KARL PEARSON'S META-ANALYSIS REVISITED: SUPPLEMENTARY REPORT

By

Art B. Owen

Technical Report No. 2009-6
May 2009

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

KARL PEARSON'S META-ANALYSIS REVISITED: SUPPLEMENTARY REPORT

By

Art B. Owen
Department of Statistics
Stanford University

Technical Report No. 2009-6
May 2009

This research was supported in part by National Science Foundation grant DMS 0604939.

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://statistics.stanford.edu

Karl Pearson's meta-analysis revisited: supplementary report

Art B. Owen*

Department of Statistics
Sequoia Hall
390 Serra Mall
Stanford CA 94305
e-mail: owen@stat.stanford.edu
url: http://stat.stanford.edu/∼owen

Abstract: This technical report contains the contents of a web supplement for the article "Karl Pearson's meta-analysis revisited" which will appear in the Annals of Statistics, Owen (2009).

1. Introduction

This technical report is a supplementary document to provide a more permanent version of a web supplement to Owen (2009).

The main paper revisits a meta-analysis method proposed by Pearson (1934) and first used by David (1934). It was thought to be inadmissible, for over fifty years, dating back to a paper of Birnbaum (1954). It turns out that the method Birnbaum analyzed is not the one that Pearson proposed. The article shows that Pearson's proposal is admissible. Because it is admissible it has better power than the standard test of Fisher (1932) at some alternatives. It has worse power at others, because Fisher's test is admissible too. Pearson's method has the advantage when all or most of the nonzero parameters share the same sign. Pearson's test has proved useful in a genomic setting, screening for age-related genes. Owen (2009) also presents an FFT based method for getting hard upper and lower bounds on the CDF of a sum of nonnegative random variables.

The figures in this supplement compare the power attained by Pearson's method to that of Fisher's method, the method of Stouffer et al. (1949), and the likelihood ratio test. For a precise description of these tests see the main article.

*This work was supported by grant DMS-0604939 from the U.S. National Science Foundation.

2. Power curves

The figures show the power of a test of the hypothesis $H_0 : \beta = (\beta_1, \ldots, \beta_m) = 0$ versus the alternative $H_A : \beta \neq 0$, using various test statistics.

An alternative is concordant if all the nonzero β_j have the same sign, either positive or negative. In the context of the main paper, the problem was to screen thousands of hypothesis tests. Concordant alternatives were considered more credible and/or more scientifically interesting and so it was of interest to have greater power for them and worthwhile to trade off less power for non-concordant alternatives.

Alternatives that are nearly concordant, because the nonzero components are overwhelmingly of one sign or the other are also of interest. In the scientific context an alternative with say 7 positive signs and 1 negative for $m = 12$ would be of interest.

Because the application is screening and because mildly non-concordant alternatives are worth detecting it is not appropriate to define the alternative as the union of two orthants. Instead we look for tests with power against all alternatives, but greater power against concordant ones. Pearson's test does this, so it has the right qualitative behavior. Left open for future work is a quantitative answer specifying exactly how much more valuable power is against alternatives in one orthant versus another. While Pearson's test lacks this quantitative precision it has an enormous simplicity.

3. Power versus concordant alternatives

Here is a description of the concordant alternatives investigated. The notation is as follows:

m	The dimension of β
k	The number of nonzero β_j
Δ	The common value for nonzero β_j
α	The level of the test
p	The power of a central $\chi^2_{(m)}$ test

In this test setting the alternative is $\beta = (\Delta, \ldots, \Delta, 0, \ldots, 0)$. Of the m components, k are nonzero. The common value of $\Delta = \Delta_k$ is chosen to give a prescribed power p (either 80% or 50%) for a chisquared test of level α (either 0.05 or 0.01) based on $\sum_{j=1}^{k} \hat{\beta}_j^2$.

4. Power versus nearly concordant alternatives

In this test setting the alternative is $\beta = (-\Delta, \Delta, \dots, \Delta, 0, \dots, 0)$. Of the m components, k are nonzero. One of them is negative and the other $k-1$ are positive. The common value of $\Delta = \Delta_k$ is chosen to give a prescribed power p (either 80% or 50%) for a chisquared test of level α (either 0.05 or 0.01) based on $\sum_{j=1}^k \hat\beta_j^2$.

5. Power curves

The power was computed using an FFT which gives an interval of finite width certain to contain the answer. It was also computed using a Monte Carlo, yielding a 99% confidence interval. The narrower interval was always chosen. Upper and lower limits are both plotted. There is usually no visible difference between them but in a few cases we can perceive two lines.

The blue line is a likelihood ratio test. It usually has the most power when there are only a few nonzeros. The red line is Pearson's test. It usually has the most power when there are many nonzeros. The black line is a Stouffer test. It usually has the most power when all or almost all of the coefficients are nonzoro. The green line is another Stouffer test that never has the most power. For details see the original article.

References

A. Birnbaum. Combining independent tests of significance. *Journal of the American Statistical Association*, 49(267):559–574, 1954.

F. N. David. On the P_{λ_n} test for randomness: remarks, futher illustration, and table of P_{λ_n} for given values of $-\log_{10}(\lambda_n)$. *Biometrika*, 26(1/2):1–11, 1934.

R. A. Fisher. *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, 4 edition, 1932.

A. B. Owen. Karl Pearson's meta-analysis revisited. *Annals of Statistics*, 37:to appear, 2009.

K. Pearson. On a new method of deternining "goodness of fit". *Biometrika*, 26(4):425–442, 1934.

S. A. Stouffer, E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. Williams Jr. *The American soldier, Volume I. Adjustment during Army life.* Princeton University Press, Princeton NJ, 1949.

