Chapter 6

EXPECTED RESISTANCES OF THE FRIEDMAN TWO-WAY ANALYSIS OF VARIANCE BY RANKS

§ 6.1 Introduction

In addition to the maximum resistance, another measure of the robustness of a statistical test is the *expected resistance*. First proposed by Coakley and Hettmansperger (1992), it is a criterion of how robust a test is on average with regards to a specified conclusion. This may be a more representative measure of a test's robustness than the maximum resistance, since the test statistic will not necessarily be in the least favorable position with regards to a specified conclusion. Simply stated, it is the ratio of the average number of contaminants necessary to break down the statistical test, given it takes at least one, divided by the total sample size. We begin by looking at the expected resistance to rejection of the Friedman test.

§ 6.2 The Expected Resistance to Rejection for the Friedman Test

The expected resistance to rejection (*ERR*) shows the proportion of bad data necessary to force rejection in the average case under the null hypothesis. By letting M_R represent the random number of contaminants to switch an acceptance to a rejection, we can mathematically represent this as

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

where N denotes the total sample size and α denotes the probability of rejecting the null hypothesis given the null hypothesis is true. To evaluate this expression, we need the distribution of M_R. For the Friedman test, this is not feasible to derive analytically. However, we can approximate the *ERR* for various block-treatment combinations through simulations. Under the null hypothesis, the ranks within each block are discrete uniform random deviates on the interval [1, t]. Knowing this, we can simply randomly assign the ranks within each of the b blocks, and obtain random rank sums for each of the t treatments. Given that the Friedman 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'. The scheme would be 'optimal' based on a set of guidelines that can easily be programmed. It should be noted that the *true* optimal contamination scheme for a random configuration may actually be somewhat intricate, and thus would not necessarily be captured in the algorithm. Having said that, the results from the simulations can be considered as upper bounds for the expected resistance to rejection. However, the set of rules in the algorithm is quite substantial and we are extremely confident in the accuracy of the simulation results. There are a total of six rules for contamination:

1. Contaminate the treatment that has its corresponding rank sum the farthest away in magnitude from the mean rank sum, and drive it farther from the mean. That is, if the farthest rank sum is below the mean, start contaminating the higher ranks within that treatment and make them small. If more than one treatment rank sum is farthest away in magnitude, then choose the treatment with the highest variance in the ranks.
2. Contaminate the treatment that has the highest variance in the ranks and drive it farther from the mean. A high variance in the ranks signifies a large potential shift of the rank sum corresponding to that treatment via contamination. If more than one treatment has the highest variance in the ranks, choose the treatment that its rank sum farthest away in magnitude.
3. Contaminate the treatment with the rank sum above the mean with the highest number of ones (1's). Since the rank sum is larger than the mean, we would like to drive it in a positive direction by increasing the rank sum. This is best accomplished by changing all the lowest ranks (1) to the highest rank (*t*).
4. Contaminate the treatment with the rank sum below the mean with the highest number of *t*'s. Since the rank sum is smaller than the mean, we would like to drive it in a negative direction by decreasing the rank sum. This is best accomplished by changing all the highest ranks (*t*) to the lowest rank (1).
5. If there are multiple treatments that are identically farthest from the mean in magnitude, contaminate the treatment(s) with the most 1's and/or *t*'s. This rule is different than (3) and (4) since this rule is conditional on being farthest from the mean. Rules (3) and (4) are unconditional; that is, a treatment that

has the most extreme ranks (either 1's or *t*'s) may not necessarily be farthest from the mean.

6. Once a treatment has had all of its extreme ranks contaminated (the 1's into *t*'s or vice-versa), then recheck with rules (1) and (2) and see if the new treatments chosen via rules (1) and (2) differ from the original treatments picked at the beginning of the algorithm.

All together, the algorithm has at most eight temporary sets of ranks that are being contaminated, one each for rules (1)-(4) and two each for (5) and (6). Thus, for a random configuration of ranks, the estimated number of contaminants necessary to force rejection is the minimum of all eight possible ways of contamination in the algorithm. The expected number of contaminants is then estimated by the average of all simulated values, and the estimated expected resistance to rejection is this average divided by the sample size and (1-α). Here, α is the simulated Type I error rate, equal to the number of initial configurations that produce a test statistic larger than the critical value, divided by the number of simulations, and not the asymptotic error rate. A typical plot of the difference between the maximum resistance to rejection (*MRR*) and expected resistance to rejection is displayed in Figure 6.1, for $t = 5$ treatments and α=0.05.

