TECHNOMETRICS, FEBRUARY 1992, VOL. 34, NO. 1

# Testing Reliability in a Stress-Strength Model When $X$ and $Y$ are Normally Distributed

**Samaradasa Weerahandi**

Bellcore
Livingston, NJ 07039

**Richard A. Johnson**

Department of Statistics
University of Wisconsin
Madison, WI 53706

We consider the stress-strength problem in which a unit of strength $X$ is subjected to environmental stress $Y$. An important problem in stress-strength reliability concerns testing hypotheses about the reliability parameter $R = P[X > Y]$. In this article, we consider situations in which $X$ and $Y$ are independent and have normal distributions or can be transformed to normality. We do not require the two population variances to be equal. Our approach leads to test statistics which are exact $p$ values that are represented as one-dimensional integrals. On the basis of the $p$ value, one can also construct approximate confidence intervals for the parameter of interest. We also present an extension of the testing procedure to the case in which both strength and stress depend on covariates. For comparative purposes, the Bayesian solution to the problem is also presented. We use data from a rocket-motor experiment to illustrate the procedure.

KEY WORDS: Confidence interval; Covariates; Generalized $p$ value; Lognormal distribution.

We treat the problem of testing reliability in the stress-strength setting when the strength of a unit or a system, $X$, has cumulative distribution (cdf) $F_1(x)$ and the ultimate stress to which it is subjected, $Y$, has cdf $F_2(y)$. Furthermore, we consider the situation in which the cdf's $F_1$ and $F_2$ are both normal (or can be transformed to being normal—e.g., lognormal). This model is of considerable importance in engineering and applies in a wide variety of circumstances.

We consider tests concerning the reliability parameter $R = P[X > Y]$, or equivalently the probability of failure of a unit $P[X < Y] = 1 - R$, when $X$ and $Y$ are distributed as $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. The population variances are unknown and not necessarily equal. For example, in the rocket-motor experiment data that we study in Section 2.2, $X$ is the motor-operating pressure, $Y$ is the motor-chamber burst strength, and an investigator is interested in determining if the probability of rocket-motor failure is very small.

Even when comparing two treatments, it may be more informative to set hypotheses concerning the *unit free* quantity, $R$, rather than in terms of $H$: $\mu_1 - \mu_2 > \delta$. Since

$$R = P[X > Y] = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right), \qquad (1)$$

where $\Phi$ is the standard normal cdf, the problem of testing $H_0 : R \le R_0$ versus $H_1 : R > R_0$ is equivalent to testing $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$, where

$$\theta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \qquad (2)$$

and $\theta_0 = \Phi^{-1}(R_0)$. In related problems, Downton (1973) derived the minimum variance unbiased estimate of $R$, and Church and Harris (1970) and Reiser and Guttman (1986) obtained approximate confidence intervals for that quantity. For other related studies on stress-strength models, see Johnson (1988).

Our procedure for testing one-sided hypotheses on $\theta$ will be presented in Sections 1 and 2, respectively, for the cases of known and unknown stress distributions. The results in these sections can be employed to construct approximate confidence intervals as well. A simulation study is carried out to compare the proposed test with a conventional approximate test in Section 1. An extension of the results in Section 2 to the case of models with covariates is presented in Section 3 and their Bayesian counterpart is given in the Appendix.

## 1. A CASE OF KNOWN STRESS DISTRIBUTION

In some applications, the stress distribution may be known to the investigator. For an example of a situation in which the stress distribution may be known, consider the case of telephone poles or support towers for power-transmission cables. In this example, usually the wind loadings are so well known that the stress distribution can be derived almost exactly. Church and Harris (1970) gave some other examples of situations in which the distribution of stress can be calculated from known quantities.

In this section, it is convenient to work with the transformed random variables $X' = (X - \mu_2)/\sigma_2$, and $Y' = (Y - \mu_2)/\sigma_2$, so that (2) can be expressed as

$$\theta = \frac{\mu}{\sqrt{1 + \sigma^2}}, \quad (3)$$

where $\mu$ is the mean and $\sigma^2$ is the variance of $X'$ and where $Y'$ has a standard normal distribution.

Let $X'_1, \ldots, X'_m$ be a random sample from the distribution of $X'$. The problem is to construct a procedure for testing (1) based on the sufficient statistics $\overline{X}'$ and $S^2 = \Sigma_{i=1}^{m} (X'_i - \overline{X}')^2/m$, where the two statistics are independent and

$$\overline{X}' \sim N(\mu, \sigma^2/m) \quad \text{and} \quad U = mS^2/\sigma^2 \sim \chi^2_{m-1}. \quad (4)$$

It is known that, for this type of hypotheses, there are no nontrivial exact $\alpha$-level (fixed-level) tests based on the minimal sufficient statistics (see Linnik 1968). One can, however, carry out a test of (3) using the concept of *generalized p value* defined explicitly by Tsui and Weerahandi (1989).

Our procedure tests $H_0$ based on the generalized $p$ value

$$\rho = 1 - E\left[\Phi\left(\frac{\overline{x}'\sqrt{U} - \theta_0\sqrt{U + ms^2}}{s}\right)\right], \quad (5)$$

where $(\overline{x}', s^2)$ is the observed value of $(\overline{X}', S^2)$ and the expectation is taken with respect to the chi-squared random variable $U$ defined in (4). The derivation of (5) is given in Section 1.2. The $p$ value—that is, value of $\rho$—in (5) is a measure (on the scale [0, 1]) of how data support $H_0$. One may also obtain an approximate fixed-level test of size $\alpha$ by rejecting the null hypothesis if the observed value of $\rho$ is less than $\alpha$.

