TESTING AND ESTIMATING LOCATION VECTORS
UNDER HETEROSKEDASTICITY

William Griffiths and George Judge

No. 36 - February, 1989
ABSTRACT

Testing and Estimating Location Vectors Under Heteroskedasticity

An exact test recently proposed by Weerahandi (1987) for testing the equality of location vectors under heteroskedasticity is evaluated. In terms of presumed size, power, computational ease, and the risk properties of resulting pre-test estimators, Weerahandi's test proves to be less attractive than a commonly used asymptotic test. A Stein-like estimator suggested as an alternative to the pre-test estimator is demonstrated to behave in a minimax way. A generalization of this estimator to a set of equations with contemporaneous correlation between the errors is proposed.

Key words: Test, size of test, power of test, pre-test, minimax estimator, squared error loss, risk, Stein-rule estimator.
1. INTRODUCTION

In a recent issue of this journal, Weerahandi (1987) considers the problem of testing the equality between two location vectors when the corresponding scale parameters are possibly unequal and proposes a simple exact test that is similar in form to that of Chow (1960). Several earlier articles also consider this problem. Toyoda (1974) and Schmidt and Sickles (1977) demonstrate how poorly the conventional Chow test can perform under heteroskedasticity. Jayatissa (1977) suggests an exact test that allows for the different variances, but it is one that has few degrees of freedom and it has been shown to lack some desirable invariance properties (Tsurumi 1984). Tsurumi and Sheffin (1985) use a Monte Carlo experiment to compare a number of asymptotic tests, emphasizing in particular an asymptotic $F$-test conditional on the posterior mean of the ratio of standard deviations. They also suggest an exact test but, like Jayatissa’s test, it has relatively few degrees of freedom and it proved less powerful than the asymptotic tests. Related work that examines the consequences of testing for heteroskedasticity on the properties of location estimators and tests is that of Greenberg (1980), Yancey, Judge and Miyazaki (1984) and Ohtani (1987).

In this paper we first focus on Weerahandi’s exact test. This test uses the magnitude of a $p$-value as the criterion for rejection or acceptance of a null hypothesis that equates location vectors. It is an unconventional test in the sense that its size is not necessarily equal to the critical $p$-value upon which the decision to accept or reject is based. A Monte Carlo sampling experiment is used to estimate the size and power of the test and to make a comparison with the properties of a conventional asymptotic $F$-test. In addition to size and power, a squared error measure is used to evaluate the sampling performance of the resulting pre-test estimators. We find that, although Weerahandi’s test is an exact one in the sense that it produces an exact $p$-value, its sampling performance in finite samples is inferior to that of a conventional asymptotic $F$-test.

We next consider an estimation rule that demonstrates under a squared error loss measure, the non-optimality of the pre-test estimator based on an asymptotic $F$-test. Some non-traditional estimation rules that make use of the test statistics for models with an unknown covariance matrix are considered and their satisfactory sampling performances are demonstrated. Finally we generalize the results for these testing-estimation rules to include three or more samples and a more
general unknown covariance matrix, that defines the seemingly unrelated statistical model. We end the paper with a discussion of the statistical implications of the results.

2. Model, Estimators and Tests

Assume we observe two \((T \times 1)\) vectors of sample observations \(y_1, y_2\) such that

\[
y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & \beta_1 \\ X_2 & \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = X \beta + e
\]

where \(X_1\) and \(X_2\) are \((T \times K)\) design matrices, \(\beta_1\) and \(\beta_2\) are \(K\) dimensional unknown location vectors and \(e_1\) and \(e_2\) are unobservable \((T \times 1)\) normal random vectors with mean vectors 0 and covariance matrix

\[
E[ee'] = \begin{bmatrix} \sigma_1^2 I_T & 0 \\ 0 & \sigma_2^2 I_T \end{bmatrix} = \Sigma
\]

Given that \(\sigma_1^2\) and \(\sigma_2^2\) are not (necessarily) equal, we are interested in testing whether the location vectors are identical \((\beta_1 = \beta_2)\), and in the sampling properties of alternative estimators for \(\beta\), given the uncertainty about the equality of \(\beta_1\) and \(\beta_2\).

