k-SAMPLE t-TEST REDUX

PATRICK S. CARMACK†

Department of Mathematics
University of Central Arkansas
201 Donaghey Avenue, Conway, AR 72035-5001, USA

JEFFREY S. SPENCE

Department of Internal Medicine, Epidemiology Division
University of Texas Southwestern Medical Center at Dallas
5323 Harry Hines Boulevard, Dallas, TX 75390-8874, USA

WILLIAM R. SCHUCANY

Department of Statistical Science
Southern Methodist University
P.O. Box 750332, Dallas, TX 75275-0332, USA

Abstract. Testing linear combinations of normal group means with unequal variances and possibly unequal samples sizes using a t-statistic is a well known problem for which the effective degrees of freedom are unknown. Satterthwaite (1946) and Welch (1947) set forth their approximations based on the ratio of combinations of fourth moments, and their methodology remains the de facto standard. We propose a new approach that extends their results by defining what we call Measure of Equivalent Exchange (MEE), which is a function of a tuning parameter, τ, and the eigenvalues of a positive semidefinite matrix. We establish a number of important properties that MEE possesses and demonstrate its efficacy with simulations involving two and three group contrasts. We conclude with some commentary about the potential applications of MEE in other areas such as spatially correlated data and general covariance cross-validation.

E-mail addresses: patrickc@uca.edu, jeffrey.spence@utsouthwestern.edu, schucany@smu.edu.

Date: January 22, 2013.

Keywords. Behrens-Fisher problem, Satterthwaite's approximation, Welch, effective degrees of freedom, eigenvalues, covariance matrices, normality.

†Corresponding author.

1. Introduction

Comparing two or more group means of Gaussian distributions with different variances and potentially different sample sizes is one of the longest standing problems in statistics. Given a sample of k groups, consider the statistic given by

$$t = \frac{\sum_{i=1}^{k} c_i \overline{x}_i}{\sqrt{\sum_{i=1}^{k} \frac{c_i^2 s_i^2}{n_i}}},$$

where n_i is the sample size, $\overline{x}_i$ is the sample mean, and s_i is the sample standard deviation of the i^{th} group, $i = 1, \ldots, k$. The vector $\mathbf{c}' = (c_1, \ldots, c_k)$ is commonly known as a contrast which is subject to the restriction $\sum_{i=1}^{k} c_i = 0$ for estimability reasons. The difficulty is that the distribution of t under the null hypothesis that all the group means are equal is, in general, analytically intractable. This is more commonly known as the Behrens-Fisher problem, which formally has its roots in Fisher (1935) who cited earlier work by Behrens (1929).

A little over a decade later, Satterthwaite (1946) and Welch (1947) addressed it by approximating the distribution of t with student's t-distribution whose degrees of freedom, ν, are approximated by the well known expression

$$\nu \approx \frac{\left(\sum_{i=1}^{k} \frac{c_i^2 s_i^2}{n_i}\right)^2}{\sum_{i=1}^{k} \frac{c_i^4 s_i^4}{n_i^2 (n_i - 1)}}.$$

Although there have certainly been other approaches proposed over the years (Kim & Cohen, 1998), the Welch-Satterthwaite (WS) equation has stood the test of time where it is included in introductory statistics texts and is the default method in many statistical analysis packages such as `R` (R Core Team, 2012) for comparing two sample means. The reason for this is quite simple. The WS approximation works well for a wide variety of combinations of different variances and sample sizes.

In this article, we introduce a function that we call the Measure of Equivalent Exchange (MEE) whose arguments are linear in the eigenvalues of a positive semidefinite matrix in combination with a nonlinear tuning parameter, τ. We demonstrate some interesting properties that MEE possess including a one-to-one correspondence between the tuning parameter τ and the cumulants of the χ^2 distribution in the k-sample case assuming that the data are independently and normally distributed. This reveals that MEE can be viewed as the logical continuation of the pioneering work of Satterthwaite and Welch. We then demonstrate MEE's efficacy via simulation of contrasts of two and three group samples comparing the new approach against the WS method. We conclude with potential applications for MEE in spatially correlated data and generalized covariance cross-validation.

2. Measure of Equivalent Exchange (MEE)

Initially, the Measure of Equivalent Exchange (MEE) function can be motivated geometrically. Covariance matrices are commonly described geometrically as ellipsoids. Even though an ellipsoid may be of a given dimension, it may not occupy that space as fully as a sphere of the same dimension and similar scale. For example, the ellipsoid associated with a 3×3 covariance matrix, in which two of the variables are highly collinear, will have one dimension that is close to vanishing, even though the matrix is technically three dimensional. As the two variables become linear combinations of one another, the ellipsoid becomes two dimensional.

2.1. **Definition.** Let $\mathbf{\Lambda}$ be a positive semidefinite $n \times n$ matrix, $\mathbf{\Lambda} \neq \varnothing$, with unique nonzero eigenvalues $\lambda_i, i = 1, \ldots, k$, each with multiplicity m_i and possibly with zero eigenvalues, $\lambda_0 = 0$, with multiplicity m_0. Then, MEE is defined as

$$\begin{aligned} \mathrm{MEE}_\tau \left[\mathbf{\Lambda}\right] &= \sqrt[\tau-1]{\frac{\left(\mathrm{tr}\left[\mathbf{\Lambda}\right]\right)^\tau}{\mathrm{tr}\left[\mathbf{\Lambda}^\tau\right]}}, \tau \geq 0 \\ &= \sqrt[\tau-1]{\frac{\left(\sum_{i=1}^{k} m_i \lambda_i\right)^\tau}{\sum_{i=1}^{k} m_i \lambda_i^\tau}}. \end{aligned}$$

We seek to capture this geometric phenomenon analytically where the loss of a dimension due to scale discrepancies and/or collinearities is modeled in a continuous fashion. Examining the definition, when τ is an integer, the numerator is proportional to the sum of all the possible τ-dimensional ellipsoid cross-sectional volumes of $\mathbf{\Lambda}$, while the denominator is proportional to the sum of the purely spherical volumes of dimension τ. In essence, the spheres provide both scale and dimensional baselines to judge the ellipsoids against. The $\tau - 1$ root simply keeps the end result on the same scale independent of τ – much the same way that one might take the square root of the area of a square or the cube root of the volume of a cube to arrive at the length of a side.

When τ is even and the data are Gaussian with covariance $\mathbf{\Lambda}$, MEE can be viewed as the ratio of traces of higher order moments of the multivariate normal, which are Kronecker products, relative to powers of the trace of the distribution's second moment. As the heteroscedasticity and/or collinearities grow, the trace of the Kronecker products shrink relative to the corresponding powers of the trace of $\mathbf{\Lambda}$.

From an analytical perspective, we can abandon the strictly geometrical interpretation of MEE and allow τ to take on any positive real value. In the context of k independent samples, we see in the properties section that τ has an intimate relationship with the cumulants of the χ^2 on $\mathrm{MEE}_\tau \left[\mathbf{\Lambda}\right]$ degrees of freedom. This connection also shows that MEE subsumes the WS approximation and is arguably the logical continuation of their work.

The cumulants also demonstrate why the $\tau - 1$ root provides the natural scale for MEE.

2.2. **Properties.** While the geometric interpretation provides an intuitive sense of how MEE behaves, its analytic properties concretely establish why it is a valid measure of dimensionality and degrees of freedom (Reviewers: Proofs of the following properties, when necessary, are in the appendix). When possible, `Mathematica` (Wolfram Research, Inc., 2010) is used to verify results, and Bernstein (2009) is used for matrix related statements throughout the article. One such property is the nonpositive partial derivative with respect to τ, namely

$$\frac{\partial \mathrm{MEE}_\tau [\mathbf{\Lambda}]}{\partial \tau} \leq 0.$$

Since MEE is a decreasing function in τ, the maximum and minimum values occur at $\tau = 0$ and as $\tau \to \infty$, respectively. The function can be directly evaluated at $\tau = 0$ giving

$$\mathrm{MEE}_0 [\mathbf{\Lambda}] = \sum_{i=1}^{k} m_i = n - m_0.$$

The asymptotic minimum value as $\tau \to \infty$ is given by

$$\lim_{\tau \to \infty} \mathrm{MEE}_\tau [\mathbf{\Lambda}] = \frac{\mathrm{tr} [\mathbf{\Lambda}]}{\lambda_1} = \frac{\sum_{i=1}^{k} m_i \lambda_i}{\lambda_1},$$

which means that the minimum degrees of freedom as defined by MEE are dictated by the largest eigenvalue or variance component. Taken together, the range of the function is then bounded by

$$m_1 \leq \frac{\sum_{i=1}^{k} m_i \lambda_i}{\lambda_1} = \mathrm{MEE}_\infty [\mathbf{\Lambda}] \leq \mathrm{MEE}_\tau [\mathbf{\Lambda}] \leq \mathrm{MEE}_0 [\mathbf{\Lambda}] = n - m_0.$$

It is also important to note that when there is only one unique nonzero eigenvalue, $\mathrm{MEE}_\tau [\mathbf{\Lambda}] = m_1 = n - m_0, \forall \tau \geq 0$, which is easily verified by direct evaluation. This means that the degrees of freedom as defined by MEE are unique and independent of τ in this special case.

