwise disjoint subsets  $S_1, S_2, \dots, S_k$ . In addition to the inequity within each group, we will have the inequity resulting from the treatment *between* the groups which we will denote by  $U_0^B$ . It seems natural to define this by simply looking at the total fair costs and actual charges within each group. That is, we take

$$U_0^B = U_0(c, d)$$

where

$$c = (\|\mathbf{a}_{S_1}\|, \|\mathbf{a}_{S_2}\|, \dots, \|\mathbf{a}_{S_k}\|) \text{ and } d = (\|\mathbf{b}_{S_1}\|, \|\mathbf{b}_{S_2}\|, \dots, \|\mathbf{b}_{S_k}\|).$$

A reasonable requirement is that the total inequity index  $U_0(\mathbf{a}, \mathbf{b})$  should depend in some natural way on the between group inequity  $U_0^B$  and the within

group inequities,  $U_o(S_i)$   $i = 1, 2, \dots, k$ . For the income inequality case, this idea was extensively investigated by Theil in [17] and was further considered by Bourguignon [3] and Shorrocks [16]. This latter work has a system of axioms concerned with such a between-within group decomposition. Applied in our context the axioms would show that up to equivalence as defined in section 3,  $U$  is given by (3.3) where the convex function  $g$  is one of the following family, indexed by a real number  $c$ .

$$\left. \begin{aligned} g_c(x) &= (x^c - 1)/[c(c-1)], \quad c \neq 0, 1 \\ g_0(x) &= -\ln(x) \\ g_1(x) &= x \ln(x) \end{aligned} \right\} \quad (5.2)$$

Exactly the same family was found by Cowell and Kuga in [4] using somewhat different axioms, which involve a weaker assumption about the between-within decomposition but additional assumptions on transfers.

The main property of this family is as follows. Suppose we are given vectors  $a$  and  $b$  and a partition as before. For each  $i = 1, 2, k$ , let

$$p_i = \|a_{S_i}\| / \|a\|, \text{ the proportion of fair costs for the } i\text{th group,}$$

$$q_i = \|b_{S_i}\| / \|b\|, \text{ the proportion of actual charges for the } i\text{th group}$$

$$w_{i,c} = p_i^{(1-c)} q_i^c.$$

Let  $U_{0,c}$  denote  $U_o$  calculated by (3.3) using  $g_c$ . (Note that in the case that  $c \leq 0$ , we must assume (3.2) since  $g_c$  is not defined at 0). Then, a straightforward calculation shows that for all  $c$ ,

$$U_{0,c}(a,b) = U_{0,c}^B + \sum_{i=1}^n w_{i,c} U_{0,c}(S_i). \quad (5.3)$$

In other words, this says that using any of the convex functions in this family, we have the total inequity index decomposed into a portion which is due to the inequity between groups plus a weighted sum of the within-group inequity indices.

The cases of particular interest are when  $c = 0$  or 1, since these are the only values giving weights which sum to unity. For  $c = 1$ , the weights are the proportion of actual charges, and for  $c = 0$ , the weights are the proportion of fair costs. Note that these two cases are dual to each other since we see immediately from (3.6) that

$$[x \ln(x)]^* = -\ln(x).$$

These cases, particularly  $c = 1$ , were considered extensively in [17]. The motivation is from information theory, where the function  $x \ln(x)$  forms the basis for Shannon's entropy. See [11]. The decomposition given in (5.3) corresponds to the information theoretic result that the entropy  $H$  of finite probability distributions satisfies

$$H(A,B) = H(A) + H(B|A).$$

Suppose that the vectors  $\mathbf{a}$  and  $\mathbf{b}$  have norm one, so they may be considered to represent finite probability distributions. In this case, the quantities  $U_{0,c}$  for  $c = 1$  and  $0$  correspond to the Kullback measures of divergence between the two distributions [13]. (Also see Brockett and Cox [2] for another application of Kullback divergence to an actuarial problem.)

Note that  $U_c$  depends continuously on the parameter  $c$ . This is not completely obvious when  $c = 0$  or  $1$  but, as indicated in [4], is easily seen by an application of L'Hopitals rule.

In the case where the weights are  $p_i$  or  $q_i$ , statement 5.3 is so strong that it almost characterizes the resulting inequity measure, in fact, without necessarily assuming (3.3) from the outset. Further research will show that, assuming only (5.3) with the weights  $w_{i,c}$  replaced by  $q_i$ , a mild continuity assumption, and the fact that  $U_0$  is always nonnegative and equal to  $0$  when the two vectors are equal, then, necessarily,  $U_0$  is given by (3.3) where  $g(x) = kx \ln(x)$  for some nonnegative constant  $k$ . The case where the weights are given by  $p_i$  similarly characterizes the function  $-\ln(x)$ . Similar results appear frequently in the literature. See [3], Foster [8], Gehrig [9].

Another fact of interest is to note that the result on duality mentioned previously generalizes to the fact that for all  $c$

$$U_{0,c}^* = U_{0,1-c},$$

which we see immediately from (3.4) and (3.6) since for all  $c \neq 0$  or  $1$

$$g_c^*(x) = g_{(1-c)}(x) - (x-1)/[c(c-1)].$$

To summarize, the class of measurement formulas resulting from the family (5.2) would seem to be an attractive one to use for the measurement of inequity. It includes most of the specific formulas which have been proposed for the measurement of income inequality and similar measurement problems.

## 6. AN EXAMPLE

We now turn to one of our basic questions. Suppose we have a group of individuals, and initially we have no information which will allow us to

distinguish between their respective fair charges. We propose to charge each the average amount. We are then given information which allows us to divide the group into two subgroups. At this point, we could charge each individual the average amount for the particular subgroup to which that person belongs. (Refer back to the restaurant situation with the large table for one concrete example.) Will this always lead to greater equity, or are there cases where we should ignore the additional information? (We are assuming here that the only goal is to maximize equity. In practice, there are certainly other factors, such as those involving market conditions, which would affect one's decision in this circumstance. We are not concerned however with these in this paper.)

We now want to give one example where, using one of the possible definitions of equity which we proposed in the last section, we should ignore the additional information. Refining the classification actually *decreases* equity.

### *Example*

Take  $g(x) = x \ln(x)$  and use (3.3). Suppose that the actual fair costs for a group of six individuals, are given by the vector

$$\mathbf{a} = (2,2,2,1,1,10).$$

Initially, all we know is that the total is 18, and without further information, we will make actual charges according to the vector

$$\mathbf{b} = (3,3,3,3,3,3).$$

Suppose now that we are given the information that the total cost of 18 is divided into 6 for individuals 1,2, and 3, and 12 for the remaining three individuals. Using this information, we could make charges according to

$$\mathbf{b}' = (2,2,2,4,4,4).$$

A direct calculation shows that  $U_0(\mathbf{a}, \mathbf{b}) = .368$ , while  $U_0(\mathbf{a}, \mathbf{b}') = .412$ .

It is instructive to analyze this in terms of the decomposition given in (5.3). Take the partition  $S = \{1,2,3\}$   $T = \{4,5,6\}$ . Using  $\mathbf{b}$  we have  $U_0^B = .059$ , while using  $\mathbf{b}'$  we obviously reduce the between-group inequity to 0. (This observation will inevitably lead some people to consider it as obvious that we must increase equity by refining the classification. However we must also consider the within-group inequity.) In both cases we have  $U_0(S) = 0$  and  $U_0(T) = .619$ . However in using  $\mathbf{b}'$ , the weight attached to group  $T$

increases from  $1/2$  to  $2/3$ , and this more than compensates for the decrease in  $U_0^B$ .

Suppose that, instead of using  $g(x) = x \ln(x)$ , we use the dual function  $g(x) = -\ln(x)$ . Now as we saw in section 5, the weights for the within-group inequities are based on the proportion of fair costs and these do not change. The within-group inequity stays the same, and the reduction in the between-group inequity causes a reduction in total. This happens not only for this, but all examples. So we have established a key fact that for the function  $-\ln(x)$ , refining the risk classification always results in additional equity.