An exact test of significance based on a $p$ value does not necessarily lead to an exact fixed-level test unless the $p$ value has a uniform distribution. According to our simulation results in Section 1.1, however, the $p$ value defined in (5) leads to an almost exact fixed-level test even for small sample sizes.

*Remark 1.* It is clear that when $\theta_0 = 0$, (5) reduces to

$$\rho = 1 - F_t\left(\frac{\overline{x}'\sqrt{m - 1}}{s}\right),$$

where $F_t$ is the cdf of Student-$t$ distribution with $m - 1$ df. This means that our proposed test is indeed an extension of the usual $t$ test, which is valid for the case $\theta_0 = 0$.

### 1.1 Performance of the Test

We are not aware of any other exact test based on a $p$ value or an exact fixed-level test for this testing problem. One can carry out an approximate test, however, using the results of Church and Harris (1970), in which the inferences on $\theta$ are based on the statistic

$$V = \frac{\overline{X}'}{\sqrt{1 + S^2}}.$$

It was shown by Church and Harris (1970) that the distribution of $V$ is asymptotically normal with mean $\theta$ and variance

$$\frac{\sigma^2}{1 + \sigma^2}\left(\frac{1}{m} + \frac{\theta^2}{2(m - 1)}\frac{\sigma^2}{1 + \sigma^2}\right). \quad (6)$$

The asymptotic variance is estimated by

$$\hat{\sigma}_v^2 = \frac{s^2}{1 + s^2}\left(\frac{1}{m} + \frac{v^2}{2(m - 1)}\frac{s^2}{1 + s^2}\right). \quad (7)$$

It can be shown that $V$ is not necessarily stochastically increasing in $\theta$ (see also Fig. 1). This is a requirement in the definition of an extreme region on which a $p$ value can be based (see Tsui and Weerahandi 1989). Even the asymptotic variance of $V$ given by (6) is an increasing function of $\theta$. Consequently, $V$ can be used only in the setting of a fixed-level test.

By means of a simulation study, we now compare the approximate fixed-level test based on the statistic $V$ with that based on the $p$ value given by (5). The simulation results reported in this section are all based on 10,000 simulation samples of size $m = 8$. The fixed-level test based on the results of Church and Harris (1970) will be referred to as the C–H test and the one based on the $p$ value will be referred to as the W–J test. At a nominal level of size $\gamma$, the C–H test rejects the null hypothesis if $v > \theta_0 + z_\gamma\hat{\sigma}_v$, where $z_\gamma$ is the $100(1 - \gamma)$th percentile of the standard normal distribution. The W–J test rejects the null hypothesis if $\rho(\theta_0) < \gamma$. When $\gamma = .05$, $\sigma = 1.0$, and for four values of $\mu$, Table 1 gives the (estimated) probabilities of Type I errors. The W–J test seems to have level nearly equal to the exact nominal size (up to two or three decimal places), but the C–H test underestimates the actual size. Further simulation revealed that the underestimation of the C–H test diminishes as $\sigma$ (and, of course, the sample size $m$) increases.

We have also verified, by simulation, that, for the values of $\mu$ given in Table 1, $\Pr(.95 < \rho) = .05$ and that $\Pr(.025 < \rho < .975) = .95$, up to two or three decimal places. Similar results were seen to be true for other nominal values of $\gamma$. This means that, based on the $p$ value defined in (5), one can carry out

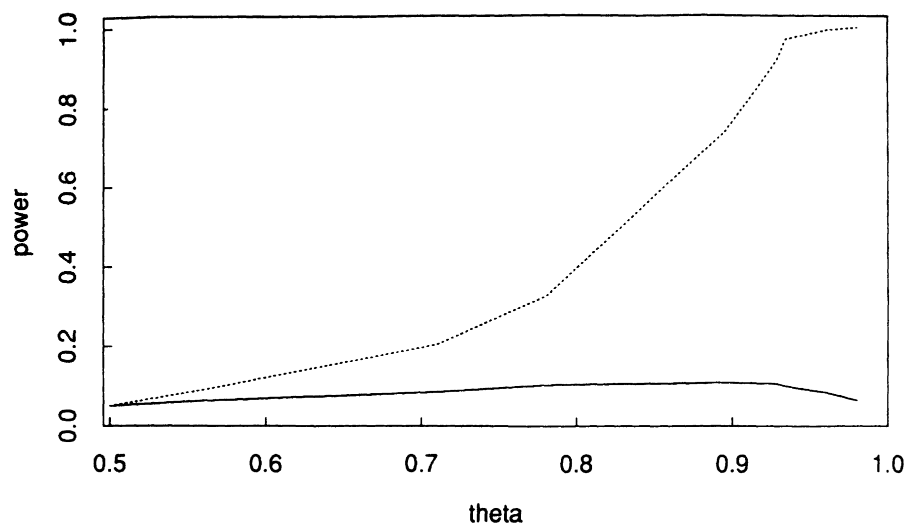


*Figure 1. Estimated Power Functions When $\theta = 1/(1 + \sigma^2)^{.5}$: ---, W–J Test; ———, C–H Test.*

almost exact one-sided or two-sided fixed-level tests on $\theta$ or on $R$.