For residuals, $\mathbf{\Lambda}$ commonly takes the form of $\mathbf{R} = (\mathbf{I} - \mathbf{H})' \mathbf{\Sigma} (\mathbf{I} - \mathbf{H})$, where $\mathbf{H}$ is the so-called hat or smoother matrix. If $\mathbf{H}$ is idempotent (i.e., $\mathbf{H}^2 = \mathbf{H}$, which implies that the eigenvalues of $\mathbf{H}$ must be 0's and 1's) and $\mathbf{\Sigma} = \sigma^2 \mathbf{I}$, as is the case for many linear models, then $\mathrm{MEE}_\tau [\mathbf{R}] = \mathrm{tr} [\mathbf{I} - \mathbf{H}] = m_1 = n - m_0, \forall \tau \geq 0$. This is due to the fact that the eigenvalues of $\mathbf{R}$ are 0 with multiplicity $m_0 = \mathrm{tr} [\mathbf{H}]$ and σ^2 with multiplicity $m_1 = n - m_0$. Hence, MEE is in agreement with the traditional definition of degrees of freedom for conventional linear models.

If $1 \leq$ dimension $r \leq \tau$ and the data are normally distributed with covariance $\mathbf{\Sigma}$ as used in Eq. (5), then

$$\kappa_r \left[Q_\tau \sum_{i=1}^{k} \frac{s_i^2}{n_i} \right] \leq \kappa_r \left[X \right],$$

where $\kappa_r [\cdot]$ denotes the r^{th} cumulant of its argument, $Q_\tau = \sqrt[\tau-1]{\sum_{i=1}^k m_i\lambda_i / \sum_{i=1}^k m_i\lambda_i^\tau}$, and $X \sim \chi^2$ on $\text{MEE}_\tau [\mathbf{V}]$ degrees of freedom with equality holding if and only if $r = 1$, $r = \tau$, or there is only one unique nonzero eigenvalue. Hence, in the k-sample case assuming normality, there is a one-to-one correspondence between τ and selecting which cumulant of the χ^2 in addition to $r = 1$ is matched, which is accomplished by factoring $\sum_{i=1}^k m_i\lambda_i = \sum_{i=1}^k \sigma_i^2/n_i$ into scale, $1/Q_\tau$, and degrees of freedom, $Q_\tau \sum_{i=1}^k m_i\lambda_i = Q_\tau \sum_{i=1}^k \sigma_i^2/n_i$. In the simplest case when there is only one non-zero eigenvalue, $1/Q_\tau = \lambda_1$ and $Q_\tau m_1 \lambda_1 = m_1$, $\forall \tau \geq 0$, which means that the partition between scale and degrees of freedom is unique in this case.

Since $\text{MEE}_\tau [a\mathbf{\Lambda}] = \text{MEE}_\tau [\mathbf{\Lambda}], a > 0$, degrees of freedom as defined by MEE are scale invariant, which is a property shared with classical of degrees of freedom.

Given the lower bound of 1 on r and τ stated in the relationship between the cumulants above and the fact that MEE cannot be directly evaluated at $\tau = 1$, the limit is given by

$$\lim_{\tau \to 1} \text{MEE}_\tau [\mathbf{\Lambda}] = \frac{\text{tr}[\mathbf{\Lambda}]}{\prod_{i=1}^k \lambda_i^{w_i}} = \frac{\sum_{i=1}^k m_i\lambda_i}{\prod_{i=1}^k \lambda_i^{w_i}},$$

where

$$w_i = \frac{m_i\lambda_i}{\sum_{j=1}^k m_j\lambda_j}.$$

Hence, the denominator, $\prod_{i=1}^k \lambda_i^{w_i}$, is a weighted geometric mean of the eigenvalues.

2.3. **Selecting τ.** While MEE appears to be a legitimate measure of degrees of freedom particularly in light of the one-to-one cumulant correspondence in the k-sample case, there is still the problem of selecting τ to arrive at a unique definition for degrees of freedom. Examining MEE on the log-scale provides some insight. Taking the natural log of MEE and multiplying through by $\tau - 1$ yields

$$(\tau - 1) \ln (\text{MEE}_\tau [\mathbf{V}]) = \tau \ln \left(\sum_{i=1}^k m_i\lambda_i \right) - \ln \left(\sum_{i=1}^k m_i\lambda_i^\tau \right).$$

Dividing the arguments of the natural log terms on the right hand side by λ_1 and $(\sum_{j=1}^{k}\lambda_j)^\tau$, respectively, and introducing a third term to maintain equality gives

$$(\tau-1)\ln\left(\mathrm{MEE}_\tau\left[\mathbf{V}\right]\right)=\tau\ln\left(\sum_{i=1}^{k}m_i\frac{\lambda_i}{\lambda_1}\right)+\tau\ln\left(\frac{\lambda_1}{\sum_{i=1}^{k}\lambda_i}\right)$$
$$-\ln\left(\sum_{i=1}^{k}m_i\left(\frac{\lambda_i}{\sum_{j=1}^{k}\lambda_j}\right)^\tau\right).$$

Next, we change the sign of the first two terms and invert the arguments of the corresponding natural logs, which results in

$$(\tau-1)\ln\left(\mathrm{MEE}_\tau\left[\mathbf{V}\right]\right)=-\tau\ln\left(\frac{\lambda_1}{\sum_{i=1}^{k}m_i\lambda_i}\right)-\tau\ln\left(\frac{\sum_{i=1}^{k}\lambda_i}{\lambda_1}\right)$$
$$-\ln\left(\sum_{i=1}^{k}m_i\left(\frac{\lambda_i}{\sum_{j=1}^{k}\lambda_j}\right)^\tau\right).$$

Finally, provided that

$$\sum_{i=1}^{k}m_i\left(\frac{\lambda_i}{\sum_{j=1}^{k}\lambda_j}\right)^\tau\approx m_1\left(\frac{\lambda_1}{\sum_{j=1}^{k}\lambda_j}\right)^\tau,$$

then

$$(\tau-1)\ln\left(\mathrm{MEE}_\tau\left[\mathbf{V}\right]\right)\approx-\tau\ln\left(\frac{\lambda_1}{\sum_{i=1}^{k}m_i\lambda_i}\right)-\ln\left(m_1\right). \tag{1}$$

Fig. 1 displays four MEE curves with the configurations indicated in the caption and their native scale approximations based on Eq. (1), which shows that the approximation is asymptotically equivalent to MEE as $\tau\to\infty$.

Statisticians are generally familiar with the expression $-p\ln(p)$, which is known as Shannon's entropy. It was originally introduced as a measure of information content in communications theory with a wide variety of applications including thermodynamics where statistical mechanics are used to model the probabilities of physical phenomena (Stowe, 2007; Volkenshtein, 2009). The $-\tau\ln(\lambda_1/\sum_{i=1}^{k}m_i\lambda_i)$ portion of Eq. (1) echoes Shannon's entropy with τ and $\lambda_1/\sum_{i=1}^{k}m_i\lambda_i$ playing the role of p. This motivates us to define τ as

$$\tau=\frac{(n-k)\,\lambda_1}{\sum_{i=1}^{k}m_i\lambda_i}, \tag{2}$$

where $n-k$ is similar to Boltzmann's constant in Gibbs entropy (Shannon's entropy scaled by Boltzmann's constant), which places the final calculation on the proper scale. It is interesting to note that τ increases as λ_1 grows or its multiplicity declines, which means that the MEE defined degrees of freedom of highly imbalanced systems approach the minimum of $\sum_{i=1}^{k} m_i \lambda_i / \lambda_1$.

Forming a right triangle with legs of length λ_1 and λ_i, the tangent of the angle formed by the leg of length λ_i and the hypothenuse is λ_1/λ_i. If m_i such triangles are formed in this manner for each eigenvalue, τ can be interpreted geometrically as the harmonic mean of the tangent values. The ratio of λ_1 and λ_i can also be viewed as an information exchange rate between the two variances. The harmonic mean is commonly and more appropriately used instead of the arithmetic or geometric mean in problems dealing with rates. As the $k = 2$ and $k = 3$ group simulations demonstrate, Definition (2) works well for a wide variety of different configurations of sample size and variance.