For comparative purposes with the maximum resistance to rejection, we simulated only those cases where the number of blocks is an integer multiple of the number of treatments. Notice in Figure 6.1 that as the number of blocks increases the maximum and expected resistance to rejection tend to zero. This is completely logical. For a consistent test, as the sample size tends to infinity, the power for rejection approaches one; thus the resistance to rejection approaches zero. As a general observation, for t = 3-10 treatments

and *b* = 1-5 times the number of treatments, the estimated *ERR*'s are in the range of about half of the *MRR*'s for a given layout. The simulated expected resistance to rejection for each *b*-*t* combination is included with the maximum resistance to rejection for various block-treatment combinations in Appendix C, Tables C1-C8.

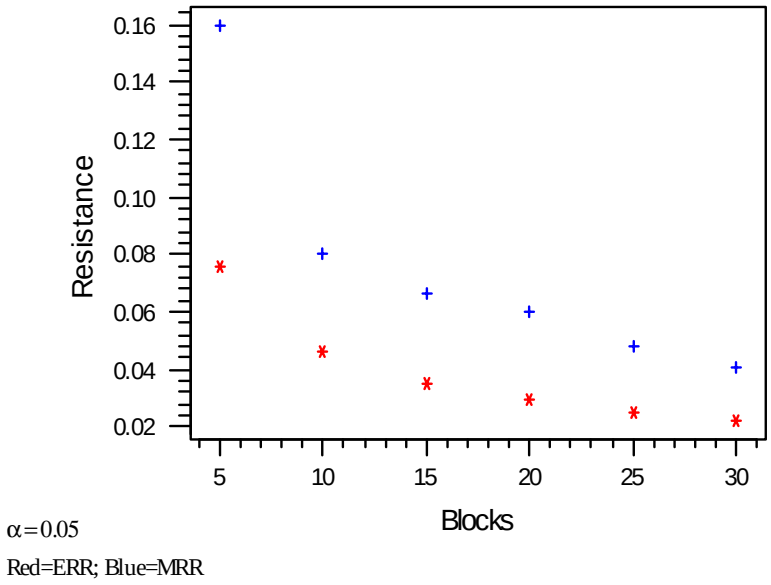

Figure 6.1. Plot of Friedman MRR and ERR (t=5)

§ 6.3 The Expected Resistance to Acceptance for the Friedman Test

The expected resistance to acceptance (*ERA*) shows the proportion of bad data necessary to force acceptance in the average case under an alternative hypothesis. By letting M_A represent the random number of contaminants necessary to switch a rejection to an acceptance, we can symbolically represent this as

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

where N denotes the total sample size and where β denotes the probability of accepting the null hypothesis given the alternative is true. Again, to evaluate this expression, we need the distribution of M_A. As with the *ERR*, we must approximate the *ERA* for various block-treatment combinations through simulations. However, within each block the ranks are *not* discrete uniform random deviates on the interval [1, t] as they were under the null hypothesis. The ranks of the observations are a function of both the fixed treatment effects and the random error term. Restating the model,

$$y_{ij} = \mu + \beta_i + \tau_j + \varepsilon_{ij},\ i = 1...b,\ \ j = 1...t$$

where y_{ij} is the observation corresponding to block i and treatment j, μ is an overall grand mean of the observations, β_i is the i^{th} block effect, τ_j is the j^{th} treatment effect, and ε_{ij} is a random error term. For simplicity in the simulations, we set $\mu = 0$. The block effects were also set to zero, since the block effects do not affect the ranks of the observations within a block. The simplified model now becomes

$$y_{ij} = \tau_j + \varepsilon_{ij},\ i = 1...b,\ \ j = 1...t\,.$$

To estimate the *ERA*, fixed treatment effects and an error distribution were specified. The treatment effects specified were 'nice' (i.e. integer form) and such that