*Remark 2.* It follows from the preceding observations that one can construct almost exact confidence intervals for $\theta$ or $R$ using (5). For instance, $[\rho^{-1}(.025), \rho^{-1}(.975)]$ is an approximate 95% (in fact, this is exact up to the nearest percent) confidence interval for $\theta$, where $\rho^{-1}(\gamma)$ is the value of $\theta$ at which the $p$ value $\rho(\theta) = \gamma$. As a particular example, a sample of size 8 from the normal distribution $N(1, 2^2)$ gave $\bar{x}' = .483$ and $s^2 = 3.470$. The 95% confidence interval for $\theta$ computed from this sample is $[-.444, .830]$. In comparison, the approximate 95% interval given by the Church and Harris (1970) procedure is $[-.451, .885]$. According to our simulations, the actual coverage of their procedure leading to this confidence interval is only about 93%. It is of interest to note that not only does the latter confidence interval overestimate the true confidence level but also is longer than the former. The related power performance is exhibited in Figures 1 and 2 and will be discussed later.

To compare the power performance of the C–H test and the W–J test, suppose that the null hypothesis $H_0 : \theta \leq .5$ is to be tested at .05 level using a random sample of size 8. According to Table 1, the actual size of the C–H test is substantially greater than .05. Therefore, for the purpose of power comparison, the critical point at .05 level, say $V_c$, was found by simulation, where $\Pr(V > V_c|\theta = .5) = .05$. Similarly, if greater accuracy on the size of the W–J test is required, its critical point $\rho_c$ can be found (by simulation) so that $\Pr(\rho(.5) < \rho_c) = .05$. Note that, subject to the constraint $\theta = .5$, $V_c$ depends on $\mu$ and $\sigma$ both. For each test, the power function $\Pi(\theta)$ is computed from the usual definition $\Pi(\theta) = \Pr(\textit{rejecting } H_0|\theta)$.

*Table 1. Estimated Sizes of Tests When the Nominal Size is .05, $m = 8$, and $\sigma = 1$*

| $\mu$ | 1.0 | 2.0 | 3.0 | 4.0 |
|---|---|---|---|---|
| C–H test | .090 | .071 | .111 | .123 |
| W–J test | .050 | .049 | .050 | .049 |

When $\theta$ varies as a function of $\sigma$, Figure 1 and Figure 2 show, respectively, for $\mu = 1.0$ and $\mu = 2.0$, the estimated power functions of the two tests. The power functions were evaluated at nine values of $\sigma$. It can be observed that the W–J test outperforms the C–H test. The difference in power of the two tests diminishes as $\mu$ (or $m$) increases. Further, notice that the power function of the C–H test is not necessarily increasing in $\theta$ when $\mu$ is close to 0, the most important situation. This is a consequence of the fact that $V$ is not necessarily stochastically increasing in $\theta$, as pointed out earlier.

### 1.2 Derivation of Equation (5)

The purpose of this section is to show that $\rho$ given by (5) is indeed a $p$ value, which serves to measure how the observed data support the null hypothesis. Since the material in the rest of the article is independent of this subsection, readers who are not interested in this derivation can proceed directly to Section 2.

Consider a null hypothesis of the form $H_0$: $\theta \leq \theta_0$. In significance testing of $H_0$, we need to define an extreme region by stochastically ordering the sample space according to the parameter of interest. However, such an extreme region will lead to a workable

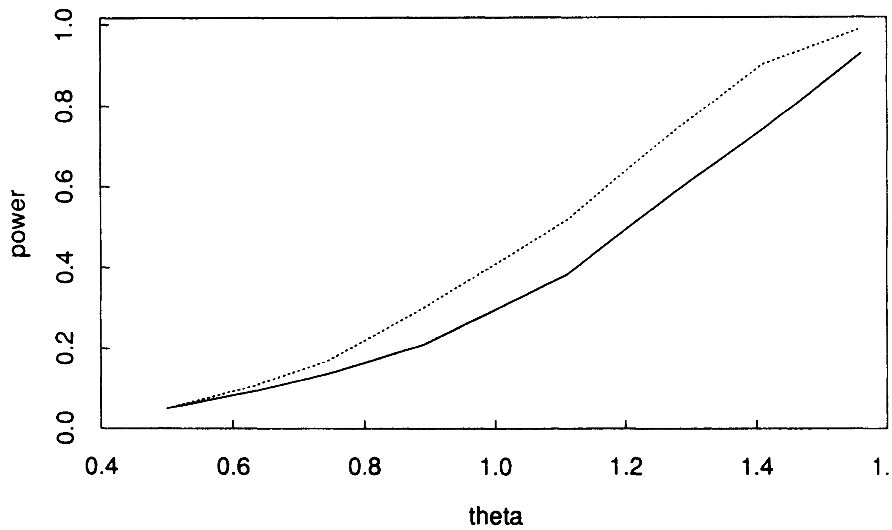


*Figure 2. Estimated Power Functions When $\theta = 2/(1 + \sigma^2)^{.5}$: ---, W–J Test; ——, C–H Test.*

solution only if the corresponding $p$ value is free of nuisance parameters. In many problems, these goals can be accomplished by the method suggested by Tsui and Weerahandi (1989). If $\mathbf{X}$ is an observable random vector with the observed value $\mathbf{x}$, their procedure is to use a *generalized test variable* of the form $T(\mathbf{X}; \mathbf{x}, \nu)$ in the definition of the extreme region, where $\nu = (\theta, \boldsymbol{\delta})$, $\theta$ is the parameter of interest, and $\boldsymbol{\delta}$ is a vector of nuisance parameters. The test variable is chosen to satisfy the requirements that the larger the value of $\theta$, the greater the probability of the extreme region, and that this probability is computable without knowing the value of the nuisance parameter $\boldsymbol{\delta}$. More precisely, the following two properties are imposed on the test variable:

1. For fixed $\mathbf{x}$, $T(\mathbf{x}; \mathbf{x}, \nu_0)$, as well as the distribution of $T(\mathbf{X}; \mathbf{x}, \nu_0)$, are free of the nuisance parameter $\boldsymbol{\delta}$, where $\nu_0 = (\theta_0, \boldsymbol{\delta})$.
2. For fixed $\mathbf{x}$, fixed $\boldsymbol{\delta}$, and for all $t$, $\Pr(T(\mathbf{X}; \mathbf{x}, \nu) > t|\theta)$ is nondecreasing in $\theta$; that is, $T$ is stochastically increasing in $\theta$.