3. Conventional iid Two-Sample Case

In the two-sample case, since MEE is scale invariant, the contrast vector, $\mathbf{c}' = (1, -1)$ is unique up to sign and thus has no further effect on our analysis. Without loss of generality, we assume that $\sigma_1^2/n_1\,(n_1 - 1) \geq \sigma_2^2/n_2\,(n_2 - 1)$, where σ_i^2 and n_i are the variance and sample size of the i^{th} group, respectively, $i = 1, 2$. We seek to generalize degrees of freedom using MEE by encoding $\mathbf{\Lambda}$ to be the matrix associated with the variance of the numerator of the test statistic, in this case $\text{Var}[\overline{x}_1 - \overline{x}_2]$, where $\overline{x}_i$ is the mean of the i^{th} sample. That is, rather than encoding $\mathbf{\Lambda}$ to be the residual matrix $\mathbf{R} = (\mathbf{I} - \mathbf{H})'\mathbf{\Sigma}(\mathbf{I} - \mathbf{H})$ introduced in section 2.2., we define a variance matrix by substituting $\mathbf{T}$ and $\mathbf{S}$, given below, for $\mathbf{I}$ and $\mathbf{H}$, respectively. In this formulation,

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \\ & \sigma_2^2 \mathbf{I}_{n_2} \end{pmatrix},$$

where $\mathbf{I}_n$ denotes the identity matrix of dimension n. Next, we define

$$\mathbf{T} = \begin{pmatrix} \frac{1}{\sqrt{n_1(n_1-1)}} \mathbf{I}_{n_1} & \\ & \frac{1}{\sqrt{n_2(n_2-1)}} \mathbf{I}_{n_2} \end{pmatrix}. \tag{3}$$

Then, the substitute for $\mathbf{H}$ is given by

$$\mathbf{S} = \begin{pmatrix} \frac{1/n_1}{\sqrt{n_1(n_1-1)}} \mathbf{J}_{n_1} & \\ & \frac{1/n_2}{\sqrt{n_2(n_2-1)}} \mathbf{J}_{n_2} \end{pmatrix}, \tag{4}$$

where $\mathbf{J}_n$ is the square matrix of ones of dimension n. Finally, this leads to the variance matrix defined by

$$\mathbf{V} = (\mathbf{T} - \mathbf{S})' \, \mathbf{\Sigma} \, (\mathbf{T} - \mathbf{S}). \tag{5}$$

We can now realize the benefit of defining the variance matrix in this manner. First, it is straightforward to verify that $\text{tr}[\mathbf{V}] = \text{Var}[\overline{x}_1 - \overline{x}_2]$. Second, the eigenvalues of $\mathbf{V}$, needed

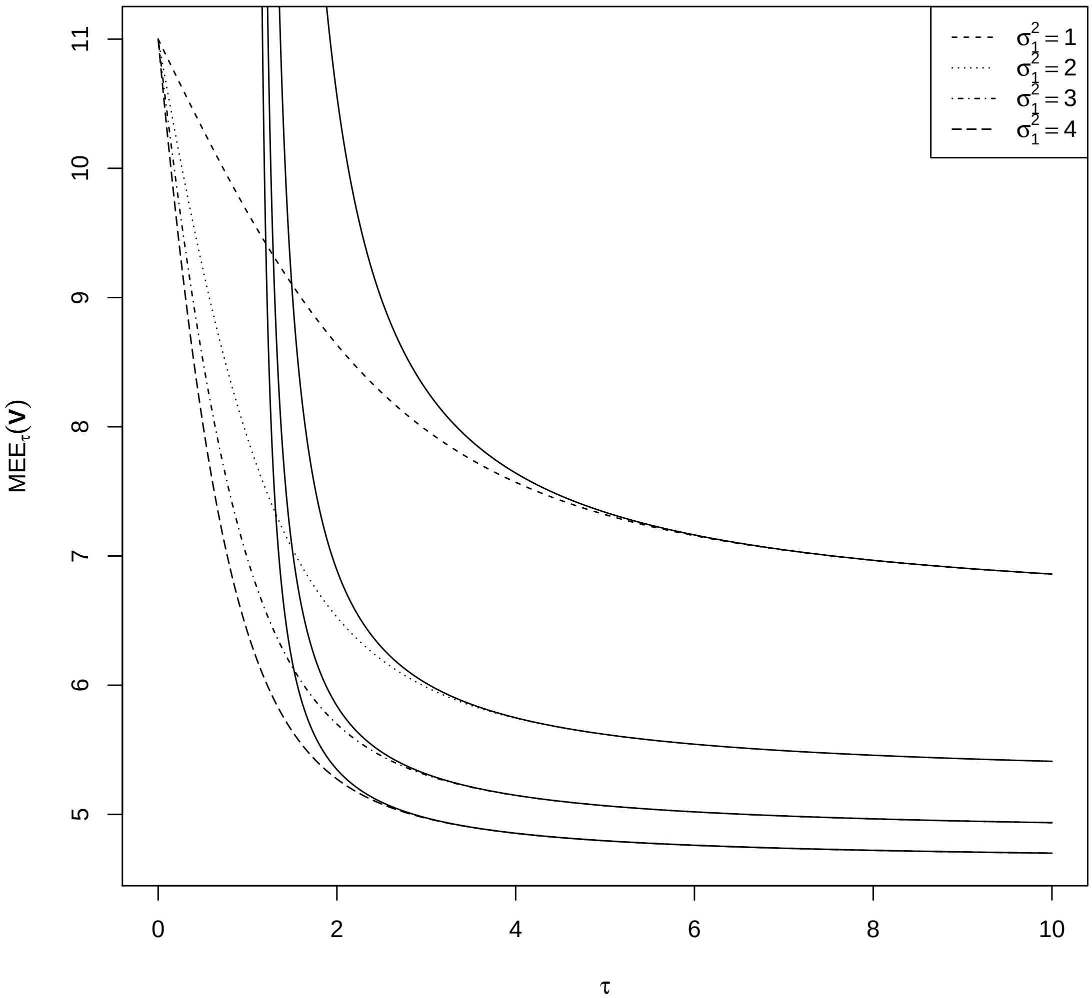

FIGURE 1. Plot of $\mathrm{MEE}_\tau\,[\mathbf{V}]$ as a function of τ with $\mathbf{V}$ as defined in Eq. (5), $n_1 = 5$, $n_2 = 8$, and $\sigma_2^2 = 1$. The legend indicates the four different curves that correspond with $\sigma_1^2 = 1, 2, 3,$ and 4, while the solid curves are the respective native scale approximations of $\mathrm{MEE}_\tau\,[\mathbf{V}]$ given by Eq. (1).

for use with MEE, are given by

$$\begin{aligned} \lambda_0 &= 0, & m_0 &= 2 \\ \lambda_1 &= \tfrac{\sigma_1^2}{n_1(n_1-1)}, & m_1 &= n_1 - 1 \\ \lambda_2 &= \tfrac{\sigma_2^2}{n_2(n_2-1)}, & m_2 &= n_2 - 1. \end{aligned}$$

Hence, MEE may take on the following values in the two-sample iid case.

$$\mathrm{MEE}_\tau\left[\mathbf{V}\right]=\begin{cases}\sqrt[\tau-1]{\dfrac{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^\tau}{(n_1-1)\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)^\tau+(n_2-1)\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)^\tau}}, & \tau\geq 0,\neq 1\\ \dfrac{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}{\left(\frac{\sigma_1^2}{n_1(n_1-1)}\right)^{w_1}\cdot\left(\frac{\sigma_2^2}{n_2(n_2-1)}\right)^{w_2}}, & \tau=1\\ \dfrac{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}{\frac{\sigma_1^2}{n_1(n_1-1)}}, & \tau\to\infty\end{cases},$$

where

$$w_i=\frac{\frac{\sigma_i^2}{n_i}}{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}, i=1,2.$$

Evaluating MEE at $\tau=2$ reveals that the WS approximation occurs as a special case.

$$\mathrm{MEE}_2\left[\mathbf{V}\right]=\frac{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)}+\frac{\sigma_2^4}{n_2^2(n_2-1)}}$$

Casella & Berger (2002) on p. 314 motivate Satterthwaite's approximation from a method of moments perspective. Given the relationship between MEE, the WS approximation, and the cumulants of the χ^2, perhaps it would be more accurately described as an application of method of cumulants. Of course, the two methods are indistinguishable for the first three moments/cumulants.

Fig. 1 shows a plot of MEE as a function of τ for four variance configurations of $k=2$ groups of size 5 and 8. All the curves start at the common maximum value of $m_1+m_2=n_1-1+n_2-1=4+7=11$, but have different minimum asymptotes due to the different variance configurations. It is also interesting to note that all the curves approach their respective asymptotes quickly, and do so at an accelerated rate as the discrepancy between the sample variances increases. Fig. 2 continues the example showing density estimates for $Q_\tau(s_1^2/n_1+s_2^2/n_2)$ in the upper panel and $t=(\overline{x}_1-\overline{x}_2)/\sqrt{s_1^2/n_1+s_2^2/n_2}$ in the lower panel. The upper panel also shows χ^2 density approximations, while the lower panel shows t density approximations where the degrees of freedom for both are calculated using τ in Eq. (2).