$\sum_j \tau_j = 0$. The error distributions used were uniform [-*u*, *u*], a 'shifted' exponential (λ), normal (0, σ^2), Laplace (0, *a*), and contaminated normal $(\varepsilon, \sigma_{CN}^2)$. For the contaminated normal distribution, a random deviate has a probability of (1-ε) to come from a normal (0, 1) and a probability of ε to come from a normal $(0, \sigma_{CN}^2)$, with $\sigma_{CN}^2 \geq 1$ and ε = 0.01, 0.02, 0.05, 0.10 for our simulations. All eight error distributions have the median equal to zero. For the 'shifted' exponential, the median is calculated through the cumulative distribution function (cdf). The cdf of the exponential can be expressed in closed form as

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

Setting $F(x)$ = ½ and solving for x yields the median as

$$\dot{M} = -\frac{\ln(0.5)}{\lambda},$$

and therefore the 'shifted' exponential is

$$f(x) = \lambda e^{-\lambda(x-\dot{M})}, \; x \in [0, \infty).$$

For all simulations the parameters were chosen such that the variances of the distributions were all equal. This is easily accomplished by first choosing a nominal σ^2, then solving for the parameters of the distribution through the variance.

For the uniform [-*u*, *u*], where

$$f(x) = \frac{1}{2u}, \; x \in [-u, u],$$

the variance can be expressed as $\sigma^2 = u^2/3$, which leads to $u = \sqrt{3}\sigma$. For the exponential (λ) where,

$$f(x) = \lambda e^{-\lambda x}, \; x \in [0, \infty),$$

the variance can be expressed as $\sigma^2 = \lambda^{-2}$, which leads to $\lambda = \sigma^{-1}$. For the Laplace (0, *a*), where,

$$f(x) = \frac{1}{2a} e^{-|x|/a}, \; x \in (-\infty, \infty),$$

the variance can be expressed as $\sigma^2 = 2a^2$, which leads to $a = \sigma/\sqrt{2}$. For the normal (0, σ^2), a transformation is not necessary since we specify σ^2 at the beginning. For contaminated normal, the variance is $\sigma^2 = (1-\varepsilon) + \varepsilon\sigma_{CN}^2$, which leads to $\sigma_{CN}^2 = \left((\sigma^2 - 1)/\varepsilon\right) + 1$.

We only simulated cases where the number of blocks is an integer multiple of the number of treatments. For comparative purposes with the expected resistance, we use the noncentrality parameter ϕ, where

$$\phi = \frac{1}{\sigma^2} \sum_{i=1}^{t} \tau_j^2 .$$

A noncentrality parameter is a standardized shift in the distribution of the test statistic under the alternative hypothesis. A noncentrality parameter equal to zero indicates that the null hypothesis is true, while a noncentrality parameter greater than zero indicates the alternative hypothesis is true. It should be noted that this ratio is not quite exactly the noncentrality parameter of the chi-square distribution, the distribution of the Friedman test statistic, but viewed more as a signal-to-noise ratio. Using this parameter will enable us to compare *ERA*s across different error distributions for a given set of treatment effects, as well as across sets of treatment effects for a specific error distribution. To limit the number of simulations, we only simulated two possible cases; one where there is only one treatment effect different from the others, and one where all treatment effects

were different. As an example, with six treatments and only one treatment difference, the set of treatment effects used was {5, -1, -1, -1, -1, -1}. For six treatments and six treatment differences, the set of effects used was {3, 2, 1, -1, -2, -3}. For the simulations, the variance for a set of treatment effects ranged from $\sigma^2 = 2^{-4}$ to $\sigma^2 = 2^8$, by increments of 1 in the exponent, plus an additional $\sigma^2 = 999$ to make $\phi \approx 0$.