Are there other functions which do this? It appears unlikely. Further research could show that given any other function in the class (5.2), we can find an example similar to that given. We conjecture that this is true for all convex functions.

This may cause some to feel that  $-\ln(x)$  is the best choice of function. There are, however, other considerations. Arguments in [5] and [16] for the income inequality case show that choosing a particular function  $g$  causes different relative effects of transfer payments depending on the income levels of the transferor and transferee. We will illustrate this in our context with a similar example.

### *Example*

Suppose that four individuals each with a fair cost of 4, are charged, respectively, 1, 3, 5, and 7. Let us compare the effect of the following:

- I. The second individual agrees to share the charges equally with the first, so each is charged 2.
- II. The fourth individual agrees to share the charges equally with the third so each is charged 6.

Consider the family of formulas given in (5.2). For  $c = 2$ , transfer II and transfer I both result in the same reduction in inequity. For  $c > 2$ , the reduction is greater for transfer II, and for  $c < 2$  the reduction is greater for transfer I. One's choice of  $c$  might then well depend on whether one feels that reduction in inequity should be more sensitive to transfers among those treated favorably or treated unfavorably. In the income equality context, some writers have advocated staying away from high values of  $c$ . As the preceding example indicates, the effect of such values is to give more weight to transfers among the "wealthy," and many are of the opinion that this should not be so. This argument would seem not to apply in our case. We

might well want to make inequity more sensitive to those who were treated unfavorably and, hence, use high values of  $c$ . Many opinions are possible, and we will not elaborate further.

7. THE GENERAL MODEL

We would like to formulate our main mathematical results in terms of a more general model. We will now consider our collection of individuals to be represented by a general probability space. The discrete model we discussed previously will turn out to be the special case where we have a space with  $n$  points, with equal probabilities.

We begin then with a probability space  $(\Omega, \Sigma, P)$  and consider two non-negative random variables:  $X$  representing the fair cost and  $Y$  representing the actual charge. By analogy to (3.1), we postulate that  $E(X) = E(Y)$ , and that  $X = 0$  implies that  $Y = 0$ .

Let  $\mu$  denote the common expectation of  $X$  and  $Y$ .

We now define a new probability measure  $P^*$  on  $(\Omega, \Sigma)$  by

$$P^*(A) = \int_A X(\omega)/\mu \, dP, \text{ for } A \text{ in } \Sigma.$$

We define the random variable  $R = Y/X$  on  $(\Omega, \Sigma, P^*)$ . This is legitimate since, by our assumption,  $R$  is defined on the set  $N = [\omega: X(\omega) \neq 0]$ , and for the complement  $N^c$ , we have  $P^*(N^c) = 0$ . The effect of this change of measure is that, as in our original model,  $R$  represents the inequity ratio per unit of fair cost. Now, letting expectations with respect to  $P^*$  be denoted by  $E^*$ , we have

$$E^*(R) = \int_N \{[Y(\omega)/X(\omega)][X(\omega)/\mu]\}dP = \mu^{-1} \int_N Y(\omega) \, dP = 1$$

since by assumption  $Y = 0$  on  $N^c$ .

We define the *inequity index*

$$U_0(X, Y) = E^*[g(R)].$$

Note that when  $P$  is the uniform measure on a finite set, the resulting  $P^*$  is precisely the probability measure considered in section 3, namely, that which gives equal probability to each unit of fair cost. Accordingly, the expectation  $E$  used in section 3 corresponds to  $E^*$  of our present section, and we see that the definition of *inequity index* agrees with that given previously when applied to the discrete model of section 3.



Let us consider the particular case that  $Y$  is the constant  $\mu$ . In this case  $R = \mu/X$ , and we will denote this random variable by  $R_X$ . We obtain information about  $R_X$  as follows.

LEMMA 7.1.

$$E^* [h(R_X)] = E [h^*(X/\mu)]$$

where  $h^*$  is defined in (3.6).

*Proof.* This follows directly from the definition of  $P^*$  since

$$\int_{\Omega} h[u/X(\omega)] dP^* = \mu^{-1} \int_{\Omega} h[u/X(\omega)]X(\omega) dP.$$

We now return to the functions  $\alpha$  and  $\beta$  defined in section 4.

LEMMA 7.2. For all  $s > 0$

$$(a) \alpha_s^* = s \beta_{1/s}$$

$$(b) \beta_s^* = s \alpha_{1/s}$$

*Proof.* These are direct calculations from the definitions. For example, suppose that  $0 < x \leq 1/s$ . Then  $1/x \geq s$  and  $\alpha_s^*(1/x) = 1/x - s$ . So  $\alpha_s^*(x) = 1 - (sx) = s(1/s - x) = s \beta_{1/s}(x)$ . The remaining calculations are similar.

THEOREM 7.3. If  $X$  is less risky than  $Y$ , then  $R_X$  is less risky than  $R_Y$ .

*Proof.* If  $X$  is less risky than  $Y$ , it is easily seen that  $X/\mu$  is less risky than  $Y/\mu$ , and applying Theorem 4.1,

$$E[\alpha_{1/s}(X/\mu)] \leq E[\alpha_{1/s}(Y/\mu)]$$

for all  $s > 0$ , so that by Lemma 7.2 (b)

$$E[\beta_s^*(X/\mu)] \leq E[\beta_s^*(Y/\mu)].$$

Now by Lemma 7.1, and the fact that  $g^{**} = g$  for all functions  $g$  defined on  $(0, \infty)$ ,

$$E^*[\beta_s(R_X)] \leq E^*[\beta_s(R_Y)],$$

and our conclusion follows directly from the definition as given in formula 4.2.

Suppose now that we have another random variable  $K$  which takes the values 1 and 2. So  $\Omega$  is partitioned into the two sets  $\{K=1\}$  and  $\{K=2\}$ . The idea here is that we have divided our set of individuals according to some observable characteristic, and we wish to analyze the resulting effect on equity. (A similar model was introduced by DeWit and VanEeghen in [7], but the emphasis in that work is somewhat different.) For simplicity, we will restrict our attention to the case that  $K$  has two values. Of course in practice, one wants to consider risk characteristics which can take on several different values (age for example). Analogous results to those following can be obtained, but the notation and formulation is much more complicated.

We first introduce some notation.

Let  $p = P(K=1)$ ,  $r = P^*(K=1)$ ,  $X_1 = (X|K=1)$ ,  $X_2 = (X|K=2)$ ,  $\mu_1 = E(X_1)$ ,  $\mu_2 = E(X_2)$ ,  $R_1 = \mu_1/X_1$  and  $R_2 = \mu_2/X_2$ . We note that  $X$  is a mixture of  $X_1$  and  $X_2$  with the weights  $p$  and  $1-p$ , respectively, and  $\mu = p\mu_1 + (1-p)\mu_2$ . Moreover

$$\begin{aligned} \mu_1 &= p^{-1} \int_{K=1} X(\omega) dP \text{ so that} \\ r &= \mu^{-1} \int_{K=1} X(\omega) dP = p \mu_1/\mu \end{aligned} \tag{7.1}$$

and similarly

$$1-r = (1-p)\mu_2/\mu. \tag{7.2}$$

Suppose now that we fix our actual charges according to the observation given by  $K$ . In other words, we take  $Y = \mu_K$ ,  $K = 1$  or  $2$ . This gives an inequity index of

$$U_o(X, \mu_K) = r E^*[g(R_1)] + (1-r) E^*[g(R_2)].$$

Suppose instead that we choose to ignore the effects of our observation and charge a constant  $\mu$ . Let  $s_1 = \mu/\mu_1$  and let  $s_2 = \mu/\mu_2$ . We see that now  $R = \mu/X$  will equal  $s_K R_K$  so that

$$\begin{aligned} U_o(X, \mu) &= E^*[g(R)] = r E^*[g(R) | K = 1] + (1-r) E^*[g(R) | K = 2] \\ &= r[E^*[g(s_1R_1)]] + (1-r) E^*[g(s_2R_2)]. \end{aligned}$$

Let  $D_K$  denote the gain in equity when the risk is refined according to the observation  $K$ . From the previous two expressions we have

$$D_K = r E^*[h_{s_1}(R_1)] + (1-r) E^*[h_{s_2}(R_2)] \quad (7.3)$$

where 
$$h_s(x) = g(sx) - g(x).$$

Consider now properties of the function  $h$  which we have just defined. In many of our previous examples of suitable  $g$ 's we can show that

A.  $h_s$  is convex for  $s > 1$  and concave for  $s < 1$   
or that (7.4)

B.  $h_s$  is concave for  $s > 1$  and convex for  $s < 1$ .

For example, if  $g(x) = x \ln(x)$ , then  $h_s(x) = xg(s) + (s-1)g(x)$  satisfies (7.4A), while if  $g(x) = x^c - 1$ , then  $h_s(x) = (s^c - 1)x^c$ . Dividing by  $c(c-1)$  shows that functions in the class given in (5.2) satisfy (7.4A) if  $c > 0$  and (7.4B) if  $c < 0$ . Naturally enough, for  $g_0(x) = \ln(x)$ ,  $h_s(x) = -\ln(s)$  is constant and trivially satisfies both A and B in (7.4).

