Pragmatic, Unifying Algorithm Gives Power Probabilities for Common F Tests of the Multivariate General Linear Hypothesis

Ralph G. O'Brien and Gwowen Shieh*

We consider the problem of computing the power of some usual F transforms of the Wilks U, Hotelling-Lawley T, and Pillai V statistics for testing H_0: $\mathbf{CBA} = \mathbf{\Theta}_0$ under the multivariate general linear model, $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$, where the rows of $\boldsymbol{\varepsilon}\mathbf{A}$ are taken as independent $N(\mathbf{0}, \mathbf{\Sigma})$ random vectors. Keeping all these matrices at full rank, let $\mathbf{C}$ be $r_C \times r_X$ and $\mathbf{A}$ be $P \times r_A$. For determining p-values, F_i ($i \in \{U, T_1, T_2, V\}$) is taken to be distributed as central $F(r_C r_A, \nu_2^{(i)})$, which is the exact distribution when $s = \min(r_C, r_A) = 1$. For determining powers, we present a pragmatic, unifying method that takes F_i to be noncentral $F(r_C r_A, \nu_2^{(i)}, \lambda_i)$, where λ_i is isomorphic to F_i. For any s, we obtain the simple form $\lambda_i = N\lambda_i^*$, where λ_i^* is not a function of the total sample size, N. We show that for $s = 1$, $F(r_C r_A, \nu_2^{(i)}, \lambda_i)$ defines the exact noncentral distribution. For $s > 1$, each F_i converges in distribution to its prescribed noncentral F distribution and numerical work supports the accuracy of all approximations for obtaining powers for all but very small N. We exploit the method to compare the powers of the various F_i statistics. Finally, we illustrate the method by computing a set of powers for a multivariate analysis of variance comparing the profiles of three correlated tests among three independent groups.

KEY WORDS: Sample-size analysis, noncentrality parameter, noncentral F distribution, MANOVA.

---

*R. G. O'Brien is Full Staff and Director, Collaborative Biostatistics Center, Department of Biostatistics/Wb4, Cleveland Clinic Foundation, Cleveland, Ohio 44195 (robrien@bio.ri.ccf.org).. G. Shieh is Associate Professor, Department of Management Science, National Chiao Tung University, Hsin Chu, Taiwan, Republic of China. This work was supported in part by grants from the US National Institutes of Health (GCRC: RR00082), the State of Florida (HRS: MQ303), and the UF Division of Sponsored Research. We thank an anonymous associate editor for sharpening our presentation. We are also indebted to colleagues Jon Shuster, Somnath Sarkar, Beiyao Zheng, and Keith Muller for reviewing earlier drafts of this work, and to Keith again for numerous "powerful" discussions over the years.

1. INTRODUCTION

Hypothesis-driven research proposals now typically include power analyses to support the chosen design and sample size. Doing so promotes an early fusion of the study's research questions, its proposed design, the specific measures to be collected, and the prescribed data analyses. Of course, the power analyses should be congruent with the stated hypotheses and their prescribed tests. It is our experience, however, that power analyses for multivariate hypotheses often use only oversimplified, univariate surrogates. For example, a proposed multivariate analysis of variance of a factorial design might be supported only by a power analysis based on some univariate, two-group t tests of the individual measures. Such incongruence weakens the statistical plan and hurts the quality of the entire proposal.

Herein we propose and assess a pragmatic strategy for computing the powers of specific Normal-theory hypothesis tests under the multivariate general linear model. Section 2 briefly reviews the noncentrality for the univariate general linear model and then extends it to specify approximate (sometimes exact) noncentral distributions for the common F transforms of the Wilks, the Hotelling-Lawley, and the Pillai statistics. Our method is a modification of and is asymptotically equivalent to the Muller-Peterson (1984) algorithm discussed in Muller, LaVange, Ramey, and Ramey (1992). But unlike their method, ours provides the exact noncentral F distribution whenever the hypothesis involves at most $s = 1$ positive eigenvalue of $\mathbf{E}^{-1}\mathbf{H}$. For $s > 1$, both methods designate approximate noncentral F distributions that converge ($N \rightarrow \infty$) well to their limiting forms. But as the Monte Carlo work in Section 3 illustrates, our method is almost always more accurate than the Muller-Peterson, and it is sufficiently accurate for performing power analyses. In Section 4 we use the method to characterize how the relative powers of the F statistics are dependent on the structure of the s eigenvalues of the population version of $\mathbf{E}^{-1}\mathbf{H}$. Section 5 focuses on the common q-group, P-variate problem, and outlines an example.

2. NONCENTRALITIES

Consider the standard (fixed-effects), full-rank multivariate general linear model, $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ is $N \times P$ of rank P; $\mathbf{X}$ is $N \times r_X$ of rank r_X; and $\mathbf{B}$ contains fixed coefficients. The rows of $\boldsymbol{\varepsilon}$ are taken to be independent P-variate Normal random vectors with mean $\mathbf{0}$ and $P \times P$ positive-definite covariance matrix $\boldsymbol{\Sigma}$. The usual estimates are $\hat{\mathbf{B}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$, and $\hat{\boldsymbol{\Sigma}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})/(N - r_X)$. The multivariate general linear hypothesis is H_0: $\mathbf{CBA} = \boldsymbol{\Theta}_0$, where $\mathbf{C}$ is $r_C \times r_X$ with full row rank, and $\mathbf{A}$ is $P \times r_A$ with full column rank; thus $r_A \le P$. $\boldsymbol{\Theta}_0$ is usually chosen to be $\mathbf{0}$. H_0 has $\nu_1 = r_C r_A$ degrees of freedom.

2.1 $r_A = 1$

If $r_A = 1$ so that $\mathbf{A} \equiv \mathbf{a}$, the problem simplifies to a univariate one with $\mathbf{y} \equiv \mathbf{Ya} = \mathbf{XBa} + \boldsymbol{\varepsilon}\mathbf{a} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The resulting estimates are $\hat{\boldsymbol{\beta}} = \hat{\mathbf{B}}\mathbf{a}$ and $\hat{\sigma}^2 = \mathbf{a}'\hat{\boldsymbol{\Sigma}}\mathbf{a}/(N - r_X)$. H_0 is tested with $F = (SSH/r_C)/\hat{\sigma}^2$, where

$$SSH = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\Theta}_0)'[\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\Theta}_0)$$

is the sums of squares for the hypothesis. It is well-known that F is distributed as an $F(\nu_1, \nu_2, \lambda)$ random variable with $\nu_1 = r_C$ and $\nu_2 = N - r_X$ degrees of freedom and noncentrality

$$\lambda = (\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\Theta}_0)'[\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}']^{-1}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\Theta}_0)/\sigma^2.$$

It is helpful to see $(\mathbf{X'X})^{-1}$ decomposed into more distinct components. Letting $\mathbf{1}$ be the $N \times 1$ vector of ones, then $\bar{\mathbf{x}} = N^{-1}\mathbf{X'1}$ is the r_X-element mean vector and $\mathbf{S}_X = N^{-1}(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')'(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')$ is the corresponding $r_X \times r_X$ covariance matrix. Then $\mathbf{X'X} = N(\mathbf{S}_X + \bar{\mathbf{x}}\bar{\mathbf{x}}') = N\boldsymbol{\Psi}$, showing that with respect to $\mathbf{X'X}$, the size of the design (N) is unrelated to $\boldsymbol{\Psi}$, which is only dependent on the means, variances, and correlations of the Xs. Thus, $[\mathbf{C}(\mathbf{X'X})^{-1}\mathbf{C}']^{-1} = N[\mathbf{C}\boldsymbol{\Psi}^{-1}\mathbf{C}']^{-1}$, so

$$\lambda = N\lambda^* = N\{(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\Theta}_0)'[\mathbf{C}\boldsymbol{\Psi}^{-1}\mathbf{C}']^{-1}(\mathbf{C}\boldsymbol{\beta} - \boldsymbol{\Theta}_0)/\sigma^2\}. \qquad (2.1)$$