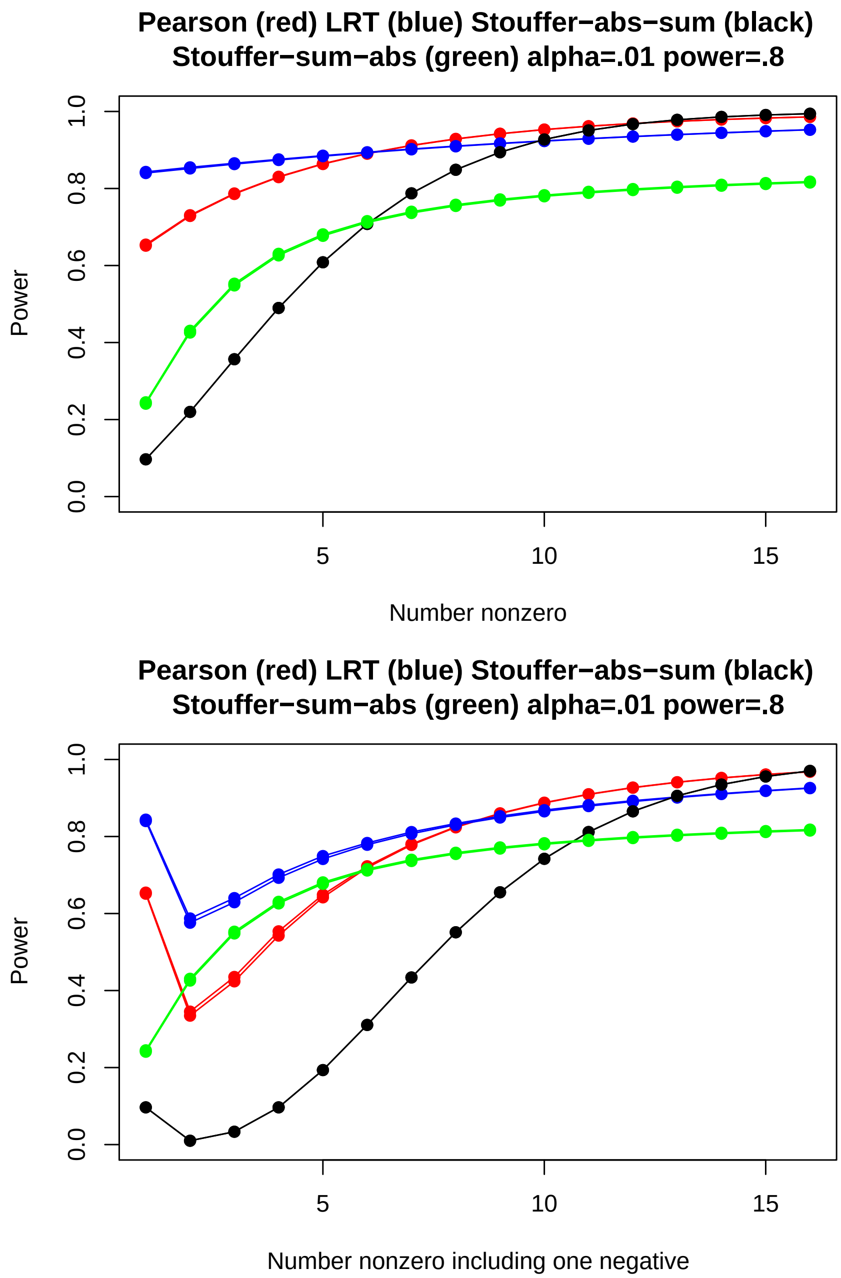
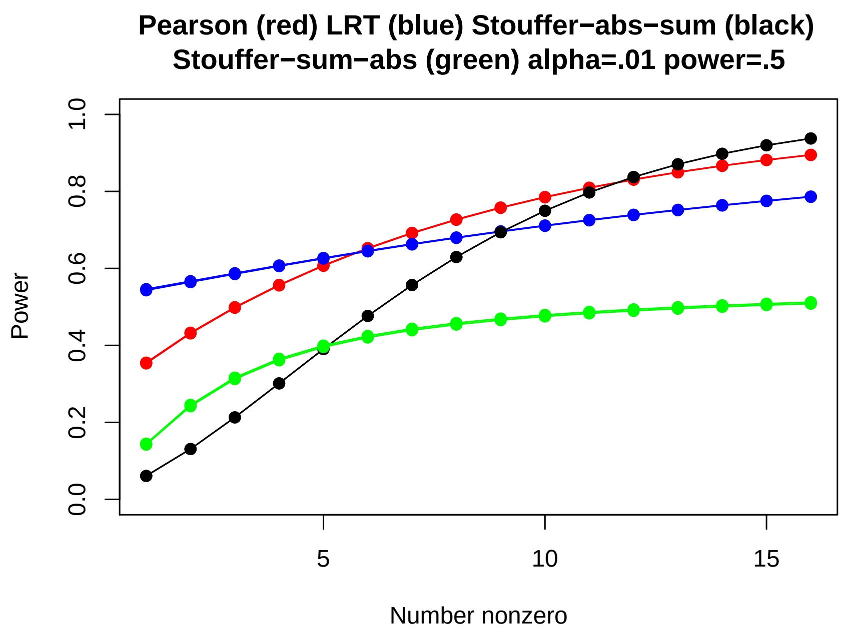


FIG 1. *The top figure shows power curves for $m = 16$ with $\alpha = 0.01$ and Δ chosen so that a chisquare test has power* 0.8. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



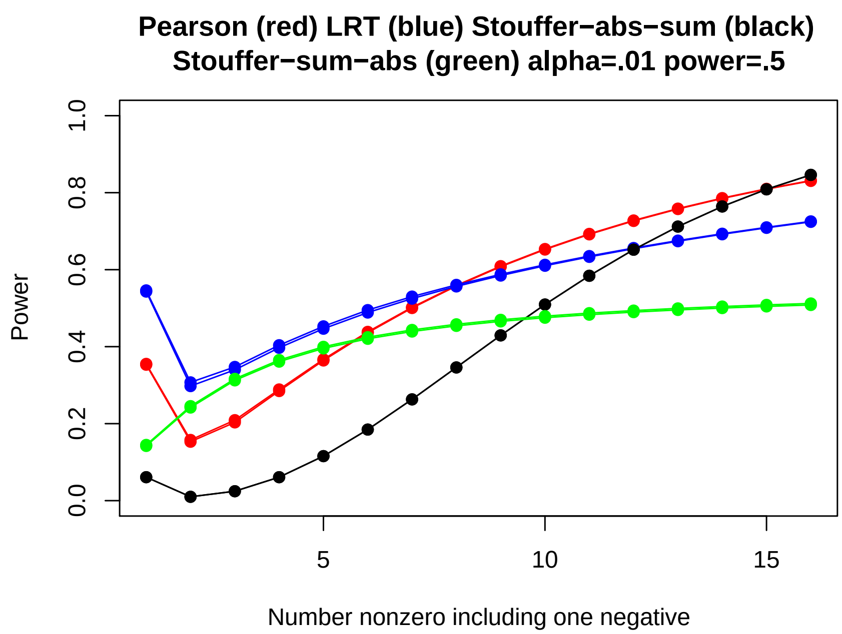


FIG 2. *The top figure shows power curves for* $m = 16$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.5. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

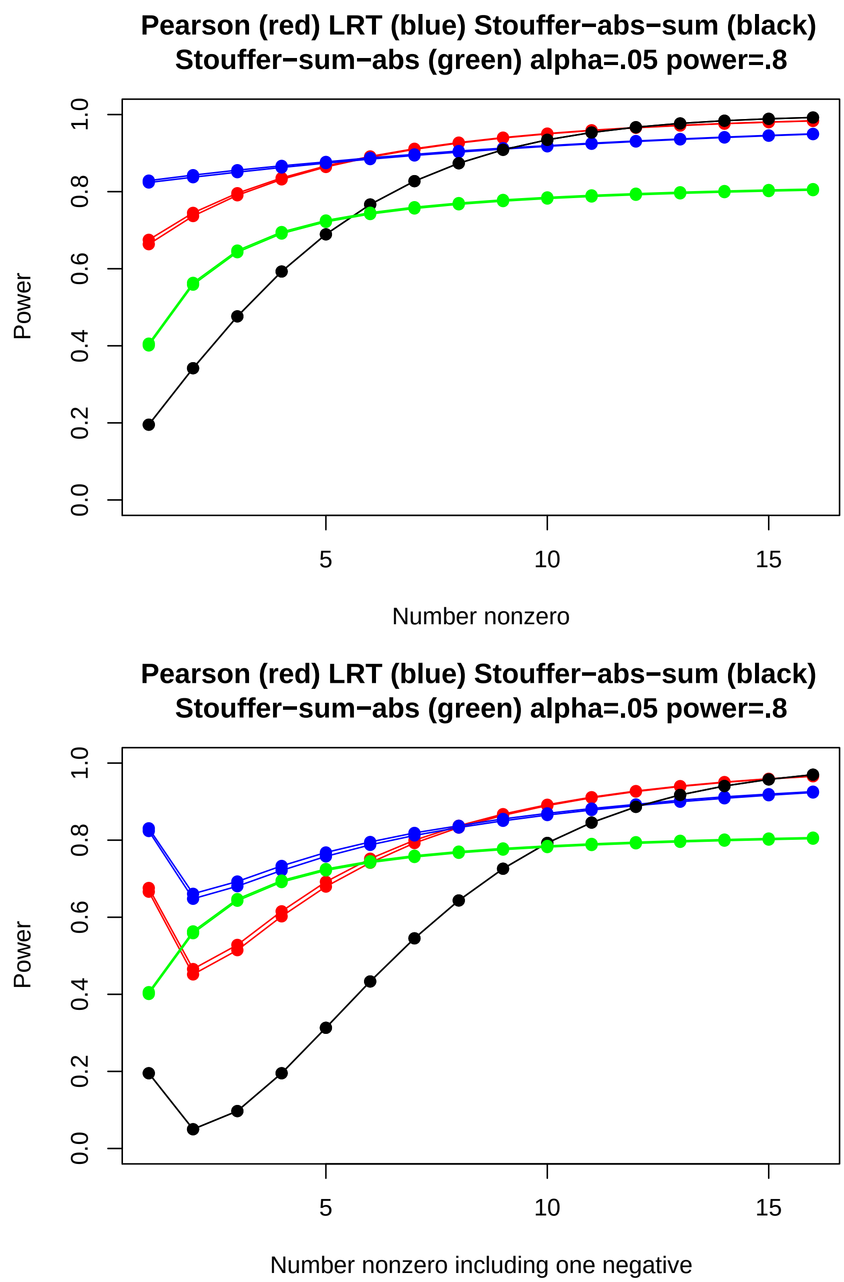


FIG 3. *The top figure shows power curves for* $m = 16$ *with* $\alpha = 0.05$ *and* Δ *chosen so that a chisquare test has power* 0.8*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