Further insight into the behavior of  $h$  can be obtained if we consider the case where  $g$  is twice differentiable. Then

$$h''(x) = s^2 g''(sx) - g''(x).$$

This shows that (7.4A) always holds in the case that  $g''(x)$  is *increasing*.

This observation regarding  $-\ln(x)$  gives us an alternate derivation of the fact that  $D_K$  must be nonnegative when this function is chosen for  $g(x)$ . In this case,  $h_s(x) = -\ln(s)$ , implying that  $E^*[h_{s_i}(R_i)] = -\ln(s_i)$  for  $i = 1$  or 2. Now substitute in (7.3), recalling that  $s_i = \mu/\mu_i$ , and using (7.1) and (7.2). If we temporarily denote the function  $x \ln(x)$  by  $k(x)$ , this yields

$$D_K = pk(\mu_1/\mu) + (1-p)k(\mu_2/\mu)$$

and by the convexity of  $k$

$$D_K \geq k([p\mu_1 + (1-p)\mu_2]/\mu) = 0.$$

We now consider more general functions for  $g$  and state a theorem which compares the effect of two subdivisions  $K$  and  $K'$ . Suppose that  $K$  leads to the  $R.V.$ 's  $X_1$  and  $X_2$  and  $K'$  leads to the  $R.V.$ 's  $X'_1$  and  $X'_2$ .

**THEOREM 7.4.** *Suppose that  $g$  satisfies (7.4). Let the subscript 1 refer to the lower mean random variable in the case that A holds or to the higher mean random variable in the case that B holds.*

(a) Suppose that  $D_K$  is nonnegative. Then if  $X_1$  is less risky than  $X'_1$  and  $X'_2$  is less risky than  $X_2$ , we must have that  $D_{K'}$  is nonnegative.

(b) Suppose that  $D_K$  is nonpositive. Then if  $X'_1$  is less risky than  $X_1$  and  $X_2$  is less risky than  $X'_2$ , we must have that  $D_{K'}$  is nonpositive.

*Proof.* Part (b) follows immediately from (a), by reversing  $K'$  and  $K$ . We now derive (a). For definiteness we consider the case where case A of (7.4) holds. The derivation is similar in case B. Since  $s_1 = \mu/\mu_1$  is greater than 1,  $h_{s_1}$  is convex. Then, since  $R_1$  is less risky than  $R'_1$ , we can invoke Theorem 4.1 to conclude that

$$E^*[h_{s_1}(R_1)] \leq E^*[h_{s_1}(R'_1)].$$

Similarly, since  $s_2$  is greater than 1 and  $R'_2$  is less risky than  $R_2$ , we conclude that

$$E^*[h_{s_2}(R_2)] \leq E^*[h_{s_2}(R'_2)].$$

Now substitute in (7.4) to see that  $D_{K'}$  is greater than or equal to  $D_K$ .

There is one case in which we can say that refining the risk must increase equity regardless of what  $g$  is.

**THEOREM 7.5.** *Suppose that  $X_2/\mu_2 = X_1/\mu_1$ . Then  $D_K$  is nonnegative. If moreover  $g$  is strictly convex and  $\mu_1 \neq \mu_2$ , then  $D_K$  is positive.*

*Proof.* In this case  $R_2 = R_1$ . Moreover  $R$  is a mixture of  $(\mu/\mu_1)R_1$  and  $(\mu/\mu_2)R_2$  with weights of  $r$  and  $1-r$ , respectively. It follows that

$$U_0(X, \mu) = E^*[g(R)] = r E^*[g(\mu/\mu_1 R_1)] + (1-r) E^*(\mu/\mu_2 R_1)$$

which by the convexity of  $g$

$$\begin{aligned} &\geq E^*[g(r\mu/\mu_1 R_1 + (1-r) \mu/\mu_2 R_2)] \\ &= E^*[g(R_1)] = U_0(X, \mu_K). \end{aligned}$$

The second statement follows immediately since the given conditions clearly imply that the inequality sign is strict.

**COROLLARY 7.6.** *Suppose that (7.5) holds and let the subscript 1 be as in Theorem 7.3. If  $X_2/\mu_2$  is less risky than  $X_1/\mu_1$ , then  $D_K$  is nonnegative.*

*Proof.* This is immediate from Theorems 7.4 and 7.5.

The preceding theorems constitute only a beginning of the theory. There are still many questions to be answered concerning necessary and sufficient conditions for refinement to increase equity.

Consider the interpretation of our theorems. Let us suppose that (7.4) holds. We will call  $X_2$  as defined in the statement of Theorem 7.3 the *dominant* random variable. Our results show that refining the risk classification does indeed increase equity provided that the dominant random variable is not too risky relative to the other random variable. Refer back to the section 6 example. The dominant random variable is the high mean one, which is just too risky when compared with the low mean one. (Of course in this case, the low mean random variable is constant and carries the minimum possible degree of riskiness.) As the amounts of riskiness of the two random variables move towards each other, there is a greater tendency for refinement to increase equity. In the extreme case given in Theorem 7.5, we could interpret the given condition as stating that the two random variables have exactly the same degree of risk, and now we always gain equity by refining.

#### 8. SUMMARY AND CONCLUSIONS

We have not reached many definite conclusions regarding the measurement of equity and the effect of refining the risk classification. Indeed it was not the intention of the paper to do so but rather to introduce the subject. We can briefly summarize our ideas and results as follows:

1. To justify an action on the basis of equity should first require some method for measuring equity. This does not imply that one must necessarily arrive at definite numbers, or even that one has to select a definite formula. One however should be aware of the theoretical background and the consequences of choosing various methods of measurement.
2. There does not appear to be any unique method of measuring equity, but the family of methods given by (3.3) seems like a reasonable choice. It is intuitively plausible, satisfies the obvious properties that one wants from such a measure, and has been extensively used by economists in similar measurement problems. There is certain evidence that we may well further restrict the choices by those given in (5.2).
3. If one wants to require that refining the risk classification should always increase equity, then it is possible to do so within the framework of (3.3) by using the function  $g(x) = -\ln(x)$ . It seems likely that this is the unique such function. In any event, it is the unique such function in the restricted family given in (5.2).
4. If one wants to admit other choices of the function  $g(x)$ , then it turns out that refining the risk classification may increase equity in some case and not in others.

This in itself is an important message. It should refute, for example, the often heard argument that, if we began to ignore certain risk characteristics, then this will inevitably lead to ignoring such factors as age in life insurance. The factor of age appears to be one where the amount of riskiness does not increase noticeably as the mean changes. The results of section 7 would then indicate that we should distinguish by age. But this need not necessarily hold for all risk factors. There may be other cases, such as indicated in the section 6 example, where, due to the high variations in expected costs among one of the classes, we may legitimately question whether or not we gain equity by refining the classification.