2.2 General Strategy for $r_A > 1$

For $r_A > 1$, SSH generalizes to the $r_A \times r_A$ sums of squares and cross products matrix for the hypothesis,

$$\mathbf{H} = N(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Theta}_0)'[\mathbf{C}\mathbf{\Psi}^{-1}\mathbf{C}']^{-1}(\mathbf{C}\hat{\mathbf{B}}\mathbf{A} - \mathbf{\Theta}_0).$$

$\hat{\sigma}^2$ generalizes to $\mathbf{A}'\hat{\mathbf{\Sigma}}\mathbf{A} = \mathbf{E}/(N - r_X)$, where $\mathbf{E} = \mathbf{A}'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})\mathbf{A}$. **H** and **E** are independent Wishart matrices, both based on $\mathbf{A}'\mathbf{\Sigma}\mathbf{A}$, and having r_C and $N - r_X$ degrees of freedom, respectively. $(SSH/r_C)/\hat{\sigma}^2$ generalizes to $\{(N - r_X)/r_C\}\mathbf{E}^{-1}\mathbf{H}$, but by tradition we work with $\mathbf{E}^{-1}\mathbf{H}$.

There is no generally optimal way to map $\mathbf{E}^{-1}\mathbf{H}$ to a univariate test statistic. The most common ones are the Wilks Likelihood Ratio (U), Hotelling-Lawley Trace (T), and Pillai Trace (V) statistics, which are reviewed below. All are based on the $s = \min(r_C, r_A)$ positive eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, denoted $\boldsymbol{\phi} = \{\phi_1, ..., \phi_s\}$, and ordered $\phi_1 > \phi_2 > \ldots > \phi_s > 0$. U, T, and V are summarized and compared by Seber (1984), Anderson (1984), and numerous other books and articles, and their critical values have been widely tabled and charted (e.g., Seber, pp. 562-564). But in practice we usually obtain p values by transforming them to F-type statistics, denoted here as F_i, $i \in \{U, T_1, T_2, V\}$. If $r_C = 1$, each F_i becomes $F = (N - r_X - r_A + 1)\phi_1/r_A$, which is also an exact F random variable, as discussed below. For $s > 1$, the F_i statistics are distinct, having different $\nu_2^{(i)}$ and λ_i.

We do not propose or study power approximations for Roy's test. For $s > 1$, no acceptable method has been developed for transforming ϕ_1 to an F or χ^2 statistic. No straightforward method exists for computing powers for Roy's statistic itself (Anderson, 1984, pp. 332), although various approximations have been developed, as reviewed by Krishnaiah (1978). Roy's statistic is fundamentally different from U, T, and V, thus its power is not accurately discerned from the power probabilities computed for F_U, F_{T_1}, F_{T_2} and F_V.

$\mathbf{E}^{-1}\mathbf{H} = (\mathbf{E}/N)^{-1}(\mathbf{H}/N) = (\mathbf{A}'\mathbf{S}\mathbf{A})^{-1}(\mathbf{H}/N)$, where **S** is the maximum likelihood estimate of $\mathbf{\Sigma}$. Whereas **E** is a central Wishart, **H** is possibly noncentral with noncentrality matrix

$$\boldsymbol{\Delta} = \mathrm{N}(\mathbf{A'\Sigma A})^{-1}(\mathbf{CBA} - \boldsymbol{\Theta}_0)'[\mathbf{C\Psi}^{-1}\mathbf{C'}]^{-1}(\mathbf{CBA} - \boldsymbol{\Theta}_0)\}$$

$$= \mathrm{N}(\mathbf{A'\Sigma A})^{-1}\mathbf{H}^* = \mathrm{N}\boldsymbol{\Delta}^*. (2.2)$$

$\mathbf{H}^*$ is the population counterpart of $\mathbf{H}$/N. Let $\boldsymbol{\phi}^* = \{\phi_1^*, ..., \phi_2^*\}$ to be the eigenvalues of $\boldsymbol{\Delta}^* = (\mathbf{A'\Sigma A})^{-1}\mathbf{H}^*$, the population counterpart of $\mathbf{E}^{-1}\mathbf{H}$. As $N \to \infty$, $\hat{\mathbf{B}} \xrightarrow{p} \mathbf{B}$, $\mathbf{H}/N \xrightarrow{p} \mathbf{H}^*$, and $\mathbf{E}/N \xrightarrow{p} \mathbf{A'\Sigma A}$, so that $N\mathbf{E}^{-1} \xrightarrow{p} (\mathbf{A'\Sigma A})^{-1}$. Thus, $\mathbf{E}^{-1}\mathbf{H} \xrightarrow{p} (\mathbf{A'\Sigma A})^{-1}\mathbf{H}^*$ and $\boldsymbol{\phi} \xrightarrow{p} \boldsymbol{\phi}^*$.

We shall specify and asymptotically justify all F distributions using a common notation and logic. First, take F_i to be an F random variable with $\nu_1 = r_C r_A$ and $\nu_2^{(i)}$ degrees of freedom and noncentrality $\lambda_i = N\lambda_i^*$, a form motivated by (2.1) and (2.2). Accordingly, $E(F_i/N) = N^{-1}(1 + \lambda_i/\nu_1)[\nu_2^{(i)}/(\nu_2^{(i)} - 2)] \to \lambda_i^*/\nu_1$, as $N \to \infty$. Also, for each F_i, $F_i/N \xrightarrow{p} f_i^*$, a constant. This leads to the approximation $\lambda_i^* = \nu_1 f_i^*$. Each f_i^* is a function of r_C, r_A and $\boldsymbol{\phi}^*$ as specified below. Finally, we cite specific theory reviewed by Anderson (1984, Section 8.6.5) to outline why for the Hotelling and Pillai statistics, $\nu_1 F_i$ converges to noncentral χ^2 distributions with the noncentralities proposed here. Likewise, the work of Kulp and Nagarsenker (1984) supports the convergence of $\nu_1 F_U$ for the noncentral Wilks statistic.

Motivated because $\mathbf{E}/(N - r_X)$ is the unbiased estimator of $\mathbf{A'\Sigma A}$, Muller and Peterson (1984) proposed extracting the eigenvalues, $\boldsymbol{\phi}^{(M)}$, of $[(\mathbf{A'\Sigma A})^{-1}/(N - r_X)][N\mathbf{H}^*]$. Thus, $\boldsymbol{\phi}^{(M)} = [N/(N - r_X)]\boldsymbol{\phi}^*$. Furthermore, they proposed making $\lambda_i^{(M)} = \nu_2^{(i)}\nu_1 f_i^{(M)}$, where $f_i^{(M)}$ uses $\boldsymbol{\phi}^{(M)}$ just as f_i^* uses $\boldsymbol{\phi}^*$. One can easily show that for $r_A = 1$ both methods lead to the exact univariate noncentrality given above. If $r_A > 1$, $\lambda_i^{(M)} < \lambda_i$, but $\lambda_i/\lambda_i^{(M)} \to 1$ as $N \to \infty$.

Muller and Barton (1989) used a similar strategy to define approximations of the non-null distributions of F statistics for univariate approaches for repeated measures analysis. O'Brien (1986) also applied the strategy to characterize the non-null distribution of the likelihood-ratio χ^2 statistic commonly used in log-linear models; c.f. Agresti (1990, Section 7.6.4).

