TRANSACTIONS OF SOCIETY OF ACTUARIES
1989 VOL. 41

AN EMPIRICAL METHOD OF COMPARING RISKS USING STOCHASTIC DOMINANCE

H. DENNIS TOLLEY AND MICHAEL R. KOSOROK*

ABSTRACT

The need to evaluate risks, in either the investment or the insurance component of business, is a primary task of the actuary. Mathematical constructs for risk assessment using expected utility theory are problematic for practical use. One reason is that the elicitation of the utility function is difficult or impossible in most situations. Second, data related to risk choices contain noise that must be accounted for in the decision process. This paper presents a method of testing second-order stochastic dominance of risks. By using this methodology, risks can be partially ordered by value, with available data. The methodology developed here requires only that the risks have yields distributed as one of the location-scale family of distributions and that the actuary can specify that family. The paper presents the simulated power of the procedure.

I. INTRODUCTION

Risk assessment has three components: (i) analysis of the contingent event, (ii) evaluation of the contractual obligation associated with the event, and (iii) determination of the premium to be received for assuming the risk of the event. Assessing the value of the risks associated with either an investment or an insurance product is a fundamental task of the actuary. One appealing method of assessing the value of a specified risk is the use of utility functions. In this case, the value of a risk is assessed based on the expected utility of income or loss associated with the risk, calculated by using the probability density function of the risk. Those risks with higher expected utility are preferred over those with lower expected utility.

As one example, consider three different proposed cost schedules for health care. Once an individual has made a claim, the claim amount and relative frequency for each schedule are as shown in Table 1. Under all three schedules, the expected cost for a claim is $36. However, depending on which schedule is used, some claims will result in different amounts of

*Michael R. Kosorok, not a member of the Society, is a graduate student at the University of Washington.

reimbursement. Choosing the "best" schedule for the insurer depends on the set of values the insurer has assigned to the distribution of different claim sizes. This set of values, which the actuary chooses, is called the utility function. Choosing (or failing to choose) a utility function provides the basis for selecting the "best" cost schedule.

TABLE 1

AMOUNT REIMBURSED AND EXPECTED FREQUENCY OF REIMBURSEMENT UNDER THREE HYPOTHETICAL HEALTH CARE SCHEDULES

Amount Reimbursed	Frequency of Reimbursement		
	Schedule I	Schedule II	Schedule III
$10	0.05	0.10	0.05
$20	0.10	0.15	0.15
$30	0.25	0.15	0.25
$40	0.40	0.25	0.25
$50	0.20	0.35	0.30

As a second example, suppose that the actuary must choose between increasing the company's whole life offering or increasing the offering of a policy that is a combination of term life and annuity income. Not including operating expenses, the risks for these policies are based on the differences between the net premiums and the actual expense. Again, depending on the value given to each dollar of profit or loss, one risk may be preferred over another.

Both examples are simple enough that the experienced actuary can readily solve each without formally appealing to sophisticated utility theory. On the other hand, it is not hard to visualize extensions of either example in which choosing the best risk is difficult. For example, at what additional premium can the whole life policy in the second example be equally preferred to the retirement income policy? Also at what premiums do payment schedules II and III in the first example become neutral relative to schedule I or to each other? To approach these questions in general entails a utility function setup.

The use of expected utility as a method of both pricing and selecting risks has been recognized as useful from a theoretical point of view, especially for risk theory (see, for example, Bowers et al. [1], Bühlmann [2], and Goovaerts et al. [6]). However, the use of expected utility in practical problems has proven problematic for two reasons. First, the actual utility function is difficult to determine in practice. Explicitly specifying the utility function under which actuaries, financial officers or stockholders are operating is impossible under most circumstances. Second, except for large volumes of business with which the company already has considerable experience, the actuary usually can specify the probability distribution of the risks only up to a set of unknown constants. The actuary may estimate these unknown constants from past experience, based on exogenous data, or with expert guess. However, in any of these cases, the uncertainty resulting from having to estimate the distribution of risks must be accounted for in evaluating a portfolio of risks.

The problem of specifying the utility function has been approached in several ways. One common approach is to choose a parametric class of utility functions that provides relatively simple expressions of expected utility (see, for example, Cass and Stiglitz [3]). A second approach, and the one of interest here, is to restrict the class of utility functions by using qualitative constraints. For example, one might require that the utility function be in the class of monotone increasing functions. This set of utility functions results in what are called first-order stochastic dominance decisions (Whitmore and Findlay [13]). The problem with such qualitative restrictions is that the risks based on the resulting class of utility functions can only be partially ordered. This means that for two different risks, one may be determined better than the other, worse than the other, or neither. The last condition of neither better nor worse is an indeterminate condition. One cannot determine the best risk with respect to the class of utility functions. The set of risks that is better than some risks and no worse than the rest is called the efficient set. Any risk in the efficient set is acceptable for the class of utility functions. Thus, the actuary should offer (or invest in) risks in the efficient set. If no efficient set exists, the offerings and investments are selected from the indeterminate set. If the premise of the partial ordering is accepted, the actuary should never offer or invest in the remaining set. To reduce the size of the efficient set of risks, the actuary must specify a smaller class of utility functions by adding further restrictions. When the only qualitative restriction is that the utility functions be monotone increasing, the efficient risks are called the first-order stochastic dominant risks.

The qualitative restrictions of interest in this paper are those utility functions that are both monotone increasing and concave downward. In essence this implies two assumptions: (i) more return is better than less return (or less loss is better than more loss); and (ii) the marginal value of an incremental increase in return decreases as the total value of the return increases (risk averse). Most will agree that these two premises, though restrictive, are usually acceptable. Under these assumptions mathematical methods of partially ordering the risks can be developed without explicit specification of a utility function. The resulting partial ordering of risks is called second-order stochastic dominance (SSD) and is recognized as a mathematical construct that is useful in determining the efficient class of choices of random variables (Whitmore and Findlay [13]). Note that the number of utility functions satisfying the second-order dominance criterion is smaller than the number satisfying the first-order criterion of restricting utility functions to be monotone increasing. Some risks in the efficient set under first order will not be so under second order. Hence, the second-order criterion produces a finer partial ordering among any pair of risks. In this sense, provided the qualitative restrictions are acceptable, second order is preferred to first-order stochastic dominance. Higher orders of stochastic dominance are created by more qualitative restrictions on the class of permissible utility functions.

The second problem in the use of expected utility is the element of uncertainty in the specification of the density functions. For example, the frequency tables in the first example (Table 1) may be estimates based on sparse, uncertain data. Decisions resting heavily on these expected frequencies may not be justifiable. When the risks are chosen based upon a parametric form of the utility function, explicit expressions of how the unknown parameters of the density function enter into the expected utility can be derived. Incorporating uncertainty into the decision process then follows standard (large sample) statistical procedures. In the case of stochastic dominance, no specific form of the expected utility function is obtained. Therefore, including uncertainty in the parameters of the distribution function is more difficult. For the case in which first-order dominance is considered, nonparametric procedures have been developed that allow selection of the efficient sets when the underlying probability density functions of the risks must be estimated (see Whitmore and Findlay [13] and Stein et al. [11]). Franck [5] has provided a likelihood ratio test applicable to both parametric and nonparametric methods.

When the criterion is second-order stochastic dominance, the solution is much less developed. By analogy to first-order results, Whitmore and Findlay [13] proposed a procedure, but it has been found to be inconsistent and biased (see Stein et al. [11]). Deshpande and Singh [4] proposed an asymptotic test for the single sample case; however, this procedure works only in restricted cases. Tolley and Pope [12] proposed an exact randomization procedure that appears to have reasonable "power." However, the computational requirement is large, even for small problems. It is unlikely that the procedure could be used in medium- and large-sample finance and actuarial problems.

The conclusion is that there is no general two-sample procedure for second-order stochastic dominance. Therefore, a method of choosing the efficient set of risks cannot be generated without explicitly assuming both the form and the parameter values of the density function of the risks. One temptation has been to assume that the estimated parameters determined from the sample distribution are, in fact, the true population parameters. This results in very little power to choose the truly efficient set of risks (Kroll and Levy [8]).

The purpose of this paper is to present an easily implemented approximate univariate statistical procedure for use in making second-order stochastic dominance decisions when the distributions of the two risks are from the same location-scale family. A location-scale family is a set of distributions that can be standardized to the same distribution by subtracting a parameter from each observation coming from the distribution and dividing the result by a second constant. The values of the two constants depend upon the particular member of the location-scale family of interest. In this paper we assume that the location-scale family is continuous over the region in which it is nonzero. This allows for a simpler development. Similar results can be obtained for discontinuous location-scale distributions. Because stochastic dominance is invariant under a fixed monotone transformation, such as the logarithm, the methodology presented is useful for risks with distributions that can be transformed to a location-scale family by such a transformation. The log-normal distribution is one such distribution. The procedure presented here is a large-sample test procedure in that the probability distribution of the test statistic is known only in the limiting case as the sample size tends to infinity. With this procedure, the efficient set can be chosen by taking all possible pairs of risks and comparing. Those that dominate some risks and are not dominated by any others would make up the efficient set.

Because the procedure considers risks restricted to the location-scale family of probability density functions, one may feel that it provides no more flexibility in choosing risks than the standard mean-variance method. Clearly the test procedure will eventually result in decisions based only on location and scale parameter estimates, as does the mean-variance procedure. However, we believe that the results presented here are useful for the following reasons:

1. The procedure is pedagogically appealing; the resulting statistical procedure resembles the classical chi-square and t-tests.
2. The procedure provides for incorporation of any skewness or kurtosis apparent in the location-scale family.