5. Obviously, as we continue to refine our risk classification, we approach the situation where all individuals in a class have the same expectation, the point of perfect equity, as we have defined it. Hence, at some point in the refining process, we must start to gain equity. This is confirmed by the theoretical results of section 7. It does not follow however that this gain in equity must occur at *all* stages. It is a common fallacy to believe that natural processes must be monotone. It could well be in a given situation that equity *decreases* up to some point in the refining process and only then starts to increase.

#### REFERENCES

1. ATKINSON, A.B. "On the Measurement of Inequality," *Journal of Economic Theory*, 2(1970): 244-263.
2. BROCKETT, P.L., AND COX, S.H. "Statistical Adjustment of Mortality Tables to Reflect Known Information," *TSA*, XXXVI (1984): 63-75.
3. BOURGUIGNON, F. "Decomposable Income Inequality Measures," *Econometrica*, 47(1979): 901-920.
4. COWELL, F.A. AND KUGA, K. "Additivity and the Entropy Concept: An Axiomatic Approach to Inequality Measurement," *Journal of Economic Theory*, 25(1981): 131-143.
5. \_\_\_\_\_. "Inequality Measurement: An Axiomatic Approach," *European Economic Review*, 15(1981): 287-305.
6. DALTON, H. "The Measurement of the Inequality of Incomes," *Economic Journal*, 30(1920): 348-361.
7. DEWIT AND VAN EEGHEN. "Rate Making and Society's Sense of Fairness," *Premium Calculations in Insurance*, Nato ASI Series, Series C: Mathematical and Physical Sciences, Vol. 2: 151-169.
8. FOSTER, J.E. "An Axiomatic Characterization of the Theil Measure of Income Inequality," *Journal of Economic Theory*, 31(1983): 105-121.
9. GEHRIG, W. "On a Characterization of the Shannon Concentration Measure," *Utilitas Mathematica*, 24(1983): 67-85.
10. GOOVAERTS, M.J., DEVYLDER, F., AND HAEZENDONCK, J. "Ordering of Risks: A Review," *Insurance; Mathematics and Economics*, 1(1982): 131-163.

11. KHINCHINE, A.J. *Mathematical Foundations of Information Theory*, New York: Dover, 1957.
12. KULLBACK, S. *Information Theory and Statistics*. New York: John Wiley & Sons, 1959.
13. MARSHALL, A.W., AND OLKIN, I. *Inequalities: Theory of Majorization and its Applications*, New York: Academic Press, 1979.
14. PIGOU, A.C. *Wealth and Welfare*, New York: Macmillan, 1912.
15. ROTHSCHILD, M., AND STIGLITZ, J.E. "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2(1970): 225-243.
16. SHORROCKS, A.F. "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48(1980): 613-621.
17. THEIL, H. *Economics and Information Theory*, Amsterdam: North Holland, 1967.





## DISCUSSION OF PRECEDING PAPER

ELIAS S. W. SHIU:

Professor Promislow is to be congratulated for this thought-provoking paper. There are many interesting ideas and concepts in it. I would like to expand on the important concept of riskiness in Section 4 and present two applications to immunization theory. Some of the results below also have been discussed in another paper by Professor Promislow [12].

Theorem 4.1 may be reformulated as follows. It is a result due to J. Karamata [8], published in 1932.

Theorem 1 ([8], [9, section XI.7], [10, p. 449]):

*Let  $\mu$  be a signed measure defined on the Borel subsets of  $(a, b)$ . Then, for all convex functions  $\phi$ ,*

$$\int_a^b \phi(t) d\mu(t) \geq 0$$

*if and only if*

$$\int_a^b d\mu(t) = 0,$$

$$\int_a^b t d\mu(t) = 0$$

*and*

$$\int_a^w \mu(a, t) dt \geq 0$$

*for all  $w \in (a, b)$ .*

The following variant of Theorem 1 is well known in the literature of stochastic dominance ([6], [7]).

Theorem 2:

*Let  $\mu$  be a signed measure defined on the Borel subsets of  $(a, b)$ . Then, for all nonincreasing convex functions  $\Psi$ ,*

$$\int_a^b \Psi(t) d\mu(t) \geq 0$$

if and only if

$$\int_a^b d\mu(t) = 0$$

and

$$\int_a^w (w - t) d\mu(t) \geq 0 \quad (1)$$

for all  $w \in (a, b)$ .

The “only if” direction in Theorem 2 is obvious as the functions  $1$ ,  $-1$ , and  $(w - t)_+ = \beta_w(t)$  are convex and nonincreasing. Also, by considering  $\Psi(t) = \alpha_w(t) = (t - w)_+$ , the condition on  $\Psi$  may be changed from nonincreasing to nondecreasing if the integral in (1) is replaced by

$$\int_w^b (t - w) d\mu(t).$$

The next result is helpful for understanding the concept of riskiness.

Theorem 3 ([19], [14], [2, p. 112], [3, p. 92]):

Let  $X$  and  $Y$  be two random variables in  $\mathbf{R}^1$  having finite means. Then,

$$E[\Psi(X)] \geq E[\Psi(Y)]$$

for each nonincreasing convex function  $\Psi$  if and only if there exists a random variable  $Z$  such that

$$\Pr(X \leq r) = \Pr(Y + Z \leq r)$$

for all  $r \in \mathbf{R}^1$  and

$$E(Z|Y) \leq 0 \quad (2)$$

almost surely.

The second half of Theorem 3 states that the random variable  $X$  has the same distribution as  $Y + Z$  where  $E(Z|Y) \leq 0$ . Thus  $X$  may be interpreted as  $Y$  plus “noise”  $Z$ ; it is therefore more risky, or more uncertain, than  $Y$ .

If in Theorem 3 the condition that the function  $\Psi$  is nonincreasing is dropped but the condition that  $E(X) = E(Y)$  is required, then (2) becomes  $E(Z|Y) = 0$  almost surely. This result was first proved by David Blackwell [1, p. 100] in the context of sufficient statistics. Also see [11, p. 108].

Below are two applications to immunization theory.

Let  $\{l_j\}$  be a stream of liability payments to occur at times  $\{s_j\}$  in the future. Both the amounts and the dates are known with certainty. Assume that the liabilities are to be funded by a stream of asset cash flows  $\{a_k\}$ . The cash inflow  $a_k$  is to occur at time  $t_k$ . Let the force of interest function be  $\delta(t)$ . Assume that assets and liabilities have the same present value,

$$\sum_k a_k \exp\left[-\int_0^{t_k} \delta(s) ds\right] = \sum_j l_j \exp\left[-\int_0^{s_j} \delta(s) ds\right].$$

Now, suppose that the interest rates change from  $\delta(t)$  to  $\delta(t) + \Delta(t)$ . What are the conditions on the cash flows and the shock  $\Delta(t)$  so that

$$\begin{aligned} \sum_k a_k \exp\left\{-\int_0^{t_k} [\delta(s) + \Delta(s)] ds\right\} \\ \geq \sum_j l_j \exp\left\{-\int_0^{s_j} [\delta(s) + \Delta(s)] ds\right\}? \quad (3) \end{aligned}$$

For simplicity, write

$$\alpha_k = a_k \exp\left[-\int_0^{t_k} \delta(s) ds\right],$$

$$\lambda_j = l_j \exp\left[-\int_0^{s_j} \delta(s) ds\right]$$

and

$$f(t) = \exp\left[-\int_0^t \Delta(s) ds\right].$$

Then (3) may be written as

$$\sum \alpha_k f(t_k) \geq \sum \lambda_j f(s_j). \quad (4)$$

The function  $f$  depends on the interest-rate shock  $\Delta$ . In the classical works of Redington [13] and Fisher and Weil [4], the shock  $\Delta$  is assumed to be constant; hence  $f$  is a convex function. By considering

$$\mu(-\infty, t] = \sum_{t_k \leq t} \alpha_k - \sum_{s_j \leq t} \lambda_j,$$

Theorem 1 and the method of linear programming can be applied to determine the assets for which (4) and (3) will hold for constant shocks.

The following is a consequence of Blackwell's Theorem.