2.3 $r_A > 1$ but $r_C = 1$ (s = 1)

It is well known that when $r_C = 1$, the U, T, and V statistics convert identically to F $= (N - r_X - r_A + 1)\phi_1/r_A$, which has r_A and $(N - r_X - r_A + 1)$ degrees of freedom. Our strategy gives $\lambda = N\phi_1^*$, where

$$\phi_1^* = (\mathbf{CBA} - \mathbf{\Theta}_0)(\mathbf{A'\Sigma A})^{-1}(\mathbf{CBA} - \mathbf{\Theta}_0)'[\mathbf{C\Psi}^{-1}\mathbf{C'}]^{-1}.$$

This characterizes the exact noncentral F distribution, a result established by first showing that $T^2 = (N - r_X)\phi_1$ is a noncentral T^2 random variable and then converting it to an exact noncentral F, as per sections 2.4.2 and 2.5.5 of Seber (1984). The result can also be established by noting that the approximation for the distribution of U given by Kulp and Nagarsenker (1984, Theorem 3.1) is exact for s = 1. Its only term then is a noncentral beta distribution function, which is transformable to the noncentral F prescribed here.

With $r_C = 1$, the Muller-Peterson (1984) method gives $\lambda^{(M)} = [(N - r_X - r_A + 1)/(N - r_X)]\lambda$. For $r_A > 1$, $\lambda^{(M)} < \lambda$. Thus whereas both the proposed method and the Muller-Peterson method specify the exact noncentral F distribution when $r_A = 1$, only the proposed method properly handles all s = 1 cases. $\lambda^{(M)}$ gives powers that are too low, leading to recommended sample sizes that are too large. This discrepancy between λ and $\lambda^{(M)}$ is important because many common situations use $r_C = 1$ and $r_A > 1$, including the one- and two-group Hotelling's T^2 tests on centroids. We shall see that this difference between λ and $\lambda^{(M)}$ extends to cases with s > 1.

2.4 $r_A > 1$ and $r_C > 1$ (s > 1)

When s >1, the U, T, and V statistics are distinct and their F transforms only lead to approximate noncentral F random variables.

Wilks (F_U). Wilks' (1932) likelihood ratio statistic is the determinant of $\mathbf{E}(\mathbf{H} + \mathbf{E})^{-1}$, or equivalently, $U = \prod_{k=1}^{s} (1 + \phi_k)^{-1}$. Rao's (1951) transformation is $F_U = \nu_2^{(U)}(U^{-1/t} - 1)/(r_C r_A)$, where

$$t = \begin{cases} 1 & r_C r_A \le 3 \\ \{[(r_C r_A)^2 - 4]/[r_C^2 + r_A^2 - 5]\}^{1/2} & r_C r_A \ge 4 \end{cases}$$

and $\nu_2^{(U)} = t[N - r_X - (r_A - r_C + 1)/2] - (r_C r_A - 2)/2$. When H_0 is true,

$F_U \sim F(\nu_1, \nu_2^{(U)}, 0)$, exactly, for s = 1 or 2; for s > 2, this is an approximation that is "adequate for practical situations" (Seber, p. 41). Using the strategy described above, $F_U/N \xrightarrow{p} f_U^* = t\{(U^*)^{-1/t} - 1\}/(r_C r_A)$; where $U^* = \prod_{k=1}^{s} (1 + \phi_k^*)^{-1}$. Thus we take $\lambda_U = N\lambda_U^*$, where $\lambda_U^* = t\{(U^*)^{-1/t} - 1\}$.

Kulp and Nagarsenker (1984) provided an approximation for the noncentral distribution of U, which quickly provides asymptotic justification for our method. Briefly: As is commonly done (c.f. Anderson, 1984, p. 330), if we take $N \rightarrow \infty$ and $\mathbf{CBA} \rightarrow \mathbf{\Theta}_0$ under a sequence of alternatives, then their Theorem 3.1 reduces to a single noncentral beta distribution function, which is transformable exactly to the noncentral F prescribed here. They noted that using the noncentral beta distribution (or, equivalently, the noncentral F) is better than using the chi-square distribution, as per Sugiura and Fujikoshi (1969), whose method is not exact under any case, even for $r_A = 1$.

For $r_A > 1$, $\lambda_U > \lambda_U^{(M)}$. Evidence heretofore that the Muller-Peterson algorithm systematically under approximates the power of F_U comes from a study by Barton and Cramer (1989). They used $\lambda_U^{(M)}$ to construct various s > 1 situations with nominal powers of .80, but reported estimated powers consistently higher than this (based on 5000 trials of each situation).

Hotelling-Lawley (F_{T_1}, F_{T_2}). Hotelling (1951) and Lawley (1938) proposed the statistic $T = \text{tr}[\mathbf{E}^{-1}\mathbf{H}] = \sum_{k=1}^{s} \phi_k$. Several F transforms have been proposed. The most commonly used one, due to Pillai and Samson (1959), is $F_{T_1} = \nu_2^{(T_1)}(T/s)/(r_C r_A)$. with $\nu_2^{(T_1)} = s(N - r_X - r_A - 1) + 2$. McKeon (1974) proposed $F_{T_2} = \nu_2^{(T_2)}(T/h)/(r_C r_A)$, with $\nu_2^{(T_2)} = 4 + (r_C r_A + 2)g$, where

$$g = \frac{(N - r_X)^2 - (N - r_X)(2r_A + 3) + r_A(r_A + 3)}{(N - r_X)(r_C + r_A + 1) - (r_C + 2r_A + r_A{}^2 - 1)} ,$$

and $h = (\nu_2^{(T_2)} - 2)/(N - r_X - r_A - 1)$. For $s \geq 2$, $F_{T_1}/F_{T_2} < 1.00$ (with $F_{T_1}/F_{T_2} \rightarrow 1$ as $N \rightarrow \infty$), but this is counterbalanced to some degree by the fact that $\nu_2^{(T_1)} > \nu_2^{(T_2)}$. We assessed the difference between F_{T_1} and F_{T_2} for 108 cases formed by crossing (r_A, r_C) = {(2, 2}, (2, 3}, (3, 2}, (3, 3)}; $N - r_X$ = {31, 66, 96}; nominal percentage points for F_{T_1} using p_{T_1} = {.005, .010, .020, .040, .050, .075, .100, .200, .400. .600, .800, .900}. We found that $F_{T_1}/F_{T_2} > 0.98$; $1.50 < \nu_2^{(T_1)}/\nu_2^{(T_2)} < 1.95$; with the ratio of the resulting p values being $0.666 < p_{T_1}/p_{T_2} < 1.026$. Seber (p. 39) stated that when H_0 is true, the F_{T_2} "approximation is surprisingly accurate and supersedes previous approximations" including F_{T_1} and another by Hughes and Saw (1972). For either F_{T_1} or F_{T_2}, $F_T/N \xrightarrow{p} f_T^*$ = $T^*/(r_C r_A)$, where $T^* = \sum_{k=1}^{s} \phi_i^*$. This gives $\lambda_T = NT^*$.

Taking F_{T_1} and F_{T_2} to be noncentral $F(r_C r_A, \nu_2^{(T_1)}, \lambda_T)$ and $F(r_C r_A, \nu_2^{(T_2)}, \lambda_T)$, respectively, is supported asymptotically by work summarized by Anderson (1984, Section 8.6.5) and Seber (1984, Section 8.6d). The asymptotic distribution of $(N - r_X)T$ is $\chi^2(r_C r_A, \text{tr}\boldsymbol{\Delta} = \lambda_T)$, again under a sequence of alternatives implying that $\mathbf{CBA} \rightarrow \boldsymbol{\Theta}_0$ as $N \rightarrow \infty$. $\chi^2(r_C r_A, \lambda_T)$ is the limiting form of $r_C r_A F(r_C r_A, \nu_2, \lambda_T)$ as $\nu_2 \rightarrow \infty$. The asymptotic distribution of F_{T_1} and F_{T_2} is established simply by noting that $(N - r_X)T$, $r_C r_A F_{T_1}$, and $r_C r_A F_{T_2}$ all have the form kT where $k/N \rightarrow 1$, as $N \rightarrow \infty$. The fact that $\nu_2^{(T_1)} > \nu_2^{(T_2)}$ implies that nominal powers computed for F_{T_1} are uniformly greater than those computed for F_{T_2}.

