Chapter 8

EXPECTED RESISTANCE COMPARISONS OF THE

BROWN-MOOD AND FRIEDMAN TESTS

§ 8.1 Introduction

After analyzing and comparing the maximum resistance of the Brown-Mood test to the Friedman test, the next natural comparison is with the expected resistance. We saw that the Brown-Mood test was able to handle a larger amount of bad data when the test statistic was in the least favorable position with regards to a specified conclusion. We now explore the differences between the two tests on average and begin with the expected resistance to rejection (*ERR*).

§ 8.2 Expected Resistance to Rejection for the Brown-Mood Test

To restate, we can mathematically express the expected resistance to rejection as

$$ERR_{H_0} = \frac{E[M_R \mid M_R > 0]}{N} = \frac{E[M_R]}{N(1-\alpha)},$$

where M_R represents the random number of contaminants to switch an acceptance to a rejection, N denotes the total sample size and α denotes the probability of rejecting the null hypothesis given the null hypothesis is true. For the Brown-Mood test, this is not feasible to derive analytically. However, we can approximate the *ERR* for various block-treatment combinations through simulations. Again, the ranks within each block are discrete uniform random deviates on the interval [1, t], under the null hypothesis. Given that the Brown-Mood test statistic accepts the null hypothesis for a random configuration of the ranks, we then contaminate the ranks in such a way that the movement of the test statistic from the acceptance region to the rejection region is 'optimal'. As with the Friedman simulations, the *true* optimal contamination scheme for a random configuration may actually be somewhat intricate, and thus would not necessarily be captured in the algorithm. Therefore, the results from the simulations can be considered as upper bounds for the expected resistance to rejection.

The 'optimal' way of contamination is much more difficult for the Brown-Mood test than the Friedman test. This is due to the fact that the scores are a function of the ranks, and contaminating the scores in an 'optimal' way really implies contaminating the ranks in an 'optimal way' such that the Brown-Mood test statistic is increasing at the fastest rate. Obviously what is optimal for the Friedman test will not necessarily be optimal for the Brown-Mood test. After studying the Brown-Mood test statistic as a function of the scores (which are a function of the ranks) for many different block-treatment combinations, only one substantial rule for contamination stood out. For a given configuration of the ranks and scores, find the treatment with the sum of scores farthest in magnitude from the mean sum of scores. Call this treatment j. (A variance

criterion, such as the one used for the Friedman contamination scheme, does not apply here since we are working with only binary digits.) For the sake of brevity, we explain only the case where the sum of scores farthest in magnitude is larger than the mean. The scheme is analogous for when the sum of scores is below the mean, except instead of optimally increasing the score sum, optimally decrease the score sum. To begin, within treatment j find all zero scores and their corresponding blocks. Assume that there are q zero scores within treatment j, thus q blocks that are associated with them. These zero scores within treatment j imply that the associated ranks are equal to or below the median rank for that block. Thus, increasing one of those ranks to the highest possible rank will change the zero score in treatment j to a one. By doing so this changes a score of one to zero for another treatment within that same block. So, within each of the q blocks, there exists a treatment j' that will have a score change from a one to a zero after contaminating treatment j. Optimally, treatment j' will have its sum of scores *below* the mean. That is, by contaminating the sum of scores for treatment j and j' are simultaneously driven farther away from the mean. The algorithm keeps track of all treatments j' that will have their sum of scores decrease by one by contaminating a rank in treatment j, and chooses the treatment j' that has the *lowest* sum of scores, preferably such that the sum of scores is *below* the mean. In summary, *given* that the contaminant is associated with treatment j, optimally place the contaminant so it increases the test statistic the most.