3.1. ***k*-Sample Case.** Unlike the two-sample case, the contrast vector, $\mathbf{c}$, for three or more groups has a role to play, though vectors such as $(1,1,-2)$ and $(1/2,1/2,-1)$ are still equivalent due to MEE's scale invariance. However, $\mathbf{c}$ is readily incorporated into the patterns established for the matrices $\mathbf{T}$ and $\mathbf{S}$ in Eqs. (3) and (4) to construct their k-sample analogues by multiplying the i^{th} block matrices in $\mathbf{T}$ and $\mathbf{S}$ by c_i. Without loss

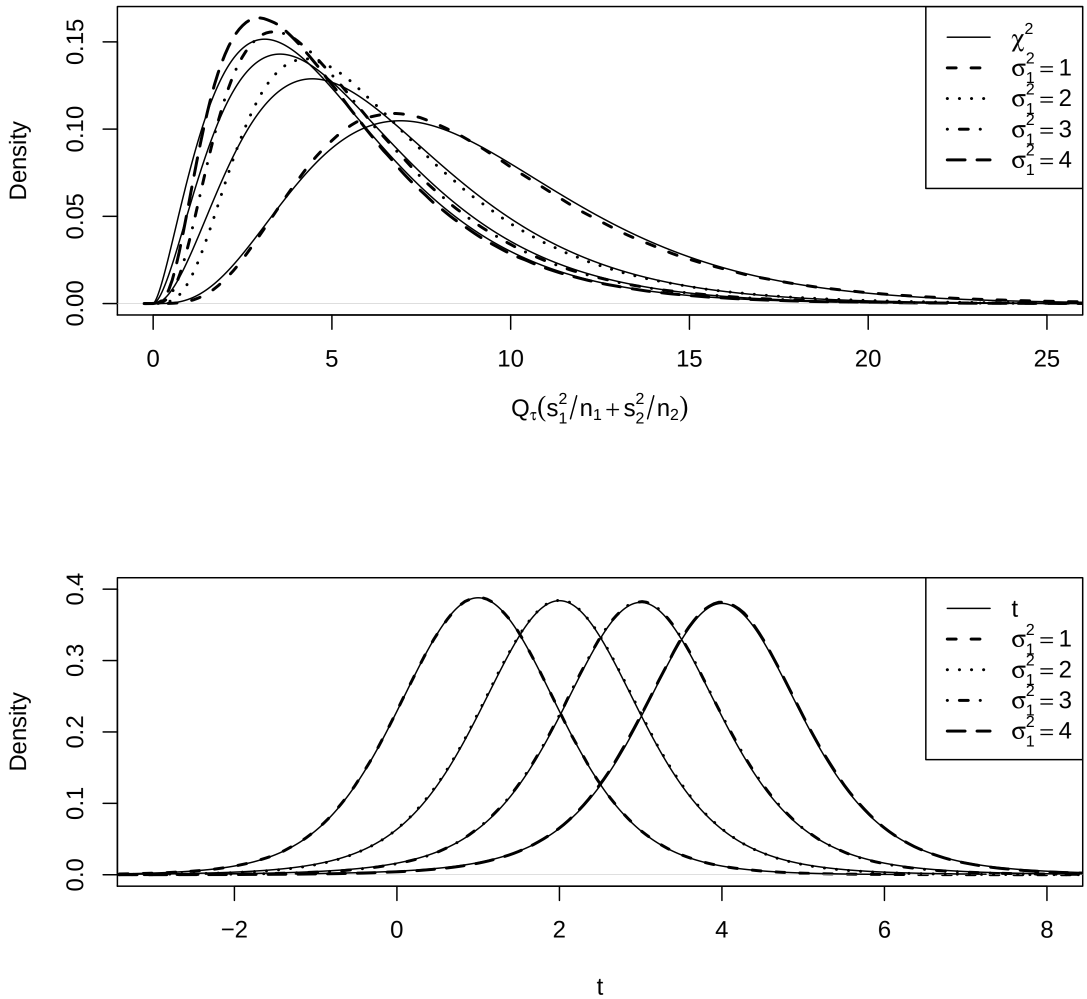

FIGURE 2. The upper panel shows the empirical density estimates for $Q_\tau(s_1^2/n_1 + s_2^2/n2)$ as various dashed outlines where $n_1 = 5$, $n_2 = 8$, $\sigma_2^2 = 1$, and $\sigma_1^2 = 1, 2, 3$, or 4 as indicated in the legend. The solid lines indicate the respective χ^2 density approximations. The lower panel shows empirical density estimates for $t = (\overline{x}_1 - \overline{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$ using the same configurations as the upper panel with approximating t-distribution densities overlaid as solid lines artificially recentered at $1, 2, 3$, and 4, respectively. The tuning parameter τ in each case was determined by Eq. (2) resulting in the degrees of freedom being calculated as $8.91, 6.45, 5.55$, and 5.12, respectively. Each density estimate is based on $b = 10^7$ replications.

of generality, we assume that $c_i^2\sigma_i^2/n_i(n_i-1) \geq c_j^2\sigma_j^2/n_j(n_j-1), 1 \leq i \leq j \leq k$. Then, the eigenvalues of $\mathbf{V}$ as defined in Eq. (5) incorporating $\mathbf{c}$ into $\mathbf{T}$ and $\mathbf{S}$ are given by

$$\begin{array}{llll} \lambda_0 & = 0, & m_0 = k & \\ \lambda_i & = \frac{c_i^2\sigma_i^2}{n_i(n_i-1)}, & m_i = n_i - 1, & i = 1, \cdots, k. \end{array}$$

Substituting these eigenvalues into MEE yields

$$\mathrm{MEE}_\tau\left[\mathbf{V}\right] = \begin{cases} \sqrt[\tau-1]{\dfrac{\left(\sum_{i=1}^k \frac{c_i^2\sigma_i^2}{n_i}\right)^\tau}{\sum_{i=1}^k (n_i-1)\left(\frac{c_i^2\sigma_i^2}{n_i(n_i-1)}\right)^\tau}}, & \tau \geq 0, \neq 1 \\ \dfrac{\sum_{i=1}^k \frac{c_i^2\sigma_i^2}{n_i}}{\prod_{i=1}^k\left(\frac{c_i^2\sigma_i^2}{n_i(n_i-1)}\right)^{w_i}}, & \tau = 1 \\ \dfrac{\sum_{i=1}^k \frac{c_i^2\sigma_i^2}{n_i}}{\frac{c_1^2\sigma_1^2}{n_1(n_1-1)}}, & \tau \to \infty \end{cases},$$

where

$$w_i = \frac{\frac{c_i^2\sigma_i^2}{n_i}}{\sum_{j=1}^k \frac{c_j^2\sigma_j^2}{n_j}}, i = 1, \ldots, k.$$

As with $k = 2$, the variance of $\sum_{i=1}^k c_i\overline{x}_i$ is encoded in the trace of $\mathbf{V}$, that is, $\mathrm{tr}\left[\mathbf{V}\right] = \mathrm{Var}\left[\sum_{i=1}^k c_i\overline{x}_i\right]$. The WS approximation still occurs as a special case when $\tau = 2$, namely

$$\mathrm{MEE}_2\left[\mathbf{V}\right] = \frac{\left(\sum_{i=1}^k \frac{c_i^2\sigma_i^2}{n_i}\right)^2}{\sum_{i=1}^k \frac{c_i^4\sigma_i^4}{n_i^2(n_i-1)}}.$$

4. Simulation Experiments

The simulation experiments consist of two or three samples to compare the efficacy of MEE using our definition of τ versus the WS approximation in `R` (R Core Team, 2012). For $k = 2$, we test the following statistic against the t-distribution, namely

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

The two normally distributed and independent samples are generated under the null hypothesis that the population means of the two groups are equal and tested against a two-tailed alternative. Four configurations of variances are considered with $\sigma_1^2 = 1, 2, 3$, or 4 and $\sigma_2^2 = 1$. The sample size of each sample is varied from 2 to 30 with each combination of n_1, n_2, σ_1^2, and σ_2^2 repeated $b = 10^8$ times.

Like the WS approximation, we use the sample variance in place of the population variance to estimate degrees of freedom using MEE. This is accomplished by estimating the eigenvalues by

$$\hat{\lambda}_i = \frac{s_i^2}{n_i\,(n_i - 1)}, i = 1, 2,$$

which leads to an estimate of our definition of τ being

$$\hat{\tau} = \frac{(n-2)\,\hat{\lambda_1}}{\sum_{i=1}^{2} m_i \hat{\lambda}_i}.$$

For the sake of simplicity, it is understood that the estimated eigenvalues are ordered from largest to smallest on a sample-by-sample basis, which is not necessarily in the same order as the groups. The same convention holds in the $k = 3$ simulations.

For each of the $b = 10^8$ repetitions, the two-sided p-value is calculated using a t-distribution with degrees of freedom estimated by the MEE and WS methods with test level estimates recorded for the commonly used $\alpha = .05$ level of significance. Even if we apply a conservative Bonferroni correction testing the difference between the test level estimates for MEE and WS, nearly all of them are statistically significantly different due to the large number of repetitions. Hence, we do not graphically indicate statistically significant differences in the simulation result figures. Fig. 3 shows filled contour plots of the empirical test level estimates for $k = 2$. The most striking feature when comparing WS in the left column to MEE on the right is that MEE either holds or comes close to holding test level for a greater portion of the included combinations. Most of the improvement comes when one sample is small coupled with a larger one. However, MEE still exhibits some difficulty, albeit to a lesser degree than WS, when the smaller sample size is associated with the larger variance as the incursions on the left side of its filled contours for $\sigma_1^2 \geq 2$ show. There is also the blue strip that becomes more pronounced at the bottom of MEE's filled contours as σ_1^2 increases. As the $k = 3$ cases also demonstrate, when MEE fails to hold level, it has a tendency to be too conservative.