Comparing λ_T to the Muller-Peterson approximation, $\lambda_{T_1}^{(M)} = [(N - r_X - r_A - 1 + 2/s)/(N - r_X)]\lambda_T < \lambda_T$ when $r_A > 1$. Applying the Muller-Peterson strategy to F_{T_2} likewise gives $\lambda_{T_2}^{(M)} < \lambda_T$.

A reviewer suggested we examine a newer F transformation, call it F_{T_3}, and the accompanying power approximation due to van der Merwe and Crowther (1984). F_{T_3} behaves much like F_{T_2}. Across the 108 cases we studied: $1.0001 < F_{T_2}/F_{T_3} < 1.0021$; $0.956 < \nu_2^{(T_2)}/\nu_2^{(T_3)} \leq 0.995$; $0.997 < p_{T_2}/p_{T_3} < 1.033$. In their power approximation, the main term is identical to using $F(r_C r_A, \nu_2^{(T_3)}, \lambda_T)$ as proposed here, and the secondary term appears to make no practical difference in computing the power. We assessed the 20 cases they evaluated in their Table 3, as well as 54 more cases arising from crossing $(r_A, r_C) = \{(2, 2), (2, 3), (3, 3)\}$; $N - r_X = \{31, 66, 96\}$; nominal power for F_{T_1} of $\pi_0 = \{.80, .90\}$; and eigenvalue structure of $\boldsymbol{\Delta} = \{E, G, X\}$ as defined below. The three largest absolute differences in nominal powers between their method for F_{T_3} and our method for F_{T_2} were 0.013, 0.012, and 0.010 (all for $r_A = 3$, $r_C = 3$, $N - r_X = 31$) with all others less than 0.010. The average (signed) error was 0.002, and the average absolute error was 0.003. Using only the primary term, their power approximation (equivalent to applying our general strategy to F_{T_3}) changed the power by at most 0.010. In light of its similarity to F_{T_2}, we see no reason to study F_{T_3} further.

Pillai (F_V). Bartlett (1939), Nanda (1950), and Pillai (1955) proposed using $V = \mathrm{tr}[(\mathbf{E} + \mathbf{H})^{-1}\mathbf{H}] = \sum_{k=1}^{s} [\phi_k/(1 + \phi_k)]$. Pillai and Mijares (1959) gave the F transform, $F_V = \nu_2^{(V)}[V/(s - V)]/(r_C r_A)$, where $\nu_2^{(V)} = s(N - r_X + s - r_A)$. $F_V/N \xrightarrow{p} f_V^* = s[V^*/(s - V^*)]/(r_C r_A)$, where $V^* = \sum_{k=1}^{s} [\phi_k^*/(1 + \phi_k^*)]$. Thus we define $\lambda_V = Ns[V^*/(s - V^*)]$.

Anderson (1984, Section 8.6.5) summarized results showing that $NV \xrightarrow{d} \chi^2(r_C r_A, N\,\mathrm{tr}\boldsymbol{\Delta}^*)$, under a sequence of alternatives in which $\mathbf{CBA} \to \boldsymbol{\Theta}_0$. To establish that $\lambda_V \to N\,\mathrm{tr}\boldsymbol{\Delta}^*$, first express $NV^* = N\,\mathrm{tr}[(\mathbf{I} + \boldsymbol{\Delta}^*)^{-1}\boldsymbol{\Delta}^*]$, where $\boldsymbol{\Delta}^* = (\mathbf{A}'\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{H}^*$. By recursively using a result in Searle (1982, pp. 151, #16g), $N V^* = N\,\mathrm{tr}[\boldsymbol{\Delta}^* - \boldsymbol{\Delta}^{*2} + \boldsymbol{\Delta}^{*3} - \boldsymbol{\Delta}^{*4} + \ldots] \to N\,\mathrm{tr}\boldsymbol{\Delta}^*$ as $\boldsymbol{\Delta}^* \to \mathbf{0}$ with $N\boldsymbol{\Delta}^*$ remaining finite. Finally, $\lambda_V = [(NV^*)^{-1} - (Ns)^{-1}]^{-1} \to NV^*$, because $(NV^*)^{-1}$ is finite under the sequence of alternatives, but $(Ns)^{-1} \to 0$. Similarly, $r_C r_A F_V \to [(NV)^{-1} - (Ns)^{-1}]^{-1} \to NV$. Thus $r_C r_A F_V \xrightarrow{d} \chi^2(r_C r_A, \lambda_V)$, and more directly, $F_V \xrightarrow{d} F(r_C r_A, \nu_2^{(V)}, \lambda_V)$.

It can be shown that $\lambda_V^{(M)} \le \lambda_V$, with equality holding if $s = r_A$ and $\phi_1^* = \phi_2^* = \ldots = \phi_s^*$.

3. ACCURACY

We showed above that when $r_C = 1$ and $r_A > 1$, the proposed method prescribes the exact noncentral distributions, whereas the Muller-Peterson algorithm prescribes distributions that are only asymptotically ($N \to \infty$) correct for this case. This difference is important because it applies to many situations found in practice. For example, $r_A = 3$, $r_C = 1$, $r_X = 2$, and $N = 30$ defines a two-group discriminant analysis with three variables and 15 cases per group. If $\lambda = 16.5$, the power is exactly 0.90. On the other hand, $\lambda^{(M)} = 15.33$, giving a power of 0.875, a 25% error relative to the exact Type II error rate.

For $s > 1$, both methods are approximate. We now assess such cases.

3.1 Comparison with Results of Lee

Lee (1971) developed an asymptotic formula to approximate the powers for U, T, and V using a complex weighted sum of noncentral χ^2-distribution functions. He assessed its accuracy relative to exact values worked out for 18 cases in which $s = r_A = 2$, $r_C = \{3, 5, 9\}$, and $N - r_X = 63$. Although these results do not strictly apply to assessing algorithms for the power of F_U, F_{T_1}, and F_V, they offer a convenient place to begin. Figure 1 gives relative errors of approximation, $(\tilde{\pi} - \pi)/[\pi(1 - \pi)]^{1/2}$, where $\tilde{\pi}$ is the approximated power of F_U, F_{T_1}, and F_V by either the Muller-Peterson or the proposed algorithm and π is the exact power of U, T, and V given by Lee. Note that because $[\pi(1 - \pi)]^{1/2} \leq 0.50$, the relative error is at least twice that of the raw error, $\tilde{\pi} - \pi$.

Figure 1. Relative errors of approximation for the Muller-Peterson (M-P) and the proposed methods for 18 cases studied by Lee (1971, Table 1). $s = r_A = 2$, $r_C = \{3, 5, 7\}$, $N - r_X = 63$.

The two F-based methods show accuracies that are quite acceptable for performing power analyses of proposed studies. For these cases, they give relative errors within ±3%, with averages much less than that. The proposed method is seen to be biased positively for U and T_1. As hypothesized above, the Muller-Peterson shows a tendency to underestimate power, but its overall accuracy seems slightly superior to the proposed method in these cases. We performed similar computations based on Lee's Table 2 and found the same pattern for $s = r_A = \{3, 4\}$ as we did for $s = r_A = 2$.

This encouraging assessment is limited in two respects. First, like other researchers of this topic, Lee was not concerned with methods for computing the power probabilities of F_U, F_{T_1}, and F_V, which are the test statistics generally used today. The powers of (say) U and F_U surely diverge for lower N, though perhaps only slightly. As a practical matter, we would prefer an algorithm that gives more accurate powers for F_U to one that gives more accurate powers for U. Second, the cases Lee created to investigate had powers much lower than are typically of interest. The largest power is 0.62, with the vast majority being less than 0.50. For problems of sample-size choice, one generally desires to have nominal powers of 0.80 or higher. We prefer 0.90.