Then, parallel to the definition of conventional extreme regions, a *generalized extreme region* is defined as $C = [\mathbf{X} : T(\mathbf{X}; \mathbf{x}, \nu) \geq T(\mathbf{x}; \mathbf{x}, \nu)]$. Under the usual desirable regularity conditions, such as unbiasedness and invariance, usually there exists only one such extreme region based on minimal sufficient statistics. Since there exist no fixed-level tests for testing (1), we propose to use this procedure to obtain tests of significance.

To proceed this way, consider the region of extreme outcomes (with observed $(\bar{x}', s^2)$ on the boundary)

$$C = \left\{ (\bar{X}', S^2) : \frac{\bar{x}' - \theta\sqrt{1 + \sigma^2 s^2/S^2}}{s} \leq \frac{\bar{X}' - \theta\sqrt{1 + \sigma^2}}{S} \right\}$$

$$= \left\{ (\bar{X}', S^2) : \frac{\sqrt{U}(\bar{x}' - \theta\sqrt{1 + ms^2/U}}{s} \leq Z \right\}, \quad (8)$$

where $Z = \sqrt{m}(\bar{X}' - \mu)/\sigma$ is distributed as standard normal and $U$, as defined in (4), has a chi-squared distribution with $m - 1$ df. It is now evident that $\rho$ given in (5) is the probability of the preceding extreme region when $\theta = \theta_0$.

As suggested by (8), consider the generalized test variable

$$T = \frac{Z}{\bar{x}'\sqrt{U} - \theta\sqrt{U + ms^2}}$$

with the value $s^{-1}$ at the observed quantities. Obviously $T$ is stochastically increasing in $\theta$ and its cdf is independent of $\sigma^2$. Hence $C$ is indeed an extreme region and its observed probability serves to measure how strongly the data support $H_0$; the smaller the probability, the greater the evidence against $H_0$.

## 2. A CASE OF UNKNOWN STRESS DISTRIBUTION

Now suppose that the reliability of a unit is to be tested using independent samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ obtained from the normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Reiser and Guttman (1986) considered this problem and gave formulas for constructing approximate confidence intervals for $R$. Kececioglu and Lamarre (1978) considered the reliability analysis of mechanical components as a situation in which observations can be obtained from both stress and strength distributions.

By sufficiency, we can confine our attention to tests based on the maximum likelihood estimators

$$\bar{X},\ S_1^2 = \sum_{i=1}^{m} (X_i - \bar{X})^2/m,\ \bar{Y}$$

$$S_2^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2/n, \tag{9}$$

where all four statistics are independent and

$$\bar{X} \sim N(\mu_1, \sigma_1^2/m), \qquad \bar{Y} \sim N(\mu_2, \sigma_2^2/n),$$

$$mS_1^2/\sigma_1^2 \sim \chi^2_{m-1}, \qquad nS_2^2/\sigma_2^2 \sim \chi^2_{n-1}. \tag{10}$$

Our test in this case is based on the generalized $p$ value

$$\rho = 1 - E\left[G_{m+n-2}\left(\frac{(\bar{x} - \bar{y})\sqrt{m+n-2}}{(s_1^2/B + s_2^2/(1-B))^{1/2}}\right)\right]. \tag{11}$$

Here $G$ is the cdf of the noncentral $t$ distribution with $m + n - 2$ df and noncentrality parameter

$$\delta_{\theta_0}(B) = \theta_0 \left(\frac{m(1 - B)s_1^2 + nBs_2^2}{(1 - B)s_1^2 + Bs_2^2}\right)^{1/2},$$

and the expectation is taken with respect to $B$, which is distributed as beta $((m - 1)/2, (n - 1)/2)$. The expectation in (11) can be evaluated by numerical integration. Notice that $\rho$ in (11) is free of nuisance parameters and thus produces an exact solution to the testing problem. The derivation of (11) is similar to that of the solution to the Behrens–Fisher problem given by Tsui and Weerahandi (1989).

## 2.1 Discussion

The generalized $p$ values in (5) and (11) can be employed in significance testing of hypotheses on the reliability parameter $R$. As in the usual definition of $p$ values, these are exact probabilities of appropriate extreme regions and serve to measure how strongly the observed data support the null hypothesis; the smaller the $p$ value, the stronger the evidence against the null hypothesis. The $p$ values do not necessarily have uniform distributions, and consequently there are no exact counterparts in nonrandomized fixed-level testing. This is the case with $p$ values in (5) and (11), and therefore they provide exact testing procedures in significance testing only. Cox (1977) discussed the distinction between significance testing using $p$ values and fixed-level testing. Even in those situations in which $p$ values provide exact fixed-level tests, significance testing is often preferred to fixed-level testing due to some drawbacks of the latter (see Kiefer 1977, Pratt 1961).