The estimated expected number of contaminants necessary to force a rejection is the average of all simulated values and the estimated expected resistance to rejection is this average divided by the sample size, $b \cdot t$. For comparative purposes, we simulated

only those cases where the number of blocks is an integer multiple of the number of treatments. Figure 8.1 displays the simulated expected resistances to rejection for both the Friedman and Brown-Mood tests for $t = 6$. This plot is typical of almost all cases studied. Specifically, the expected resistance to rejection is higher for the Brown-Mood test for when there are a small number of blocks relative to the number of treatments. The *ERR* values start becoming almost indistinguishable when the number of blocks approaches as little as twice the number of treatments. The one and only exception is when $b = t = 3$. The *ERR* for both tests was done by enumeration and is *exact*. At $\alpha = 0.05$, the Friedman test has a much higher *ERR* at 0.184 than the Brown-Mood at 0.139. Tables C1-C8 in Appendix C show all *ERR* calculations for $t = 3\text{-}10$ and $b = nt$, $n = 1\text{-}5$. There do exist some cases where the simulated *ERR* for the Friedman test are higher than the Brown-Mood, but these differences are extremely small and are probably due to the Monte Carlo sampling.
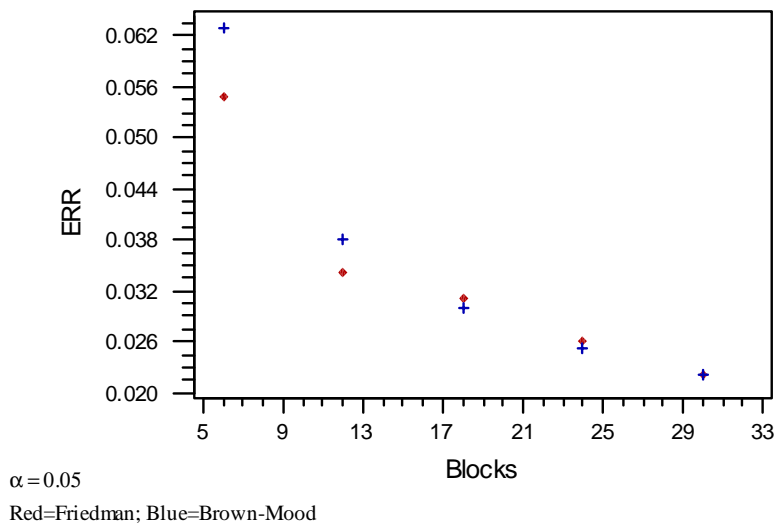

Figure 8.1. ERR of the Brown-Mood and Friedman Tests

§ 8.3 The Expected Resistance to Acceptance for the Brown-Mood Test

To restate, we can mathematically express the expected resistance to acceptance as

$$ERA_{H_A} = \frac{E[M_A \mid M_A > 0]}{N} = \frac{E[M_A]}{N(1-\beta)},$$

where M_A represent the random number of contaminants to switch a rejection to an acceptance, N denotes the total sample size and β denotes the probability of failing to reject the null hypothesis given the alternative hypothesis is true. For the Brown-Mood test, this is not feasible to derive analytically. Thus we simulate as before with the Friedman test, employing the same eight error distributions. The contamination scheme used is very similar to the scheme described in Section 8.2. The difference now is that we search for the treatment with a sum of scores farthest from the mean and bring it closer in the most 'optimal' manner. It is optimal in the sense that given we are contaminating in treatment j, the change in the test statistic is the largest. We begin by examining the four non-contaminated normal distributions and make comparisons to the Friedman test.

§ 8.3.1 Brown-Mood ERA with Uniform, Exponential, Normal and Laplace Errors

Figure 8.2 shows the *ERA* vs. ϕ for the uniform, exponential, normal, and Laplace error distributions for $t = 6$, $b = 24$, $\alpha = 0.05$, and six treatment effects. For the Brown-Mood test, these figures are quite representative of any block-treatment combination. Thus, only one combination will be shown, and the conclusions drawn are for general *b-t* combinations. As with the Friedman expected resistance to acceptance, the longer tailed distributions have an effect on the *ERA* for when $1.0 < \phi < 10.0$. That is, the *ERA*s are