3. The procedure provides the basis for selecting between several portfolios of convex combinations of random variables.

II. THE MAIN RESULT

Assume that two independent risks are to be compared. Each may be an insurable risk or an investment. Each may be a single entity or a combination or portfolio of entities. In any case let $\bar{X}_1$ and $\bar{X}_2$ represent the estimated average yield of these two risks, with S_1 and S_2 the estimated standard deviations of the yield, respectively. Here yield includes the premium or risk loading. We assume that the estimates of $\bar{X}$ and S are calculated from a set of data in the usual manner. Explicitly, let $X_j^{(1)}$, $j = 1, \ldots n_1$ and $X_j^{(2)}$, $j = 1, \ldots n_2$ be samples of size n_1 and n_2 from populations 1 and 2, respectively. Let $N = n_1 + n_2$ denote the total sample size. Then

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_j^{(i)}$$

and

$$S_i = \left\{ \frac{1}{n_i} \sum_{j=1}^{n_i} (X_j^{(i)} - \bar{X}_i)^2 \right\}^{1/2}$$

are estimates of location and scale for population i. Similarly define estimates of

$$c_i = \frac{E(X_i - \mu_i)^3}{2\sigma_i}$$

and

$$v_i = \frac{E(X_i - \mu_i)^4 - \sigma_i^4}{4\sigma_i^2}$$

as $\hat{c}_i$ and $\hat{v}_i$, respectively. Let $f_0(x)$ denote the probability density function with location at zero and a scale of unity. We assume that $f_0(x)$ is continuous and nonzero over its domain of x values. In other words, there is not a region in the "middle" where the function $f_0(x)$ is zero and nonzero on either side of the region. We choose $f_0(x)$ such that a location of zero corresponds to a mean of zero. We assume X_1 and X_2 are random variables with density $f_1(x)$ and $f_2(x)$, respectively, where

$$f_1(x) = \frac{1}{\sigma_1} f_0 \left(\frac{x - \mu_1}{\sigma_1}\right) \quad (1)$$

$$f_2(x) = \frac{1}{\sigma_2} f_0 \left(\frac{x - \mu_2}{\sigma_2}\right) \quad (2)$$

In these expressions, μ_1, μ_2, σ_1, and σ_2 are unknown location and scale parameters. In other words, $f_0(x)$ represents the standard form of the distribution of yields for the risks considered. Both risks are assumed to have this same form, although each has its own location (mean yield) and scale (standard deviation of the yield). For example, $f_0(x)$ may be exponential, uniform, normal, or any of a large set of location-scale families (see Lehmann [9] for examples and further details).

Similarly, we make the following definitions of notation:

z_α is the upper 100α percentile of the standard normal distribution

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_2 - S_1}$$

$$\hat{K}_1 = \int_{-\infty}^{Z_0} f_0(z)\, dz$$

$$\hat{K}_2 = \int_{-\infty}^{Z_0} z f_0(z)\, dz$$

$$t_1 = (\bar{X}_2 - \bar{X}_1)\, \hat{K}_1 + (S_2 - S_1)\, \hat{K}_2$$

$$t_2 = (\bar{X}_2 - \bar{X}_1)$$

$$f_i = (n_1 + n_2)/n_i$$

$$\hat{\gamma}_1 = \hat{K}_1^2 (f_1 S_1^2 + f_2 S_2^2) + \hat{K}_2^2 (f_1 \hat{v}_1 + f_2 \hat{v}_2) + 2 \hat{K}_1 \hat{K}_2 (f_1 \hat{c}_1 + f_2 \hat{c}_2)$$

$$\hat{\gamma}_2 = \hat{K}_1 (f_1 S_1^2 + f_2 S_2^2) + \hat{K}_2 (f_1 \hat{c}_1 + f_2 \hat{c}_2)$$

$$\hat{\gamma}_3 = f_1 S_1^2 + f_2 S_2^2$$

To determine whether, statistically, risk 1 is better than risk 2 by using the criterion of SSD, the following steps are taken:

1. Choose an α-level or type I error level.

2. Determine whether the product $(\hat{\gamma}_1 \hat{\gamma}_3)/N^2$ is approximately the same as $\hat{\gamma}_2^2/N^2$ (within 10^{-16}). If $\hat{\gamma}_1 \hat{\gamma}_3 \approx \hat{\gamma}_2^2$, then form

$$d^2 = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

and

$$t = t_2/d; \tag{3}$$

otherwise skip to step 4.
3. Determine which of events E_2, E_4, or E_5 has occurred.
If $t > z_{\alpha/2}$, event E_2 has occurred.
If $t < -z_{\alpha/2}$, event E_4 has occurred.
If $-z_{\alpha/2} \leq t \leq z_{\alpha/2}$, event E_5 has occurred.
Skip to step 6.
4. Form the statistic D as

$$D = N \frac{t_1^2 \hat{\gamma}_3 + t_2^2 \hat{\gamma}_1 - 2t_1 t_2 \hat{\gamma}_2}{\hat{\gamma}_3 \hat{\gamma}_1 - \hat{\gamma}_2^2}, \tag{4}$$

and check to determine whether $D \leq -2 \log 2\alpha$. If so, then conclude that event E_5 has occurred and skip to step 6.
5. If $D > -2 \log 2\alpha$, then determine which of the following events has occurred:

$$E_1 \text{ if } \left\{ \frac{\sqrt{N}\, t_1}{\sqrt{\hat{\gamma}_1}} > z_{\alpha/2} \text{ and } \frac{\sqrt{N}\, t_2}{\sqrt{\hat{\gamma}_3}} \not< -z_{\alpha/2} \right\}$$

$$E_2 \text{ if } \left\{ \frac{\sqrt{N}\, t_2}{\sqrt{\hat{\gamma}_3}} > z_{\alpha/2} \text{ and } \frac{\sqrt{N}\, t_1}{\sqrt{\hat{\gamma}_1}} \not< -z_{\alpha/2} \right\}$$

$$E_3 \text{ if } \left\{ \frac{\sqrt{N}\, t_1}{\sqrt{\hat{\gamma}_1}} < -z_{\alpha/2} \text{ and } \frac{\sqrt{N}\, t_2}{\sqrt{\hat{\gamma}_3}} \not> z_{\alpha/2} \right\}$$

$$E_4 \text{ if } \left\{ \frac{\sqrt{N}\, t_2}{\sqrt{\hat{\gamma}_3}} < -z_{\alpha/2} \text{ and } \frac{\sqrt{N}\, t_1}{\sqrt{\hat{\gamma}_1}} \not> z_{\alpha/2} \right\}$$

E_5 if {all other possibilities}.

6. If E_1 and/or E_2 has occurred, conclude that there is empirical evidence that risk 2 is better than risk 1 using SSD. If E_3 and/or E_4 has occurred, there is empirical evidence that risk 1 is better than risk 2. If E_5 has occurred, there is insufficient empirical evidence to favor one risk over another without either a more specific utility function or more accurate estimates (more extensive data).

Note that although the terms 2α and $\alpha/2$ appear in various steps in the procedure, the size of the test is α. By size we mean the probability of concluding that one risk is second-order stochastic dominant over another when actually both have the *same* distribution. As noted previously, two risks can have unequal probability distributions and yet have no dominance present. If this condition is prevalent among the risks being evaluated, the value of α in the procedure should be increased to avoid mistaking this form of inequality as second-order stochastic dominance.

III. EXAMPLE

To fix ideas, this section illustrates the use of the second-order stochastic dominance algorithm to select between two contingencies. For this example, consider two investments. From the past 100 investments of the first type, the mean profit has been $\mu_1 = 0.3$ with a standard deviation of $s_1 = 1.0$. For the second type of investment, data from 50 policies yield a profit of $\mu_2 = 0.3$ and a standard deviation, s_2, of 3.0. We wish to determine whether there is sufficient evidence to prefer one investment over another based upon the data.

We will assume that the yield is normally distributed. Under this assumption, $c = 0$. Therefore, we will estimate $\hat{c}$ as zero. For the normal distribution recall that

$$\nu = \frac{1}{4}\left[E\,\frac{(X-\mu)^4}{\sigma^2} - \sigma^2\right] = \frac{\sigma^2}{2}.$$

An estimate of ν_i to be used is $\hat{\nu}_i = \dfrac{s_i^2}{2}$.

From this information we can now put values to the definitions in the previous sections. Explicitly,

$$Z_0 = \frac{0.3 - 0.3}{3.0 - 1.0} = 0.0$$

$$\hat{K}_1 = \int_{-\infty}^{0} f_0(z)\, dz = \Phi\,(0) = 0.5$$

$$\hat{K}_2 = \int_{-\infty}^{0} z\, f_0(z)\, dz = -\,\phi\,(0) = -\,0.399$$

where $\Phi(x)$ and $\phi(x)$ are the cumulative distribution and the density of the standard normal distribution, respectively.

$$\begin{aligned}
t_1 &= (0.3 - 0.3)\,(0.5) + (3.0 - 1.0)\,(-0.399) = -0.798\\
f_1 &= (50 + 100)/100 = 1.5\\
f_2 &= (50 + 100)/50 = 3.0\\
\hat{\gamma}_1 &= (0.5)^2\,[3 \cdot (1.0)^2 + 1.5\,(3.0)^2]\\
&\quad + (-0.399)^2\,[3\,(0.5) + (1.5)\,(4.5)]\\
&\quad + 2\,(0.5)\,(-0.399)\,[3\,(0) + 1.5\,(0)]\\
&= 5.438\\
\hat{\gamma}_2 &= (0.5)\,[3\,(1.0)^2 + 1.5\,(3.0)^2]\\
&\quad + (-0.399)\,[3\,(0) + 1.5\,(0)]\\
&= 8.25\\
\hat{\gamma}_3 &= 3\,(1.0)^2 + (1.5)\,(3.0)^2\\
&= 16.5
\end{aligned}$$

The steps in Section II proceed as follows:

1. Select $\alpha = 0.05 \Rightarrow z_{\alpha/2} = 1.96$
2. $N = 150$

$$\frac{\hat{\gamma}_1\hat{\gamma}_3}{N^2} = 0.00399$$

$$\frac{\hat{\gamma}_2^2}{N^2} = 0.00303$$

Thus $\frac{\hat{\gamma}_1\hat{\gamma}_3}{N^2} \neq \frac{\hat{\gamma}_2^2}{N^2}$; therefore skip to step 4.