It can be shown (see Appendix) that the generalized significance value (11) is numerically (not logically, of course) equivalent to the Bayesian posterior probability that $H_0$ is true under the noninformative prior $d\mu_1 d\mu_2 \sigma_1^{-1} d\sigma_1 \sigma_2^{-1} d\sigma_2$. As is usually the case in the normal theory, it is also equivalent to the fiducial probability that $\theta \le \theta_0$. Nevertheless, it should be emphasized that the $\rho$ given in (11) is an exact probability of a subset (extreme region) of the sample space obtained by treating $\theta$ as a fixed unknown constant rather than a random variable. Since the testing procedure is generalized Bayes, we can expect a test based on the generalized $p$ value to have good properties. In a related study, Weerahandi (1987) gave a generalized $p$ value for comparing two normal population means [i.e., $p$ value given by (11) when $\theta_0 = 0$] or two regression models. Thursby (1990) used Monte Carlo analyses to show that this test was best among competitors whose actual size is less than or equal to the nominal size.

Since $\Pr(X > Y) = \Pr(\log X > \log Y)$, (5) and (11) also apply to the case in which $X$ and $Y$ are lognormally distributed; the formulas are to be applied after taking the logarithms of the observations. This is, of course, more generally true if there is a monotonic transformation that can transform both distributions into normal distributions. If $X$ and $Y$ are exponentially distributed, the null hypothesis $H_0$ can be tested by the usual $F$ test (see Lawless 1982, p. 112). For other distributions such as Weibull and gamma, no exact tests are available.

Since the $p$ value given by (11) is not uniformly distributed, it does not lead to exact confidence-interval procedures. As in the case of known stress distribution, however, one can obtain approximate confidence intervals by treating the $p$ value evaluated at $\theta$ as a random variable with an approximate uniform distribution. Notice that for large sample sizes $B$ in (11) becomes a degenerate random variable and the second term on the right side of (11) becomes the cdf of a continuous random variable. Consequently, $\rho$ in (11) is uniformly distributed and hence the assumption is asymptotically true. To study the accuracy of the approximation for small sample sizes, we have carried out a simulation study when $m = 10$ and $n = 10$ and when $m = 30$ and $n = 40$. Notice that for the purpose of simulation we can assume without loss of generality that $\mu_2 = 0$ and $\sigma_2 = 1$ and set values of $\mu_1$ and $\sigma_1$ relative to these parameters. Typically, $\mu_1$ is greater than $\mu_2$ (consequently, $\sigma_1$ tends to be greater than $\sigma_2$), and therefore we have chosen just one negative value and three positive values of $\mu_1$. The accuracy of the approximation when applied to the problem of constructing 95% confidence intervals is carried out using 5,000 simulated samples from each distribution. The probability $\Pr(.025 \le \rho \le .975)$ is estimated by the sample proportion of $\rho$ between .025 and .975. Table 2 shows the estimated probabilities as a percentage for a wide

*Table 2. Estimated Percent Coverage of Approximate 95% Confidence Interval*

| $\mu_1$ = | −1.0 | .0 | 1.0 | 2.0 | 4.0 |
|---|---|---|---|---|---|
| | *When m = 10 and n = 10* | | | | |
| $\sigma_1$ = .5 | 95.6 | 96.0 | 96.0 | 95.9 | 95.8 |
| $\sigma_1$ = 1.0 | 96.2 | 96.4 | 96.3 | 96.2 | 95.9 |
| $\sigma_1$ = 2.0 | 96.0 | 95.9 | 95.9 | 95.5 | 95.7 |
| $\sigma_1$ = 3.0 | 95.6 | 95.6 | 95.5 | 95.4 | 95.5 |
| | *When m = 30 and n = 40* | | | | |
| $\sigma_1$ = .5 | 95.2 | 95.5 | 95.6 | 95.9 | 95.8 |
| $\sigma_1$ = 1.0 | 95.6 | 95.5 | 95.6 | 95.8 | 95.9 |
| $\sigma_1$ = 2.0 | 95.4 | 95.5 | 95.5 | 95.5 | 95.3 |
| $\sigma_1$ = 3.0 | 95.1 | 95.3 | 95.5 | 95.4 | 95.5 |

variety of values of $\mu_1$ and $\sigma_1$, including some values that can be considered extreme. According to our simulation, for small sample sizes such as $m = 10$ and $n = 10$, $\Pr(.025 \le \rho \le .975) = .96$ in almost all of the values of parameters we considered. For moderate sample sizes such as $m = 30$ and $n = 40$, this probability ranges from .95 to .96. When each sample size is about 100, the probabilities of .96 appearing in Table 2 were found to reduce to .95, correct up to two decimal places. We have also observed that if an investigator applies (11) in a fixed-level testing, the actual size of the test is somewhat less than the nominal size. However, for reasons outlined previously, we do not recommend fixed-level testing when one can carry out significance testing using exact $p$ values.

The $p$ value in (11) can be evaluated by numerical integration. The expected value appearing in (11) can also be approximated using random samples from the underlying distribution with respect to which the expectation is to be taken. For the problem of testing $H_0$: $\mu_1 - \mu_2 \le (\sigma_1^2 + \sigma_2^2)^{1/2}$—that is, when $\theta_0 = 1$—Table 3 displays the $p$ values computed from (11) for a range of values of $D = (\bar{x} - \bar{y})/(s_1^2 + s_2^2)^{1/2}$, $R = s_1^2/(s_1^2 + s_2^2)$, and two sets of values of $m$ and $n$. The FORTRAN code, which employs two IMSL subroutines, used in this numerical integration is available from the first author. The underlying integrand is so well behaved that it took less than a minute to compute all 50 $p$ values shown in Table 3 correct up to four decimal places. Table 4 shows the approximate $p$ values obtained using 200 random numbers from the beta distribution beta $((m - 1)/2, (n - 1)/2)$. It is evident that, although the approximation is pretty good, its accuracy decreases as $R$ increases.