Theorem 4 [15]:

Let the real numbers  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_m, t_1, t_2, t_3, \dots, t_m, \lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n, s_1, s_2, s_3, \dots, s_n$  be given. Assume that all  $\alpha_k$  and  $\lambda_j$  are non-negative. Then,

$$\sum_{k=1}^m \alpha_k \phi(t_k) \geq \sum_{j=1}^n \lambda_j \phi(s_j)$$

for all continuous convex function  $\phi$  if and only if there exists a nonnegative  $m$  by  $n$  matrix  $B$  such that the sum of each of its  $m$  rows is 1,

$$(\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_m)B = (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$$

and

the  $n$  columns of the matrix  $B$  partition the asset cash flows into  $n$  streams,

In the context of immunization, the second half of Theorem 4 says that the  $n$  columns of the matrix  $B$  partition the asset cash flows into  $n$  streams, each with the same present value and duration as one of the  $n$  liability outflows. Thus, a necessary and sufficient condition for the immunization of multiple liabilities is the separate immunization of each liability outflow. This result has been stated without proof on page 138 of [5].

For more detailed discussions on the immunization theory of multiple liabilities, see [16], [17] and [18].

#### REFERENCES

1. BLACKWELL, D. "Comparison of Experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neyman, Berkeley: University of California Press, 1951:93-102.
2. BRUMELLE, S.L. AND VICKSON, R.G. "A Unified Approach to Stochastic Dominance," *Stochastic Optimization Models in Finance*, edited by W.T. Ziemba and R.G. Vickson, New York: Academic Press, 1975:101-113.
3. FISHBURN, P.C. AND VICKSON, R.G. "Theoretical Foundations of Stochastic Dominance," *Stochastic Dominance: An Approach to Decision-Making Under Risk*, edited by G.A. Whitmore and M.C. Findlay, Lexington, Massachusetts: Lexington Books, 1978:39-113.
4. FISHER, L. AND WEIL, R.L. "Coping with the Risk of Interest-Rate Fluctuations: Returns to Bondholders from Naïve and Optimal Strategies," *Journal of Business*, 44 (1971):408-431.
5. FONG, H.G. AND FABOZZI, F.J. *Fixed Income Portfolio Management*, Homewood, Illinois: Dow Jones-Irwin, 1985.

6. HANOCH, G. AND LEVY, H. "The Efficiency Analysis of Choices Involving Risk," *The Review of Economic Studies*, 36 (1969):335-346. Reprinted in *Stochastic Optimization Models in Finance*, edited by W.T. Ziemba and R.G. Vickson, New York: Academic Press, 1975:89-100.
7. HARDAR, J. AND RUSSELL, W.R. "Rules for Ordering Uncertain Prospects," *The American Economic Review*, 49 (1969):25-34.
8. KARAMATA, J. "Sur une inégalité relative aux fonctions convexes," *Publications Mathématiques de l'Université de Belgrade*, 1 (1932):145-148.
9. KARLIN, S.J. AND STUDDEN, W.J. *Tchebycheff Systems: With Applications in Analysis and Statistics*, New York: Wiley, 1966.
10. MARSHALL, A.W. AND OLKIN, I. *Inequalities: Theory of Majorization and Its Applications*, New York: Academic Press, 1979.
11. PHELPS, R.R. *Lectures on Choquet's Theorem*, Princeton: Van Nostrand, 1966.
12. PROMISLOW, D. "Comparing Risks," *Advances in the Statistical Sciences*, VI; *Actuarial Science*, edited by I.B. MacNeill and G.J. Umphrey, Dordrecht, Holland: D. Reidel Publishing Co., 1987:79-93.
13. REDINGTON, F.M. "Review of the Principles of Life-Office Valuations," *JIA*, 78 (1952):286-315.
14. ROTHCHILD, M. AND STIGLITZ, J.E. "Increasing Risk: I. A Definition," *Journal of Economic Theory*, 2 (1970):225-243.
15. SHERMAN, S. "On a Theorem of Hardy, Littlewood, Pólya, and Blackwell," *Proceedings of the National Academy of Sciences of the United States of America*, 37 (1951):826-831.
16. SHIU, E.S.W. Discussion of "Measuring the Interest Rate Risk," *TSA*, 37 (1985):275-279.
17. SHIU, E.S.W. "A Generalization of Redington's Theory of Immunization," *ARCH*, 1986.2:69-81.
18. SHIU, E.S.W. "Immunization — The Matching of Assets and Liabilities," *Advances in the Statistical Sciences*, VI; *Actuarial Science*, edited by I.B. MacNeill and G.J. Umphrey, Dordrecht, Holland: D. Reidel Publishing Co., 1987:145-156.
19. STRASSEN, V. "The Existence of Probability Measures with Given Marginals," *Annals of Mathematical Statistics*, 36 (1965):423-439.

JAMES C. HICKMAN:

Professor Promislow's unfairness function attains its minimum value of zero when  $\mathbf{a} = \mathbf{b}$ , that is, when the vector of actual charges denoted by  $\mathbf{b}$  is equal to the vector of fair costs denoted by  $\mathbf{a}$ . The Introduction makes it clear that fair costs are the actuarial present values of corresponding future benefit payments. The unfairness function  $U(\mathbf{a}, \mathbf{b})$  measures a special kind of distance between  $\mathbf{a}$  and  $\mathbf{b}$ . The unfairness function has certain properties,

described in Section 5, that are selected to correspond to some of our common-sense notions of the properties of unfairness.

In the Introduction, it is stated that the

...sole concern in this paper is the inequity that arises from making use of less than perfect information in determining the expectation of loss.

I will examine this issue from a somewhat different viewpoint using elements of statistical decision theory. The development will be based on a simple model employing the following notation:

$X$  will denote the random variable that can be interpreted as the present value of future claims on a policy.

$\theta$  will denote a vector of parameters that determine the distribution of  $X$ . In life insurance the elements of  $\theta$  might be the numerical values representing age, sex, height, weight, and so on.

$I$  will denote an information gathering process for learning about  $\theta$ . One could think of  $I$  as a set of classification procedures. The process  $I$  will not necessarily provide perfect information about  $\theta$ .

Before the classification information is collected by process  $I$ , the risk about  $\theta$ .

Before the classification information is collected by process  $I$ , the risk parameters have a distribution which can be interpreted as the prior (before the collection of information) distribution of risk parameters in the population of insurance applicants. After the information specified by process  $I$  is collected, the risk parameters have a posterior distribution. Because information is not collected on all risk parameters and some information collected is subject to error, the posterior distribution of  $\theta$ , given  $I$ , is not a degenerate distribution at the true value of  $\theta$ .

These ideas lead to the following results:

$$\Pi(I) = E_{X|I} [X|I] = E_{\theta|I} E_{X|\theta} [X|\theta] = E_{\theta|I} [\Pi(\theta)]$$

$$\text{Var}(X|I) = \text{Var}_{\theta|I} [E(X|\theta)] + E_{\theta|I} [\text{Var}(X|\theta)].$$

If  $I$  yields perfect information about  $\theta$ ,

$$\Pi(I) = \Pi(\theta)$$

$$\text{Var}_{\theta|I} (E[X|\theta]) = 0$$

$$E_{\theta|I} [\text{Var}(X|\theta)] = \text{Var}(X|\theta)$$

$$\text{Var}(X|I) = \text{Var}(X|\theta).$$

To simplify the subsequent development, we will assume that the economic consequences of incomplete knowledge about the value of  $X$  can be measured by  $\sqrt{\text{Var}(X|I)}$  and that the cost of information process  $I$  is  $C(I) \geq 0$ . Then the gross premium, denoted by  $G$ , for a policy paying benefits of  $X$  would be

$$G = \Pi(I) + k \sqrt{\text{Var}(X|I)} + C(I) + L$$

where  $k \geq 0$ ,  $k \sqrt{\text{Var}(X|I)}$  is a premium component associated with risk, and  $L$  is the expense loading. The decision theory approach is to examine a set of feasible information processes to minimize  $k \sqrt{\text{Var}(X|I)} + C(I)$ .