3. (Skipped)
4. $D = 150\,[(-0.798)^2\,(16.5) + (0)^2\,(5.438)$
   $-\,2\,(-0.798)\,(0.0)\,(8.250)]/(16.5)\,(5.438) - (8.250)^2$
   $= 72.750$
   $-\,2 \log 2\alpha = 4.605$

Since $72.750 > 4.605$, do not skip to step 6.

5. $$\frac{\sqrt{N}\, t_1}{\sqrt{\hat{\gamma}_1}} = \frac{\sqrt{150}\,(-0.798)}{\sqrt{5.438}}$$
$$= -4.191$$

$$\frac{\sqrt{N}\, t_2}{\sqrt{\hat{\gamma}_2}} = \frac{\sqrt{150}\,(0.0)}{\sqrt{16.5}}$$
$$= 0.0$$

Since $-4.191 < -z_{\alpha/2}$ and $0.0 \not> z_{\alpha/2}$, conclude E_3 has occurred.

6. From the conclusion in step 5, we conclude that investment 1 is better than investment 2 relative to the class of increasing concave downward utility functions.

IV. PROOF OF MAJOR RESULT

As noted in the introduction, we wish to choose between two risks by using an expected utility approach in which the utility function is unspecified except that it is required to be positive and concave downward. A risk that has higher expected utility than another risk for all utility functions of this type is said to be second-order stochastic dominant over the second risk. To place this idea in mathematical terms, define $G(y)$ as

$$G(y) = \int_{-\infty}^{y} \left\{ \int_{-\infty}^{t} [f_1(x) - f_2(x)]\, dxdt \right\}. \tag{5}$$

Note that by the definition of $f_1(x)$ and $f_2(x)$, $G(-\infty) = 0$. Therefore, either $G(y) \equiv 0$ for all y or $G(y)$ has at least one maximum or minimum, although it may not be attained except at $y = +\infty$.

Definition

Population 1 is said to be second-order stochastic dominant (SSD) over population 2, if $G(y) \le 0$ for every y and $G(y_0) < 0$ for at least one y_0.

This definition is the same as given by Gooverts et al. [6], except that we require strict inequality for at least one value of y. It can be shown that this definition is equivalent to the two qualitative restrictions considered in the Introduction (see Whitmore and Findlay [14]). The hypothesis of interest is H_0: Pop.1 SSD over Pop. 2.

Lemma 1

If $\sigma_1 \neq \sigma_2$, then $G(y)$ takes its extreme value (either a global maximum or global minimum) at

$$y_0 = \frac{\sigma_2\mu_1 - \sigma_1\mu_2}{\sigma_2 - \sigma_1}$$

if

$$f_0\left(\frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_2}\right) \neq 0.$$

If

$$f_0\left(\frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_2}\right) = 0,$$

then $G(y)$ takes in extreme value as $y \to \infty$.

Proof

Rearranging the order of integration, we can write

$$G(y) = \int_{-\infty}^{y}\left(\int_{x}^{y}[f_1(x) - f_2(x)]\,dt\right)dx$$

$$= \int_{-\infty}^{y}(y - x)\,[f_1(x) - f_2(x)]\,dx$$

$$= y[F_1(y) - F_2(y)] - \int_{-\infty}^{y} x\,[f_1(x) - f_2(x)]\,dx \tag{6}$$

Differentiating with respect to y, we have

$$G'(y) = F_1(y) - F_2(y)$$

$$= F_0\left(\frac{y - \mu_1}{\sigma_1}\right) - F_0\left(\frac{y - \mu_2}{\sigma_2}\right) \tag{7}$$

Since $f_0(x)$ is assumed to be continuous and nonzero, $F_0(x)$ is monotone increasing. Thus Equation (7) implies that if $G'(y_0)$ is zero (and only if), then it is at the root y_0 of the equation

$$\frac{y - \mu_1}{\sigma_1} = \frac{y - \mu_2}{\sigma_2}.$$

If $\sigma_1 \neq \sigma_2$

$$\begin{aligned} y_0 &= \left(\frac{\mu_1}{\sigma_1} - \frac{\mu_2}{\sigma_2}\right) \frac{\sigma_1\sigma_2}{\sigma_2 - \sigma_1} \\ &= \frac{\sigma_2\mu_1 - \sigma_1\mu_2}{\sigma_2 - \sigma_1}. \end{aligned} \tag{8}$$

For this value of y_0, $G'(y)$ has either a local maximum, local minimum or an inflection point. To determine which, we examine the second derivative of $G(y)$.

$$\begin{aligned} G''(y) &= f_1(y) - f_2(y) \\ &= \frac{1}{\sigma_1} f_0\left(\frac{y - \mu_1}{\sigma_1}\right) - \frac{1}{\sigma_2} f_0\left(\frac{y - \mu_2}{\sigma_2}\right). \end{aligned}$$

For $y_0 = \dfrac{\sigma_2\mu_1 - \sigma_1\mu_2}{\sigma_2 - \sigma_1}$ the arguments become

$$\begin{aligned} \frac{y_0 - \mu_1}{\sigma_1} &= \frac{\sigma_2\mu_1 - \sigma_1\mu_2 - \mu_1\sigma_2 + \mu_1\sigma_1}{\sigma_1(\sigma_2 - \sigma_1)} \\ &= \frac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1}. \end{aligned}$$

and

$$\frac{y_0 - \mu_2}{\sigma_2} = \frac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1}.$$

Setting $z_0 = \dfrac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1}$ we see that

$$G''(y_0) = f_0(z_0)\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right)$$

which is positive or negative as $\sigma_2 > \sigma_1$ or $\sigma_2 < \sigma_1$. Thus, if $f(z_0) \neq 0$, for $\sigma_1 \neq \sigma_2$, there is exactly one point, y_0, at which $G(y)$ reaches its maximum or its minimum, and the result follows. If $f(z_0) = 0$, then there is no maximum value of $G(y)$. In this case the sign of $G'(y)$ remains constant, and therefore the extreme value of $G(y)$ is attained as $y \to \infty$.

From the proof of lemma 1, $G(y)$ can be evaluated at two important points, y_0 and ∞. Since these two values are used in the sequel, we will determine here the values of $G(y_0)$ and $G(\infty)$. From Equation (6), replacing y by

$$y_0 = \frac{\sigma_2\mu_1 - \sigma_1\mu_2}{\sigma_2 - \sigma_1}$$

we get

$$G(y_0) = y_0 \left[F_1(y_0) - F_2(y_0)\right] - \int_{-\infty}^{y_0} x \left[f_1(x) - f_2(x)\right] dx.$$

Recall that

$$F_1(y_0) = F_0\left(\frac{y_0 - \mu_1}{\sigma_1}\right) = F_0\left(\frac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1}\right).$$

Similarly

$$F_2(y_0) = F_0\left(\frac{y_0 - \mu_2}{\sigma_2}\right) = F_0\left(\frac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1}\right).$$

Also,

$$\int_{-\infty}^{y_0} x f_1(x)\, dx = \int_{-\infty}^{y_0} \frac{x}{\sigma_2} f_0\left(\frac{x - \mu_1}{\sigma_1}\right) dx.$$

Making the substitution of variables $z = (x - \mu_1)/\sigma_1$, we note that the range of x from $-\infty$ to y_0 implies the range of z is from $-\infty$ to $(\mu_1 - \mu_2)/(\sigma_2 - \sigma_1)$. This substitution yields

$$\int_{-\infty}^{y_0} x f_1(x)\, dx = \int_{-\infty}^{z_0^*} (\sigma_1 z - \mu_1) f_0(z)\, dz,$$

where $z_0^* = (\mu_1 - \mu_2)/(\sigma_2 - \sigma_1)$. Similarly we have

$$\int_{-\infty}^{y_0} x f_2(x)\, dx = \int_{-\infty}^{z_0} (\sigma_2 z - \mu_2) f_0(z)\, dz.$$

In summary, if $\sigma_1 \neq \sigma_2$, then $G(y)$ takes the extreme value of

$$G(y_0) = (\mu_2 - \mu_1) \int_{-\infty}^{z_0^*} f_0(z)\, dz + (\sigma_2 - \sigma_1) \int_{-\infty}^{z_0^*} z f_0(z)\, dz.$$

From Equation (6) we have

$$\lim_{y \to \infty} G(y) = \int_{-\infty}^{\infty} x f_2(x)\, dx - \int_{-\infty}^{\infty} x f_1(x)\, dx + \lim_{y \to \infty} y[F_1(y) - F_2(y)],$$

provided that the expected value of X_1 and X_2 both exist. Note that

$$|F_1(y) - F_2(y)| < |1 - F_1(y)| + |1 - F_2(y)| = \text{Prob}\left(Z > \frac{y - \mu_1}{\sigma_1}\right) + \text{Prob}\left(Z > \frac{y - \mu_2}{\sigma_2}\right),$$

where Z is a random variable with density function $f_0(x)$, with mean zero and with variance 1. From Chebychev's inequality, this last expression is bounded as

$$\text{Prob}\left(|Z| > \frac{y - \mu_1}{\sigma_1}\right) < \frac{\sigma_1^2}{(y - \mu_1)^2}$$

and

$$\text{Prob}\left(|Z| > \frac{y - \mu_2}{\sigma_2}\right) < \frac{\sigma_2^2}{(y - \mu_2)^2}.$$

Therefore

$$y|F_1(y) - F_2(y)| < \frac{y\,\sigma_1^2}{(y - \mu_1)^2} + \frac{y\,\sigma_2^2}{(y - \mu_2)^2}.$$

This last expression goes to zero as $y \to \infty$. Therefore

$$G(\infty) = \lim_{y \to \infty} G(y) = EX_2 - EX_1 = \mu_2 - \mu_1$$

Lemma 2

If $\sigma_1 = \sigma_2$, then $G(y)$ reaches its extreme at $y = \infty$.