3.2 Monte Carlo Study

For the reasons just stated, we performed a Monte Carlo study to assess how well we can compute the power of F_U, F_{T_1}, F_{T_2}, and F_V directly, for situations with meaningful

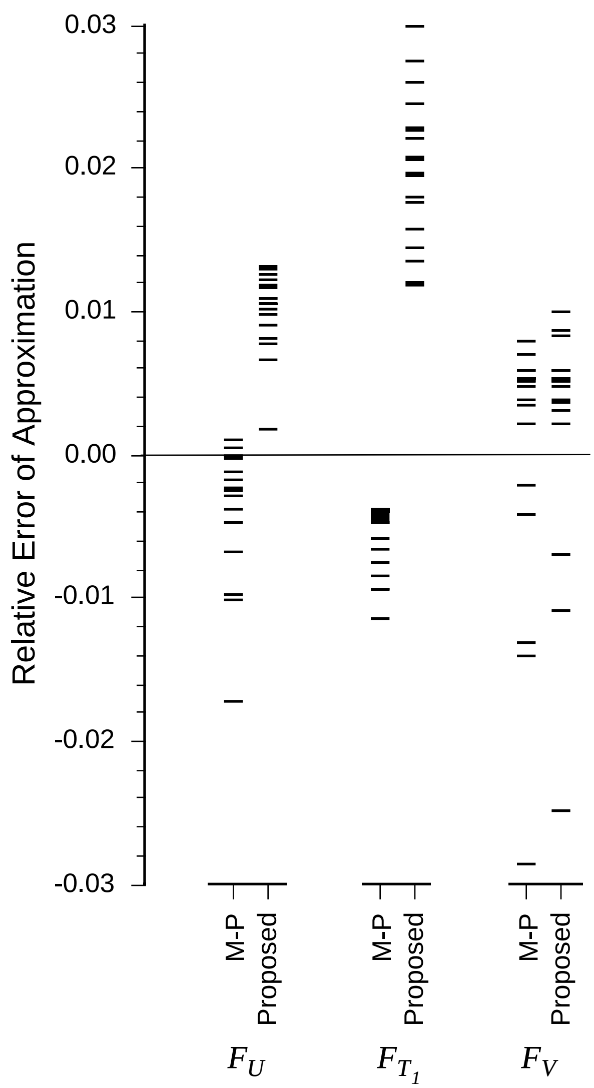

Relative Error of Approximation
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
M-P
Proposed
M-P
Proposed
M-P
Proposed
F_U
F_{T_1}
F_V

Figure 1.

power, $\pi \geq .80$.

Method and design. Because F_{T_1} has the simplest proposed nominal noncentral distribution, we used this to specify the particular non-null cases to simulate. For a given N, r_X, r_C, and r_A, and a given nominal power, $\pi_{T_1} = \pi_0$, we employed the SAS® functions FINV and FNONCT to find λ_0 such that Prob[$F(r_C r_A, \nu_2^{(T_1)}, \lambda_0) > F_{.05}(r_C r_A, \nu_2^{(T_1)}, 0)$] = π_0. For s > 1, we defined four different structures for the first s eigenvalues of $\boldsymbol{\Delta}^*$, $\boldsymbol{\phi}^* = \{\phi_1^*, \phi_2^*, ..., \phi_s^*\}$:

$$\text{Equal (E):} \quad \boldsymbol{\phi}^* = (\lambda_0/N)s^{-1}\{1, 1, ..., 1\}$$

$$\text{Linear (L):} \quad \boldsymbol{\phi}^* = (\lambda_0/N)[s(s+1)/2]^{-1}\{s, s-1, ..., 1\}$$

$$\text{Geometric (G):} \quad \boldsymbol{\phi}^* = (\lambda_0/N)(2^s - 1)^{-1}\{2^{s-1}, 2^{s-2}, ..., 1\}$$

$$\text{Extreme (X):} \quad \boldsymbol{\phi}^* = (\lambda_0/N)\{1, 0, ..., 0\}.$$

The sum of the roots, $T^* = \lambda_0/N$, is the same for all four structures; thus, $\lambda_T = \lambda_0$, and the nominal power for F_{T_1} is π_0. Though lower than π_0, the nominal power for F_{T_2} is also the same under all four structures. On the other hand, λ_U and λ_V are affected by the structure of the eigenvalues. To keep π_0 fixed, the elements in $\boldsymbol{\phi}^*$ must decrease with increasing N.

We examined all four F statistics under the E, L, G, and X structures for N = {10, 20, 30, 40, 50, 100}, $r_X = 4$, $r_C = \{2, 3\}$, $r_A = 3$, and $\pi_0 = \{.80, 90\}$. This produced sets of s = 2 and s = 3 cases that blanket a wide range of situations that are practical for power analyses.

The experiment was performed using the IML® matrix language of SAS. Each trial of a particular case proceeded as follows. Without loss of generality, $\boldsymbol{\Sigma} \equiv \mathbf{I}$. $\mathbf{H}$, an $r_A \times r_A$ noncentral Wishart matrix with r_C degrees of freedom, was formed by first using the NORMAL function to generate an $r_C \times r_A$ matrix $\mathbf{Z}$ having independent rows that are $N(\mathbf{0}, \mathbf{I})$. Let $\mathbf{M}$ be the $r_C \times r_A$ matrix having elements $m_{kk} = (\phi_k^*)^{1/2}$, k = 1 to s, and $m_{kk'} = 0$ for $k \neq k'$. Then $\mathbf{H} = (\mathbf{Z} + \mathbf{M})'(\mathbf{Z} + \mathbf{M})$ is Wishart with noncentrality $\mathbf{M}'\mathbf{M}$, a diagonal matrix with elements $\{\phi_1^*, \phi_2^*, ..., \phi_s^*, 0, ...0\}$. $\mathbf{E}$, the $r_A \times r_A$ central Wishart matrix with $N - r_X$ degrees of freedom, was formed by generating an $(N - r_X) \times r_A$ matrix $\mathbf{Z}$ having independent rows that are $N(\mathbf{0}, \mathbf{I})$ and computing $\mathbf{E} = \mathbf{Z}'\mathbf{Z}$. The roots of $\mathbf{E}^{-1}\mathbf{H}$ were then obtained and used to form F_U, F_{T_1}, F_{T_2}, F_V, which were compared to their respective nominal .05-level critical values. 5000 trials of each case where run, giving standard errors for each estimated true power of 0.0042 for π = 0.90 and 0.0057 for π = 0.80. We checked the accuracy of our programming and of IML's NORMAL random number generator by simulating a full slate of $s = r_C = 1$, $r_A = 3$ cases and verifying that the obtained results fell within reasonable sampling errors of the known exact powers.

Figure 2. Proposed (−) and Muller-Peterson (×) approximations and estimated true powers (⊙) as a function of eigenvalue structure and total sample size. $\hat{\alpha}$ is the estimated true Type-I error rate. $r_X = 4$; $r_A = 3$; $s = r_C = \{2, 3\}$.

The results for $\pi_0 = .90$ are presented in Figure 2. The results for the linear structure for $\boldsymbol{\phi}^*$ are not shown as they are identical to those of the geometric structure when s = 2 and virtually the same when s = 3. Also, the pattern of results for $\pi_0 = 0.80$ were in complete accord with those for $\pi_0 = 0.90$. $\hat{\alpha}$ is the estimated true Type I error rate, the "power" at $\boldsymbol{\phi}^* = \mathbf{0}$. Note only that they are too low for F_V when $N \leq 20$.

The power results reflect our theoretical conclusions and show a pattern consistent with those involving Lee's tables. The Muller-Peterson values are generally too low, whereas those obtained by the proposed method are somewhat high in some cases. The Muller-Peterson is better for F_V with low N, but this is a "lucky" consequence of having the deflated $\hat{\alpha}$ values suppress the true power. In general, there is a clear tendency for the proposed method to give more accurate results.