2.1 Testing

Let \(b_i = (X_i'X_i)^{-1}X_i'y_i\), and \(\hat{\sigma}_i^2 = (y_i-X_i b_i)'(y_i-X_i b_i)/(T-K)(i = 1,2)\) be the usual least squares location and scale estimates. It is well known that
(2.3) \[(b_1 - b_2) \sim N \left[ \left( \bar{\beta}_1 - \bar{\beta}_2 \right), \left( \sigma_1^2 (X_1' X_1)^{-1} + \sigma_2^2 (X_2' X_2)^{-1} \right)^{-1} \right] \]

(2.4) \[w_1 = (b_1 - b_2) \left[ \sigma_1^2 (X_1' X_1)^{-1} + \sigma_2^2 (X_2' X_2)^{-1} \right]^{-1} \sim \chi^2_{(r - K)} \]

(2.5) \[w_2 = \frac{(T - K) \sigma_1^2}{\sigma_1^2} + \frac{(T - K) \sigma_2^2}{\sigma_2^2} \sim \chi^2_{[2(T - K)]} \]

and that the quantities \(w_1\) and \(w_2\) in (2.4) and (2.5) are independent. Taking the ratio of \(w_1\) to \(w_2\), with each divided by its degrees of freedom, yields the F-statistic

(2.6) \[F_1 = \frac{(b_1 - b_2) \left[ \sigma_1^2 (X_1' X_1)^{-1} + \sigma_2^2 (X_2' X_2)^{-1} \right]^{-1}}{(\tilde{\sigma}_1^2 \sigma_1^2 + \tilde{\sigma}_2^2 \sigma_2^2)/2} \sim F_{[K, \chi^2_{(T - K)}]} \]

If \(\sigma_1^2 = \sigma_2^2\) then \(F_1\) becomes the usual Chow statistic for testing the equality \(\bar{\beta}_1 = \bar{\beta}_2\) under the assumption of equal variances.

When \(\sigma_1^2 \neq \sigma_2^2\), these unknown parameters remain in \(F_1\) and an alternative strategy is necessary. One way to proceed is to replace \(\sigma_1^2\) and \(\sigma_2^2\) by consistent estimators \(\tilde{\sigma}_1^2\) and \(\tilde{\sigma}_2^2\), in which case \(F_1\) becomes

(2.7) \[F_A = (b_1 - b_2) \left[ \tilde{\sigma}_1^2 (X_1' X_1)^{-1} + \tilde{\sigma}_2^2 (X_2' X_2)^{-1} \right]^{-1} \sim \frac{(b_1 - b_2)}{\tilde{\sigma}_1^2 \sigma_1^2 \sigma_2^2 / 2} \sim F_{[(T - K)]} \]

Under the usual assumptions about the limiting behavior of \(X_1\) and \(X_2\), we have \(F_A \sim F_{[\chi^2_{(T - K)}]}\). Thus, a commonly used approximate large sample test for testing \(\bar{\beta}_1 = \bar{\beta}_2\) under heteroskedasticity is that based on the statistic \(F_A\). We shall refer to this test as the asymptotic F-test. Other asymptotic alternatives do exist. The statistic \(F_A\) is equal to the Wald statistic divided by its degrees of freedom \(K\). The Wald test by itself could be used, as could the Lagrange multiplier or likelihood ratio tests. However, it has been argued that the F-test version is likely to be better in terms of the accuracy with which the actual size approximates the nominal size. See, for example, Woodland (1986).
Another alternative to $F_A$ when $\sigma_1^2 \neq \sigma_2^2$ is the procedure suggested by Weerahandi. To outline his test we note that

$$B = \frac{\hat{\sigma}_1^2 / \sigma_1^2 - \hat{\sigma}_2^2 / \sigma_2^2}{\hat{\sigma}_1^2 / \sigma_1^2 + \hat{\sigma}_2^2 / \sigma_2^2} \sim \text{beta} \left[ \frac{(T - K)}{2}, \frac{(T - K)}{2} \right]$$

and, furthermore, that $B$ is independent of $w_1$ and $w_2$. Substituting $B$ into (2.6) yields

$$F_w = 2(b_1 - b_2) \left[ \frac{\hat{\sigma}_1^2}{B} \left( X_1' X_1 \right)^{-1} + \frac{\hat{\sigma}_2^2}{1 - B} \left( X_2' X_2 \right)^{-1} \right]^{-1} (b_1 - b_2) / K$$