## 2.2 Example

As a specific example of an application in which data from both stress and strength distributions are available, consider the rocket-motor experiment data reported by Guttman, Johnson, Bhattacharyya, and Reiser (1988). Suppose that one is interested in testing the reliability of the rocket motor at the highest operating temperature—namely, 59 degrees centigrade—at which the operating pressure ($X$) distribution tends to be closest to the chamber burst strength ($Y$) distribution. As in the work of Guttman et al. (1988), each distribution is assumed to be normal; a quantile-quantile plot of data has supported this assumption. Shown in Table 5 are some observed values of $Y$ from 17 motor cases and a sample size 24 from the operating pressure distribution.

Suppose that an investigator wishes to establish that there is less than one in a million chances that the pressure may exceed the rocket-motor strength; that is, it is of interest to test the null hypothesis $H_0$: $R \le .999999$ against the alternative hypothesis $H_1$: $R > .999999$, or equivalently $H_1$: $\theta > 4.75059$. The observed summary statistics are $\bar{x} = 16.485$, $\bar{y} = 7.789$, $s_1^2 = .3409$, and $s_2^2 = .05414$, and the corresponding $p$ value computed at $\theta = 4.75059$ is $\rho = .0000042$. Now it is evident that the observed data have provided very strong evidence in favor of the desired alternative hypothesis $H_1$. An approximate 95% confidence interval for $\theta$ obtained using (11) is [9.52, 17.67]. In this application, although the investigator can use this parametric analysis to make a strong case in favor of the claim made, it is advisable to further support the validity of the claim by means of a nonparametric test.

*Table 3. Exact p Values by Numerical Integration*

| | m = 10, n = 10 | | | | | m = 10, n = 20 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| D | .1 | .2 | .3 | .4 | .5 | .1 | .2 | .3 | .4 | .5 |
| .5 | .951 | .954 | .956 | .957 | .957 | .983 | .980 | .978 | .975 | .973 |
| 1.0 | .603 | .617 | .626 | .632 | .633 | .576 | .596 | .608 | .615 | .619 |
| 1.5 | .208 | .210 | .212 | .214 | .214 | .091 | .107 | .124 | .141 | .158 |
| 2.0 | .055 | .052 | .049 | .048 | .048 | .006 | .009 | .013 | .019 | .026 |
| 2.5 | .015 | .013 | .011 | .010 | .010 | .000 | .001 | .001 | .003 | .005 |

*Table 4. Approximate p Values Using 200 Beta Random Numbers*

| | $m = 10, n = 10$ | | | | | $m = 10, n = 20$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *D* | *.1* | *.2* | *.3* | *.4* | *.5* | *.1* | *.2* | *.3* | *.4* | *.5* |
| .5 | .951 | .955 | .954 | .957 | .958 | .982 | .980 | .978 | .975 | .973 |
| 1.0 | .602 | .625 | .616 | .632 | .642 | .567 | .589 | .604 | .605 | .615 |
| 1.5 | .205 | .219 | .203 | .215 | .227 | .083 | .102 | .120 | .128 | .156 |
| 2.0 | .051 | .055 | .047 | .048 | .055 | .005 | .008 | .012 | .013 | .027 |
| 2.5 | .012 | .014 | .011 | .009 | .013 | .000 | .000 | .001 | .001 | .006 |

## 3. THE STRESS–STRENGTH MODEL WITH COVARIATES

An important extension of the stress-strength model allows the strength $X$ and the stress $Y$ to depend on some covariates, where $\mathbf{z}_1$ is a $p \times 1$ vector and $\mathbf{z}_2$ is a $q \times 1$ vector, respectively (see Guttman et al. 1988). In particular, under a linear model structure, $X = \boldsymbol{\beta}_1'\mathbf{z}_1 + \epsilon_1$, where $\epsilon_1 \sim N(0, \sigma_1^2)$, $Y = \boldsymbol{\beta}_2'\mathbf{z}_2 + \epsilon_2$, where $\epsilon_2 \sim N(0, \sigma_2^2)$, and the errors $\epsilon_1$ and $\epsilon_2$ are independent.

Let $(\mathbf{X}, \mathbf{Z}_1)$ be the matrices formed by $m$ observations $(x_j, \mathbf{z}_{1j})$, $j = 1, \ldots, m$, on strength, and let $(\mathbf{Y}, \mathbf{Z}_2)$ be the matrices formed by $n$ observations $(y_j, \mathbf{z}_{2j})$, $j = 1, \ldots, n$, on stress. The least squares estimators $\hat{\boldsymbol{\beta}}_1 = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1\mathbf{X}$ and $\hat{\boldsymbol{\beta}}_2 = (\mathbf{Z}_2'\mathbf{Z}_2)^{-1}\mathbf{Z}_2\mathbf{Y}$ and residual sums of squares $(m - p)S_1^2 = \sum_{i=1}^m \hat{\epsilon}_{1i}^2$ and $(n - q)S_2^2 = \sum_{j=1}^n \hat{\epsilon}_{2j}^2$ have distributions

$$\hat{\boldsymbol{\beta}}_i \sim N(\boldsymbol{\beta}_i, \sigma_i^2(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}), \qquad i = 1, 2,$$
$$(m - p)S_1^2/\sigma_1^2 \sim \chi^2_{m-p};$$
$$(n - q)S_2^2/\sigma_2^2 \sim \chi^2_{n-q}. \tag{12}$$