For $k = 3$, we test the following statistic,

$$t = \frac{\overline{x}_1 + \overline{x}_2 - 2\overline{x}_3}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} + 4\frac{s_3^2}{n_3}}},$$

against the t-distribution using degrees of freedom estimated by MEE and the WS approximation. The three samples are independently and normally distributed under the null hypothesis that all three group means are equal. Every combination of $\sigma_1^2 \in \{1, 2, 3, 4\}$, $\sigma_2^2 \in \{1, 2, 3, 4\}$, and $\sigma_3^2 = 1$ is included with the sample sizes allowed to vary between 2 and 16. Due to symmetry, $\sigma_1^2 \leq \sigma_2^2$, and each configuration is repeated $b = 10^8$ times. As with the two-sample simulations, the p-value of the two-sided test calculated using the t-distibution on MEE and WS estimated degrees of freedom and test level estimates

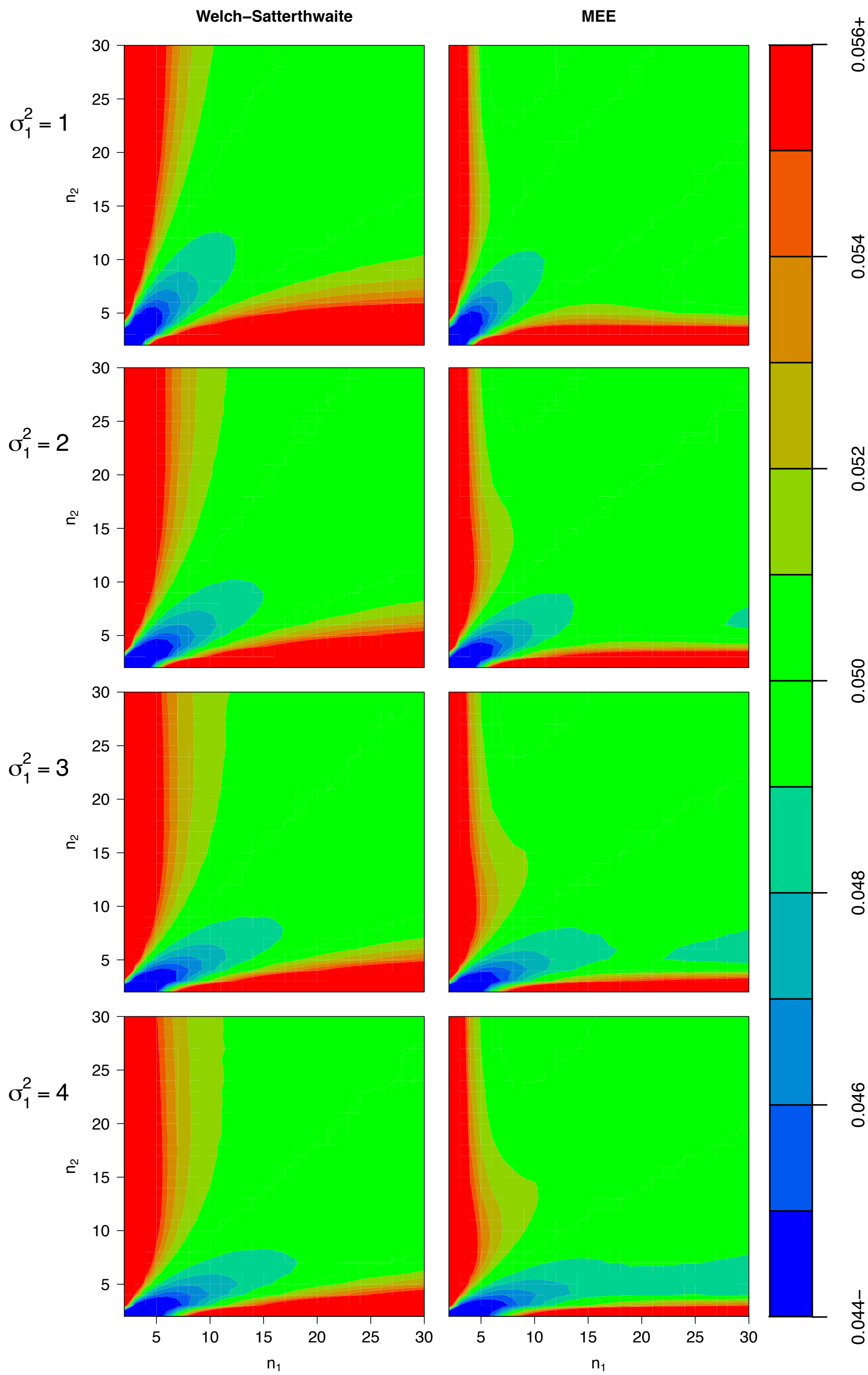

FIGURE 3. Filled contour plots of the $k = 2$ test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1 and n_2 at the $\alpha = .05$ level of significance. $\sigma_2^2 = 1$ for all the plots, while the rows correspond to $\sigma_1^2 = 1, 2, 3,$ and 4, respectively. Each combination of n_1, n_2, σ_1^2, and σ_2^2 is replicated $b = 10^8$ times.

are recorded at several different levels of significance with the $\alpha = .05$ results discussed here.

Since we are dealing with a three group contrast, that must be taken into consideration for both MEE and WS. The eigenvalue estimates are given by $\hat{\lambda}_i = c_i^2 s_i^2 / n_i(n_i - 1), i = 1, 2, 3$, and $\hat{\tau} = (n-3)\hat{\lambda}_1 / \sum_{i=1}^{3} m_i \hat{\lambda}_i$, where $\mathbf{c}' = (1, 1, -2)$. This has the effect that even though $\sigma_3^2 = 1$, the variance contribution of the third group to the numerator of the test statistic is $c_3^2 = 4$ times larger. Hence, the configurations where $\sigma_1^2 = \sigma_2^2 = 4$, in a sense, act as the equal variance cases.

Fig. 4 shows the filled contour plots of the empirical test level estimates at $\alpha = .05$ with WS in the left column and MEE in the right for the selected values of $n_3 \in \{4, 8, 12, 16\}$ and variances $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$ (Reviewers: Please see the additional figures in the appendix.). This particular figure is included due to its combination of equal and unequal variances. As in the $k = 2$ simulations, a perusal of the plots generally shows that MEE holds level for a larger portion of the plots than WS. Many of the gains occur when two of the sample sizes are small while the remaining one is not. The notable exception occurs in the $n_3 = 4$ row where MEE performs much better than WS in the upper right portion of the plot, but becomes too conservative otherwise. The other three rows demonstrate that when MEE fails to hold level, it tends to be too conservative as in the $k = 2$ simulations.

The blue vertical streak on the left side of the MEE plots is associated with smaller values of n_1. Recall that $\sigma_1^2 = 1$ is effectively the smallest of the three sample variances, but that small n_1 makes the eigenvalue associated with this group large. The simulations also capture the estimated values of τ and empirically estimated degrees of freedom that are obtained by maximum likelihood applied to the simulated t-values. A comparison of the two suggests that the estimated values of τ are too large in these cases, which leads MEE to underestimate degrees of freedom in these cases. The same phenomenon occurs in general for all the factor combinations when MEE is too conservative.

The nine other variance configurations for $k = 3$ exhibit similar patterns as the cases discussed above where MEE generally expands the number of cases that approximately hold test level when compared with WS. They also show that when MEE fails to hold test level it is generally too conservative. Overall, extremely small sample sizes of 2, 3, and 4 generally result in extremely liberal test level estimates for both MEE and WS suggesting that they should be avoided when possible in practice.