Consider a simple example where  $\theta$  consists of one element which can take on values  $w_1$  and  $w_2$ . The prior distribution is  $\Pr[\theta = w_1] = p$  and  $\Pr[\theta = w_2] = 1 - p$ . Suppose information process  $I$  provides no information about  $\theta$ . With perfect information, we would have premiums  $\Pi(w_1)$  and  $\pi(w_2)$ . With process  $I$ , we obtain

$$\begin{aligned} \min [\pi(w_1), \pi(w_2)] &\leq \pi(I) = p \pi(w_1) + (1 - p) \pi(w_2) \\ &\leq \max [\pi(w_1), \pi(w_2)] \end{aligned}$$

$$\begin{aligned} \text{Var}(X|I) &= p \text{Var}(X|w_1) + (1 - p) \text{Var}(X|w_2) \\ &\quad + p(1 - p) [\pi(w_1) - \pi(w_2)]^2. \end{aligned}$$

A simple numerical illustration is summarized in the following two tables. In the first table the basic parameter values are given and in the second the results are illustrated.

	$w_1$	$w_2$
$E_{x \theta} [X \theta]$	13	17
$\sqrt{\text{Var}(X \theta)}$	8	8

	$P$				
	0	0.25	0.50	0.75	1.00
$\pi(X I)$	17	16	15	14	13
$\sqrt{\text{Var}(X I)}$	8	8.19	8.25	8.19	8

This example was selected to be appropriate for a blended mortality table combining male and female experience. The numerical values correspond roughly with the values of a life annuity at age 65 with zero interest rate.

The point of this discussion is to suggest the possibility of a decision theory approach to the problem addressed in the paper. Measuring the inequality distance between a vector of actuarial present values and a vector of actual charges is an important problem. However, it seems that the practical problem of which parameters can be measured will be settled by law. Decision theory can help determine which parameters should be measured and the accuracy that should be specified, taking into consideration the cost of information and the reduction in variance achieved by acquiring it.

STUART A. KLUGMAN:

Professor Promislow has presented a novel way (at least to actuaries) of looking at the risk classification problem. However, it would be uncharitable to claim that the problem of measuring equity has not been previously considered. Two examples that come to mind are Norberg [3] and Woll [4]. As far as the present discussion is concerned, I would like to augment the paper in three directions. First, Professor Promislow's paper proceeds from a simple case (a fixed set of true and assigned net premiums) directly to an entirely general case (based on measure theoretic concepts). For the benefit of readers who are not comfortable with this setting, I will present an intermediate version of the model based on the Society of Actuaries Course 110 level mathematical statistics. This model will be seen to be sufficient for describing the situation typically faced by an insurer. It has the added benefit of making the operational meaning of the criteria clear. Next, I will introduce a different criterion: squared error. Finally, I will look at these criteria from a perspective, not mentioned by Professor Promislow, that may aid in their interpretation. Along the way I will provide an argument for  $g(x) = -\ln(x)$  as the function of choice for Professor Promislow's approach.

### *A Model for Risk Classification*

Let us begin by assuming that each insured has an intrinsic, unknown, parameter  $\theta$ . Our knowledge of human behavior is not perfect, so knowledge of  $\theta$  does not imply that we know what will happen to the insured. We do assume that knowledge of  $\theta$  completely determines the probability distribution of the loss,  $W$ . Assume  $W$  is a continuous random variable with p.d.f.  $f(w|\theta)$ . Consistent with Professor Promislow's statement, if we know the



value of  $\theta$  we will charge the expected loss as the net premium. That is, we charge  $\mu_\theta = E(W|\theta) = \int_w f(w|\theta)dw$ .

Next, assume there are  $k$  risk classes. Each insured will be assigned to one of the classes  $1, \dots, k$  on the basis of some observable characteristics. Each policyholder can then be characterized by the pair  $(J, \theta)$  drawn from the joint density  $f(j, \theta)$ . The density will be discrete with respect to  $J$  and continuous with respect to  $\theta$ . The nature of this density will depend on the rule for assigning individuals to risk classes.

Each insured in risk class  $j$  will be charged a net premium of  $\pi_j$ . For the moment no restrictions are placed on the choice of these net premiums. The success of a classification system will depend on how well the assigned net premium  $\pi_j$  approximates the correct net premium  $\mu_\theta$ . Let the function  $L(\pi_j, \mu_\theta)$  measure the penalty incurred. Then a reasonable choice for a measure of riskiness of the classification system is the expected penalty incurred for a randomly selected insured. That is,

$$U = \sum_{j=1}^k \int L(\pi_j, \mu_\theta) f(j, \theta) d\theta.$$

Keep in mind that there are three discretionary items:

1. the classification system (represented by  $f(j, \theta)$ ),
2. the premium structure (represented by  $\pi_j$ ), and
3. the loss function (represented by  $L$ ).

I believe this framework is sufficient to describe the situation faced by the typical insurer. In Professor Promislow's notation, I am using  $\mu_\theta$  in place of his  $x$  and  $\pi_j$  in place of his  $Y$ . In the next two sections we look at two loss functions and the restrictions, if any, that need to be placed on the premium structure.

### *A Second Loss Function*

Professor Promislow restricts attention to loss functions of the form  $L(\pi_j, \mu_\theta) = \mu_\theta g(\pi_j/\mu_\theta)$ . Norberg [3] addresses the same problem and looks only at squared error,  $L(\pi_j, \mu_\theta) = (\pi_j - \mu_\theta)^2$ . Let  $U_N$  ( $N$  for Norberg) refer to the unfairness measure when using this criterion. Let  $U_L$  refer to the unfairness measure when using Professor Promislow's criterion with  $g(x) = -\ln(x)$ .

If we continue to let the net premium  $\pi_j$  be unspecified, we can ask the question: "For a given loss function and classification system, what set of net premiums will produce the smallest amount of unfairness?" For the Norberg criterion this is easy to find. Setting the derivative of  $U_N$  with respect

to  $\pi_j$  equal to zero (and assuming the derivative can be taken inside the integral) gives

$$0 = \int 2(\pi_j - \mu_\theta) f(j, \theta) d\theta$$

$$\pi_j = \int \mu_\theta f(j, \theta) d\theta / f(j) = \int \mu_\theta f(\theta|j) d\theta$$

where  $f(j) = \int f(j, \theta) d\theta$  is the marginal probability of being assigned to class  $j$ . A little bit of algebra will show this premium can also be written as

$$\pi_j = E(W|j),$$

and so it really is the net premium for those assigned to class  $j$ . It would appear that this is a desirable property for any classification evaluation system, that the optimal net premium is indeed *the* net premium for the class.

For Professor Promislow's class of loss functions we need an additional restriction. To solve the minimization problem, it is necessary to restrict attention to sets of net premiums that satisfy

$$\sum_{j=1}^k f(j) \pi_j = E(W).$$

This was assumed from the start by Professor Promislow. It guarantees that the total premiums we expect to collect equal the total expected payout. Using Lagrange multipliers we need the derivative of

$$U_L + \lambda[E(W) - \sum_{j=1}^k f(j) \pi_j].$$

It is

$$\pi_j = \int \mu_\theta f(j, \theta) d\theta / \lambda f(j),$$

This yields

$$\pi_j = \int \mu_\theta f(j, \theta) d\theta / \lambda f(j),$$

and it is easy to see that with  $\lambda = 1$  we get the same result as with squared error. So this is another argument in favor of using  $g(x) = -\ln(x)$  from Professor Promislow's collection of possibilities. It is equally easy to show that the other members of Professor Promislow's family of equations (5.2) do not yield the net premium as the optimal choice. In particular, with  $g(x) = x \ln(x)$  the result is

$$\pi_j = \lambda \exp[\int \ln(\mu_\theta) f(j, \theta) d\theta / f(j)]$$

with  $\lambda$  selected to make the net premiums add up to the desired amount.