Moreover, all four of these random variables are independent. In the present context, the parameter of interest is

$$R = P[X > Y] = \Phi\left(\frac{\boldsymbol{\beta}_1'\mathbf{z}_1 - \boldsymbol{\beta}_2'\mathbf{z}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right). \tag{13}$$

where $\mathbf{z}_1$ and $\mathbf{z}_2$ are given values for the predictor variables. Testing $H_0$: $R \le R_0$ versus $H_1$: $R \ge R_0$ is equivalent to testing $H_0$: $\theta \le \theta_0$ versus $H_1$: $\theta \ge \theta_0$, where

$$\theta = \frac{\boldsymbol{\beta}_1'\mathbf{z}_1 - \boldsymbol{\beta}_2'\mathbf{z}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \tag{14}$$

In view of the similarity between the present problem and the simpler case in Section 2, and the parallel between (10) and (12), we can immediately deduce that $H_0$ can be tested on the basis of the generalized $p$ value

$$p = 1 - E\left[G^{\eta}\left(\frac{(\hat{\boldsymbol{\beta}}_1'\mathbf{z}_1 - \hat{\boldsymbol{\beta}}_2'\mathbf{z}_2)\sqrt{m + n - p - q}}{\sqrt{(m - p)s_1^2 A_1/B + (n - q)s_2^2 A_2/(1 - B)}}\right)\right], \tag{15}$$

where $G^{\eta}$ is the cdf of the noncentral $t$ distribution with $m + n - p - q$ df and noncentrality parameter

$$\eta = \theta_0\left(\frac{(m - p)s_1^2(1 - B) + (n - q)s_2^2 B}{(m - p)s_1^2(1 - B)A_1 + (n - q)s_2^2 B A_2}\right),$$

where $A_i = \mathbf{z}_i'(\mathbf{Z}_i'\mathbf{Z}_i)^{-1}\mathbf{z}_i$, $i = 1, 2$, and the expectation is taken with respect to the random variable $B$, which is distributed as beta $[(m - p)/2, (n - q)/2]$.

*Remark 3.* In a Bayesian setting with noninformative priors on the unknown parameters, $p$ in (15) also gives the posterior probability that $H_0$: $\theta \le \theta_0$ is true.

## ACKNOWLEDGMENTS

We thank the editor, the associate editor, and the referees for their helpful comments on an earlier version of this article. The research of the second author was supported by the Air Force Office of Scientific Research under Grant AFOSR-87-0256.

*Table 5. Rocket-Motor Experiment Data*

| *Chamber burst strength (X)* | *Operating pressure (Y)* |
|---|---|
| 15.30 | 7.74010 |
| 17.10 | 7.77490 |
| 16.30 | 7.72270 |
| 16.05 | 7.77925 |
| 16.75 | 7.96195 |
| 16.60 | 7.44720 |
| 17.10 | 8.07070 |
| 17.50 | 7.89525 |
| 16.10 | 8.07360 |
| 16.10 | 7.49650 |
| 16.00 | 7.57190 |
| 16.75 | 7.79810 |
| 17.50 | 7.87640 |
| 16.50 | 8.19250 |
| 16.40 | 8.01705 |
| 16.00 | 7.94310 |
| 16.20 | 7.71835 |
| | 7.87785 |
| | 7.29040 |
| | 7.75750 |
| | 7.31960 |
| | 7.63570 |
| | 8.06055 |
| | 7.91120 |

## APPENDIX: A BAYESIAN ANALYSIS

Motivated by the structure of our solution to the testing problem in Section 2, we develop here the closely related Bayesian approach for making inferences about $R$ or, equivalently, $\theta = (\mu_1 - \mu_2)/(\sigma_1^2 + \sigma_2^2)^{1/2}$. Enis and Geisser (1971) gave the solution to this problem when $\sigma_1 = \sigma_2$. Here we drop the assumption of equal variances.

We assume independent conjugate prior distributions for $(\mu_1, \sigma_1)$ and $(\mu_2, \sigma_2)$, of the form

$$\mu_i|\sigma_i^2 \sim N(\nu_i, \sigma_i^2/\lambda_i)$$

$$\tau_i = \sigma_i^{-2} \sim G(\alpha_i, \beta_i)$$

for $i = 1, 2$, where $\tau$ denotes the precision of the normal distribution and $G(\alpha, \beta)$ is the gamma distribution with parameters $\alpha$ and $\beta$. The posterior distributions are then given by (see DeGroot 1985)

$$\mu_1|\sigma_1^2 \sim N(\nu_1', \sigma_1^2/m'),$$

$$\text{where} \quad \nu_1' = \frac{\lambda_1\nu_1 + m\bar{x}}{\lambda_1 + m}, \qquad m' = m + \lambda_1$$

$$\mu_2|\sigma_2^2 \sim N(\nu_2', \sigma_2^2/n'),$$

$$\text{where} \; \nu_2' = \frac{\lambda_2\nu_2 + n\bar{y}}{\lambda_2 + n}, \qquad n' = n + \lambda_2$$

$$\tau_i = \sigma_i^{-2} \sim G(\alpha_i', \beta_i'), \qquad i = 1, 2, \tag{A.1}$$

where $\alpha_1' = \alpha_1 + m/2$, $\alpha_2' = \alpha_2 + n/2$, $\beta_1' = \beta_1 + (ms_1^2/2) + [m\lambda_1(\bar{x} - \nu_1)^2/2(\lambda_1 + m)]$, and $\beta_2' = \beta_2 + (ns_2^2/2) + n\lambda_2(\bar{y} - \nu_2)^2/2(\lambda_2 + n)]$.