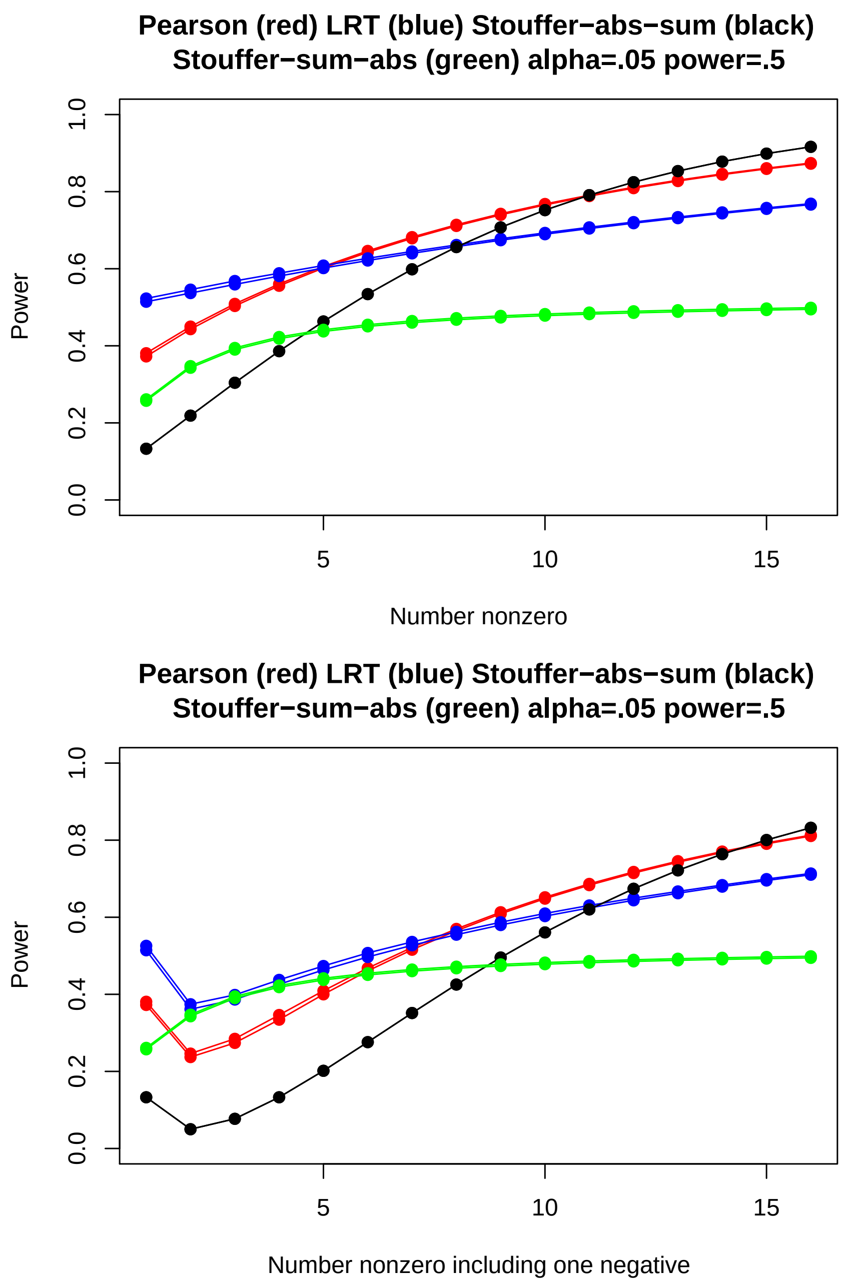


FIG 4. *The top figure shows power curves for* $m = 16$ *with* $\alpha = 0.05$ *and* Δ *chosen so that a chisquare test has power* 0.5. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

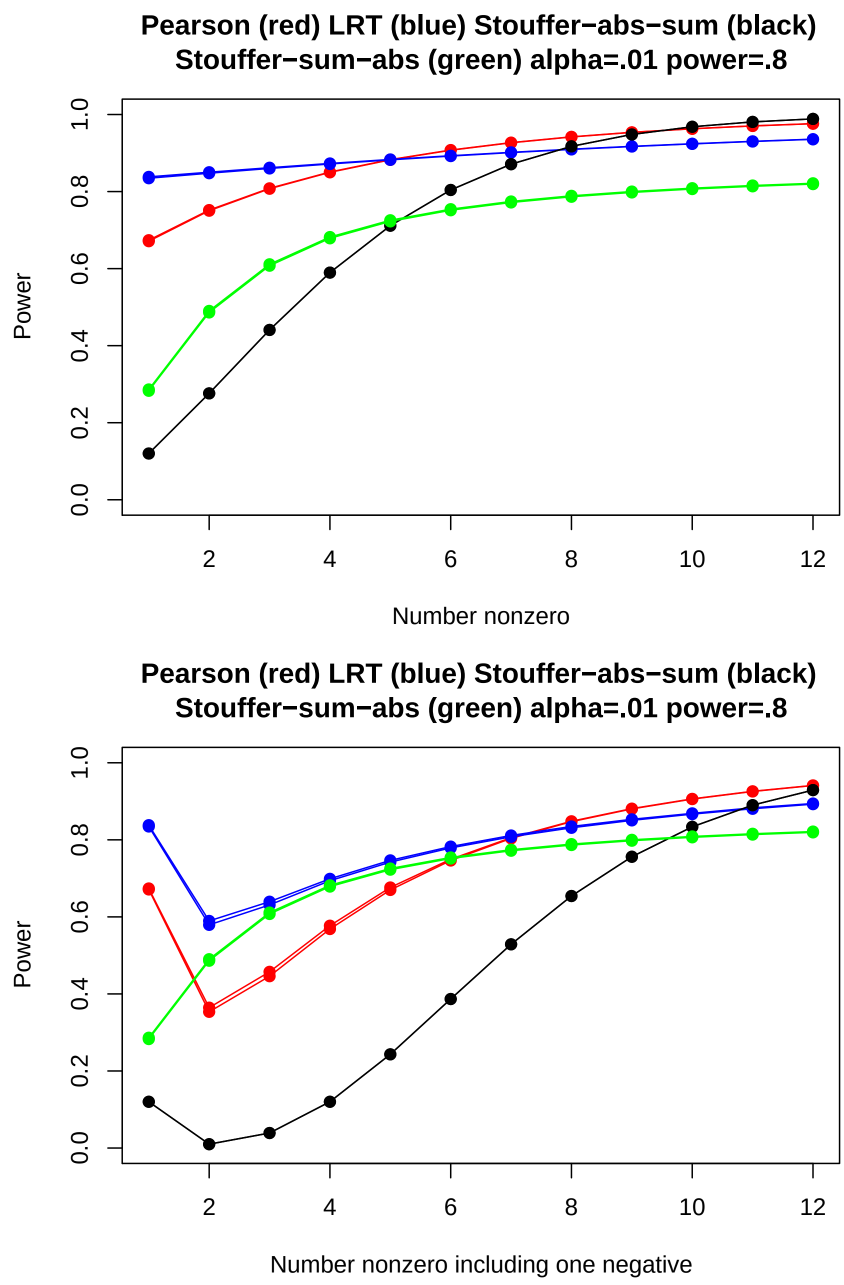
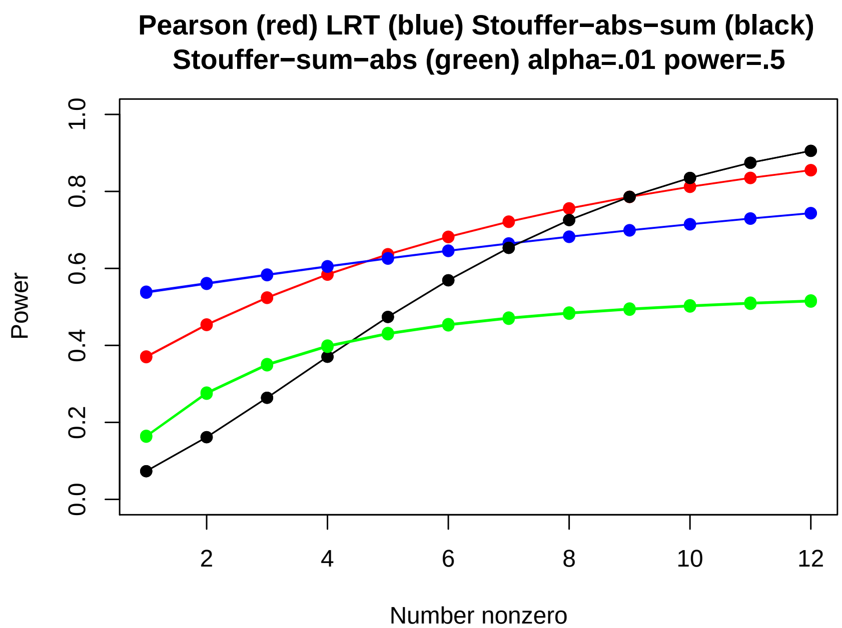