§ 6.3.1 Friedman ERA with Uniform, Exponential, Normal and Laplace Errors

We first examine the four distributions not associated with the contaminated normal. After looking at the differences between the error distributions with regards to the *ERA*, we see that there is very little discrepancy across most of the values of ϕ. As an example, Figure 6.2 and Figure 6.3 show the *ERA* vs. ϕ for the uniform, exponential, normal, and Laplace error distributions when $t = 6$ and $b = 24$ and $\alpha = 0.05$. (These figures are representative of any block-treatment combination. Thus, only one combination will be shown, and the conclusions drawn are for general *b*-*t*). The six treatment, twenty-four block example was chosen simply because it is a 'good size' design. Tables of all simulated *ERA*s can be found in Appendix C, Table C.11). Figure 6.2 has all six treatments different while Figure 6.3 has only one treatment difference. All of those four error distributions for both sets of treatment effects exhibit an 'S'-shape pattern (much like a statistical power curve) in the expected resistance as the noncentrality parameter, ϕ, increases (in $\log_{10}$ units). Naturally, as ϕ increases, the expected resistance to acceptance increases, since larger ϕ values indicate the father away the acceptance region is. For this example, the absolute *MRA* is $24/144 \approx 0.1667$, but it should be noted that this is *only* achievable when all six treatment effects are different

(see Figure 6.2). For the case with only one treatment difference, the *MRA* is much less at about 0.0535 (see Figure 6.3). Therefore, as ϕ approaches infinity, the *ERA* converges to a 'relative' maximum resistance to acceptance for that number of treatment differences. Figure 6.4 graphically shows this difference as ϕ increases (using normal errors). The two functions represent the extreme cases; one treatment difference and all treatments different. For cases that are in-between, the *ERA*s would also fall in-between.

Figure 6.5 shows the *ERA*s for each error distribution using this same example, with the six treatment differences. It represents each curve in Figure 6.2, now superimposed on each other. Again, there is very little discrepancy between the distributions, except possibly where $1.0 < \phi < 10.0$. Outside of this range, either the treatment effects are insignificant relative to the variance ($\phi \approx 0$) or the treatment effects are dominating the variance (ϕ 'large'), and the *ERA*s are very similar. However, in the range $1.0 < \phi < 10.0$, it appears that the longer tailed distributions (exponential, Laplace) have more effect on the resistance than the shorter tailed distributions (normal, uniform). Specifically, the *ERA* is somewhat higher for the longer tailed distributions. This is due to the longer tail distributions having a higher density of 'smaller' errors than the shorter tailed distributions. That is, the longer tailed distributions are more compact about the median of zero. Because of this, the ranks have a higher tendency to mimic the treatment effects. As a small example, consider $y_{i1} = 1 + \varepsilon_{i1}$ and $y_{21} = 2 + \varepsilon_{i2}$, where $\tau_1 = 1$ and $\tau_2 = 2$ in this case. Obviously, the rank of y_{i1} should be less than the rank of y_{i2}. But, these ranks will change if the errors are large enough in the proper direction. The shorter tailed distributions have a higher probability of this occurring, thus rearranging the initial structure of the treatment effects. This starts the test statistic closer to the acceptance

region, thus lowers the number of contaminants to force acceptance and equivalently lowers the *ERA*.

Finally, we contrast the *ERA*s across blocks for a fixed number of treatment effects and error distribution. Figure 6.6 displays the curves $t = 6$ treatments and 6 treatment effects, normal errors, and blocks of 6, 18 and 30. As we can see, for very small values of ϕ, the smaller designs have a larger *ERA*. This is due to the fact that for very small ϕ, irrespective of block size, it takes on average only a couple of contaminants to force an acceptance. The discrepancy in the *ERA*s comes once this average is standardized by the sample size. On the other hand, for large values of ϕ, the test statistic computed is starting at or near a maximum value, and this maximum value depends on b and t. Now, the critical value is fixed and independent of the number of blocks, but for those large values of ϕ, the higher b is, the farther into the rejection region the test statistic is, thus the higher the average number of contaminants necessary to force acceptance. As the number of blocks increases, this average number of contaminants increases faster, implying that even after standardizing by the sample size, the *ERA*s for larger designs are higher. We should mention that for a given number of blocks, b^*, this case will fall in-between the two functions in Figure 6.6 for which those corresponding block numbers bound b^*. For example, the functions for $b = 18$ and $b = 30$ '*sandwich*' the function for $b^* \in [19,29]$.