To examine Weerahandi’s proposal for implementing a test based on this statistic let us momentarily return to the test based on $F_A$. As is well known, to implement this test we can proceed in one of two ways. Assuming a fixed significance level of 5%, we can find the observed value of $F_A$ and reject the null hypothesis ($H_0: \beta_1 = \beta_2$) if this observed value is greater than the 5% critical value. Alternatively, we can find (under $H_0$) the probability of the test statistic $F_A$ exceeding its observed value, and reject $H_0$ if this probability ($p$-value) is less than 0.05. Both procedures are equivalent and, in a sufficiently large sample, the $p$-value has a uniform distribution in repeated sampling in the sense that $P(p < .05) = .05$. That is, a 5% significance level implies a correct null is rejected 5% of the time.

With the Weerahandi statistic $F_w$, it is not possible to calculate an observed value of the test statistic because $B$ is unknown. Thus, the procedure of rejecting $H_0$ when an observed $F_w$ exceeds some critical value is not possible. However, it is possible to replace the statistics $b_1$, $b_2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ in $F_w$ by their observed values, to form a partially observed statistic $F_w'$. Then, recognizing that $B$ has a beta distribution and that it is independent of $b_1$, $b_2$, $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$, it is possible to compute the probability of obtaining an $F_w$ greater than the partially observed statistic $F_w'$. As Weerahandi demonstrates, this probability is given by

$$p = 1 - E_B \left[ F_{\chi^2(T-K)} (F_w') \right]$$
where $F_{(K,2(T-K))}(\cdot)$ is the distribution function value from the $F$-distribution with $K$ and $2(T-K)$ degrees of freedom, and $E_B[\cdot]$ represents the expectation of this value with respect to the distribution of $B$. The integral given by this expectation can be evaluated numerically. Weerahandi recommends rejecting $H_0$ if this value is less than some prescribed value, say 0.05.

An intuitive explanation of the procedure is as follows: For each value of $B$ between 0 and 1 an observed value of $F_w$ (say $f_\omega$) is computed; then, for each $f_\omega$, we find $P[F_w > f_\omega]$. These probabilities are averaged with the beta distribution providing the weights for the averaging process. If the average probability is less than 0.05, the hypothesis is rejected. This test has intuitive appeal, but it is not a conventional test with a fixed level of significance. The $p$-value in (2.10) does not have a uniform distribution in the sense that $P[p < .05] = .05$. Thus, in repeated sampling, with a rule that says reject if $p < .05$, $H_0$ will not necessarily be rejected 5% of the time and the size of the test will not necessarily be .05.

This characteristic raises the question of whether, using conventional criteria, the Weerahandi test which involves the use of a numerical integration program is preferable to more commonly used, computationally easier asymptotic tests such as the asymptotic $F$-test. The reason for looking for alternatives to $F_A$ is, presumably, to obtain a test that is more powerful in finite samples, and that has a finite sample size corresponding more closely to the presumed size. Also, if estimation of $\beta_1$ and $\beta_2$ is the ultimate objective, then, within a decision theoretic context, the new test is useful if it leads to a pre-test estimator that is risk superior to the pre-test estimator resulting from the existing test. It is these questions that we investigate in Section 3 of the paper. In particular, we estimate the power and size of the asymptotic $F$ and Weerahandi tests as well as compare the risks of the pre-test estimators for $\beta$ that each of these tests generate. In Section 4 the sampling characteristics of two Stein-like estimators are also investigated, one based on $F_A$ and one based on an average value for $F_w$.