Proof

Without loss of generality, we may assume in this case that $\sigma_1 = \sigma_2 = 1$. From the proof of lemma 1, $G'(y) = F_0(y - \mu_1) - F_0(y - \mu_2)$ has either a positive or a negative sign for all y depending on whether $\mu_1 < \mu_2$ or $\mu_1 > \mu_2$. Since $G(-\infty) = 0$, the result follows.

From the results of these two lemmas and the facts that $G(y)$ is continuous and $G(-\infty) = 0$, we can conclude in the case of $\sigma_1 \neq \sigma_2$ that if SSD exists, then one of the following four cases exists:

1. $G(y_0) > 0$ and $G(+\infty) \geq 0$.
2. $G(y_0) < 0$ and $G(+\infty) \leq 0$.
3. $G(+\infty) > 0$ and $G(y_0) \geq 0$.
4. $G(+\infty) < 0$ and $G(y_0) \leq 0$.

In the case where $\sigma_1 = \sigma_2$, then SSD implies one of the following two cases:

5. $G(+\infty) > 0$.
6. $G(+\infty) < 0$.

If any of the cases 1, 3 or 5 hold, then risk 2 is better than risk 1 in the sense that risk 2 is SSD over risk 1. Similarly, if any of the cases 2, 4, or 6 hold, then risk 1 is SSD over risk 2. Consequently, the vector $\mathbf{g}' = [G(y_0), G(+\infty)]$ can be used as the basis for forming a test statistic for the hypotheses. For simplicity, we denote $G(y_0)$ as g_1 and $G(+\infty)$ as g_2. A test statistic can be formed by replacing μ_1, μ_2, σ_1, and σ_2 in the definition of $\mathbf{g}$ with the estimates of $\bar{X}_1$, $\bar{X}_2$, S_1, and S_2. Denote the estimate $\hat{\mathbf{g}} = [\hat{G}(y_0), \hat{G}(+\infty)]$. To determine the large sample distribution of $\hat{\mathbf{g}}$, the following well-known theorem is essential and is included here for completeness.

Theorem

Suppose that $\mathbf{Y}_n = (Y_{1n}, \ldots, Y_{rn})'$ is a random vector depending on n such that

$$[\sqrt{n}\,(Y_{1n} - \theta_1), \ldots, \sqrt{n}\,(Y_{rn} - \theta_r)]$$

tends in law to the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V}$, and suppose that $h_1, \ldots, h_r$ are r real-valued functions of $\theta = (\theta_1, \ldots, \theta_r)$, defined and continuously differentiable in a neighborhood of the true parameter θ and such that the matrix $\mathbf{B} = \{\partial h_i/\partial\theta_j\}$ of partial derivatives is nonsingular in a neighborhood of the true parameter θ. Then

$$\{\sqrt{n}\,[h_1(\mathbf{Y}_n) - f_1(\theta)], \ldots, \sqrt{n}\,[h_r(\mathbf{Y}_n) - f_r(\theta)]\}'$$

tends in law to the multivariate normal distribution with mean vector $\mathbf{0}$ and with covariance matrix $\mathbf{B}\ \mathbf{V}\ \mathbf{B}'$.

Proof

See Lehmann [9].

Collorary 1

If the random vector

$$\mathbf{Y}_i = \begin{pmatrix} \bar{X}_i \\ S_i \end{pmatrix}$$

is derived from a sample of n_i independent and identically distributed random variables from a location-scale distribution and

$$\theta_i = \begin{pmatrix} \mu_i \\ \sigma_i \end{pmatrix},$$

then $\sqrt{n_i}\,(\mathbf{Y}_i - \theta_i)$ is distributed asymptotically as a bivariate normal random variable with mean vector $\mathbf{0}$ and covariance matrix

$$\mathbf{V}_i = \begin{pmatrix} \sigma_i^2 & c_i \\ c_i & v_i \end{pmatrix}.$$

Proof

Let $\mathbf{Y}_i^* = [X_i - \mu_i, (X_i - \mu_i)^2]' = (Y_{1i}, Y_{2i})'$. Then $E\mathbf{Y}_i^* = (0,\sigma_i^2)' = \theta_i^*$ and

$$\operatorname{var}(\mathbf{Y}_i) = \begin{bmatrix} \sigma_i^2 & E(X_i - \mu_i)^3 \\ E(X_i - \mu_i)^3 & E(X_i - \mu_i)^4 - \sigma_i^4 \end{bmatrix} = \mathbf{V}_i.$$

Let

$$\mathbf{Y}_{in}^{*\prime} = \left[\frac{\sum_{j=1}^{n_i} (X_{ij} - \mu_i)}{n_i}, \frac{\sum_{j=1}^{n_i} (X_{ij} - \mu_i)^2}{n_i}\right] = (Y_{1n}^{(i)}, Y_{2n}^{(i)}).$$

By the central limit theorem $\sqrt{n_i}\,(\mathbf{Y}_{in}^* - \theta_i^*)$ tends in law to the multivariate normal distribution with mean vector $\mathbf{0}$ and with covariance matrix $\mathbf{V}_i$.

Setting

$$h_1(\mathbf{Y}_{in}^*) = Y_{i1n} = \bar{X}_i \text{ and } h_2(\mathbf{Y}_{in}^*) = Y_{i2n} - Y_{i1n}^2 = \frac{\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{n_i} = S_i^2,$$

we apply the preceding theorem with

i. $\mathbf{h}(\theta_i^*) = (0, \sigma_i^2)$

ii. $\mathbf{B} = \{\partial h_k / \partial \theta_{i1}^*\} = \begin{bmatrix} 1 & 0 \\ -2y_{i1n} & 1 \end{bmatrix}.$

Evaluated at $\mathbf{y}_{in} = \theta^*$, we have $\mathbf{B} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$, and $\mathbf{B}\,\mathbf{V}_i\mathbf{B}' = \mathbf{V}_i$. Setting $\mathbf{Y}_{in} = (\bar{X}_i, S_i^2)' = (y_{i1n}, y_{i2n})'$, $h_1(\mathbf{y}_{in}) = y_{i1n}$, and $h_2(\mathbf{y}_{in}) = \sqrt{y_{i2n}}$, we have

$$\mathbf{h}(\theta_i) = (\mu_i, \sigma_i)' \text{ and } \mathbf{B} = \{\partial h_L / \partial \theta_{iK}\} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{2\sqrt{y_{i2n}}} \end{bmatrix}.$$

Evaluated at $\mathbf{y}_{in} = \theta_i$,

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & \dfrac{1}{2\sigma_i} \end{bmatrix},$$

and

$$\mathbf{B}\,\mathbf{V}_i\,\mathbf{B}' = \begin{bmatrix} \sigma_i^2 & \dfrac{E(x_i - \mu_i)^3}{2\sigma_i} \\ \dfrac{E(x_i - \mu_i)^3}{2\sigma_i} & \dfrac{E(x_i - \mu_i)^4 - \sigma_i^4}{4\sigma_i^2} \end{bmatrix} = \begin{bmatrix} \sigma_i^2 & c_i \\ c_i & v_i \end{bmatrix} = \Sigma_i.$$

Therefore, applying the theorem

$$\sqrt{n_i}\left\{\begin{bmatrix} \bar{X}_i \\ S_i \end{bmatrix} - \begin{bmatrix} \mu_i \\ \sigma_i \end{bmatrix}\right\} = \sqrt{n_i}\,(z_i - \tau_i)$$

tends in law to the multivariate normal distribution with mean vector **0** and covariance matrix

$$\Sigma_i = \begin{bmatrix} \sigma_i^2 & c_i \\ c_i & v_i \end{bmatrix}.$$

Let $N = n_1 + n_2$ be the total number of observations on which the estimates $\bar{X}_1$, S_1, $\bar{X}_2$, and S_2 are based. Assume that the relative proportion in each sample tends to a constant as the total sample increases, that is, $\lim_{N \to \infty} (N/n_i) = f_i$. Let η denote the large sample expectation of $\hat{\mathbf{g}}$, that is, $\lim_{N \to \infty} E\hat{\mathbf{g}} = \eta$. Similar to Section II, define the following,

$$\Gamma_i = \begin{pmatrix} \gamma_1^{(i)} & \gamma_2^{(i)} \\ \gamma_2^{(i)} & \gamma_3^{(i)} \end{pmatrix},$$

where

$$\gamma_1^{(i)} = K_1^2\,\sigma_i^2 + K_2^2\,v_i + 2K_1\,K_2\,c_i,$$
$$\gamma_2^{(i)} = K_1\,\sigma_i^2 + K_2\,c_i,$$
$$\gamma_3^{(i)} = \sigma_i^2.$$

Similarly define $\hat{\Gamma}_i$. Here the hat, "^", indicates that consistent estimates of μ_i, σ_i, c_i, and v_i are used in place of the true parameters, and the lack of the hat indicates true parameter values are used.