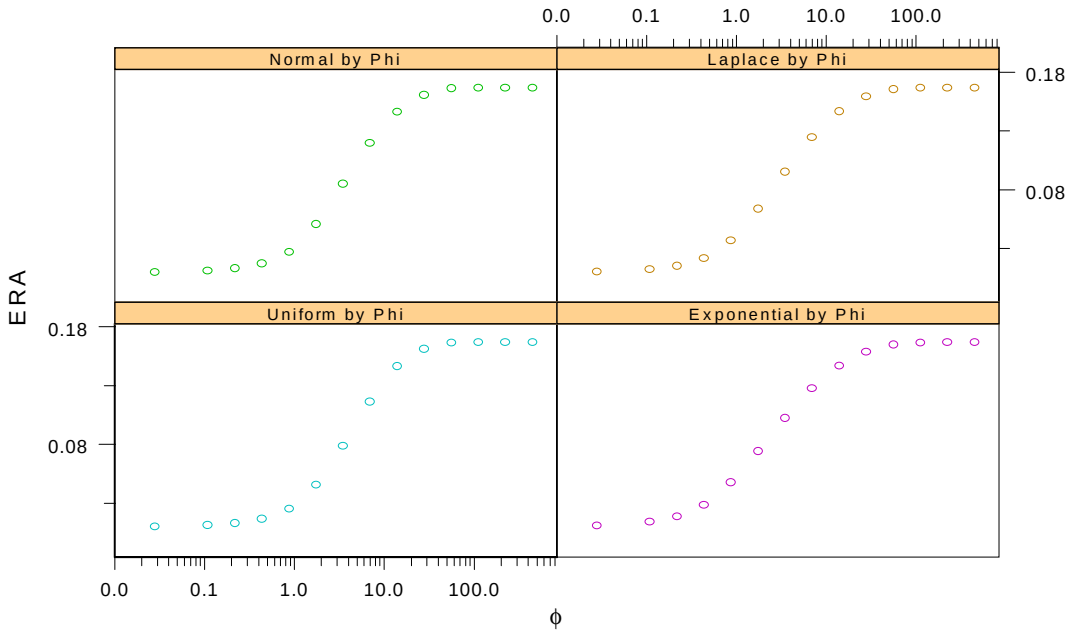

Figure 6.2. Simulated ERA with Six Treatment Effects (b = 24, t = 6)

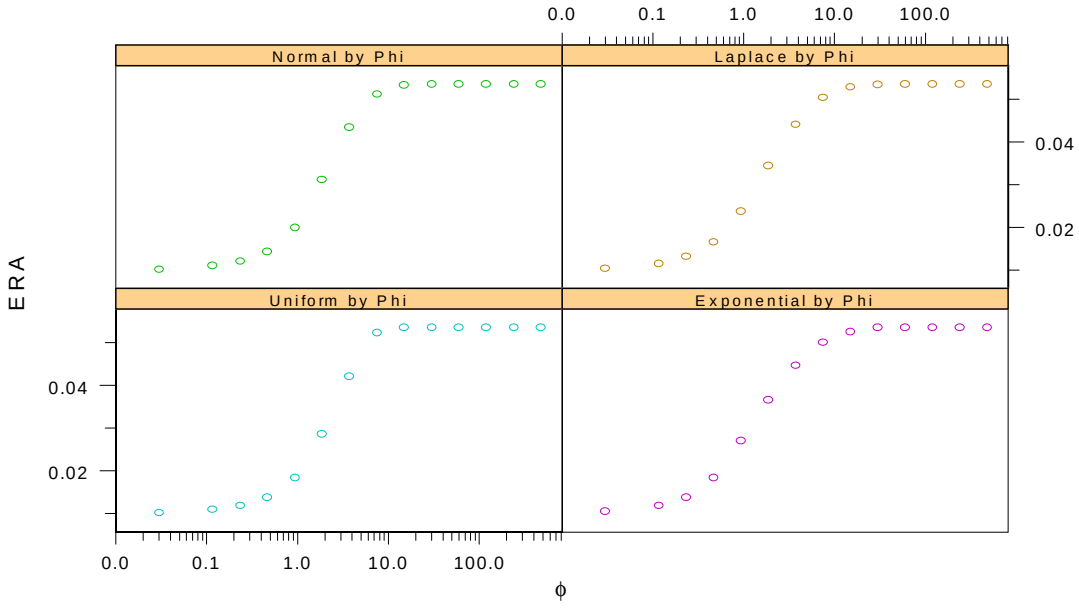

Figure 6.3. Simulated ERA with One Treatment Effect (b = 24, t = 6)