Some insights into the relationship between the two tests can be obtained by noting that, when $B = 0.5$, $F_A = F_w$. Furthermore, a value of $B = 0.5$ implies that $\sigma_1^2/\sigma_1^2 = \sigma_2^2/\sigma_2^2$. Thus, when $\sigma_1^2$ and $\sigma_2^2$ are both under or overestimated by the same proportion, we would expect the performance of the two tests to be similar. However, since $\sigma_1^2$ and $\sigma_2^2$ are independent, there is no reason to believe there will be any relationship between $\sigma_1^2/\sigma_1^2$ and $\sigma_2^2/\sigma_2^2$, except that, as $T$ becomes large, both ratios will approach unity. These observations suggest that the tests will be similar when $T$ is large, but that the Weerahandi test may be better in small samples since it places some weight on possible values of $B$ other than 0.05.
2.2 Estimators

The estimators that are relevant in defining the pre-test estimators are as follows: First the least squares estimator \( \hat{b} = (b_1', b_2')' \) is best unbiased when \( \beta_1 \neq \beta_2 \). Secondly, the assumption of identical location vectors but different variances is reflected by the restricted two-stage Aitken estimator

\[
(2.11) \quad \hat{g}(\hat{\Sigma}) = \left[ \frac{X_1'X_1}{\hat{\sigma}_1^2} + \frac{X_2'X_2}{\hat{\sigma}_2^2} \right] \left[ \frac{X_1'X_1}{\hat{\sigma}_1^2} \hat{b}_1 + \frac{X_2'X_2}{\hat{\sigma}_2^2} \hat{b}_2 \right]
\]

The finite sample properties of this estimator have been investigated by Taylor (1977; 1978) and Kariya (1981). The corresponding pre-test estimators that are of interest are

\[
(2.12) \quad \hat{\beta}_{AP} = I_{(0, s)}(F_A) \left[ \begin{array}{c} \hat{g} \\ \hat{g} \end{array} \right] + I_{(s, \infty)}(F_A) \left[ \begin{array}{c} \hat{b}_1 \\ \hat{b}_2 \end{array} \right]
\]

when use is made of the asymptotic \( F \)-test and

\[
(2.13) \quad \hat{\beta}_w = I_{(0, 0.05)}(P) \left[ \begin{array}{c} \hat{b}_1 \\ \hat{b}_2 \end{array} \right] + I_{(0.05, 1)}(P) \left[ \begin{array}{c} \hat{g} \\ \hat{g} \end{array} \right]
\]

that uses the Weerahandi test statistic. In (2.12) and (2.13) \( I_0(\cdot) \) is a zero-one indicator function and \( c \) is the relevant critical value.

The sampling performance of the estimators \( \delta(\hat{\beta}_{\lambda}) \) will be evaluated and compared by their risk \( \rho(\beta, \delta(\hat{\beta}_{\lambda})) = E[L(\beta, \delta(\hat{\beta}_{\lambda}))] \). In the sampling experiments to follow we use as a measure of performance the mean squared prediction error \( \rho(\beta, \delta(\hat{\beta}_{\lambda})) = E[(\delta(\hat{\beta}_{\lambda}) - \beta)'X'X(\delta(\hat{\beta}_{\lambda}) - \beta)] \).

3. SAMPLING EXPERIMENT RESULTS

To evaluate the performance of the tests and the resulting pre-test estimators we make use of Monte Carlo sampling procedures.
3.1 Sampling Experiment

The experiment was conducted using three sample sizes, \( T = 8, 20 \) and \( 40 \). In each case and in each sample we set the dimension of the location vector at \( K = 4 \) and examined the problem of estimating \( K \) means, with \( n \) replications on each mean process. Under these circumstances \( X'_1X_1 = X'_2X_2 = nI_K \). The values for \( n \) were \( n = 2, 5 \) and \( 10 \), for \( T = 8, 20 \) and \( 40 \), respectively.

The sum of the variances was kept constant throughout, \( \sigma_1^2 + \sigma_2^2 = 10 \), but three different variance ratios \( \gamma = \sigma_2^2/\sigma_1^2 = 1, 9 \) and \( 25 \) were considered. For the location vectors we set \( \beta_1 = (1,1,1,1)' \) and \( \beta_2 = a\beta_1 \), where \( a \) is a scalar that controls the extent to which \( \beta_2 \) differs from \( \beta_1 \). A large number of values of \( a \) were considered. In the reporting of the risk for the estimators and the test results, the "difference" between \( \beta_2 \) and \( \beta_1 \) was measured by the noncentrality parameter

\[
\lambda = (\beta_1 - \beta_2)' \left[ \sigma_1^2(X'_1X_1)^{-1} + \sigma_2^2(X'_2X_2)^{-1} \right]^{-1} (\beta_1 - \beta_2)
\]

For each parameter setting 500 samples were generated.