Figure 3 re-plots the results from the $s = r_A = r_C = 3$ cases to show the relative errors of approximation, $(\tilde{\pi}_i - \hat{\pi}_i)/[\hat{\pi}_i(1 - \hat{\pi}_i)]^{1/2}$, where $\tilde{\pi}_i$ is the proposed approximated power for a given F_i statistic and $\hat{\pi}_i$ is the Monte Carlo estimate of F_i 's true power. Cases were chosen to set $\tilde{\pi}_{T_1} = .90$. The results show that F_U has the most error-free approximations. For the equal and geometric (and linear) structures, the approximations for F_{T_1}, F_U, and F_V are positively biased, while those for F_{T_2} are negatively biased. Structure X induces the most error, with the proposed method performing best for F_U and F_{T_2}.

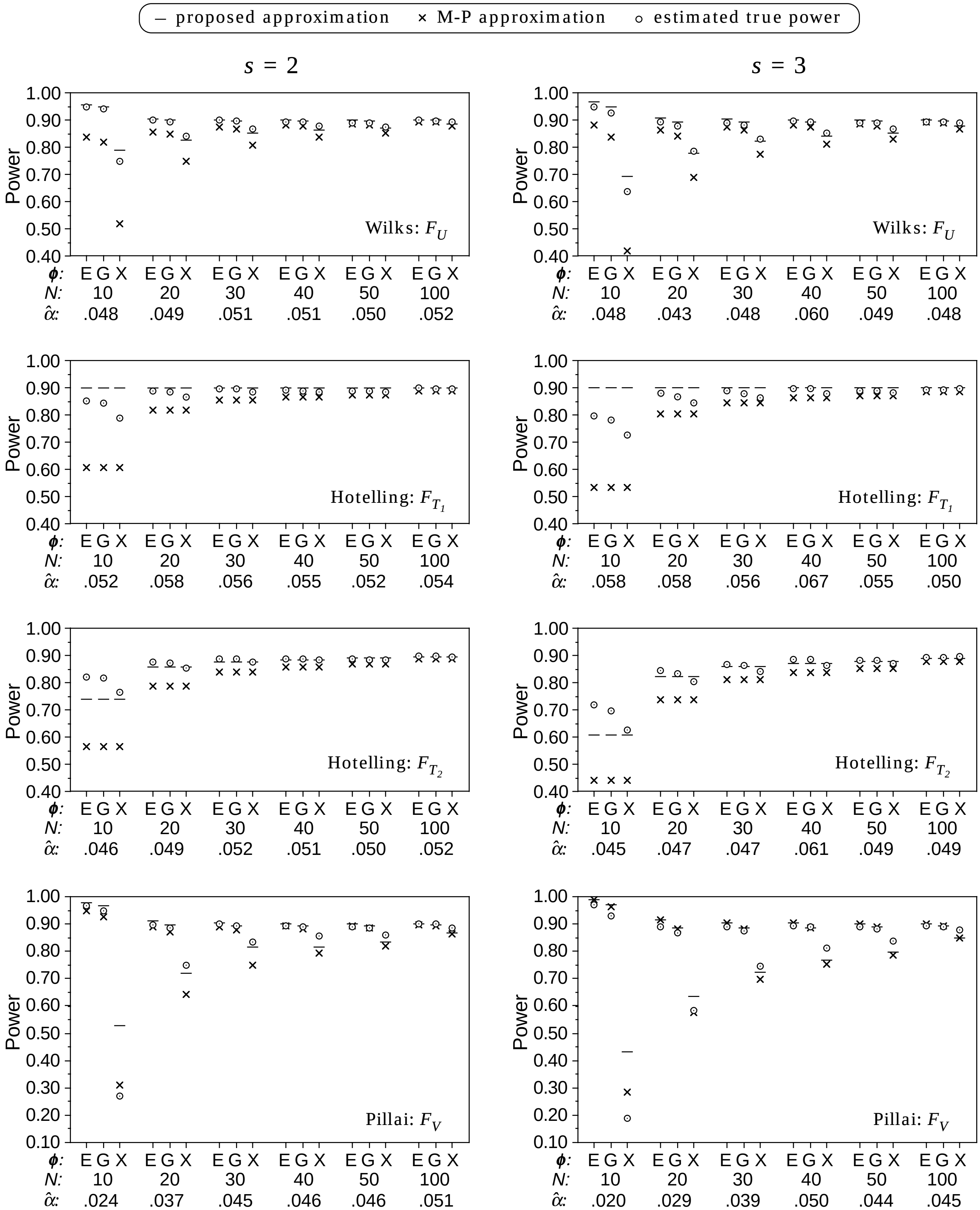

– proposed approximation × M-P approximation ○ estimated true power
s = 2
s = 3
Power
Wilks: F_U
Hotelling: F_{T_1}
Hotelling: F_{T_2}
Pillai: F_V
ϕ: E G X E G X E G X E G X E G X E G X
N: 10 20 30 40 50 100
α̂: .048 .049 .051 .051 .050 .052
α̂: .048 .043 .048 .060 .049 .048
α̂: .052 .058 .056 .055 .052 .054
α̂: .058 .058 .056 .067 .055 .050
α̂: .046 .049 .052 .051 .050 .052
α̂: .045 .047 .047 .061 .049 .049
α̂: .024 .037 .045 .046 .046 .051
α̂: .020 .029 .039 .050 .044 .045

Figure 2.

Figure 3. Estimated relative errors of approximation as a function of sample size and eigenvalue structure . $s = r_A = r_C = 3$; $r_X = 4$; nominal power of .90 for F_{T_1}.

In conclusion, only when N is quite small ($N \le 20$) do we see any worrisome breakdown in accuracy. It is exceptional that this single, straightforward scheme performs so well over these four test statistics and four eigenvalue structures.

4. COMPARING POWERS OF F_U, F_{T_1}, F_{T_2}, and F_V

The proposed algorithm provides a convenient way to compare the powers of F_U, F_{T_1}, F_{T_2}, and F_V. Following Anderson (1984, p. 332), we frame our view by employing the coefficient of variation of $\boldsymbol{\phi}^*$,

$$CV = \frac{\left[\sum_{k=1}^{s} (\phi_k^* - \bar{\phi}^*)^2 / s\right]^{1/2}}{\bar{\phi}^*} ,$$

where $\bar{\phi}^*$ is their average. Figure 4 plots CV against the approximate powers computed using the proposed algorithm and the estimated true powers for the case of N = 50, $r_X = 4$, and $r_C = r_A = s = 3$.   The lines were constructed by computing the powers using $\boldsymbol{\phi}^* = \{T^*c, T^*(1-c)/2, T^*(1-c)/2\}$ where T^* gives a power of $\pi = 0.90$ for F_{T_1} using $\alpha = .05$. c varies between c = 1/3 (Structure E: CV = 0) and c = 1.0 (Structure X: $CV = \sqrt{2}$). Although the L and G structures do not have a pattern of $\{T_e c, T_e(1-c)/2, T_e(1-c)/2\}$, their approximated powers fall right in line with those that do, as illustrated. Figure 4 demonstrates that these F_i statistics have the same power relations that Anderson described for T, U, and V: As CV increases, F_{T_1} and F_{T_2} become more powerful than F_U, which becomes more powerful than F_V. Quite similar images appear when graphing the powers for all other combinations of $r_C = s = \{2, 3\}$, $\pi = \{.80, .90\}$, and N = {50, 100}. It can be shown that under Structure E, $\lambda_T = \lambda_V$ whenever $s = r_A$. In addition, if s > 1, then $\nu_2^{(T_2)} < \nu_2^{(T_1)} < \nu_2^{(V)}$, so it follows that the nominal power of F_V exceeds that for F_{T_1}, which exceeds that for F_{T_2} in this case. We also note from the work of Schatzoff (1966) and Olson (1974) that Roy's statistic has greater power than these F_i statistics under