FIG 5. *The top figure shows power curves for* $m = 12$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.8. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



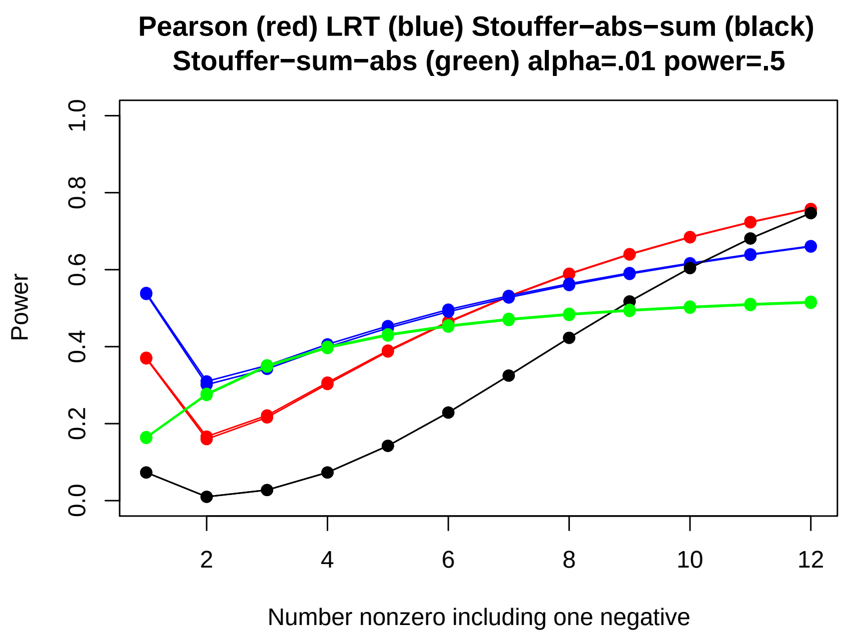
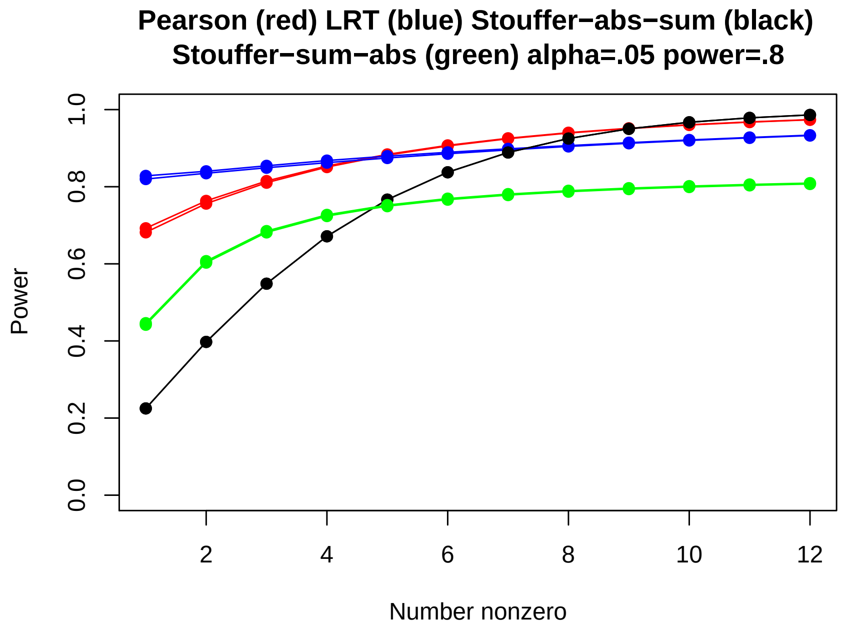


FIG 6. *The top figure shows power curves for* $m = 12$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.5*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



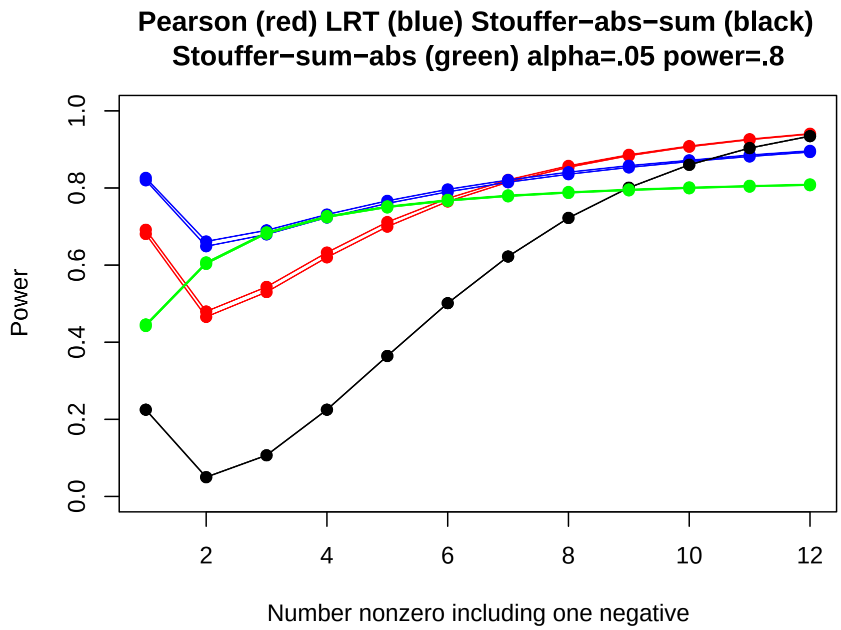


FIG 7. *The top figure shows power curves for* $m = 12$ *with* $\alpha = 0.05$ *and* Δ *chosen so that a chisquare test has power* 0.8. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