3.2 Size and Power Results

The estimated sizes of both tests for each variance ratio, \( \gamma = \sigma_2^2/\sigma_1^2 \), and each sample size \( T \), are given in Table I. A 5% significance level was used for the asymptotic \( F \)-test and a "critical p-value" of 0.05 was used for the Weerahandi test. When \( \gamma = 1 \), the asymptotic \( F \)-test is valid in small as well as large samples because of our experimental setup. Thus, we would expect all the \( F \)-test sizes for \( \gamma = 1 \) to be within a reasonable sampling error of 0.05. Such is indeed the case; with 500 samples and \( p = .05 \) the standard deviation of an estimate for \( p \) is .0097. The sizes of the asymptotic \( F \)-test are also reasonable for \( \gamma = 9 \) and 25, providing one of the larger sample sizes (\( T = 20 \) or \( 40 \)) is employed. For \( T = 8 \), however, the actual sizes (0.088 and 0.106) clearly exceed the nominal size of 0.05, indicating that the asymptotic theory is not yet relevant.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$\gamma = \sigma_2^2/\sigma_1^2 = 1$</th>
<th>$\gamma = \sigma_2^3/\sigma_1^3 = 9$</th>
<th>$\gamma = \sigma_2^2/\sigma_1^2 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T=8</td>
<td>.046</td>
<td>.012</td>
<td>.088</td>
</tr>
<tr>
<td>T=20</td>
<td>.050</td>
<td>.044</td>
<td>.058</td>
</tr>
<tr>
<td>T=40</td>
<td>.060</td>
<td>.060</td>
<td>.060</td>
</tr>
</tbody>
</table>

For the Weerahandi test we first note that its size is never greater than that of the asymptotic $F$-test. It is approximately the same as the $F$-test and approximately "correct" for the large sample size ($T = 40$), but it becomes considerably less as $T$ decreases to 8. An evaluation of the two tests on the basis of how well actual size approximates presumed size suggests there is little difference between the tests for large $T$ and that, for small $T$, the asymptotic $F$-test rejects too frequently and Weerahandi's test does not reject often enough.

With respect to the performance of the tests when the location vectors are not equal, we find that the power of the asymptotic $F$-test is always greater than that of Weerahandi's test. Furthermore, there is no evidence that the Weerahandi test would be as powerful or more powerful even if it was size corrected. When a size discrepancy between the 2 tests exists, the discrepancy first grows as the noncentrality parameter increases and then declines as both power functions approach unity. When there is no size discrepancy (e.g., at $T = 40$ and $\gamma = 25$ where both sizes are 0.056), the power of the Weerahandi test falls fractionally below that of the asymptotic $F$-test. The estimated power functions graphed in Figure 1 for $\gamma = 25$ and for $T = 40$ and 8 are typical of the results.
3.3 Risk of the Pre-test Estimators

We next turn to a comparison of the tests on the basis of the risks of the pre-test estimators \( \hat{\beta}_F \) and \( \hat{\beta}_w \) that result from using the asymptotic F-test (2.12) and Weerahandi's test (2.13). Usually, the objective of testing is to determine whether or not the samples should be pooled for estimation purposes; the risk function of the pre-test estimators is, therefore, most relevant. In Figure 2, for \( T = 20 \) and \( \gamma = 9 \), we have graphed the risks of the various estimators relative to the risk from separate least squares estimation of each equation \([b = (b_1', b_2')]\). When an estimator's risk falls below one, it is risk superior to \( b \), and the converse is true when it is greater than one.