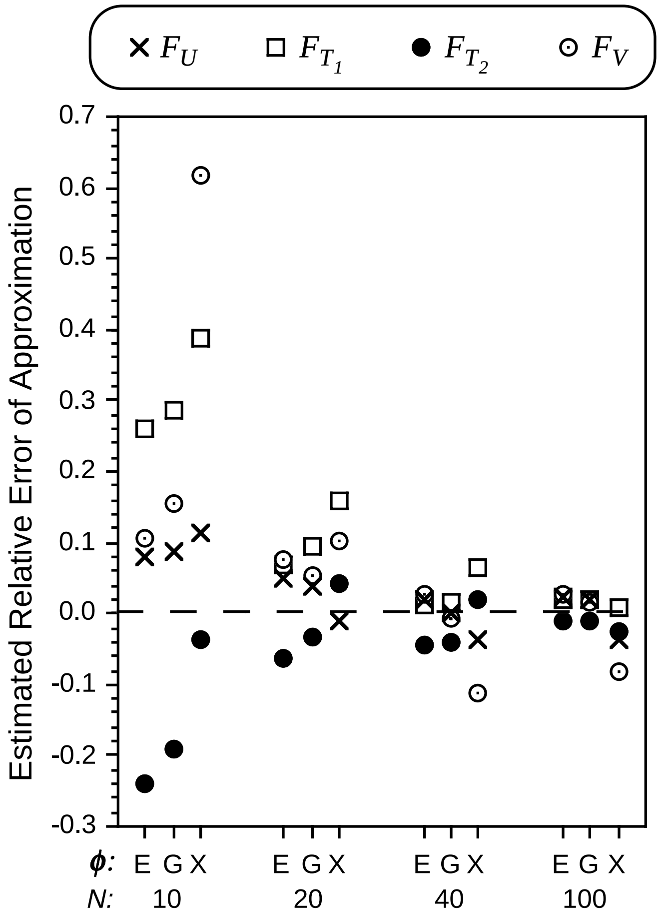

F_U
F_{T_1}
F_{T_2}
F_V
Estimated Relative Error of Approximation
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
ϕ: E G X E G X E G X E G X
N: 10 20 40 100

Figure 3.

Structure X, but in general it has less power otherwise.

Figure 4. Nominal (lines) and estimated (points) powers of F_V, $F_{T_1,}$ F_{T_2}, and F_V as a function of the coefficient of variation of ϕ_1^*, ϕ_2^*, ϕ_3^*. N= 50; s = r_A = r_C = 3; r_X = 4; nominal power of .90 for F_{T_1}.

---

5. THE q-GROUP PROBLEM, WITH EXAMPLE

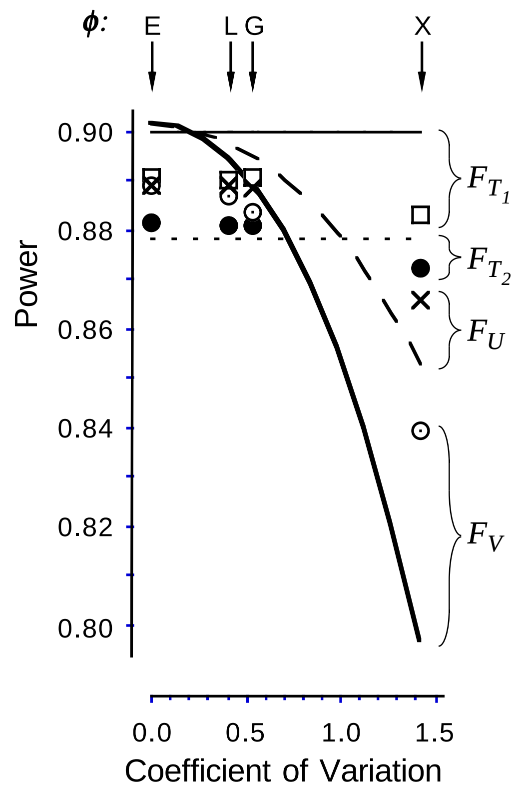
Most applications of the proposed method will be for power analyses in which q independent groups are compared with respect to P correlated measurements. This problem has a specific structure that we now exploit. An example follows.

Let $\ddot{\mathbf{X}}$ be the $q \times r_X$ *essence model matrix* formed by assembling the q unique rows of $\mathbf{X}$ ($N \times r_X$; $r_X \le q$). $\ddot{\mathbf{X}}$ is the collection of the q unique design points (e.g. the q groups) for the proposed study. If n_j of the rows of $\mathbf{X}$ are equal to the jth row of $\ddot{\mathbf{X}}$, define $\mathbf{W}$ to be the $q \times q$ diagonal matrix containing weights $w_j = n_j/N$. Thus $(\mathbf{X'X}) = N(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})$, so that

$$\mathbf{H}^* = (\mathbf{CBA} - \boldsymbol{\Theta}_0)'[\mathbf{C}(\ddot{\mathbf{X}}'\mathbf{W}\ddot{\mathbf{X}})^{-1}\mathbf{C}']^{-1}(\mathbf{CBA} - \boldsymbol{\Theta}_0).$$

Thus the eigenvalues of $\boldsymbol{\Delta}^* = (\mathbf{A}'\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{H}^*$, and hence the λ_i^*, are based on the q defined design points ($\ddot{\mathbf{X}}$), the q sample-sizes weights ($\mathbf{W}$); the conjectured values for the unknown effects ($\mathbf{B}$), the conjectured common covariance matrix ($\boldsymbol{\Sigma}$), and the specification of the hypothesis ($\mathbf{C}$, $\boldsymbol{\Theta}_0$). Importantly, λ_i^* is not related to N.

To illustrate the method briefly, consider a profile analysis arising from crossing a 3-level between-subjects factor, “Group,” with a 3-level within-subjects (repeated measures) factor, “Test.” The ith subject will provide three observations, $\mathbf{y} = [y_{i1}\ y_{i2}\ y_{i3}] = [Test_{i1}\ Test_{i2}\ Test_{i3}]$. With N subjects total and taking the ith subject to be in the second group, the cell means formulation of $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$ is



φ:
E
L
G
X
0.90
0.88
0.86
0.84
0.82
0.80
Power
F_{T_1}
F_{T_2}
F_U
F_V
0.0
0.5
1.0
1.5
Coefficient of Variation

Figure 4.

$$\begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ \vdots & \vdots & \vdots \\ y_{i1} & y_{i2} & y_{i3} \\ \vdots & \vdots & \vdots \\ y_{N1} & y_{N2} & y_{N3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_{33} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \vdots & \vdots & \vdots \\ \varepsilon_{i1} & \varepsilon_{i2} & \varepsilon_{i3} \\ \vdots & \vdots & \vdots \\ \varepsilon_{N1} & \varepsilon_{N2} & \varepsilon_{N3} \end{bmatrix}.$$

Thus, $\ddot{\mathbf{X}}$ is the 3×3 identity matrix. The sample size weights are to be $w_1 = .250$, $w_2 = .375$, and $w_3 = .375$, giving the elements of the diagonal matrix **W**. Two scenarios for **B** are

$$\mathbf{B}_{(1)} = \begin{bmatrix} 97 & 110 & 97 \\ 95 & 100 & 110 \\ 102 & 95 & 105 \end{bmatrix} \text{ and } \mathbf{B}_{(2)} = \begin{bmatrix} 97 & 110 & 97 \\ 100 & 100 & 100 \\ 102 & 95 & 105 \end{bmatrix}.$$

The conjectured within-group standard deviations for y_{i1}, y_{i2}, and y_{i3} are 15, 20, and 15. The within-group correlations are $\rho_{12} = 0.30$, $\rho_{13} = 0.60$, $\rho_{23} = 0.30$. Thus

$$\boldsymbol{\Sigma} = \begin{bmatrix} 15 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 1 & .30 & .60 \\ .30 & 1 & .30 \\ .60 & .30 & 1 \end{bmatrix} \begin{bmatrix} 15 & 0 & 0 \\ 0 & 20 & 0 \\ 0 & 0 & 15 \end{bmatrix} = \begin{bmatrix} 225 & 90 & 135 \\ 90 & 400 & 90 \\ 135 & 90 & 225 \end{bmatrix}.$$