Lemma 3

$\sqrt{N}$ ($\hat{\mathbf{g}}$ − η) is distributed asymptotically as a normal random variable with mean zero and covariance matrix $f_1 \Gamma_1 + f_2 \Gamma_2$, provided the covariance matrix is nonsingular.

Proof

By assumption $X_i^{(1)}$ and $X_i^{(2)}$ are two independent series of independent random variables with densities $f_i(x)$. The Taylor series expansion gives

$$\sqrt{N}\,(\hat{\mathbf{g}} - \eta) = \sqrt{N}\,\mathbf{B}_1 \begin{pmatrix} \bar{X}_1 - \mu_1 \\ S_1 - \sigma_1 \end{pmatrix} + \sqrt{N}\,\mathbf{B}_2 \begin{pmatrix} \bar{X}_2 - \mu_2 \\ S_2 - \sigma_2 \end{pmatrix} + o_p\,(1),$$

where

$$\mathbf{B}_1 = \begin{pmatrix} \dfrac{\partial \hat{g}_1}{\partial \bar{X}_1} & \dfrac{\partial \hat{g}_1}{\partial S_1} \\[2ex] \dfrac{\partial \hat{g}_2}{\partial \bar{X}_1} & \dfrac{\partial \hat{g}_2}{\partial S_1} \end{pmatrix}$$

and

$$\mathbf{B}_2 = \begin{pmatrix} \dfrac{\partial \hat{g}_1}{\partial \bar{X}_2} & \dfrac{\partial \hat{g}_1}{\partial S_2} \\[2ex] \dfrac{\partial \hat{g}_2}{\partial \bar{X}_2} & \dfrac{\partial \hat{g}_2}{\partial S_2} \end{pmatrix}$$

From the corollary to the theorem, we know

$$\sqrt{n_i}\,\mathbf{B}_i \begin{pmatrix} \overline{X}_i - \mu_i \\ S_i - \sigma_i \end{pmatrix}$$

is distributed asymptotically as a bivariate normal random vector with mean **0** and covariance $\mathbf{B}_i\,\mathbf{V}_i\,\mathbf{B}_i'$. Since, by assmption $(N/n_i) \rightarrow f_i$, we may conclude that

$$\sqrt{N}\,\frac{\sqrt{n_i}}{\sqrt{n_i}} \begin{pmatrix} \overline{X}_i - \mu_i \\ S_i - \sigma_i \end{pmatrix} \approx \sqrt{f_i}\,\sqrt{n_i} \begin{pmatrix} \overline{X}_i - \mu_i \\ S_i - \sigma_i \end{pmatrix}.$$

Thus, following the theorem, to establish the result we need only to determine values for the matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. Focusing first on $\hat{g}_1 = (\overline{X}_2 - \overline{X}_1)\,\hat{K}_1 + (S_2 - S_1)\,\hat{K}_2$, we differentiate with respect to $\overline{X}_I$ to get

$$\frac{\partial \hat{g}_1}{\partial \overline{X}_1} = -\hat{K}_1 + (\overline{X}_2 - \overline{X}_1)\frac{\partial \hat{K}_1}{\partial \overline{X}_1} + (S_2 - S_1)\frac{\partial \hat{K}_2}{\partial \hat{X}_I}$$

Recall that

$$\hat{K}_1 = \int_{-\infty}^{z_0} f_0(z)\,dz$$

and

$$\hat{K}_2 = \int_{-\infty}^{z_0} z\,f_0(z)\,dz,$$

where $Z_0 = (\overline{X}_1 - \overline{X}_2)/(S_2 - S_1)$. Therefore

$$\frac{\partial \hat{K}_I}{\partial \overline{X}_1} = \frac{1}{S_2 - S_1}\,f_0\,(Z_0)$$

$$\frac{\partial \hat{K}_2}{\partial \overline{X}_1} = \frac{1}{S_2 - S_1}\,Z_0\,f(Z_0)$$

Hence,

$$\frac{\partial \hat{g}_1}{\partial \overline{X}_1} = -\hat{K}_1 + \frac{(\overline{X}_2 - \overline{X}_1)}{S_2 - S_1} f(Z_0) + \frac{(S_2 - S_1)\, Z_0 f(Z_0)}{(S_2 - S_1)}$$
$$= -\hat{K}_1.$$

Similarly, one can show the following

$$\frac{\partial \hat{g}_1}{\partial S_1} = -\hat{K}_2$$
$$\frac{\partial \hat{g}_1}{\partial \overline{X}_2} = \hat{K}_1$$
$$\frac{\partial \hat{g}_1}{\partial S_2} = \hat{K}_2.$$

Therefore,

$$\hat{\mathbf{B}}_1 = \begin{pmatrix} -\hat{K}_1 & -\hat{K}_2 \\ -1 & 0 \end{pmatrix}.$$

Similarly one can show, after some algebra, that

$$\hat{\mathbf{B}}_2 = \begin{pmatrix} \hat{K}_1 & \hat{K}_2 \\ 1 & 0 \end{pmatrix}.$$

We note that the values of $\mathbf{B}_1$ and $\mathbf{B}_2$ are defined similarly by replacing the estimated values of location and scale with their true values: μ_1, μ_2, σ_1, σ_2.

Note that if $\gamma_j = f_1\gamma_j^{(1)} + f_2\gamma_j^{(2)}$, then the covariance matrix is nonsingular provided that $\gamma_1\gamma_3 - \gamma_2^2 \neq 0$. Consequently, evaluating $\mathbf{B}_i$ at $\overline{X}_i = \mu_i$ and $S_i = \sigma_i$ and performing the matrix operations, we have the result.

Corollary 2

The statistic $D = \mathbf{N}\hat{\mathbf{g}}'\ (f_1\hat{\Gamma}_1 + f_2\,\hat{\Gamma}_2)^{-1}\,\hat{\mathbf{g}}$ is distributed asymptotically as a chi-square with 2 degrees of freedom and noncentrality parameter $N\eta'$ $(f_1\,\Gamma_1 + f_2\,\Gamma_2)^{-1}\,\eta$ provided that $\gamma_1\gamma_3 - \gamma_2^2 \neq 0$.

Proof

The proof follows directly from lemma 3.

One can see from the definition of g_1 and g_2 that as $\sigma_2^2 \rightarrow \sigma_1^2$, then $g_1 \rightarrow g_2$ and the matrices $\mathbf{B}_i$ both become singular. In this case, $\gamma_1\gamma_3 - \gamma_2^2 \rightarrow 0$. Therefore multiplying the matrices together in corollary 2 gives the results listed in Section II of this paper for the case $\sigma_1^2 \neq \sigma_2^2$.

From lemma 2, comparing risks using the stochastic dominance in the case where $\sigma_1^2 = \sigma_2^2$ reduces to comparing the means, or location parameters. This comparison follows immediately from the theorem by recalling the central limit theorem for averages of independent random variables.

IV. POWER OF THE TEST PROCEDURE

The test procedure described in the proofs leaves one unsure as to how to determine whether $\gamma_1\gamma_3 - \gamma_2^2$ is zero. If this determinant is zero, then one of the eigenvalues of $f_1\Gamma_1 + f_2\Gamma_2$ will be zero and the test procedure resembles the comparison of means. If, however, neither eigenvalue is zero, the determinant is nonzero and the test is a two-step procedure, the first step being an overall chi-square test and the second step, performed only in the case where the first produced a significant result, consisting of the comparison of two statistics. The problem of testing is made more difficult by the fact that we use estimates of the covariance. Therefore, $\hat{\gamma}_1\hat{\gamma}_3 - \hat{\gamma}_2$ may be close to zero by chance alone, even though $f_1\Gamma_1 + f_2\Gamma_2$ is of full rank.

Regarding the determinant of the true covariance matrix, some statisticians believe that testing for singularity in the covariance matrix makes no sense (see, for example, Mardia et al. [10]). Pursuing this line, the comparison of means approach would be used only if the *calculated* determinant $(\hat{\gamma}_1\hat{\gamma}_3/N^2) - (\hat{\gamma}_2^2/N^2)$ was equal to zero within the computational error of the computer. In the simulations described below, this rule was used.

The two-step procedure of testing stochastic dominance in the case where the determinant is not zero might appear to result in an inflation of the α-level. However, since the second step in the procedure is undertaken only in the case where the first step is significant, the α-level is protected. Examples of a similar two-step multivariate procedure, given by Hummel and Sligo [7], indicate that the true α-level is in fact slightly *below* the nominal level. Thus, in our case, we would expect the test procedure to be conservative, asymptotically. Recall that the α-level is the probability of concluding stochastic dominance when, in fact, $F_A = F_B$.

To evaluate the power of the procedure as described above for medium and small sample sizes, a series of simulations was run. Three location-scale families were used in these simulations: the normal, uniform, and exponential densities. These densities are expressed as

(Normal)

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty$$

(Uniform)

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{12\sigma^2}}, \qquad \mu - \sigma\sqrt{3} < x < \mu + \sigma\sqrt{3}$$

(Exponential)

$$f(x;\mu,\sigma) = \frac{1}{e\sigma} \exp\left(-\frac{(x-\mu)}{\sigma}\right), \qquad \mu - \sigma < x < \infty$$

For these three densities, the estimates of c_i and v_i are as given in Table 2. For each density, two sets of location-scale parameters were selected for comparison. Table 3 lists these parameter sets. Five hundred simulations were then performed for each set of parameters for samples of size 10, 25 and 75 for each of the two samples in the comparison. The number of times the null hypothesis of no stochastic dominance was rejected in favor of SSD (one-sided test) was tabulated. Figures 1 through 3 are plots of the observed proportion of time the hypothesis was rejected using the procedure given in Section II. The x-axis in these plots is the "distance" between the two underlying distributions. This distance is measured as the noncentrality parameter corrected for degrees of freedom.