![Figure 1. Power of the Tests](image-url)
The sampling results suggest that \( \hat{\beta}_w \) (labelled WEER pre-test) is slightly risk superior to \( \hat{\beta}_{AF} \) (labelled Asy. F pre-test) for values of \( \lambda \) less than 4 and risk inferior when \( \lambda > 4 \). These results are a reflection of the fact that the asymptotic F-test rejects \( \beta_1 = \beta_2 \) more frequently. More frequent rejection is desirable when \( \lambda \) is large because \( b \) has smaller risk than \( \hat{g} \) (labelled EGLS); less frequent rejection is desirable when \( \lambda \) is small since \( \hat{g} \) is risk superior to \( b \). A choice between \( \hat{\beta}_w \) and \( \hat{\beta}_{AF} \) is not possible without the assignment of a prior distribution to \( \lambda \), or without changing the loss function. However, it is clear that the computationally more expensive "exact" test of Weerahandi has no obvious advantages over the conventional asymptotic F-test. The Stein estimator that also appears in Figure 2 will be discussed in Section 4.

The graphs for other settings of \( \gamma \) and \( T \) yield similar conclusions, although the magnitudes of the discrepancies change. For \( T = 40 \) there is little difference between the risks for \( \hat{\beta}_w \) and \( \hat{\beta}_{AF} \), reflecting the similar performance of the two tests. When \( T = 8 \) and the asymptotic F-test rejects far more frequently, the risks for \( \hat{\beta}_w \) and \( \hat{\beta}_{AF} \) differ considerably. For example, at \( \lambda = 28.8 \) and \( \gamma = 9 \), the relative risks for \( \hat{\beta}_w \) and \( \hat{\beta}_{AF} \) are respectively 2.482 and 1.456.

### 3.4 Non-optimality of the Pre-test Estimators

It is interesting to note that the Sclove, Morris and Radhakrishnan (1972) squared error loss inadmissibility result for conventional single sample pre-test estimators also holds for the two sample case we consider. If for expository purposes we consider the Stein-rule estimator for the
orthonormal linear statistical model, the logic of the result is as follows: The Stein positive-rule estimators (Baranchik 1970)

\[ \delta_1^*(b_i) = 1 - \min \left\{ 1, \frac{c^* (T - K) \sigma_i^2}{b_i'} \right\} b_i \]

(3.2)

dominate under squared error loss the maximum likelihood estimators \( b_i \), for \( i = 1, 2 \), respectively and for \( 0 < c^* < 2(K-2)/(T-K+2) \). Therefore, replacing \( b_i \) in (2.12) by the risk superior estimators \( \delta_1^*(b_i) \), we have the modified pre-test estimator (Judge and Bock 1978, p.190)

\[
\hat{\beta}'_{AF} = I_{(\sigma^2)}(F_A) \left[ \begin{array}{c} \delta_1^*(b_1) \\ \delta_2^*(b_2) \end{array} \right] + I_{(\sigma^2)}(F_A) \left[ \begin{array}{c} \delta_1^*(b_1) \\ \delta_2^*(b_2) \end{array} \right]
\]

(3.3)

This estimator dominates the conventional pre-test estimator \( \hat{\beta}_{AF} \) and hence demonstrates the latter's inadmissibility. An idea of the characteristics of the risk functions for \( \hat{\beta}_{AF} \) and \( \hat{\beta}'_{AF} \) is provided in Figure 3 for a \( c \) value at the .05 significance level and for the settings \( c^* = (K-2)/(T-K+2) \), \( T = 20 \) and \( \gamma = 9 \). Although the differences are small, at each level of the noncentrality parameter \( \lambda \) the risk of \( \hat{\beta}_{AF} \) is equal to or greater than that of the modified pre-test estimator \( \hat{\beta}'_{AF} \).
4. AN ALTERNATIVE TO THE PRE-TEST ESTIMATOR

Pre-test estimators are discontinuous functions of the data and we have in Section 3 noted their inferior sampling performance, for statistical model (2.1). Given that, for many statistical models and loss functions, discontinuous functions of the data result in non-optimal decision rules, we consider the question as to whether smoothed or continuous estimators may offer an attractive and perhaps minimax decision rule. The Stein estimator considered in Section 3.3 is one illustration of a continuous estimator for the linear statistical model that is, under squared error loss, minimax but inadmissible (Bock 1987).