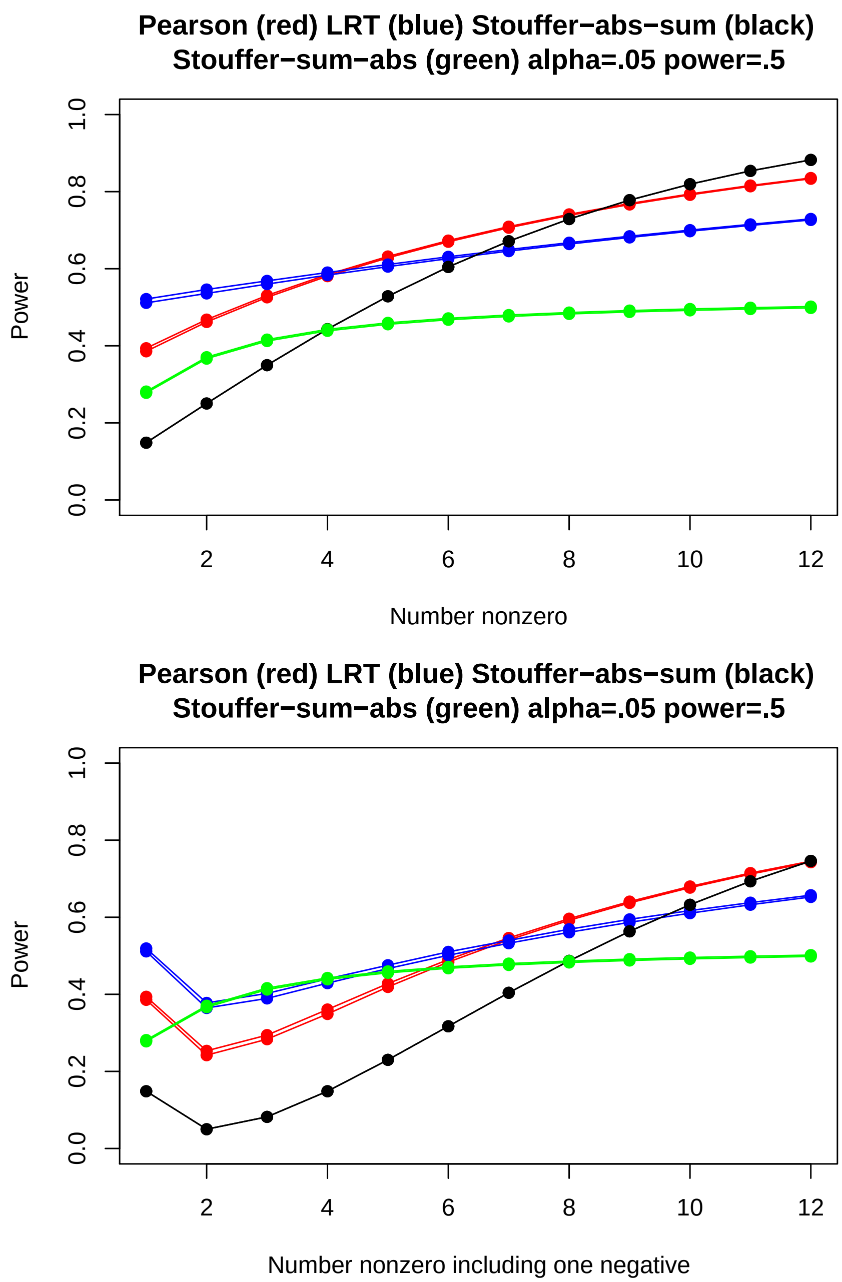
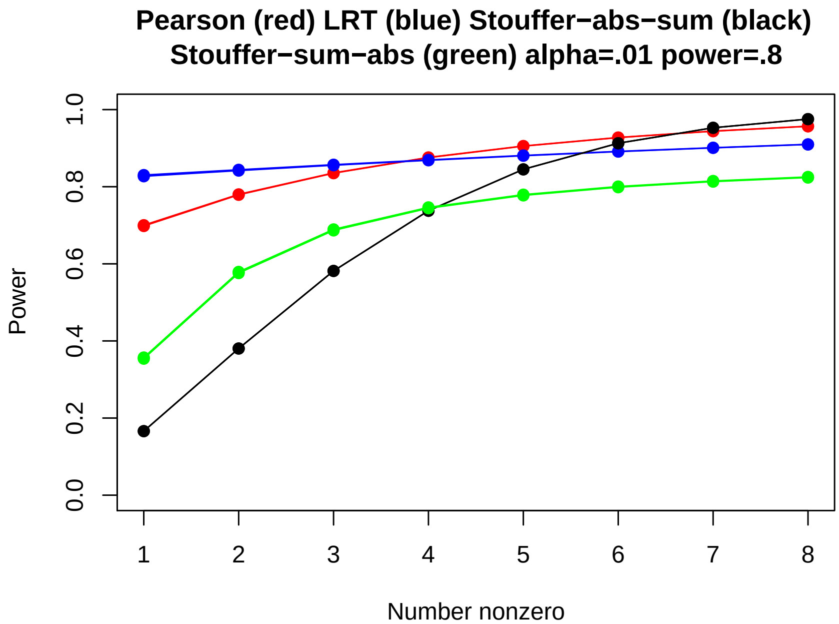


FIG 8. *The top figure shows power curves for $m = 12$ with $\alpha = 0.05$ and Δ chosen so that a chisquare test has power* 0.5*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



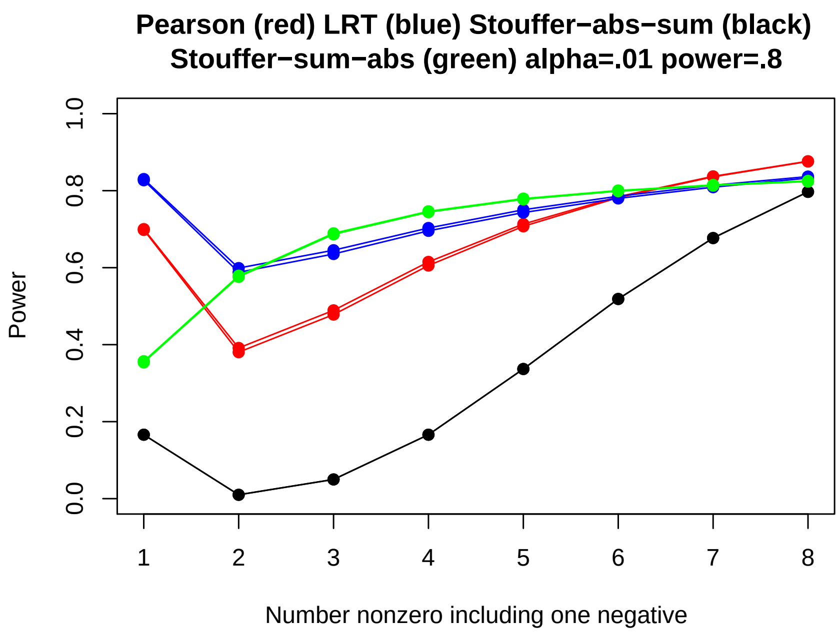
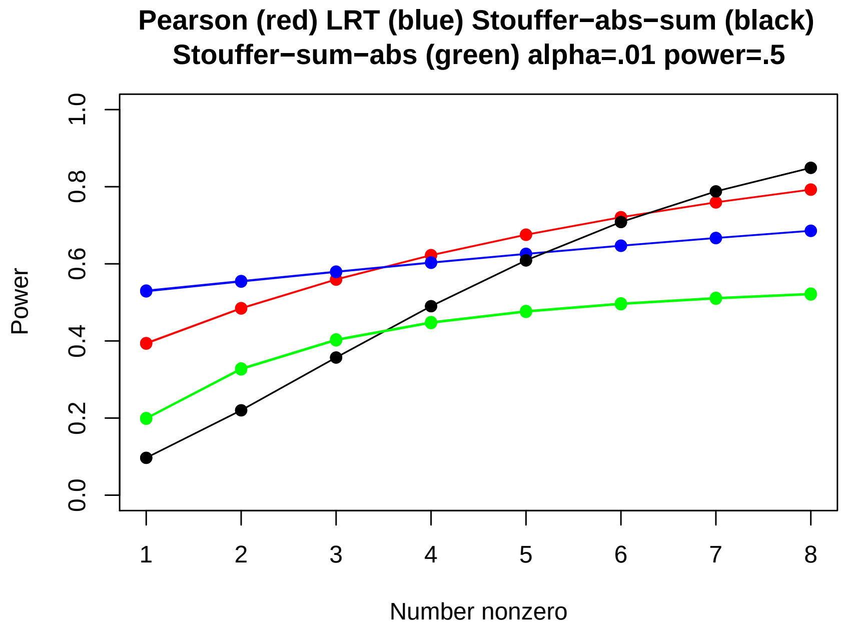