TABLE 2

ESTIMATES OF c_i AND v_i FOR THE NORMAL EXPONENTIAL AND UNIFORM LOCATION-SCALE FAMILIES

	Estimates of	
Family	c_i	v_i
Normal	0	$S_i^2/2$
Uniform	0	$S_i^2/5$
Exponential	S_i^2	$2S_i^2$

TABLE 3

LIST OF THE PARAMETERS OF EACH SIMULATED CASE FOR THE THREE LOCATION-SCALE FAMILIES

Case	Population A		Population B	
	μ	σ	μ	σ
1	0.00	1.00	0.00	1.00
2	0.00	1.00	0.00	1.25
3	0.00	1.00	0.00	1.50
4	0.00	1.00	0.00	1.75
5	0.00	1.00	0.00	2.00
6	0.00	1.00	−0.25	1.00
7	0.00	1.00	−0.25	1.25
8	0.00	1.00	−0.25	1.50
9	0.00	1.00	−0.25	1.75
10	0.00	1.00	−0.25	2.00
11	0.00	1.00	−0.50	1.00
12	0.00	1.00	−0.50	1.25
13	0.00	1.00	−0.50	1.50
14	0.00	1.00	−0.50	1.75
15	0.00	1.00	−0.50	2.00
16	0.00	1.00	−0.75	1.00
17	0.00	1.00	−0.75	1.25
18	0.00	1.00	−0.75	1.50
19	0.00	1.00	−0.75	1.75
20	0.00	1.00	−0.75	2.00
21	0.00	1.00	−1.00	1.00
22	0.00	1.00	−1.00	1.25
23	0.00	1.00	−1.00	1.50
24	0.00	1.00	−1.00	1.75
25	0.00	1.00	−1.00	2.00
26	0.00	1.00	−1.50	1.00
27	0.00	1.00	−1.50	2.00
28	0.00	1.00	0.00	3.00
29	0.00	1.00	−1.00	3.00
30	0.00	1.00	−1.50	3.00

As shown by the plots, the ability to detect one risk as being better than another when, in fact, the first risk is preferred using the SSD criterion increases as the favorability of the first risk increases. Notice also that the simulated power of detecting the better risk is irregular, as demonstrated by the plots. There are two reasons for this. First, because each point in each plot is simulated using 500 replications, some noise is expected in the plots. Second, one risk can be preferred over another because of differences in either location, scale, or both. Therefore, the power plots should be three-dimensional plots with separate axes for location, scale, and power. We have compressed such a three-dimensional plot into two dimensions, based on the asymptotic argument associated with $G(y_0)$. To the extent the sample

FIGURE 1

Plot of the Observed Proportion of Time the Null Hypothesis of Equal Distributions Was Rejected in Favor of SSD Versus the "Distance" between the Two Underlying Distributions. (The hypothesis tests were performed for equal sample sizes of 10, 25, and 75 drawn for sample from a normal probability distribution; alpha level = 0.05).

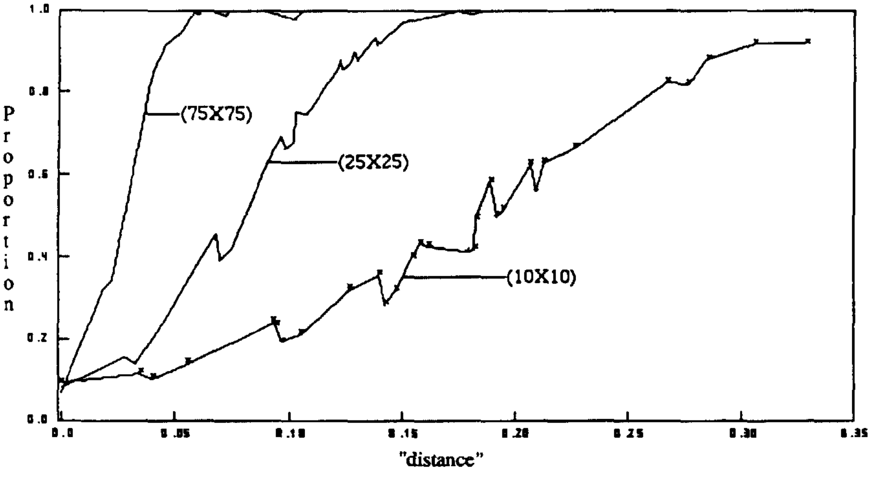


FIGURE 2

Plot of the Observed Proportion of Time the Null Hypothesis of Equal Distributions Was Rejected in Favor of SSD Versus the "Distance" between the Two Underlying Distributions. (The hypothesis tests were performed for equal sample sizes of 10, 25, and 75 drawn for sample from a uniform probability distribution; alpha level = 0.05.)

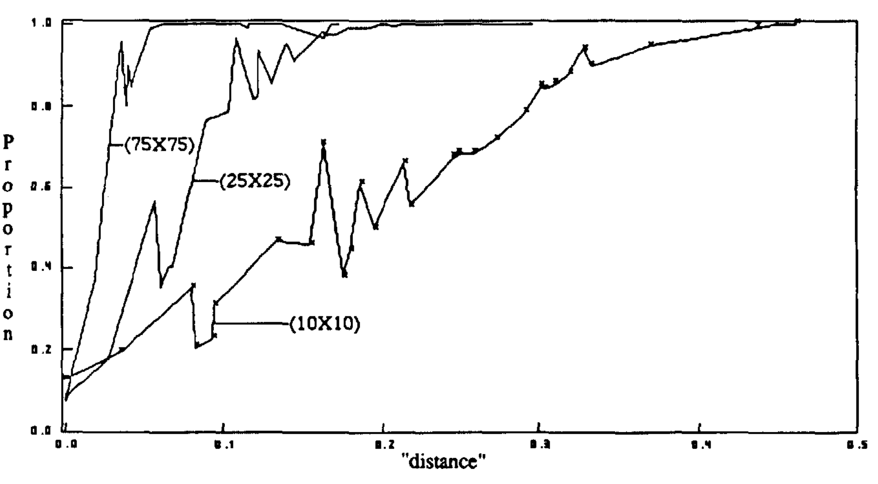
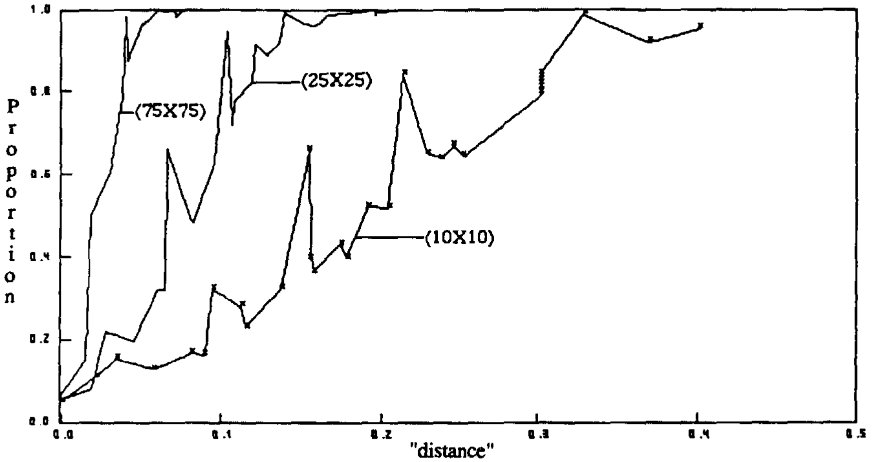


FIGURE 3

PLOT OF THE OBSERVED PROPORTION OF TIME THE NULL HYPOTHESES OF EQUAL DISTRIBUTIONS WAS REJECTED IN FAVOR OF SSD VERSUS THE "DISTANCE" BETWEEN THE TWO UNDERLYING DISTRIBUTIONS. (THE HYPOTHESIS TESTS WERE PERFORMED FOR EQUAL SAMPLES SIZES OF SIZE 10, 25, and 75 DRAWN FOR SAMPLE FROM AN EXPONENTIAL PROBABILITY DISTRIBUTION; ALPHA LEVEL = 0.05.)



sizes are insufficient to use the first-order approximation of $G(y_0)$, the compression from three to two dimensions will produce irregularities in the plots of power.

V. SUMMARY

In this paper we have presented a method of comparing risks by using the criterion of second-order stochastic dominance (SSD). The SSD criterion is more restrictive than first-order stochastic dominance. Consequently, the partial ordering is finer. However, the qualitative restrictions giving rise to second-order stochastic dominance are often justifiable. Therefore, the SSD criterion would seem to be an effective way of ordering risks. This paper contributes to the literature on choosing efficient sets of risks by providing a large-sample statistical criterion that incorporates the uncertainty inherent in data based estimates of the densities with the SSD criterion.

With the methodology presented in this paper, risks can be divided into three sets by using the SSD criterion: (i) those that are dominated by one or more risks, (ii) those that are dominant over some and dominated by none, and (iii) all other risks. In any investment or risk management strategy, choosing the first group seems always imprudent. This is because, based upon the data available, one can conclude, statistically, that there are no utility functions in the class of monotone increasing, concave downward functions under which the risk would be a good choice.