slightly higher than the shorter tailed distributions. Outside of that region, the *ERA*s are almost identical, indicating no distributional differences.
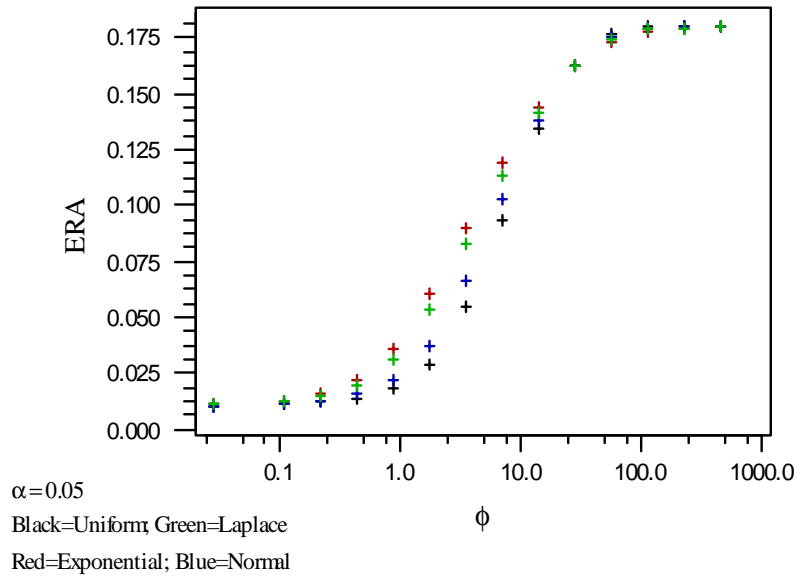

α=0.05
Black=Uniform; Green=Laplace
Red=Exponential; Blue=Normal

Figure 8.2. Brown-Mood ERA Comparisons Between Distributions

A more interesting comparison is directly against the Friedman test. We saw in Chapter 7 that the Brown-Mood test had a higher maximum resistance to acceptance at every value of *t* for a fixed *n*, and in Section 8.2 that the Brown-Mood test had a slightly higher expected resistance to rejection. However, Figure 8.3 and Figure 8.4 display sharp contrasts to the previous comparisons. Figure 8.3 shows that for all six treatment effects, on average the Friedman test has a higher resistance than the Brown-Mood for low to moderately high values of ϕ. For very large values of ϕ, the test statistic is near its maximum and we know from Chapter 7 the Brown-Mood test is more resistant than the

Friedman test if all the treatment effects are different (recalling that the maximum resistance to acceptance is achievable only when *all* treatment effects are different).
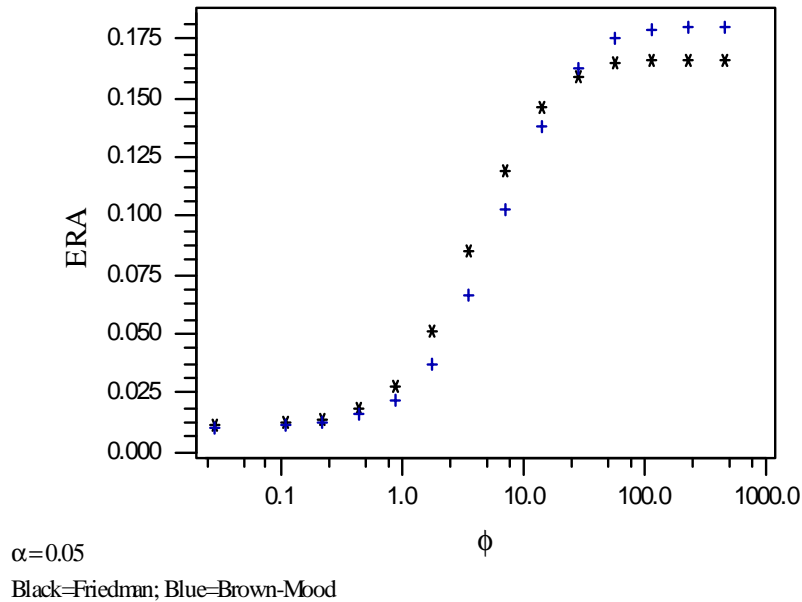