FIG 9. *The top figure shows power curves for $m = 8$ with $\alpha = 0.01$ and Δ chosen so that a chisquare test has power* 0.8. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



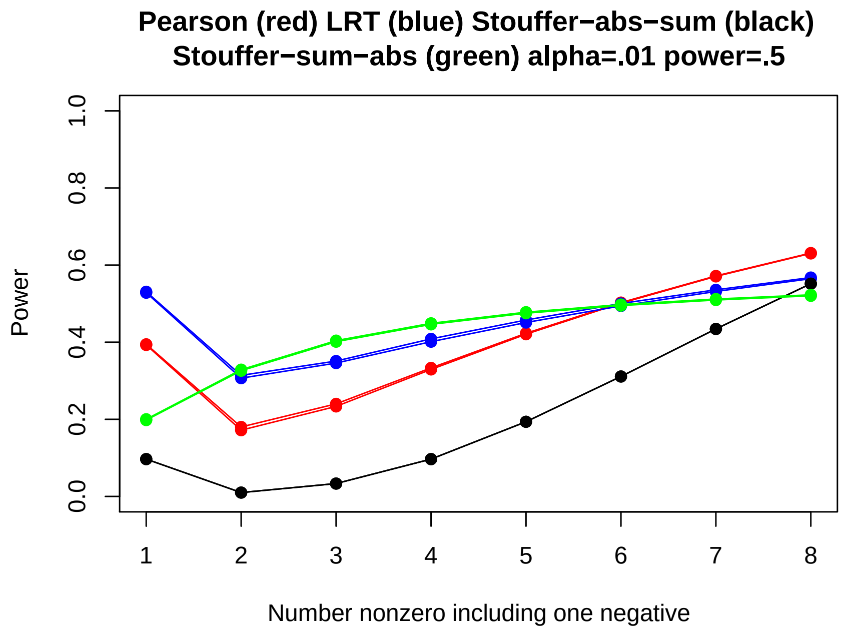
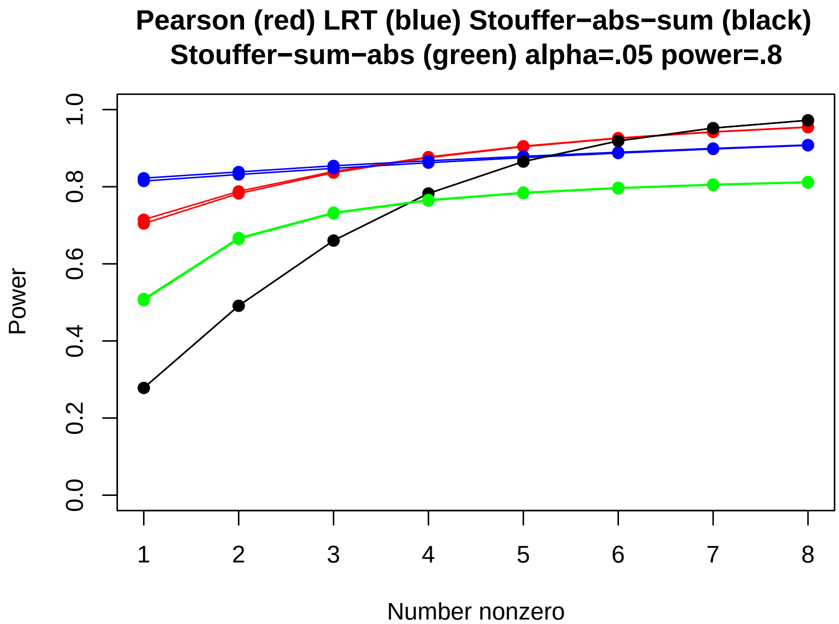


FIG 10. *The top figure shows power curves for* $m = 8$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.5*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



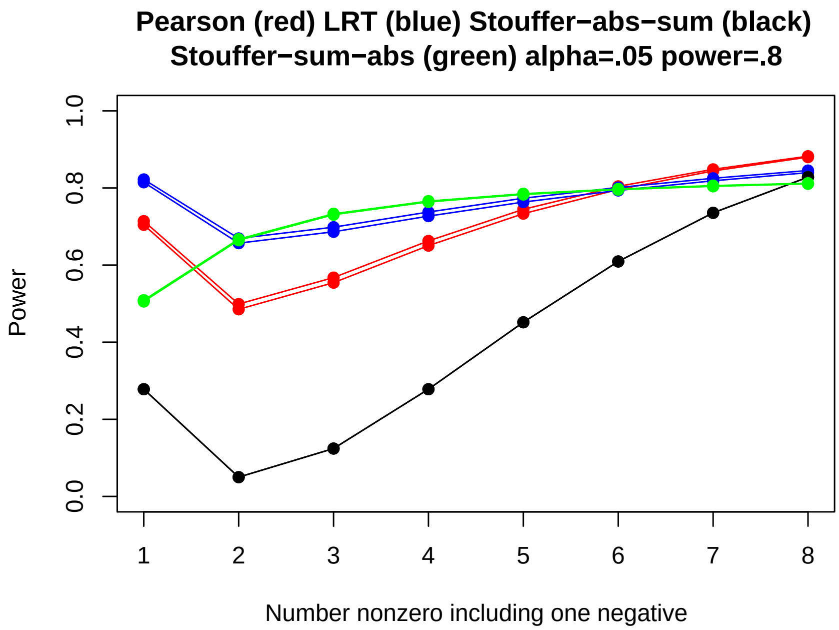


FIG 11. *The top figure shows power curves for* $m = 8$ *with* $\alpha = 0.05$ *and* Δ *chosen so that a chisquare test has power* 0.8. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