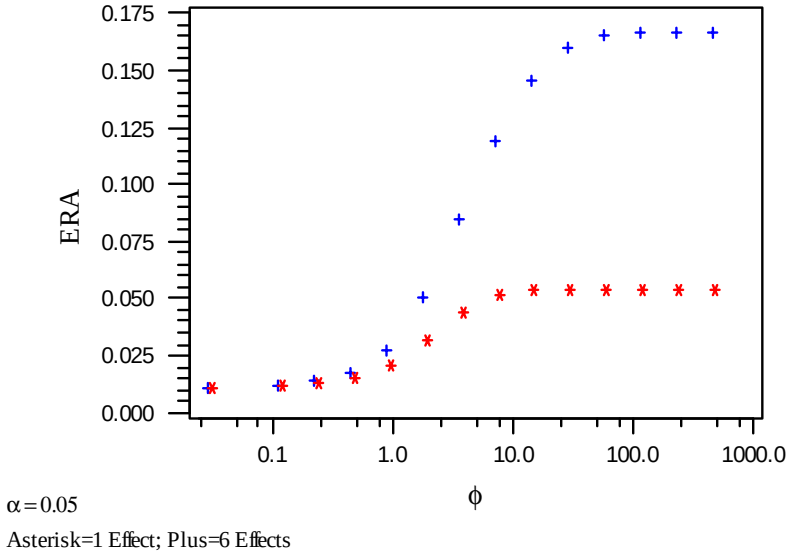

Figure 6.4. ERA Comparison Between One and Six Effects (b=24, t=6)

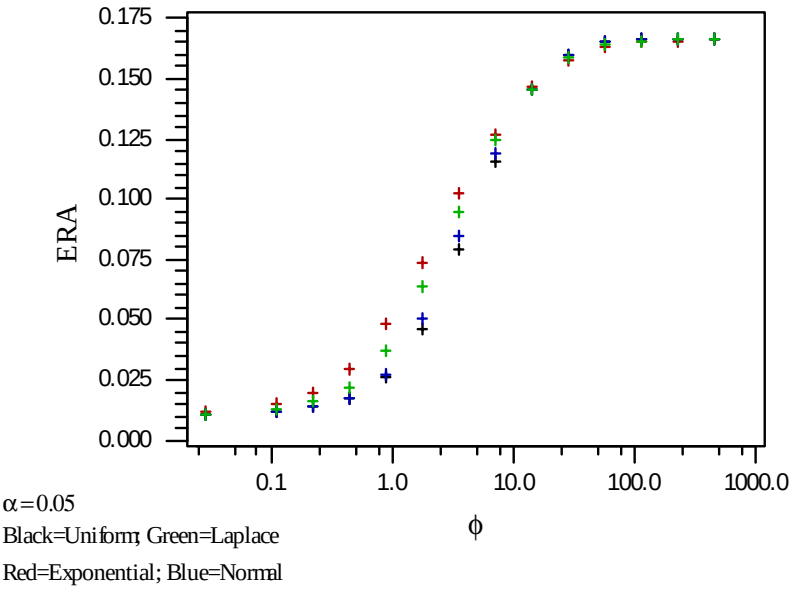

Figure 6.5. ERA Comparisons Between Distributions with Six Effects (b=24, t=6)