α=0.05
Black=Friedman; Blue=Brown-Mood

Figure 8.3. Brown-Mood and Friedman ERA with Six Effects

Up until now, the differences between the two tests with regards to the *MRR*, *MRA*, *ERR*, and *ERA* with *t* treatment effects, have been fairly marginal. Figure 8.4 paints a vastly different picture. Based on the simulation results, the *ERA* for the Friedman test is not only uniformly higher than the Brown-Mood test across all values of ϕ, but the discrepancy is *unbelievably large* for even moderately small values of ϕ. This can be attributed to the finer partition of the sample space for the Friedman test statistic. Since the Brown-Mood test statistic is based on the sums of binary digits, the possible values of the test statistic are much fewer in number than for the Friedman test. In other

words, there exist many combinations of ranks that will lead to the same Brown-Mood test statistic, but these same combinations will lead to *many* different Friedman test statistics. Because of this, as ϕ increases, the Friedman test statistic becomes larger (farther in the rejection region), while the Brown-Mood test statistic remains relatively unchanged.
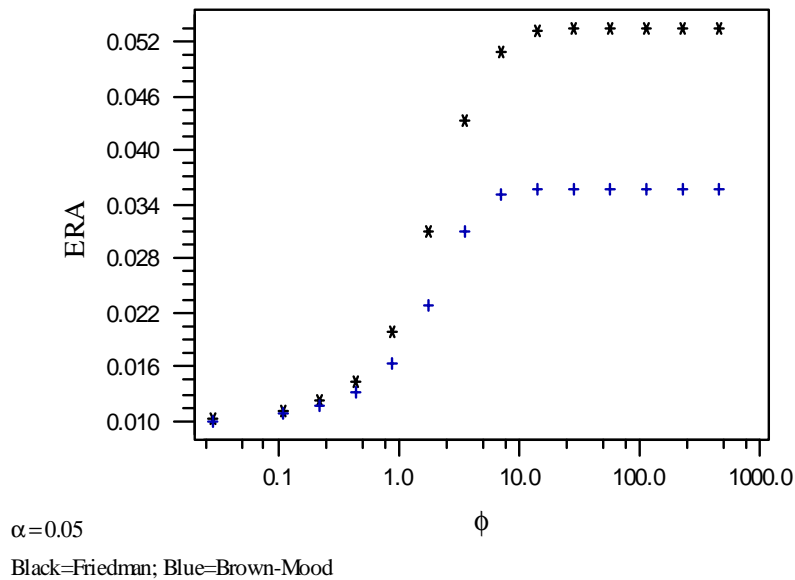

$\alpha = 0.05$

Black=Friedman; Blue=Brown-Mood

Figure 8.4. Brown-Mood and Friedman ERA with One Effect

§ 8.3.2 Brown-Mood ERA with Contaminated Normal Errors

We compare and contrast by highlighting the same six-treatment, twenty-four block case, and conclusions drawn are for general *b-t* combinations. The patterns of the Brown-Mood *ERA*s for the contaminated normal parallel those of the Friedman *ERA*s, and thus will not be shown. Therefore, we will jump directly into the comparisons