The choice between risks in the second and third groups is more complex. Clearly, risks in the second group are statistically better than risks in the first group. Hence choice of these risks appears defensible. However, a risk in the third group may be classified as an indeterminate risk simply because of lack of data or too much noise in the data. On the other hand, risks in the third group might actually be members of the first group, but the statistical methodology failed to identify them as such for the same data reasons. The third option for risks in the third group is that they truly are indeterminate. This happens when the risks are preferred under some utility functions and not preferred under others. In such cases, perfect knowledge of the density functions associated with the risk would not give any more information. The choice of these risks requires further specification of the utility function. Note that in certain situations there may be no members of the second group; this means that there may be no dominant risks. A market perfectly efficient with respect to second-order stochastic dominance would result in no risks that are dominant or dominated.

REFERENCES

1. BOWERS, N.L., GERBER, H.V., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathemtics*. Itasca, Ill.: Society of Actuaries, 1986.
2. BÜHLMANN, H. *Mathematical Methods in Risk Theory*. New York: Springer-Verlag, 1970.
3. CASS, D. AND STIGLITZ, J.E. "The Structure of Investor Preferences and Asset Returns, and Separability in Portfolio Allocation: A Contribution of the Pure Theory of Mutual Funds," *Journal of Economic Theory* 2 (1970): 122–60.
4. DESHPANDE, J.V. AND SINGH, H. "Testing for Second Order Stochastic Dominance," *Communications in Statistics. Part A: Theory and Methods* 14 (1985): 887–97.
5. FRANCK, W.E. "A Likelihood Ratio Test for Stochastic Ordering," *Journal of the American Statistical Association* 79 (1984): 686–91.
6. GOOVAERTS, M.J., DEVYLDER, F. AND HAEZENDONCK, J. *Insurance Premiums: Theory and Applications*. New York: North-Holland, 1984.
7. HUMMEL, T.J. AND SLIGO, J.R. "Empirical Comparison of Univariate and Multivariate Analysis of Variance Procedures," *Psychological Bulletin* 76 (1971): 49–57.
8. KROLL, Y. AND LEVY, H. "Sampling Errors and Portfolio Efficient Analysis," *Journal of Financial and Quantitative Analysis* 15 (1980): 655–89.

9. LEHMANN, E.L. *Theory of Point Estimation.* New York: John Wiley and Sons, 1984.
10. MARDIA, K.V., KENT, J.T. AND BIBBY, J.M. *Multivariate Analysis*. New York: Academic Press, 1979, p. 235.
11. STEIN, W., PFAFFENBERGER, R. AND FRENCH, D. "Sampling Errors in Stochastic Dominance," *Journal of Financial Research* 10 (1987): 259–68.
12. TOLLEY, H.D. AND POPE, R.D. "Testing for Stochastic Dominance," *American Journal of Agricultural Economics* 70 (1988): 693–700.
13. WHITMORE, G.A. AND FINDLAY, M.C. *Stochastic Dominance*. Lexington, Mass.: D.C. Heath and Company, 1978.

DISCUSSION OF PRECEDING PAPER

S. DAVID PROMISLOW:

To test for stochastic dominance, when one does not know the underlying distributions but relies instead on observed data, is an interesting and challenging statistical problem. The authors have made a contribution to the literature on this subject. I believe that a full understanding of such statistical testing requires knowledge of the theory of dominance. This has been discussed a few times in actuarial literature. In addition to the paper by Goovaerts, DeVylder, and Haezendonck, referenced by the authors, the same writers have an extensive discussion in Goovaerts et al. [1]. Elias Shiu has made frequent applications of stochastic dominance to immunization theory; see [3], for example. Moreover, see Promislow [2] for a summary of some aspects of dominance together with actuarial applications.

Despite these references, I would surmise that most actuaries are not very familiar with the concept. A reader who relies on the highly oversimplified example of Section III, as an indication of what testing for dominance can do, is likely to be misled. (In fairness to the authors, this example is presumably designed to illustrate the computational aspects of the algorithm and seems fine for that purpose.) Anyone planning to use dominance testing in a practical decision-making process should carefully note the fact, outlined in the introduction to the paper, that dominance does not help one to make final decisions when comparing risks. It can only serve to eliminate some possibilities, leaving a smaller set of choices. In selecting investments, for example, the concept of dominance cannot help us to solve the problem that we are all frequently faced with, namely, how to choose from two possibilities, when one has a higher return but also carries more risk. In this situation, it is usually the case that neither distribution dominates the other.

In addition, anyone planning to make use of second-order dominance should be aware of the peculiar behavior of this partial ordering. I illustrate this with some examples. (These are based on well-known facts about dominance. For completeness we will give formal proofs of all the statements later on in the discussion.) Throughout the discussion I use *dominance* to mean *second-order dominance.*

Example 1

Suppose you have to choose between the following two types of investments. Type 1 has an annual yield uniformly distributed over the interval 10 to 11 percent. Type 2 has a yield uniformly distributed over the interval

9.99999 to 30 percent. (We assume here that these actual distributions are known, so there is no statistical problem.) It is almost impossible to imagine that there is someone who would *not* choose type 2. And yet, type 2 will *not* dominate over type 1, and this is true no matter how many 9's we add to the decimal point or how large we make the 30. It is well-known that if one distribution is bounded below by a number a, then any distribution that has a positive probability of getting a value less than a cannot dominate the first. We can see this intuitively in the above example because it *is* possible to imagine a scenario, rather far-fetched, in which someone prefers the type 1 investment. For example, you have gambling debts of $110,000, but your total capital is only $100,000. Your bookie gives you one year to pay, but warns that if you are even one penny short at that time, you are in danger of being shot. Hence, you invest your $100,000 in the type 1 investment, not willing to risk even the very slight chance of obtaining a yield less than 10 percent and coming up short of the required amount.

Example 2

Let's get even more bizarre. Compare a normal with a mean of 1 million and standard deviation of 1 to a normal with a mean of minus 1 million and a standard deviation of 0.9999. Will everybody prefer the first distribution in this case? The answer is no!, since it is well-known that for two normals, a necessary condition for dominance is a *lower* standard deviation.

Example 3

I am sure that all of us at one time or other have chosen a distribution with a lower mean than another, because the alternative carried more risk than we desired. We exhibit this behavior whenever we choose a government bond over a stock, or a blue chip stock over a more speculative one. However, a distribution with a lower mean can never *dominate*, regardless of how much extra risk the higher mean distribution carries.

These examples may be extreme, but they still indicate that, in practical situations, it can be difficult to find distributions that dominate others. The point is that the dominant one must be preferred by *all* risk averse individuals, no matter how extreme their degree of risk aversity is, as indicated by the first two examples, or no matter how slight their degree of risk aversity is, as shown by example 3. The authors' statement on page 509 that "any risk in the efficient set is acceptable for the class of utility functions" needs some qualification in my opinion. It is possible to have a set that is theoretically efficient with respect to dominance but that contains risks that no reasonable decision-maker is likely to want.

Referring to the final paragraph of the paper, I am not sure why the authors want to distinguish between risks in the second and third groups. In both cases (assuming no type 2 errors) we have risks that are *not* dominated by any others and hence are potential candidates for a final choice. For example, take the case of comparing three risks, X, Y, and Z. Suppose Y dominates X, but Z is incomparable with both X and Y. This means that some individuals prefer Z to Y. Hence, I fail to see why Z should be excluded from the efficient set. To use the language of mathematics rather than economics, we are simply seeking the maximal elements of a certain partially ordered set, and Z is certainly one of those. The authors claim that risks in the third group may have been incorrectly classified. True, but certainly this can also be the case for risks in the second group.

It is not clear to me what control the proposed test has over the error of concluding dominance for incomparable distributions, which are likely to be quite prevalent, as we have noted above. The authors allude to this at the end of Section II and claim that one should *increase* the value of α used, to avoid such an error. I don't follow this remark. Will not this result in even *more* chance of rejecting the hypothesis and incorrectly asserting dominance in this case? The way I see things, the null hypothesis asserts that the distributions are the same, so presumably the test allows one to be confident (by choosing α low enough) that one will not incorrectly identify two *equal* distributions as showing dominance. But how likely is this error in the *incomparable* case? In certain extreme cases, like my Example 1 above, I would think that this or any other test would be almost certain to conclude dominance of the type 2 distribution. In this instance, even though not theoretically correct, it would seem to be a good conclusion to reach from a practical viewpoint. However, consider the situation with different endpoints. For example, let X be uniform on $(4,5)$ and let Y be uniform on $(1,9)$. I don't believe that one can point to either of these as being superior. They are genuinely incomparable, and the choice will depend on the particular decision-maker. We can show this conclusively. Consider two examples of typical utility functions, $u(x) = \ln(x)$, and $v(x) = x^{1/2}$. Then $E[y(X)] = 1.502$, while $E[u(Y)] = 1.472$, so the person with utility given by u would prefer X. On the other hand, $E[v(X)] = 2.120$, while $E[v(Y)] = 2.167$, so the person with utility given by v would prefer Y. In this case we would not want our test to incorrectly conclude dominance and eliminate from the efficient set a risk that should be there. One can make an even more serious error. Suppose that one risk really does dominate another, but the sample is such that dominance is concluded *the wrong way*. (This is certainly possible, and is discussed by Kroll and Levy, ref. [8] in the paper.) I therefore raise the following question. In the Tolley-Kosorok test, is there any way to estimate the probability of such errors for a given choice of α? If not, then I don't see how one can be confident that the test is doing the job it is designed for.

The authors do *not* point out the theoretical conditions for second-order dominance in the three examples they use. For example, it is well-known that one normal dominates another if and only if it has a higher (or equal) mean and a lower (or equal) standard deviation, while one translated exponential dominates another if and only if it has a higher (or equal) mean and a higher (or equal) lower bound. A good reference for this and many other such relationships is Stoyan [4], in which the term *concave ordering* is used for second-order dominance. I don't think it is clear to the uninitiated reader that in all the cases given in Table 3, population A really does dominate population B.