Comparing the test profiles across the three groups (the Group $\times$ Test interaction) can be specified with H_0: **CBA** = **0** where

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

For $\mathbf{B}_{(1)}$, these specifications lead to $\boldsymbol{\phi}^* = \{.278, .134\}$, which is almost an linear/geometric eigenvalue structure. The primary noncentralities are $\lambda_U^* = .407$, $\lambda_T^* = .412$, $\lambda_V^* = .403$. For $\alpha = .05$ and N = 48, the approximate powers are $\pi_U = .949$, $\pi_{T_1} = .951$, $\pi_{T_2} = .943$, and $\pi_V = .947$. For $\mathbf{B}_{(2)}$, $\boldsymbol{\phi}^* = \{.181, .004\}$, which is less than for $\mathbf{B}_{(1)}$ and close to being an extreme eigenvalue structure. Accordingly, the primary noncentralities and powers (N = 48) are now lower and more varied: $\lambda_U^* = .178$, $\lambda_T^* = .185$, $\lambda_V^* = .171$; $\pi_U = .610$, $\pi_{T_1} = .630$, $\pi_{T_2} = .612$, and $\pi_V = .590$. For N = 96, the powers are $\pi_U = .923$, $\pi_{T_1} = .937$, $\pi_{T_2} = .929$, and $\pi_V = .911$. Note that the nominal Type II error rate for F_V is 41% greater than for F_{T_1}, which should become the prescribed statistic for the written protocol if $\mathbf{B}_{(2)}$ were the primary scenario. Other factors, especially robustness, might mitigate against such a decision. For instance, Olson (1974) concluded that V was the more robust statistic.

6. CONCLUSION

The value of the proposed strategy stems from several factors. (1) This is an all-in-one method that relates directly and simply to the familiar F transforms of the U, T, and V statistics. These Fs are taken to be noncentral Fs with their usual degrees of freedom. Computing their noncentrality parameters is isomorphic to computing the F statistics on population values instead of sample values. As described fully in O'Brien and Muller (1993), exploiting the correspondence between a familiar test statistic and its noncentrality parameter gives intuition and pragmatism to power analysis. This is not the case for the more abstruse methods that have been proposed over the years for numerous special cases involving U, T, and V directly (see Krishnaiah, 1978). (2) The method gives exact powers when applied to any test having $s = 1$. (3) For $s > 1$, the method affords a unifying set of asymptotic results that support its large-sample correctness across the various F transforms. (4) For $s > 1$, but with smaller sample sizes, the empirical work summarized and graphed above shows the method to be sufficiently accurate for almost all applied work. Although methods that are more exact have been developed for special situations involving U, T and V, such work does not extend to the approximate F_U, F_{T_1}, F_{T_2}, and F_V statistics being used so predominately today.

In conclusion, we offer a way to compute power probabilities for the multivariate general linear hypothesis that is simple, general, and accurate. We hope these qualities motivate statistical planners to perform power analyses that are more congruent with the multivariate hypotheses they propose to test.

References

Agresti, A. (1990), *Categorical Data Analysis* , New York: John Wiley.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis* (2nd ed.), New York: Wiley.

Barton, C. N. and Cramer, E. C. (1989), “Hypothesis Testing in Multivariate Linear Models with Randomly Missing Data,” *Communications in Statistics—Simulation and Computations,* 18, 875-895.

Bartlett, M. S. (1939), “A Note of Tests of Significance in Multivariate Analysis,” *Proceedings of the Cambridge Philosophical Society,* 35, 180-185.

Hughes, J. T. and Saw, J. G. (1972), “Approximating the Percentage Points of Hotelling's Generalized T_0^2 Statistic,” *Biometrika,* 59, 224-226.

Krishnaiah, P. R. (1978), “Some Recent Developments on Real Multivariate Distributions,” in *Developments in Statistics* (Vol.1), ed. P. R. Krishnaiah, New York: Academic Press, pp. 135-169.

Kulp, R. W. and Nagarsenker, B. N. (1984), “An Asymptotic Expansion of the NonNull Distribution of Wilks Criterion for Testing the Multivariate Linear Hypothesis,” *Annals of Statistics,* 12, 1576-1583.

Lee, Y. S. (1971), “Asymptotic Formulae for the Distribution of a Multivariate Test Statistic: Power Comparisons of Certain Multivariate Tests,” *Biometrika,* 58, 647-651.

McKeon, J. J. (1974), “F Approximations to the Distribution of Hotelling's T_0^2,” *Biometrika,* 61, 381-383.

Muller, K. E. and Barton, C. N. (1989), “Approximate Power for Repeated-Measures ANOVA Lacking Sphericity,” *Journal of the American Statistical Association,* 84, 549-555.

Muller, K. E., LaVange, L. M., Ramey, S. L., and Ramey, C. T. (1992), “Power Calculations for General Linear Multivariate Models Including Repeated Measures Applications,” *Journal of the American Statistical Association*, 87, 1209-1226.

Muller, K. E. and Peterson, B. L. (1984), “Practical Methods for Computing Power in Testing the Multivariate General Linear Hypothesis,” *Computational Statistics & Data Analysis*, 2, 143-158.

Nanda, D. N. (1950), “Distribution of the Sum of Roots of a Determinantal Equation Under a Certain Condition,” *Annals of Mathematical Statistics,* 21, 432-439.

O'Brien, R. G. (1986), "Using the SAS System to Perform Power Analysis for Log-Linear Models," in *Proceedings of the Eleventh Annual SAS Users Group International (SUGI),* Cary, NC: SAS institute, 778-784.

O'Brien, R. G. and Muller, K. E. (1993), "Unified Power Analysis for t-Tests through Multivariate Hypotheses," in Edwards, L. K. (ed.), *Applied Analysis of Variance in Behavioral Science,* New York: Marcel Dekker, 297-344.

Olson, C. L. (1974), "Comparative Robustness of Six Tests in Multivariate Analysis of Variance," *Journal of the American Statistical Association,* 69, 894-908.

Pillai, K. C. S. (1955), "Some New Test Criteria in Multivariate Analysis," *Annals of Mathematical Statistics,* 26, 117-121.

Pillai, K. C. S. and Mijares, T. A. (1959), "On the Moments of the Trace of a Matrix and Approximations to its Distribution," *Annals of Mathematical Statistics,* 30, 1135-1140.

Pillai, K. C. S. and Samson, P., Jr. (1959), "On Hotelling's Generalization of T^2," *Biometrika,* 46, 160-168.

Rao, C. R. (1951), "An Asymptotic Expansion of the Distribution of Wilks's Criterion, *Bulletin of the International Statistical Institute,* 33, 177-180.

Searle, S. R. (1971), *Linear Models,* New York: John Wiley.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics,* New York: John Wiley.

Seber, G. A. F. (1984), *Multivariate Observations,* New York: John Wiley.

Schatzoff, M. (1966), "Sensitivity Comparisons Among Tests of the General Linear Hypothesis," *Journal of the American Statistical Association,* 61, 415-435.

Sugiura, N. and Fujikoshi, Y. (1969), "Asymptotic Expansions of the Non-Null Distributions of the Likelihood Ratio Criteria for the Multivariate Linear Hypothesis and Independence," *Annals of Mathematical Statistics,* 40, 942-952.

van der Merwe, L. and Crowther, N. A. S. (1984), "An Approximation to the Distribution of Hotelling's Generalized T_0^2 -Statistic," *South African Statistical Journal,* 18, 68-90.

Wilks, S. S. (1932), "Certain Generalizations in the Analysis of Variance," *Biometrika,* 24, 471-494.