*Another View*

From the insurer's perspective, the risk involved is the relationship between the amount of money collected and the amount paid. In the setting described here, the dollar loss on a randomly selected policy is  $W - \pi_j$ . The major difference between this loss random variable and the one introduced in Bowers et al. [1] is that here the amount of premium collected is also random, due to the unknown (prior to issue) class to which the insured will be assigned. Restricting attention to net premiums for the  $\pi_j$ , the expected loss is zero and so the variance of the loss is

$$E[(W - \pi_j)^2].$$

A small amount of algebraic manipulation will show that this is equal to

$$U_N + \int \text{Var}(w|\theta) f(\theta) d\theta$$

and since the second term does not depend on the risk classification elements (items 1-3 in Section 1), the criterion presented here is the same as the  $U_N$  criterion.

We have, then, that the  $U_N$  criterion relates to a well-accepted measure of risk, the variance of the excess of loss over revenue. We next wonder if the  $U_L$  measure has a similar relationship. The loss with respect to premiums and revenue is

$$-E[W \ln(\pi_j/W)] = U_L + E[\mu_\theta \ln(\mu_\theta)] - E[W \ln(W)]$$

where once again the second term does not depend on the risk classification scheme.

So, both measures can be related to an expected discrepancy between revenue and expenditures. From its use in statistics we are familiar with squared error loss, but we are perhaps less familiar with the logarithmic loss introduced by Professor Promislow. The following expansion may provide some insight:

$$\begin{aligned} U_L &= -E[\mu_\theta \ln(\pi_j/\mu_\theta)] \\ &= -E\{\mu_\theta \ln[1 + (\pi_j - \mu_\theta)/\mu_\theta]\} \\ &= -E[(\pi_j - \mu_\theta) - (\pi_j - \mu_\theta)^2/2\mu_\theta] \\ &= E[(\pi_j - \mu_\theta)^2/\mu_\theta]/2. \end{aligned}$$

This looks very much like a chi-square goodness-of-fit measure. The numerator is the squared difference between the observed and expected values

while the denominator is the expected value. This indicates a potential drawback of the  $U_L$  measure. The presence of  $\mu_\theta$  in the denominator means that low-cost risks will receive the highest weight when computing the expected loss, while the highest cost risks will receive the lowest weight.

Professor Promislow has given us a lot to think about in the evaluation of risk classification schemes. I hope that someone will take the logical next step — to apply these concepts to some real classification problems. Unfortunately, this may be very difficult to do. Application of these methods requires knowledge of  $f(j, \theta)$  when all we can observe is  $W$  given  $j$ . Cummins et al. [2] provide several examples to indicate how difficult this can be.

#### REFERENCES

1. BOWERS, N.J., JR. et al. *Actuarial Mathematics*, Itasca, IL: Society of Actuaries, 1986.
2. CUMMINS, J.D., et al. *Risk Classification in Life Insurance*. Boston: Kluwer-Nijhoff, 1983.
3. NORBERG, R. "A Credibility Theory for Automobile Bonus Systems," *Scandinavian Actuarial Journal*, 1976:92-107.
4. WOLL, R. "A Study of Risk Assessment," *Proceedings of the Casualty Actuarial Society*, 1979:84-136.

#### (AUTHOR'S REVIEW OF DISCUSSION)

S. DAVID PROMISLOW:

I would like to thank Professors Shiu, Hickman, and Klugman for their discussions. Each of them has provided interesting and valuable additions to the paper.

Professor Shiu shows how far reaching the idea of riskiness can be and adds to some of his previous work involving the application of this concept to immunization theory.

Professor Hickman produces an interesting optimization problem by balancing the cost of obtaining additional information with the resulting increased equity.

I am grateful to Professor Klugman for pointing out other literature where equity is discussed. These references in turn have led me to discover many others, some of which should be acknowledged. Tryfos [4] uses absolute deviation, which is equivalent to formula (3.3) of the paper with  $g(x) = |x - 1|$ , to compare two automobile insurance classification schemes. I would also like to thank Professor Tryfos for bringing my attention to a paper by Schmalensee [2]. This work considers a model somewhat similar to mine

and also discusses some effects on equity of suppressing information. The most extensive work in this area that I have discovered is that of Ferreira [1], in a study done for the Massachusetts State Rating Bureau. His motivation and objectives are very close to those of my paper. His methods, criteria, and resulting formulas are somewhat different however, which further indicates the difficulties involved in arriving at a suitable way to measure equity.

### *Squared Error*

I would like to discuss at length the important idea of squared error introduced by Professor Klugman.

This of course is a common method for measuring the "closeness" of one distribution to another. It is perhaps most familiar in the case of variance where we use squared error to obtain a measure of distance from a distribution to the degenerate distribution concentrated at its mean. Hence, it also forms the basis of Professor Hickman's suggested measure of unfairness.

One of the most useful properties of squared error is that, as in the case of the formulas given by (3.2) in the paper, we get a nice within-group and between-group decomposition. For the purpose of comparing with formula (5.3), I will illustrate with reference to the discrete model given in Section 3. The result remains true in the general setting. Suppose that given the vectors  $\mathbf{a}$  and  $\mathbf{b}$  we define

$$U(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n |a_i - b_i|^2$$

and the corresponding normalized quantity

$$U_0(\mathbf{a}, \mathbf{b}) = \frac{1}{n} \sum_{i=1}^n |a_i - b_i|^2.$$

For this discrete case,  $U_0$  corresponds to Professor Klugman's  $U_N$ .

Given any partition of  $\{1, 2, \dots, n\}$  into subsets,  $S_1, S_2, \dots, S_k$  where there are  $n_i$  elements in  $S_i$ , we let  $\mathbf{a}'$  and  $\mathbf{b}'$  be obtained from  $\mathbf{a}$  and  $\mathbf{b}$ , respectively, by replacing each entry by the average entry of the subset containing it. Let  $\mathbf{a}'' = \mathbf{a} - \mathbf{a}'$ ,  $\mathbf{b}'' = \mathbf{b} - \mathbf{b}'$ . Then the between set inequity is given by

$$U_0^B = U_0(\mathbf{a}', \mathbf{b}')$$

and the inequity due to the  $i$ -th subset is given by

$$U_0(S_i) = \frac{1}{n_i} \sum_{i \in S_i} |a''_i - b''_i|^2,$$

leading to the decomposition

$$U_0(\mathbf{a}, \mathbf{b}) = U_0^p + \sum_{i=1}^k \frac{n_i}{n} U_0(S_i).$$

There is a simple derivation of this decomposition formula for those familiar with elementary Hilbert space theory. In this case it is easier to look at the corresponding formula for  $U$  rather than  $U_0$ . Then the between-group term is the norm squared of the orthogonal projection of the vector  $\mathbf{a} - \mathbf{b}$  on the subspace consisting of vectors which are constant on the sets of the partition, and the within-group term is the norm squared of the projection on the orthogonal complement.

With the given definition of  $U_0$  we then have a decomposition analogous to formula (5.3). Note that the weights depend neither on the fair costs nor the actual charges, but on the number of individuals in the group. As with the case of  $g(x) = -\ln(x)$ , the within-group portion is independent of the classification. Referring to the model in the form given by Professor Klugman, this means that the within-group inequity does not depend on the premium structure. Hence, for both of these measurement formulas, the minimum equity must be obtained when the between-group inequity is reduced to zero by charging each class the net premium for that particular class. This gives an alternative derivation of Professor Klugman's results.

This concept of optimality of net premiums is discussed by Schmalensee [2], referred to as "competitive equitability." In fact, Schmalensee takes this concept as one of the main postulates in his axiomatic approach to justify squared error.