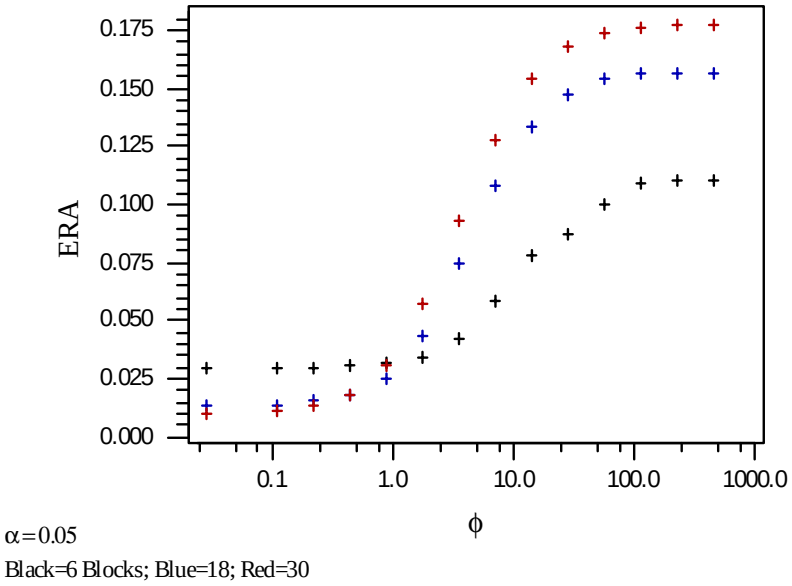

Figure 6.6. ERA Comparison Between 6, 18 and 30 Blocks (t=6)

§ 6.3.2 Friedman ERA with Contaminated Normal Errors

We now examine the effect on the *ERA* with contaminated normal errors. As mentioned previously, the contaminated normal distribution is a hybrid of two normal distributions and for this study, ε = 0.01, 0.02, 0.05, 0.10 and $\sigma_{CN}^2 \geq 1$. We compare by highlighting the same six-treatment, twenty-four block case (and conclusions drawn are for general *b*-*t*). In Figure 6.7 we can see that as ϕ decreases, the *ERA*s of each contaminated normal decrease, as we expect, but as ϕ approaches zero, the *ERA*s for each contaminated normal appear to stabilize at a much higher value than for any of the other four error distributions. Specifically, for ε = 0.01, the *ERA* stabilizes at 0.156, while for ε = 0.10, the *ERA* stabilizes at 0.116. This makes sense since the contaminated normal is

still a standard normal (1-ε)100% of the time. So, even while σ_{CN}^2 is adjusted to equate the variance of the contaminated normal to σ^2 (and for large values of σ^2, σ_{CN}^2 can become extremely large), the small values of ε keep the *ERA*s similar to the *ERA* with normal errors and $\sigma^2 = 1$. Notice also that at the highest value of ϕ, the *ERA*s for each contaminated normal distribution are identical. This is because at that value of ϕ, σ_{CN}^2 = 1 for each contaminated normal, implying each contaminated normal is just a standard normal.

Figure 6.8 shows the difference for the six-treatment, twenty-four-block case between one treatment effect and six treatment effects. Unlike in Figure 6.4, the discrepancy in the *ERA*s is across *all* values of ϕ, not just at higher values. With six treatment effects, the *ERA*s are much higher that those with only one treatment effect. Also, when the number of treatment effects is in-between these two extreme cases, then that function will be bounded by the functions in Figure 6.8.

Finally, we contrast the *ERA*s across blocks for a fixed number of treatment effects and error distribution. Figure 6.9 displays the curves t = 6 treatments and 6 treatment effects, contaminated normal errors with ε = 0.10, and blocks of 6, 18 and 30. As one can see, the *ERA* functions are extremely similar in shape, with the only difference being the number of blocks. Unlike with the other four error distributions, the *ERA*s for a larger number of blocks are uniformly higher across all values of ϕ. Again, this is reflecting that the contaminated normal distribution is still mostly a standard normal distribution, even at extremely small values of ϕ, and the difference in *ERA*s is through the difference in the number of blocks.