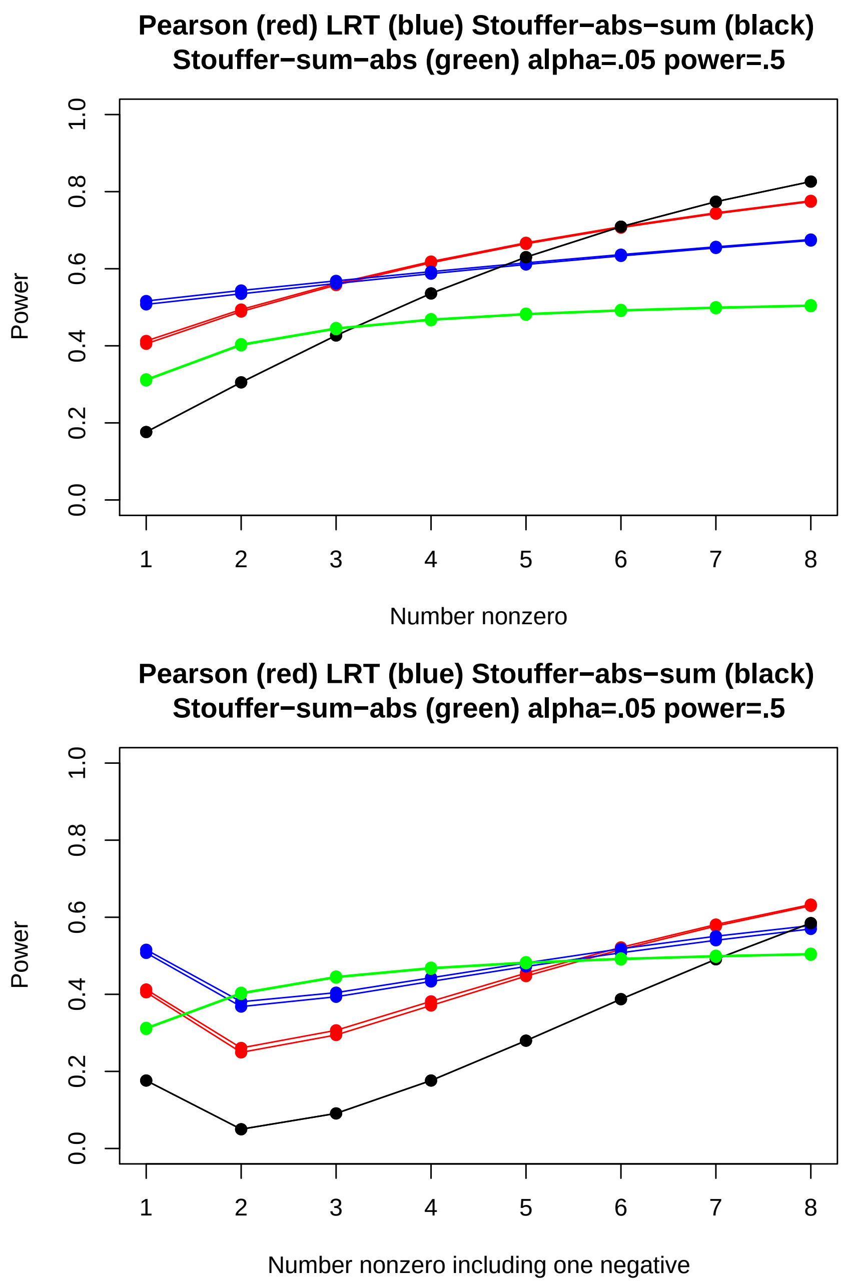
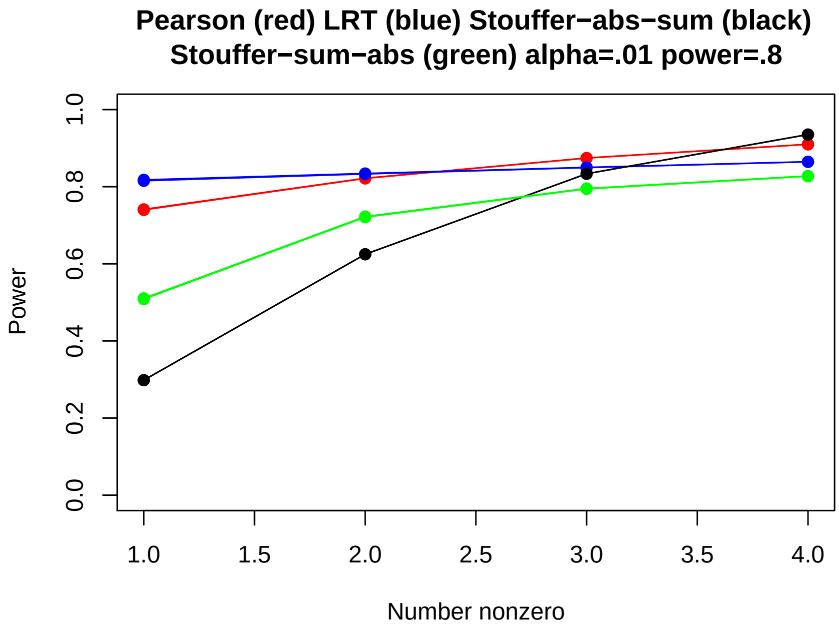


FIG 12. *The top figure shows power curves for* $m = 8$ *with* $\alpha = 0.05$ *and* Δ *chosen so that a chisquare test has power* 0.5. *The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



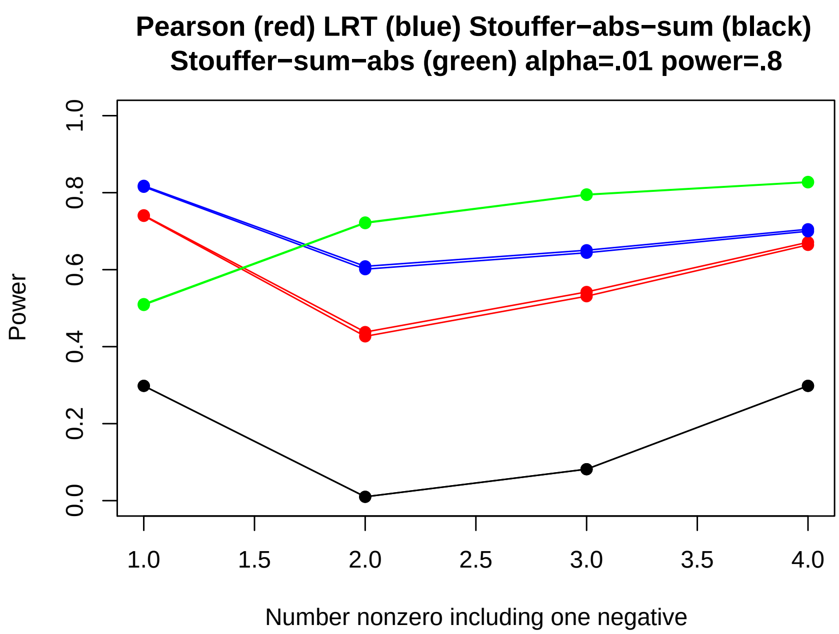


FIG 13. *The top figure shows power curves for* $m = 4$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.8*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*

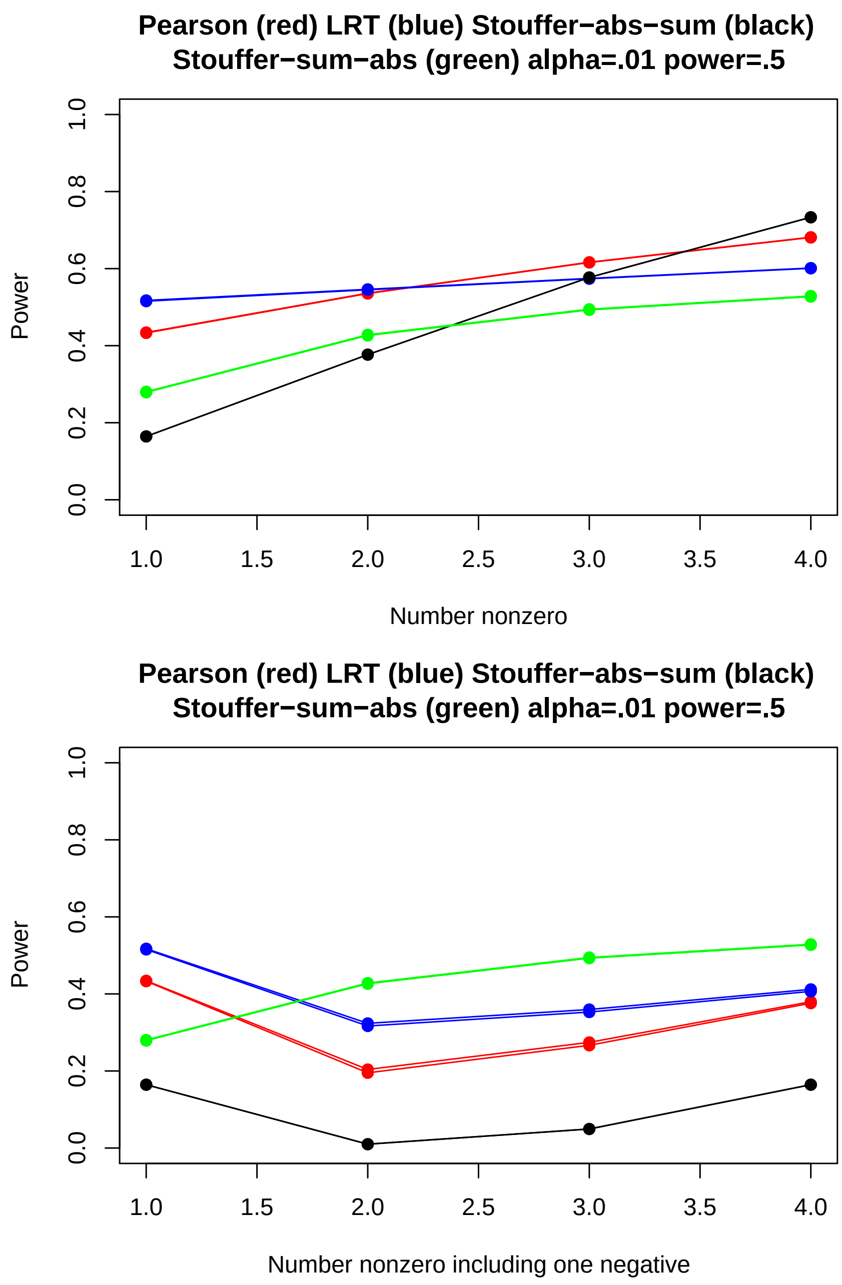
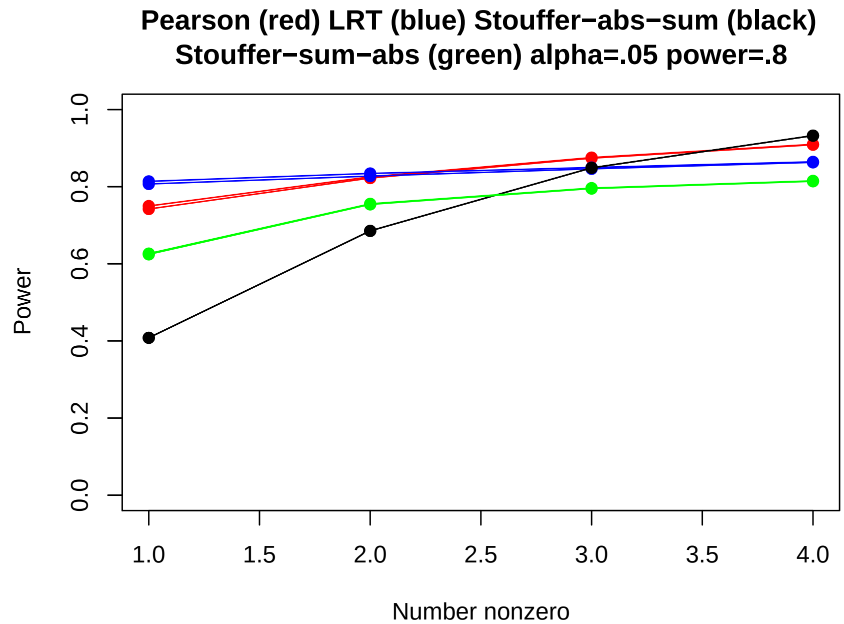