Professor Klugman's remark that this principle is a "desirable property for any classification evaluation system" needs qualification, in my opinion. I agree that given a method of evaluation we think is sound, any method of classification should result in the optimality of net premiums. But if this does not occur, inefficiency of classification may be the cause rather than a faulty evaluation method. To illustrate, consider the example in Section 6 of the paper. As shown in the paper, when we evaluate by using formula (3.3) with  $g(x) = x \ln(x)$ , the net premiums are not optimal. In my opinion, this fact does not detract from the appropriateness of the method. Given that we are going to divide the six individuals into two groups of three and charge

each group the same premium, then there is no way to avoid subsidizing the high risk individual with a cost of 10. We might question, however, if it seems equitable to have the entire burden of the subsidy fall on the two very best risks simply because they happen, through some observable characteristic, to be placed in the same class as this high cost individual. This occurs using the net premiums for the two classes. Using the optimal allocation for  $g(x) = x \ln(x)$ , some of the subsidy is born by the intermediate risks, those with a fair cost of 2. (Admittedly the total subsidy is larger in this case.)

There is another way to compare the squared error with the formulas given by (3.3). Using the squared error we can write the above formula for  $U$  as

$$U(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^n a_i^2 g(r_i)$$

where  $g(r) = (r - 1)^2$  and  $r_i = b_i/a_i$ .

From this point of view, we have a formula similar to that suggested in the paper, but which uses as weights the squares of the fair costs rather than the costs themselves. The effect is to give higher weight to the high cost entries. We could, as Professor Klugman suggests, look upon the presence of  $\mu_\theta$  in the denominator of the formula for  $U_L$  as giving highest weight to the lowest cost risks. On the other hand, we could look upon it as a correction to the overly low weights given to these risks by the use of squared error. It depends on one's point of view.

There are some possible difficulties with the use of squared error to measure equity. For example, the transfer principle, as given in Section 5, no longer holds. Take

$$\mathbf{a} = (2, 5, 11), \quad \mathbf{b} = (7, 10, 1), \quad \mathbf{b}' = (6, 11, 1).$$

For any  $g$  used in formula (3.3) we will get

$$U(\mathbf{a}, \mathbf{b}) \geq U(\mathbf{a}, \mathbf{b}')$$

since the change from  $\mathbf{b}$  to  $\mathbf{b}'$  involves transfer of charge from an individual who was charged 3.5 times his fair cost to one who was only charged 2 times, and the amount transferred is small enough to preserve the order of ratios (3 as opposed to 2.2 after the transfer). Using  $U_N$  however we do not reduce inequity, since

$$U_N(\mathbf{a}, \mathbf{b}) = 150, \quad U_N(\mathbf{a}, \mathbf{b}') = 152.$$

The validity of the transfer principle as we have stated it depends on the principle that the ratio of actual charges to fair costs is a reasonable way to

gauge inequity for a particular individual. There are possible alternative viewpoints. Some, for example, may measure this by the absolute difference between actual charges and fair costs. In such a case, using squared error would reduce inequity under the resulting modification of the transfer principle. To my mind it seems that someone who has a fair cost charge of 1 and is charged 10 is treated more inequitably than one who has a fair cost of 100 and is charged 109.

For another example, based on the same idea, suppose a number of individuals have exactly the same risk of loss but, on account of an inefficient classification system, are charged different amounts. This will produce some unfairness. Suppose now that the insurer must increase premiums and does so by charging each individual a constant amount, which is the equitable method since the risks are the same. Intuitively, it appears that this should have a leveling effect and reduce inequity. Indeed it does, using the formulas of the paper with a strictly convex function  $g$ , for instance, any of those in (5.3), as shown by the subadditivity principle given in Section 5. Take  $\mathbf{a}' = \mathbf{b}'$  to be the vector with constant entries equal to the increase so that  $U_0(\mathbf{a}', \mathbf{b}') = 0$ . But this does not happen for the squared error formula, since adding the same constant to all entries obviously leaves inequity unchanged.

### *Claim Costs*

Professors Klugman and Hickman have both made a worthy contribution by bringing another "layer" into the model, namely, the actual cost of the risk (Professor Hickman's  $X$  and Professor Klugman's  $W$ ). I did not need this quantity when computing inequity, but paradoxically it provides a reason for introducing it. It is important to clarify what should be included and what should not be, as there is often much confusion on this point. Obviously, there will be individuals who have exactly the same risk parameters but who will incur different claims because of random fluctuations. Indeed, this fact forms the very foundation of insurance and it is not these deviations that we wish to consider as inequity.

Professor Klugman notices that with both  $U_N$  and  $U_L$  we can obtain an additive decomposition of the total variation between claims and premiums into that attributable to the inequity inherent in the classification and that attributable to randomness. This is another consequence of the additive decomposition mentioned above. Consider the partition whereby two individuals are in the same subset if they have the same risk parameters and are charged the same premium, that is, the partition determined by the pair  $(j, \theta)$



in Professor Klugman's notation. Then apply the particular formula to measure the difference between claim costs and premiums collected. The between-group term gives the portion due to inequity, and the within-group term gives that portion of the difference which is due to randomness and could not have been predicted beforehand from our knowledge of  $\theta$ . This will not work when using a function from the family (5.2) other than  $-\ln x$ . We do get the decomposition as given by formula (5.3), but the weights in the within-group portion depend on the premiums. Consequently, the second term in the decomposition will depend on the classification scheme.

There is an alternative way to arrive at the split into the inequity and randomness portions for the two given formulas. The randomness component is found by simply applying the postulated method to measure the "distance" between  $W$  and  $\mu_\theta$ .

### *Conclusion*

The works on equity referred to in the discussions (by Professor Klugman and this review) for the most part deal directly with the case of automobile insurance, where problems of risk classification have caused the most controversy. However, life and pension actuaries certainly face similar questions. Debates over unisex pricing furnish such an example. I think that all Society of Actuaries members could benefit from looking at some of the casualty actuarial literature presenting views that depart somewhat from traditional philosophy. For an interesting, completely nontechnical work, I would particularly recommend Stone [3]. This publication consists of excerpts from the final report of Commissioner J.M. Stone and includes his justification for some controversial recommendations, such as the elimination of age and sex as rating variables for auto insurance. Many of his ideas were influenced by Ferreira [1], mentioned above. The conclusion of the paper by Woll, a reference given by Professor Klugman, contains a summary of some of this work.

A major theme of these writers is the idea that if a class is extremely heterogeneous, then the mean expected loss is not representative of the expected claims for any individual and, hence, may not be suitable as a net premium. Referring again to the example in Section 6, one may view things in this way. The net premium of 4 is really not representative for a class with the fair costs of (1,1,10). Of course, such a principle involves the subjective notion of extreme heterogeneity. The theorems I proved in Section 7 involving the riskiness concept could be viewed as one attempt to provide a more precise definition for this notion.

These observations provide an important psychological reason why many actuaries tend to reject new ideas on risk classification. In this, as well as in other actuarial problems, we often put too much emphasis on mean values. We reason, for example, that the average expected cost for a class must necessarily give us the appropriate net premium. Our traditional actuarial education sometimes leads us to sweep stochastic difficulties under the rug by replacing distributions by their expected values and relying on the law of large numbers to make everything right. For many purposes such a point of view turns out to be a brilliant idea. It allows us to derive complicated formulas and make calculations in an easy manner. But when we move from the computational to the conceptual level, we must recognize that features of a distribution other than its expected value have to be considered. We can expect this perspective to be more prevalent in the future as it forms a basic theme in the new actuarial textbook.

In conclusion, I would like to again thank the three discussants of my paper.

#### REFERENCES

1. FERREIRA, J. "Identifying Equitable Insurance Premiums for Risk Classes: An Alternative to the Classical Approach," *Automobile Insurance Risk Classification: Equity & Accuracy*, Boston: Division of Insurance, Commonwealth of Massachusetts, 1978.
2. SCHMALENSEE, R. "Imperfect Information and the Equitability of Competitive Prices," *Quarterly Journal of Economics*, 99 (1984):441-460.
3. STONE, J.M. "Excerpt from the Opinion, Findings and Decision on 1978 Automobile Insurance Rates," *Automobile Insurance Risk Classification: Equity & Accuracy*, Boston: Division of Insurance, Commonwealth of Massachusetts, 1978.
4. TRYFOS, P. "The Equity of Classification Systems in Automobile Insurance," *Journal of Risk and Insurance*, LIV (September 1987):569-581.