One of the most interesting parts of the paper is the authors' analysis of the function G. This can be extended to generalize the above facts and establish necessary and sufficient conditions for dominance in almost all cases of continuous distributions from the same location-scale family.

To illustrate this, we will first want to write the function G in the form

$$G(y) = \int_{-\infty}^{y} F_1(t) - F_2(t)\, dt, \tag{1}$$

which follows by evaluating the inner integral in the authors' formula (5). (This is a frequently used form for the function G; see [1], for example.) It makes for a much more direct proof than given in the paper of the well-known fact that

$$G(\infty) = \mu_2 - \mu_1 \tag{2}$$

since this follows immediately from (1) when combined with

$$\mu = \int_{-\infty}^{0} F(t)dt + \int_{0}^{\infty} 1 - F(t)dt \tag{3}$$

(Formula (3) is a familiar result, but for those who want a reference, simply apply Theorem 3.1 of the text *Actuarial Mathematics* by Bowers et al., ref. [1] in the paper, writing the random variable X as $X^+ - X^-$ and noting that $P(X^- > t) = P(X < -t)$ for all $t \geq 0$.)

Note that (2) is true in all cases in which both expectations exist. We don't need the same location-scale family nor the existence of variances as in the authors' derivation.

An immediate consequence of (2) is the fact, mentioned above, that:

Lemma 1. If a distribution with mean μ_1 dominates one with mean μ_2, we must have $\mu_1 \geq \mu_2$.

Proof. If not, from (2), $G(\infty)$ would be positive, and G could not be less than or equal to zero over its entire range.

This also can be seen directly from the expected utility definition by using as a utility function just $u(x) = x$, which is increasing and concave. (Even if one wants to take just "strictly concave" functions, we can still suitably approximate the identity function as closely as we like by such, and use a limiting argument.)

We can also use (1) to prove the fact, noted above, that:

Lemma 2. Suppose that distribution 2 is bounded below by a and distribution 1 has a positive probability of taking values less than a. Then distribution 1 cannot dominate.

Proof. The conditions imply that $F_2(t) = 0$, for $t < a$ but that $F_1(b) > 0$, for some $b < a$. Hence

$$G(a) \geq \int_b^a F_1(t) - F_2(t)dt = \int_b^a F_1(t)dt > 0.$$

For an alternative proof using the utility theory criterion, we can choose a monotone concave function u such that $u(x)$ is 0 for $x \geq a$, and negative for $x < a$. The second distribution gives an expected utility of zero, while the first gives negative expected utility.

Lemmas 1 and 2 are perfectly general and apply to any distributions. We now assume the same conditions as the authors. That is, we postulate a standard distribution with mean 0 and standard deviation 1, which has a density function f_0, which is positive and continuous on an interval (a, b) and which is equal to 0 outside of this interval. (Since it is irrelevant how

f_0 is defined at the end points, a and b, we take these values equal to 0, which is convenient for our development.) Of course, f_0 can be positive on the whole line as in the normal, in which case $a = -\infty$ and $b = \infty$.

We will make use of the fact that since f_0 has mean zero we must have

$$a < 0, \text{ and } b > 0. \tag{4}$$

We then consider two distinct distributions from the same location-scale family. For $i = 1, 2$, distribution i has the density function

$$f_i(x) = \frac{1}{\sigma_i} f_0\left(\frac{x - \mu_i}{\sigma_i}\right),$$

and has mean μ_i and standard deviation σ_i. We want to deduce conditions for dominance of distribution 1 over distribution 2.

We know that that f_i is positive on $(\mu_i + a\sigma_i,\ \mu_i + b\sigma_i)$. Let

$$a' = \min\{\mu_i + a\sigma_i, i = 1, 2\}$$

and

$$b' = \max\{\mu_i + b\sigma_i, i = 1, 2\}.$$

Note then that G is necessarily zero on $(-\infty, a')$ and takes the constant value of $(\mu_2 - \mu_1)$ on (b', ∞). Therefore, some of the statements made on page 519 of the paper (for example, the second line) are not really precise when a and or b are finite. It is best to distinguish two possible cases from the beginning.

Case 1

The point

$$z_0 = \frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_1}$$

is in the interval (a, b), which will imply that

$$y_0 = \mu_1 + z_0\sigma_1 = \mu_2 + z_0\sigma_2$$

is in the interval (a', b'), and it is the only possible critical point in this interval. Since $f(z_0) > 0$, G has a local max at y_0 if and only if $\sigma_1 > \sigma_2$.

Case 2

The point z_0 is $\leq a$, or $\geq b$, so y_0 is outside (a',b'), in which case G is monotone. This corresponds to the case $f(z_0)=0$. (Note that this case always occurs when $\sigma_1=\sigma_2$, since we then define z_0 by a limiting value as $|\sigma_1-\sigma_2|$ approaches 0, and this will be $-\infty$ or ∞ as μ_1 is $>$ or $<$ μ_2. Because the risks are distinct, we can't have equality.

We can now arrive at the conditions for dominance. There are two cases to consider, depending on whether the distributions are bounded below.

Theorem 1

Suppose that $a=-\infty$. Then distribution 1 is SSD over distribution 2 if and only if

(i) $\mu_1 \geq \mu_2$

(ii) $\sigma_1 \leq \sigma_2$.

Proof

Suppose that (i) and (ii) hold. Then $G(-\infty)$ is 0, and from (i) and formula (2), $G(\infty)$ is ≤ 0. Regardless of whether we are in Case 1 or 2 above, it follows from (ii) that it is not possible for G to have a local maximum in (a',b'). Since G has at most one critical point in this interval, it must be that $G \leq 0$ throughout, and SSD occurs.

Lemma 1 shows that (i) is necessary.

To show that (ii) is necessary, suppose that (i) holds but $\sigma_1 > \sigma_2$. Since $b>0$,

$$\mu_1 + b\sigma_1 > \mu_2 + b\sigma_2.$$

Now, if $b<\infty$, then

$$\mu_1 - \mu_2 > b\sigma_2 - b\sigma_1$$

implying that

$$z_0 = \frac{\mu_1 - \mu_2}{\sigma_2 - \sigma_1} < b.$$

Of course this conclusion is automatic if $b=\infty$.

Since $a=-\infty$, we are in case 1 above and the one critical point in (a', b') is a local maximum. Hence G must become positive and dominance does not occur.

Theorem 2

Suppose that the lower bound a is finite. Then X_1 is SSD over X_2 if

(i) $\mu_1 \geq \mu_2$,

(ii) $\mu_1 + a\sigma_1 \geq \mu_2 + a\sigma_2$.

Remark. Since $a<0$, conditions (i) and (ii) of Theorem 1 imply condition (ii′) of Theorem 2, but (ii) is no longer necessary. This is a fact of considerable interest. It shows that in the case where the distributions are bounded below, a distribution with a *larger* variance can *still* dominate, provided that its mean is sufficiently larger than the other.

Proof

Condition (i) is necessary as in Theorem 1. Condition (ii′) is equivalent to saying that the lower bound of distribution 1 is higher than that of distribution 2, so it is necessary by Lemma 2.

Now assume that (i) and (ii′) hold. Condition (ii) implies immediately that

$$\mu_1 - \mu_2 \geq a(\sigma_2 - \sigma_1)$$

If $\sigma_1 \leq \sigma_2$, dominance follows exactly as in the proof given above in Theorem 1. If $\sigma_1 > \sigma_2$, we have that $z_0 < a$. Hence we are in Case 2, so G is monotone and we again have dominance.

Note that the conditions for dominance in all cases are stated in terms of simple inequalities concerning means and variances and depend on the lower bound point a, but on *no* other feature of f_0.

This latter observation leads me to wonder whether there are perhaps simpler and more natural tests for dominance based on standard procedures for comparing means and variances. Of course testing the point $G(\infty)$ reduces, via (2), to the comparison of means. Is it possible that the authors' somewhat involved procedure of estimating $G(y_0)$ could be simplified with a statistic based on conditions (ii) and (ii′) of the above two theorems?

REFERENCES

1. Goovaerts, M.J., DeVylder, F., and Haezendonck, J. "Ordering of Risks: A Review," *Insurance: Mathematics and Economics* 1 (1982): 131–61.
2. Promislow, S.D. "Comparing Risks," *Advances in the Statistical Sciences VI: Actuarial Science*. Edited by I.B. MacNeill and G.J. Umphry. Dordrecht, Holland: D. Reidel Publishing Co., 1987: 79–93.

3. 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. Umphry. Dordrecht, Holland: D. Reidel Publishing Co., 1987: 145–56.
4. STOYAN D. *Comparison Methods for Queues and Other Stochastic Models.* New York: Wiley and Sons, 1983.

(AUTHORS' REVIEW OF DISCUSSION)

H. DENNIS TOLLEY AND MICHAEL R. KOSOROK:

Mr. Promislow's comments help to clear up several fuzzy points in the paper as well as add to the theoretical development. The examples on second-order stochastic dominance help to clarify the fact that because no utility function is specified, one must pay some price for the decision criterion. In this case, there are reasonable situations in which the dominance criterion will not be fruitful. Thus, the lack of dominance can appear in several ways where one would intuitively expect some decision to be possible. In the body of the paper, this problem requires the analyst to specify two probability levels when forming a test for dominance. Changing the relative values of these two levels indirectly adjusts the empirical decision process for "near dominant" situations in which no theoretical dominance is present. As a consequence, the dominance-no dominance decision is really a multiple decision problem.