FIG 14. *The top figure shows power curves for* $m = 4$ *with* $\alpha = 0.01$ *and* Δ *chosen so that a chisquare test has power* 0.5*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



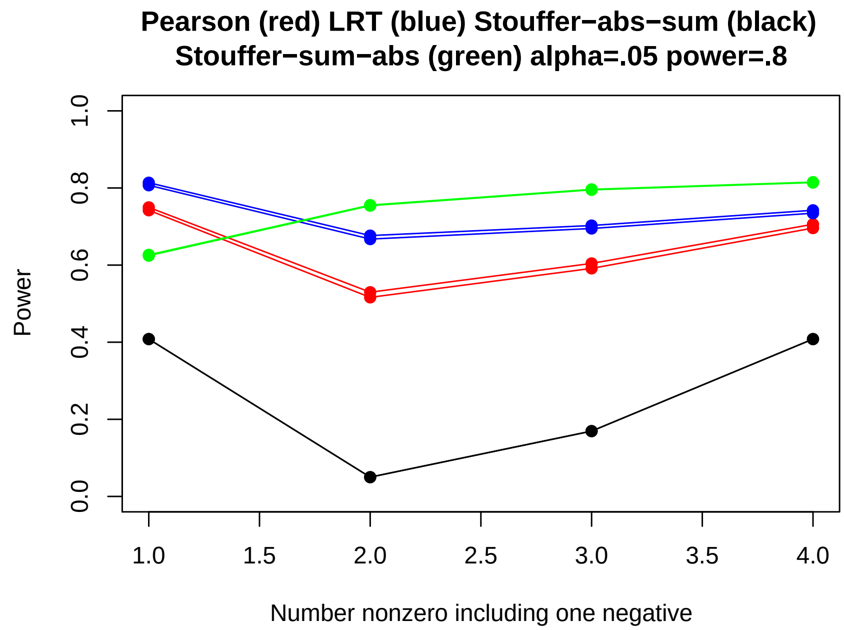
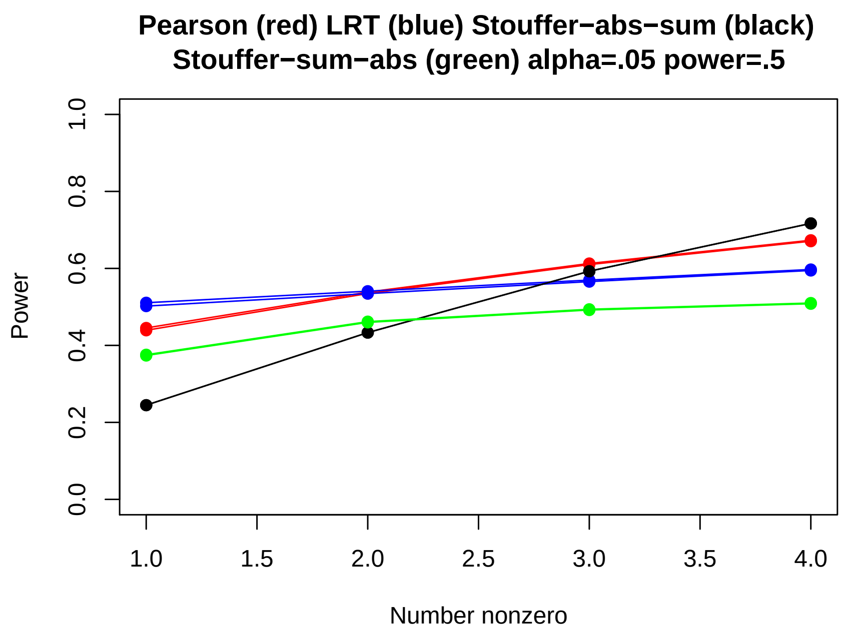


FIG 15. *The top figure shows power curves for $m = 4$ with $\alpha = 0.05$ and Δ chosen so that a chisquare test has power 0.8. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*



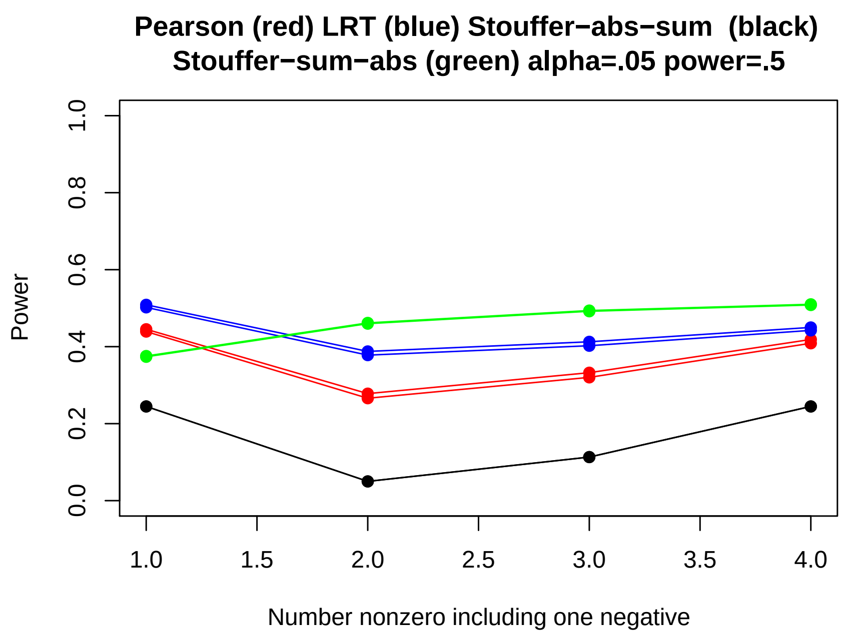


FIG 16. *The top figure shows power curves for $m = 4$ with $\alpha = 0.05$ and Δ chosen so that a chisquare test has power* 0.5*. The alternative is concordant. The bottom figure shows the nearly concordant alternative.*