5. Discussion

As the simulations demonstrate, MEE, in combination with the entropy motivated definition for τ, appears to generally improve upon and continue the work of Satterthwaite and Welch, at least for the cases considered here. The relationship between MEE and the cumulants of the χ^2 in the k-sample case serves to confirm its validity as a measure of

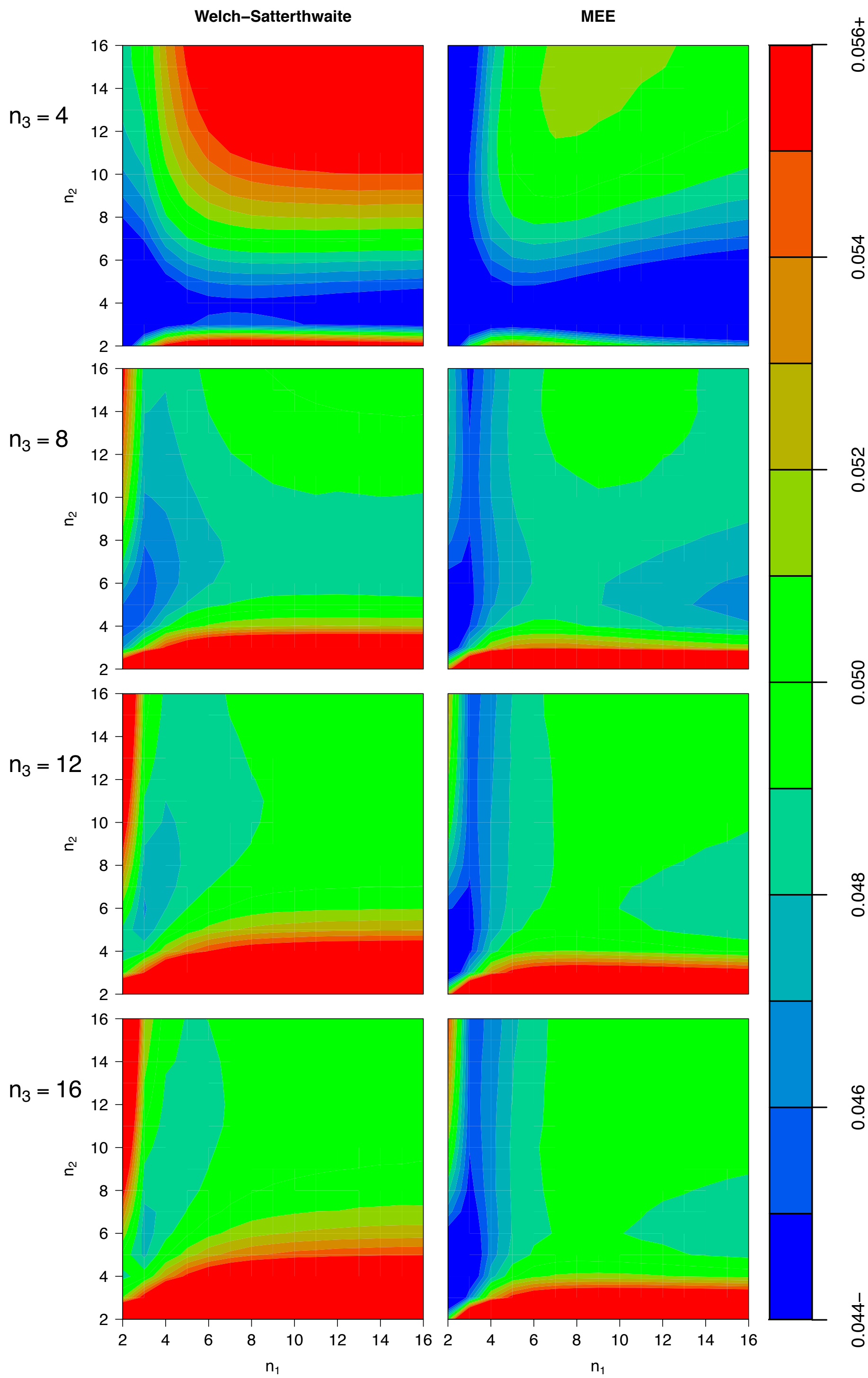

FIGURE 4. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^8$ times.

degrees of freedom and establishes the WS approximation as a special case when $\tau = 2$. At the same time, MEE agrees with traditional degrees of freedom for linear models with idempotent linear smoothers. The definition of τ also connects MEE to the well established bodies of entropy and information theory.

Even though we did not discuss the pooled variance method in the simulation section, we did collect test level estimates for it. We omitted pooled variance due to its abysmal performance in the unequal variance cases. Despite some historical understanding the dangers of preliminary conditional testing to decide between pooled variance and WS (Moser & Stevens, 1992), one occasionally still encounters new textbooks recommending this antiquated approach, particularly in statistics texts written for other disciplines. It is our sincere hope that the improvements that MEE brings over WS will hasten the end of this questionable preliminary testing procedure.

While the present definition of τ works well for a wide variety of variance and sample size combinations, we will continue to conduct research into improving it. Specifically, we want to discover why the present definition tends to produce estimates that are too large in the cases where the test level is too conservative. The improvement will likely come from a deeper understanding of the two terms, $\tau \ln(\sum_{i=1}^{k} \lambda_i/\lambda_1)$ and $\ln(\sum_{i=1}^{k} m_i(\lambda_i/\sum_{j=1}^{k} \lambda_j)^{\tau})$, which largely cancel one another out.

When conducting the basic research for MEE, a seemingly mundane detail turned out to be crucial in the k-sample t-test case. We learned that the variance of the contrast must be encoded directly into the structure of $\mathbf{V}$ instead of using the more traditional residual variance matrix representation. More specifically, the trace of $\mathbf{V}$ is the variance of the contrast. As we apply MEE in a other settings, properly encoding $\mathbf{V}$ will likely continue to play a pivotal role.

Our experience with $\mathbf{V}$ has also made us aware that the traditional meaning of balance needs to be addressed in a more holistic manner using the eigenvalues, which involve the variances and sizes of the samples as well as the contrast. For example, in a design setting with pilot data available, sample sizes can be selected so that the eigenvalues are approximately equal, which makes the task of estimating degrees of freedom far simpler. Of course, this does not address designs with multiple contrasts, which will require further research.

We were careful to define MEE as a general framework separate from the specific context of the k-sample t-test to emphasize that, while it plays a central role in the present context estimating the effective degrees of freedom, MEE can be used in a variety of situations. We are currently conducting research where MEE is applied in the context of spatially correlated data to arrive at an effective sample size. We are simultaneously exploring its use in the context of correlated cross-validation where we hope to derive a criterion that works for general covariance structures.

Acknowledgements

Research reported in this publication was supported by the National Institute Of Biomedical Imaging And Bioengineering of the National Institutes of Health under Award Number R21EB014563. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Behrens, W. (1929). Ein beitrag zur fehlerberechnung bei wenigen beobachtungen. *Landwirtschaftliche Jahrbücher*, 68, 807–837.

Bernstein, D. (2009). *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton University Press.

Casella, G. & Berger, R. (2002). *Statistical Inference*. Belmont, California: Brooks/Cole Publishing Company, 2nd edition.

Fisher, R. (1935). The fiducial argument in statistical inference. *Annals of Eugenics*, 8, 391–398.

Kim, S. & Cohen, A. (1998). On the Behrens-Fisher problem: A review. *Journal of Educational and Behavioral Statistics*, 23(4), 356–377.

Moser, B. & Stevens, G. (1992). Homogeneity of variance in the two-sample means test. *The American Statistician*, 46(1), 19–21.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Satterthwaite, F. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110–114.

Stowe, K. (2007). *An Introduction to Thermodynamics and Statistical Mechanics*. New York: Cambridge University Press.

Volkenshtein, M. (2009). *Entropy and Information*. New York: Springer.

Welch, B. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika*, 34, 28–35.

Wolfram Research, Inc. (2010). *Mathematica, Version 8.0*. Champaign, Illinois: Wolfram Research, Inc.

Appendix

Lemma 1. *$\lim_{\tau\to 1} MEE_\tau[\mathbf{\Lambda}] = \frac{\sum_{i=1}^k m_i\lambda_i}{\prod_{i=1}^k \lambda_i^{w_i}} = \frac{tr[\mathbf{\Lambda}]}{\prod_{i=1}^k \lambda_i^{w_i}}$, where $w_i = \frac{m_i\lambda_i}{\sum_{j=1}^k m_j\lambda_j}$.*

Proof.

$$\begin{aligned}
\lim_{\tau\to 1} \sqrt[\tau-1]{\frac{\left(\sum_{i=1}^k m_i\lambda_i\right)^\tau}{\sum_{i=1}^k m_i\lambda_i^\tau}} &= \lim_{\tau\to 1} \exp\left(\frac{\tau\ln\left(\sum_{i=1}^k m_i\lambda_i\right)}{\tau-1} - \frac{\ln\left(\sum_{i=1}^k m_i\lambda_i^\tau\right)}{\tau-1}\right) \\
&= \exp\left(\lim_{\tau\to 1}\left(\frac{(\tau-1)\ln\left(\sum_{i=1}^k m_i\lambda_i\right)}{\tau-1} + \right.\right. \\
&\qquad \left.\left.\frac{\ln\left(\sum_{i=1}^k m_i\lambda_i\right)}{\tau-1} - \frac{\ln\left(\sum_{i=1}^k m_i\lambda_i^\tau\right)}{\tau-1}\right)\right) \\
&= \exp\left(\ln\left(\sum_{i=1}^k m_i\lambda_i\right) - \lim_{\tau\to 1}\frac{\ln\left(\frac{\sum_{i=1}^k m_i\lambda_i^\tau}{\sum_{i=1}^k m_i\lambda_i}\right)}{\tau-1}\right) \\
&= \frac{\sum_{i=1}^k m_i\lambda_i}{\exp\left(\lim_{\tau\to 1}\frac{\ln\left(\frac{\sum_{i=1}^k m_i\lambda_i^\tau}{\sum_{i=1}^k m_i\lambda_i}\right)}{\tau-1}\right)} \\
&= \frac{\mathrm{tr}\left[\mathbf{\Lambda}\right]}{\exp\left(\lim_{\tau\to 1}\frac{\ln\left(\frac{\sum_{i=1}^k m_i\lambda_i^\tau}{\sum_{i=1}^k m_i\lambda_i}\right)}{\tau-1}\right)}
\end{aligned}$$