against the Friedman test and mention that all individual *ERA* calculations for the Brown-Mood test can be found in Appendix D. We highlight the contaminated normal distribution with $\varepsilon = 0.10$. Figure 8.5 shows that for all six treatment effects, the Brown-Mood test *ERA* is higher than the Friedman test *ERA*. This is consistent with the values obtained from Figure 8.3. The specific treatment effects used for this example were {3, 2, 1, -1, -2, -3}, and thus $\phi = 28$ for a distribution with variance 1 (i.e. a standard normal). From Figure 8.3, $\phi = 28$ is where the Brown-Mood test starts to become more resistant than the Friedman test. Since the contaminated normal distribution is still a standard normal with probability $(1-\varepsilon)$, it seems intuitive that the *ERA* for the Brown-Mood test is slightly higher than for the Friedman test.
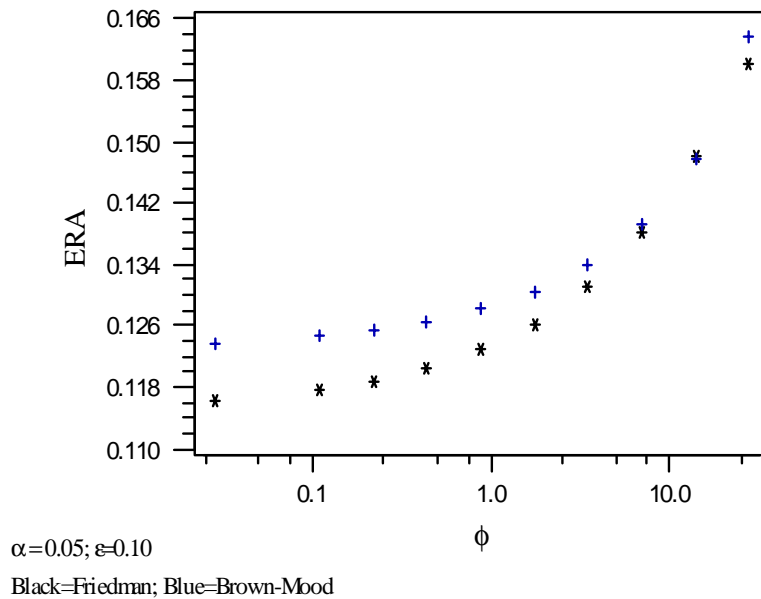
ERA

0.166
0.158
0.150
0.142
0.134
0.126
0.118
0.110

0.1   1.0   10.0

ϕ

$\alpha = 0.05$; $\varepsilon = 0.10$

Black=Friedman; Blue=Brown-Mood

Figure 8.5. Brown-Mood and Friedman ERA with Six Effects and CN Errors

Figure 8.6 displays the drastic difference in *ERA* between the two tests, similar to the differences seen in Figure 8.4. As with the case of normal errors and only one treatment effect, the Friedman test has uniformly higher *ERA*s across all values of φ and the difference is vast. This again can be attributed to the finer partition of the sample space using the Friedman test statistic. The Brown-Mood statistic cannot discriminate between what are extremely high/low ranks, moderately high/low ranks, and so on. In essence the Brown-Mood test takes a 'middle of the road' stance. It is this neutrality in the test statistic that is detrimental to the expected resistance to acceptance.
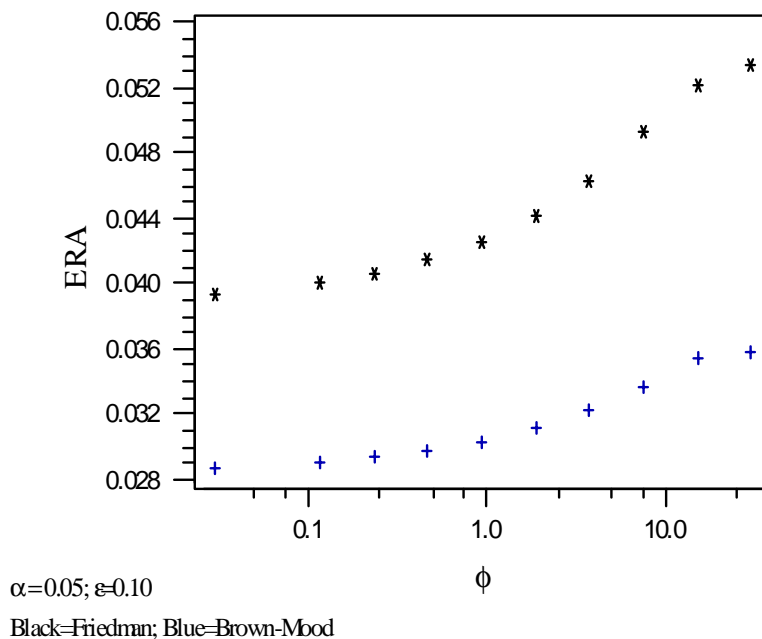

Figure 8.6. Brown-Mood and Friedman ERA with One Effect and CN Errors