Another possibility is to make use of the asymptotic $F$ or Weerahandi test statistics to formulate an estimator that is a continuous function of the data. Although many possible continuous estimators exist for our statistical model, one attractive possibility is the following variant of the positive Stein-rule estimator (Judge and Bock 1978, p.179; Yancey, Judge, and Miyazaki 1984):
where $E \left[ F_w \right] = \int_0^1 F_w f(B) dB$, and $f(B)$ is the beta density function for $B$. The Stein-like estimator $\hat{\beta}_{AF}^+$ is a natural extension of the Stein estimator for a set of linear equations under heteroskedasticity where $\hat{g}$ is the shrinkage vector. The estimator given in (4.2) is a Stein-like estimator that uses the average value of $F_w$, the average being taken over all values of $B$, using the beta density as the weighting function.

The empirical risk function for $\hat{\beta}_{AF}^+$ for $T = 20$ and $y = 9$ is given in Figure 2 (labelled Stein). The risk for (4.2) that makes use of the Weerahandi test statistic is not included since its empirical risk function is virtually identical to that of (4.1).

The empirical risk of the Stein-like estimators (4.1) and (4.2) is considerably less than that of the pre-test estimators (2.12) and (2.13) over a large range of the $\lambda$-parameter space and is only slightly higher than the pre-test risk over a small range of the parameter space when $\lambda$ is close to zero. The restriction $\lambda_1 - \lambda_2 = 0$, provides a natural shrinkage direction for the Stein-like estimators. Although theoretical analysis of these estimators is difficult due to the complicated dependencies, it is interesting to note that in all cases considered for $y$ and $T$ the Stein-like estimators $\hat{\beta}_{AF}^+$ dominated $\beta$; this provides strong evidence that they behave in a minimax way. The empirical risks for other values of $y$ and $T$ change in predictable ways and thus yield similar conclusions. For example, for $T = 40$, at the origin when $\lambda = 0$, the relative risks of $\hat{\beta}_{AF}^+$ for $\gamma = 1, 9$ and $25$ are $0.690, 0.476$ and $0.405$, respectively. For $T = 8$, for $\lambda = 0$ and $\gamma = 1, 9$ and $25$, the relative risks for $\hat{\beta}_{AF}^+$ are $0.731, 0.556$ and $0.496$ respectively. In each case for $T$ and $\gamma$, as $\lambda$ increased the relative risks of the Stein-like estimators approached one.
5. GENERALIZATIONS

Given the estimation and test results for the two sample problem it seems reasonable to extend the methodology to a more general case where there are more than two equations and where, in addition to heteroskedasticity, there can be contemporaneous correlation between the errors in different equations. This statistical model is the so-called seemingly unrelated regressions model (Zellner 1962). It can be written as

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_M
\end{bmatrix} =
\begin{bmatrix}
  X_1 \\
  X_2 \\
  \vdots \\
  X_M
\end{bmatrix}
\begin{bmatrix}
  \beta_1 \\
  \beta_2 \\
  \vdots \\
  \beta_M
\end{bmatrix}
+ 
\begin{bmatrix}
  e_1 \\
  e_2 \\
  \vdots \\
  e_M
\end{bmatrix}
\]

or, more compactly, the \( M \) error related equations may be written as

\[
y = X\beta + e
\]

where \( E[e] = 0 \) and \( E[ee'] = \Sigma \delta_I \).

The hypothesis of interest is that which Zellner (1962) considered under the heading "testing for aggregation bias," namely

\[
H_0: \beta_1 = \beta_2 = \ldots = \beta_M
\]

When this hypothesis is false, and \( \Sigma \) is known, the generalized least squares estimator \( \hat{\beta} = CX'(\Sigma^{-1}\delta_I)X'y \), where \( C = [X'(\Sigma^{-1}\delta_I)X]^{-1} \), is best linear unbiased. Let \( R \) be an \( [(M-1)K \times MK] \) matrix constructed so that the null hypothesis (5.3) can be written as \( R\beta = 0 \). Then, assuming \( H_0 \) is true,

\[
R\hat{\beta} \sim N \left( 0, RCR' \right)
\]