To obtain the posterior distribution of $\theta = (\mu_1 - \mu_2)/(\sigma_1^2 + \sigma_2^2)^{1/2}$, notice that

$$\theta|\sigma_1^2, \sigma_2^2 \sim N\left(\frac{\nu_1' - \nu_2'}{\sqrt{\sigma_1^2 + \sigma_2^2}}, \frac{\sigma_1^2/m' + \sigma_2^2/n'}{\sigma_1^2 + \sigma_2^2}\right), \tag{A.2}$$

where $\nu_i'$ denotes the posterior mean of $\mu_i|\sigma_i^2$. Hence the posterior cdf of $\theta$ can be expressed as

$$F(\theta) = E\left[P\left(Z \le \frac{\theta\sqrt{\sigma_1^2 + \sigma_2^2} - (\nu_1' - \nu_2')}{\sqrt{\sigma_1^2/m' + \sigma_2^2/n'}}\right)\right]. \tag{A.3}$$

where $Z$ denotes the standard normal random variable and the expectation is taken with respect to $\sigma_1^2$ and $\sigma_2^2$.

Since $W_i = \beta_i'/\sigma_i^2$ and are distributed as $G(\alpha_i', 1)$ for $i = 1, 2$, are independent, we have that $S = W_1 + W_2 \sim G(\alpha_1' + \alpha_2', 1)$ and $R = W_1/S \sim$ beta $(\alpha_1', \alpha_2')$. Thus we can express $F(\theta)$ as

$$F(\theta) = E\left[P\left(Z \le \frac{\theta(\beta_1'/W_1 + \beta_2'/W_2)^{1/2} - (\nu_1' - \nu_2')}{(\beta_1'/(m'W_1) + \beta_2'/(n'W_2))^{1/2}} W_1, W_2\right)\right]$$

$$= E\left[\Phi\left(\frac{\theta(\beta_1'/R + \beta_2'/(1 - R))^{1/2} - (\nu_1' - \nu_2')S^{1/2}}{(\beta_1'/(m'R) + \beta_2'/(n'(1 - R)))^{1/2}}\right)\right], \tag{A.4}$$

where $\Phi$ is the standard normal cdf and the expectation is taken with respect to the independent random variables $R$ and $S$. The posterior distribution (A.4) provides the basis for inferences concerning $\theta$.

*Remark A.1.* In practice, probabilities determined by (A.4) could be evaluated by performing two-dimensional numerical integrations correct up to a desired number of decimal places. It could also be approximated by generating a large number of random digits from the distribution of $(R, S)$ and taking the average value of the integrated in (A.4).

*Remark A.2.* If we begin with a noninformative prior distribution $d\mu_1 d\mu_2 \sigma_1^{-1} d\sigma_1 \sigma_2^{-1} d\sigma_2$ in place of the conjugate prior (A.1), then it can be deduced from (A.4) or shown directly that (A.4) reduces to

$$F(\theta) = 1 - E\left[G_{m+n-2}^{\theta}\left(\frac{(\bar{x} - \bar{y}\sqrt{m + n - 2}}{\sqrt{s_1^2/B + s_2^2/(1 - B)}}\right)\right],$$

where the expectation is taken with respect to $B$, which is distributed as beta$((m - 1)/2, (n - 1)/2)$ and $G_{m+n-2}^{\theta}$ is the cdf of the noncentral $t$ distribution with $m + n - 2$ df and noncentrality parameter

$$\theta\left(\frac{m(1 - B)s_1^2 + nBs_2^2}{(1 - B)s_1^2 + Bs_2^2}\right)^{1/2}.$$

Now, it is evident that the $p$ value given in (11) is numerically the same as the posterior probability that the null hypothesis $H_0$ is true.

*[Received February 1990. Revised April 1991.]*

## REFERENCES

Church, J. D., and Harris, B. (1970), "The Estimation of Reliability From Stress–Strength Relationships," *Technometrics*, 12, 49–54.

Cox, D. R. (1977), "The Role of Significance Tests," *Scandinavian Journal of Statistics*, 4, 49–62.

DeGroot, M. (1985), *Probability and Statistics*, Reading, MA: Addison-Wesley.

Downton, F. (1973), "The Estimation of the Pr($Y < X$) in the Normal Case," *Technometrics*, 15, 551–558.

Enis, P., and Geisser, S. (1971), "Estimation of the Probability That $X < Y$," *Journal of the American Statistical Association*, 66, 162–168.

Guttman, I., Johnson, R. A., Bhattacharyya, G. K., and Reiser, B. (1988), "Confidence Limits for Stress-Strength Models With Explanatory Variables," *Technometrics*, 30, 161–168.

Johnson, R. A. (1988), "Stress–Strength Models for Reliability," *Handbook of Statistics* (Vol. 7), eds. P. R. Krishnaiah and C. R. Rao, Amsterdam: North-Holland.

Kececioglu, D., and Lamarre, G. (1978), "Mechanical Reliability Confidence Limits," *Journal of Mechanical Design*, 100, 607–612.

Kiefer, J. (1977), "Conditional Confidence Statements and Confidence Estimators," *Journal of American Statistical Association*, 72, 789–808.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley.

Linnik, Y. (1968), *Statistical Problems With Nuisance Parameters* (translation of mathematical monograph 20), New York: American Mathematical Society.

Pratt, J. W. (1961), "Review of Lehmann's Testing Statistical Hypotheses," *Journal of the American Statistical Association*, 56, 163–166.

Reiser, B., and Guttman, I. (1986), "Statistical Inference for Pr($Y < X$)," *Technometrics*, 28, 253–257.

Thursby, J. G. (1990), "A Comparison of Several Exact and Approximate Tests for Structural Shift Under Heteroscedasticity," working paper, Purdue University, Dept. of Economics.

Tsui, K., and Weerahandi, S. (1989), "Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, 84, 602–607.

Weerahandi, S. (1987), "Testing Regression Equality With Unequal Variances," *Econometrica*, 55, 1211–1215.