And now for the denominator of the last expression.

$$\begin{aligned}
\exp\left(\lim_{\tau\to 1}\frac{\ln\left(\frac{\sum_{i=1}^k m_i\lambda_i^\tau}{\sum_{i=1}^k m_i\lambda_i}\right)}{\tau-1}\right) &\overset{*}{=} \exp\left(\lim_{\tau\to 1}\frac{\sum_{i=1}^k m_i\lambda_i^\tau\ln(\lambda_i)}{\sum_{i=1}^k m_i\lambda_i^\tau}\right) \text{ by L'Hôpital's rule} \\
&= \exp\left(\frac{\sum_{i=1}^k m_i\lambda_i\ln(\lambda_i)}{\sum_{i=1}^k m_i\lambda_i}\right) \\
&= \exp\left(\sum_{i=1}^k w_i\ln(\lambda_i)\right)
\end{aligned}$$

$$= \prod_{i=1}^{k} \lambda_i^{w_i}$$

Combining the two yields the desired result. $\square$

Lemma 2. $\lim_{\tau\to\infty} MEE_\tau[\mathbf{\Lambda}] = \frac{\sum_{i=1}^k m_i\lambda_i}{\lambda_1} = \frac{tr[\mathbf{\Lambda}]}{\lambda_1}$

Proof. We deal with the numerator first.

$$\begin{aligned}
\lim_{\tau\to\infty} \sqrt[\tau-1]{\sum_{i=1}^{k} m_i\lambda_i} &= \lim_{\tau\to\infty} \exp\left(\frac{\tau \ln\left(\sum_{i=1}^k m_i\lambda_i\right)}{\tau-1}\right) \\
&= \exp\left(\lim_{\tau\to\infty} \frac{\tau \ln\left(\sum_{i=1}^k m_i\lambda_i\right)}{\tau-1}\right) \\
&\stackrel{*}{=} \exp\left(\lim_{\tau\to\infty} \ln\left(\sum_{i=1}^{k} m_i\lambda_i\right)\right) \text{ by L'Hôpital's rule} \\
&= \sum_{i=1}^{k} m_i\lambda_i \\
&= \operatorname{tr}[\mathbf{\Lambda}]
\end{aligned}$$

Now for the denominator.

$$\begin{aligned}
\lim_{\tau\to\infty} \sqrt[\tau-1]{\sum_{i=1}^{k} m_i\lambda_i^\tau} &= \lim_{\tau\to\infty} \exp\left(\frac{\ln\left(\sum_{i=1}^k m_i\lambda_i^\tau\right)}{\tau-1}\right) \\
&= \exp\left(\lim_{\tau\to\infty} \frac{\ln\left(m_1\lambda_1^\tau \cdot \frac{\sum_{i=1}^k m_i\lambda_i^\tau}{m_1\lambda_1^\tau}\right)}{\tau-1}\right) \\
&= \exp\left(\lim_{\tau\to\infty}\left(\frac{\ln\left(m_1\lambda_1^\tau\right)}{\tau-1} + \frac{\ln\left(\frac{\sum_{i=1}^k m_i\lambda_i^\tau}{m_1\lambda_1^\tau}\right)}{\tau-1}\right)\right) \\
&= \exp\left(\ln\left(\lambda_1\right) + 0\right) \\
&= \lambda_1
\end{aligned}$$

$\square$

Lemma 3. *If $1 \leq r \leq \tau$ and the data are normally distributed with covariance $\mathbf{V}$ as defined in Eq. (5), then*

$$\kappa_r\left[Q_\tau \sum_{i=1}^{k} \frac{s_i^2}{n_i}\right] \leq \kappa_r\left[X\right],$$

where $\kappa_r[\cdot]$ denotes the r^{th} cumulant of its argument, $Q_\tau = \sqrt[\tau-1]{\sum_{i=1}^{k} m_i\lambda_i / \sum_{i=1}^{k} m_i\lambda_i^\tau}$, and $X \sim \chi^2$ on $MEE_\tau[\mathbf{V}]$ degrees of freedom with equality holding if and only if $r = 1$, $r = \tau$, or there is only one unique nonzero eigenvalue.

Proof. By Hölder's inequality,

$$\sum_{i=1}^{k} m_i \lambda_i^{\frac{\tau-r}{\tau-1}} \lambda_i^{\frac{\tau(r-1)}{\tau-1}} \leq \left(\sum_{i=1}^{k} m_i \left(\lambda_i^{\frac{\tau-r}{\tau-1}}\right)^{\frac{\tau-1}{\tau-r}}\right)^{\frac{\tau-r}{\tau-1}} \left(\sum_{i=1}^{k} m_i \left(\lambda_i^{\frac{\tau(r-1)}{\tau-1}}\right)^{\frac{\tau-1}{r-1}}\right)^{\frac{r-1}{\tau-1}}$$

$$\sum_{i=1}^{k} m_i \lambda_i^r \leq \left(\sum_{i=1}^{k} m_i \lambda_i\right)^{\frac{\tau-r}{\tau-1}} \left(\sum_{i=1}^{k} m_i \lambda_i^\tau\right)^{\frac{r-1}{\tau-1}}$$

$$\left(\frac{\sum_{i=1}^{k} m_i\lambda_i}{\sum_{i=1}^{k} m_i\lambda_i^\tau}\right)^{\frac{r}{\tau-1}} \sum_{i=1}^{k} m_i \lambda_i^r \leq \left(\frac{\sum_{i=1}^{k} m_i\lambda_i}{\sum_{i=1}^{k} m_i\lambda_i^\tau}\right)^{\frac{\tau}{\tau-1}} \sum_{i=1}^{k} m_i \lambda_i^\tau$$

$$2^{r-1}\Gamma(r)\, Q_\tau^r \sum_{i=1}^{k} m_i \lambda_i^r \leq 2^{r-1}\Gamma(r)\, \mathrm{MEE}_\tau[\mathbf{V}]$$

$$\kappa_r\left[Q_\tau \sum_{i=1}^{k} \frac{s_i^2}{n_i}\right] \leq \kappa_r[X], X \sim \chi^2_{\mathrm{MEE}_\tau[\mathbf{V}]}$$ □

Similarly, if $r > \tau$, then $\kappa_r[Q_\tau \sum_{i=1}^{k} \frac{s_i^2}{n_i}] \geq \kappa_r[X]$, with equality holding if and only if there is only one unique nonzero eigenvalue.