and
Furthermore,  
\[
(y - X \hat{\beta})'(\Sigma^{-1} \otimes I_T)(y - X \hat{\beta}) - X_{(M - 1)K}^2
\]  
and the statistics in (5.5) and (5.6) are independent. Thus  
\[
F_2 = \frac{\hat{\beta}' R'(R \hat{C} R')^{-1} R \hat{\beta}}{(y - X \hat{\beta})'(\Sigma^{-1} \otimes I_T)(y - X \hat{\beta})/M(T - K)} \sim F_{[(M - 1)K,M(T - K)]}
\]
is a possible statistic for testing the null hypothesis if \(\Sigma\) is known. To overcome the problem of unknown \(\Sigma\) we can replace the unknown elements with the estimates \(\hat{\sigma}_y = (y_i - X_i \hat{b}_i)'(y_i - X_i \hat{b}_i)/(T-K)\). Under suitable assumptions about the limiting behavior of the explanatory variables, the new \(F_2\) will have the same asymptotic distribution (Zellner 1962). A further simplification can be made by noting that the denominator in (5.7) converges in probability to one. Indeed, if maximum likelihood likelihood estimators for \(\beta\) and \(\Sigma\) were used in the denominator (except that \((T-K)\) rather than \(T\) is used in the divisor for estimation of \(\Sigma\)), the denominator in (5.7) would be identically one. Using these results we have  
\[
F_A = \frac{\hat{\beta}' R'(R \hat{C} R')^{-1} R \hat{\beta}}{(y - X \hat{\beta})'(\hat{\Sigma}^{-1} \otimes I_T)(y - X \hat{\beta})/M(T - K)} \sim F_{[(M - 1)K,M(T - K)]}
\]
where  
\[
\hat{C} = [X'(\hat{\Sigma}^{-1} \otimes I_T)X]^{-1} \quad \text{and} \quad \hat{\beta} = \hat{C}X'(\hat{\Sigma}^{-1} \otimes I_T)y.
\]
Consequently, the statistic \(F_A\) can form the basis for specifying a testing mechanism and for specifying the corresponding pre-test estimator or a Stein-like estimator that combines \(\hat{\beta}\) and a suitably defined restricted estimator. This restricted estimator is given by  
\[
\hat{\theta} = \left[Z'(\hat{\Sigma}^{-1} \otimes I_T)Z\right]^{-1}Z'(\hat{\Sigma}^{-1} \otimes I_T)y
\]
where  
\[
Z' = (X_1', X_2', \ldots, X_M').
\]  
A corresponding positive-part Stein estimator is
If it is assumed that there is no contemporaneous correlation, each $\hat{\beta}_i$ is replaced by the least squares estimator $b_i$. Although the topic needs further investigation, our results suggest that the Stein-like estimator in (5.10) is uniformly superior to the seemingly unrelated regression estimator and better than its corresponding pre-test estimator for a large part of the parameter space.

Certainly, the pre-test estimator that uses $F_A$ has risk characteristics consistent with those of the earlier pre-test estimators (Figure 2). In any event, the estimator in (5.10) is an appealing alternative to other Stein-like estimators that have been suggested for this statistical model [see Chapter 6 of Srivastava and Giles (1987) and references therein].

6. A FINAL COMMENT

The problem of finding an exact finite sample test for the equality of two location vectors in the presence of different scale parameters is an old one. Weerahandi has suggested a novel and appealing procedure for filling this void. However, if criteria such as known finite sample size, power, and risk of pre-test estimation are regarded as important, our sampling results indicate that the Weerahandi test is no better and in many ways inferior to a more conventional, computationally easier asymptotic $F$-test. It is important to report this finding because, without it, practitioners are likely to unnecessarily adopt the more complicated approach. If proposers of new tests would provide a statistical evaluation of their tests and consequent pre-test estimators, practitioners could make better decisions about the choice of their testing and estimation techniques.

The question of whether or not to pool two or more samples of data is frequently encountered in applied work. Continuous Stein-like estimators that make use of the asymptotic $F$ statistic may offer a risk superior alternative to conventional pre-test estimators that are normally used when the pooling question arises.
7. REFERENCES


A Note on A Bayesian Estimator in an Autocorrelated Error Model. William Griffiths and Dan Dao, No. 3 - April 1979.


Bayesian Econometrics and How to Get Rid of Those Wrong Signs. William E. Griffiths, No.31 - November 1987.