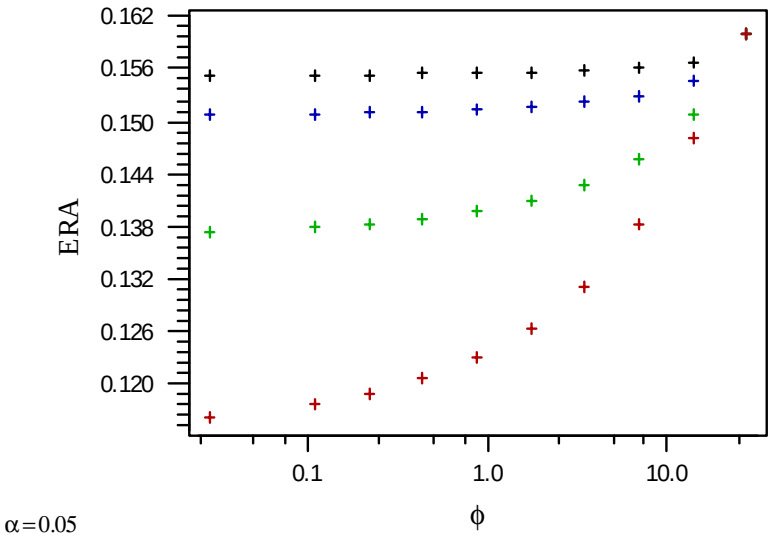

Figure 6.7. ERA Comparisons Between Contaminated Normal Errors

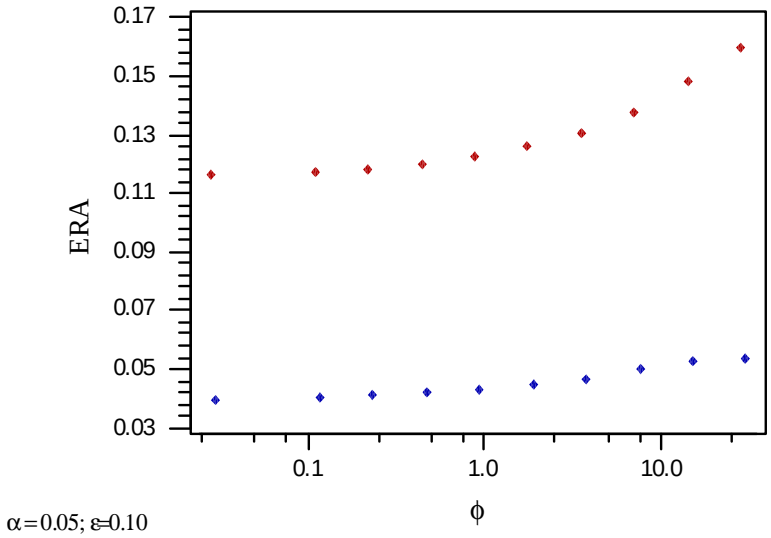

Figure 6.8. ERA Comparison Between One and Six Effects with CN Errors

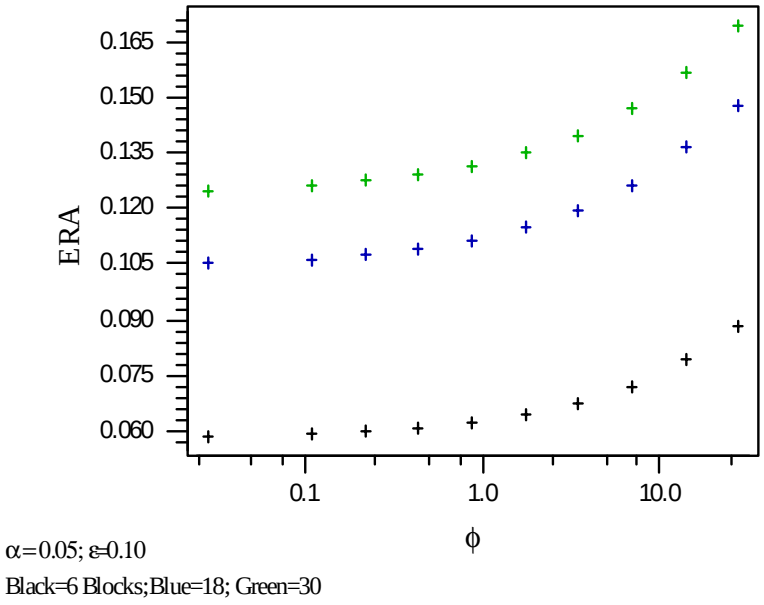

Figure 6.9. ERA Comparisons Between 6, 18 and 30 Blocks with CN Errors