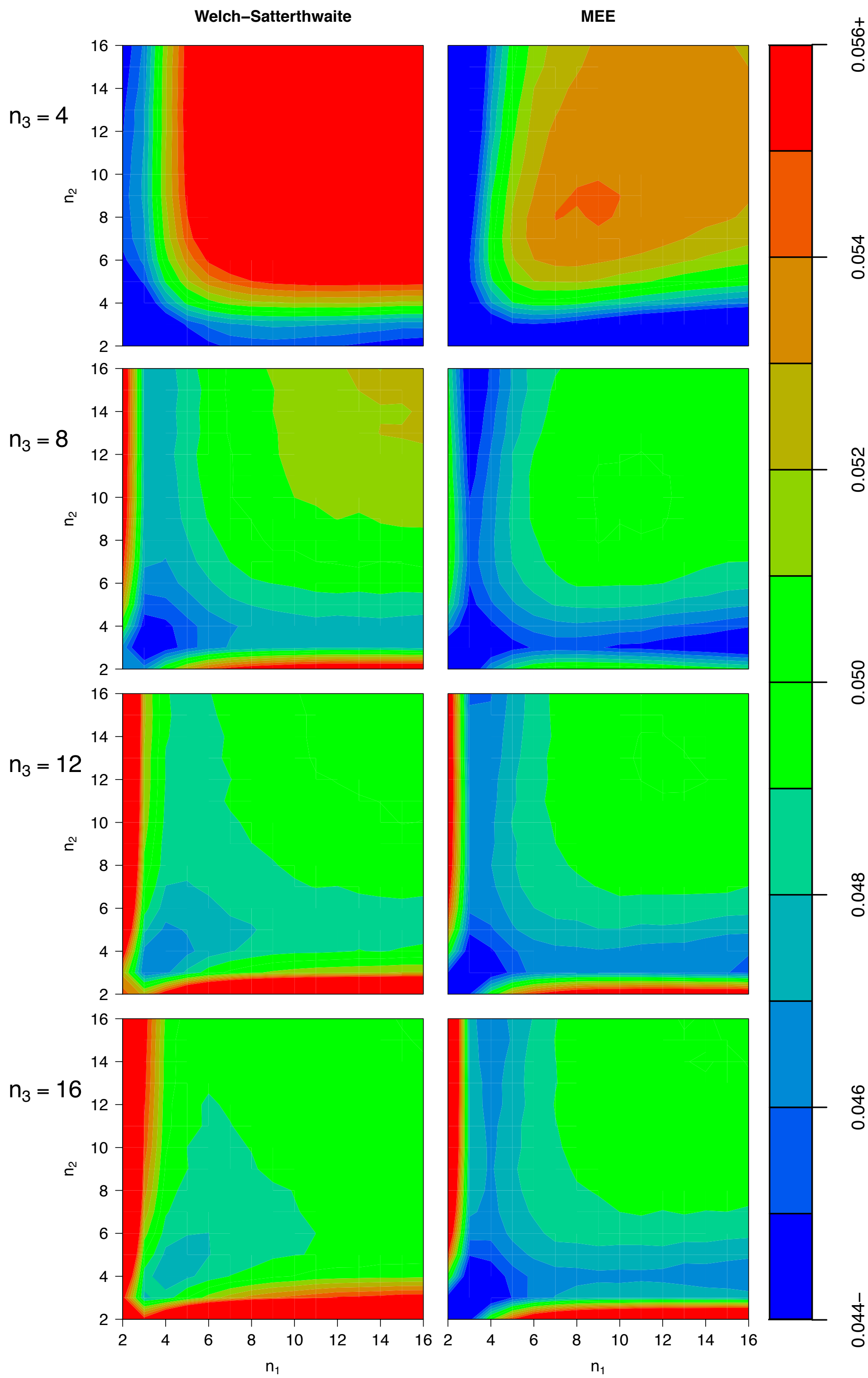

FIGURE 5. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

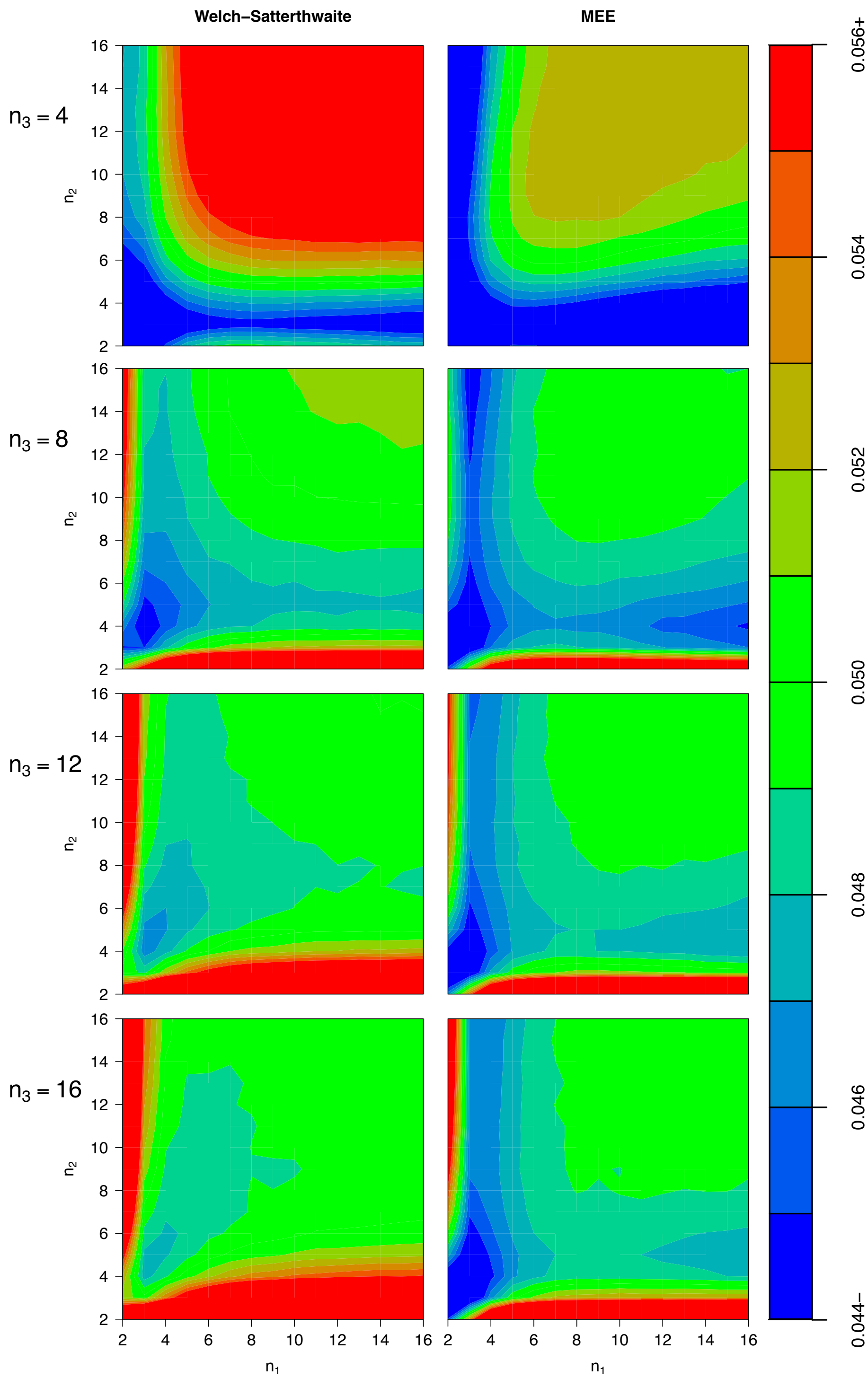

FIGURE 6. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 1$, $\sigma_2^2 = 2$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

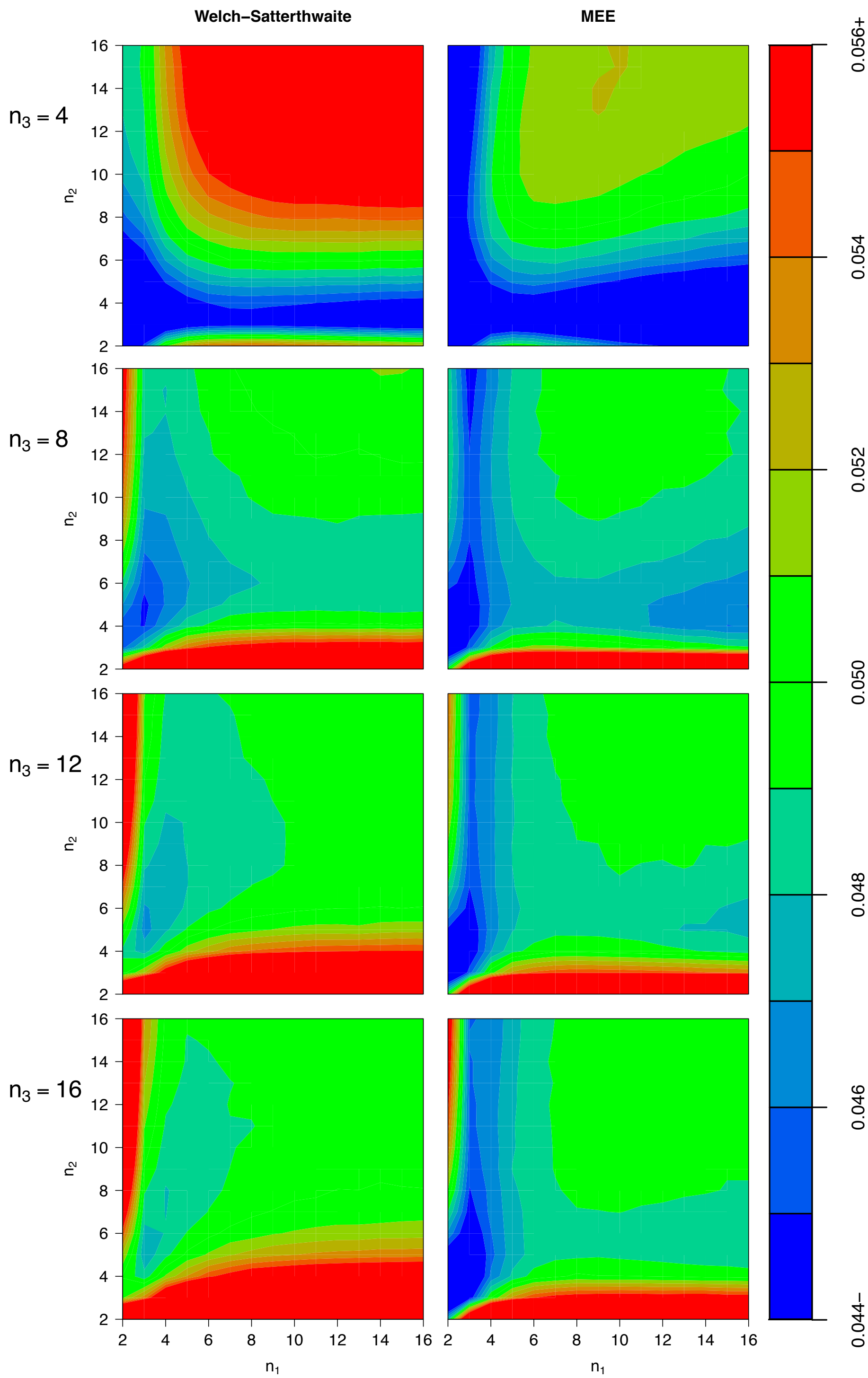

FIGURE 7. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 1$, $\sigma_2^2 = 3$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

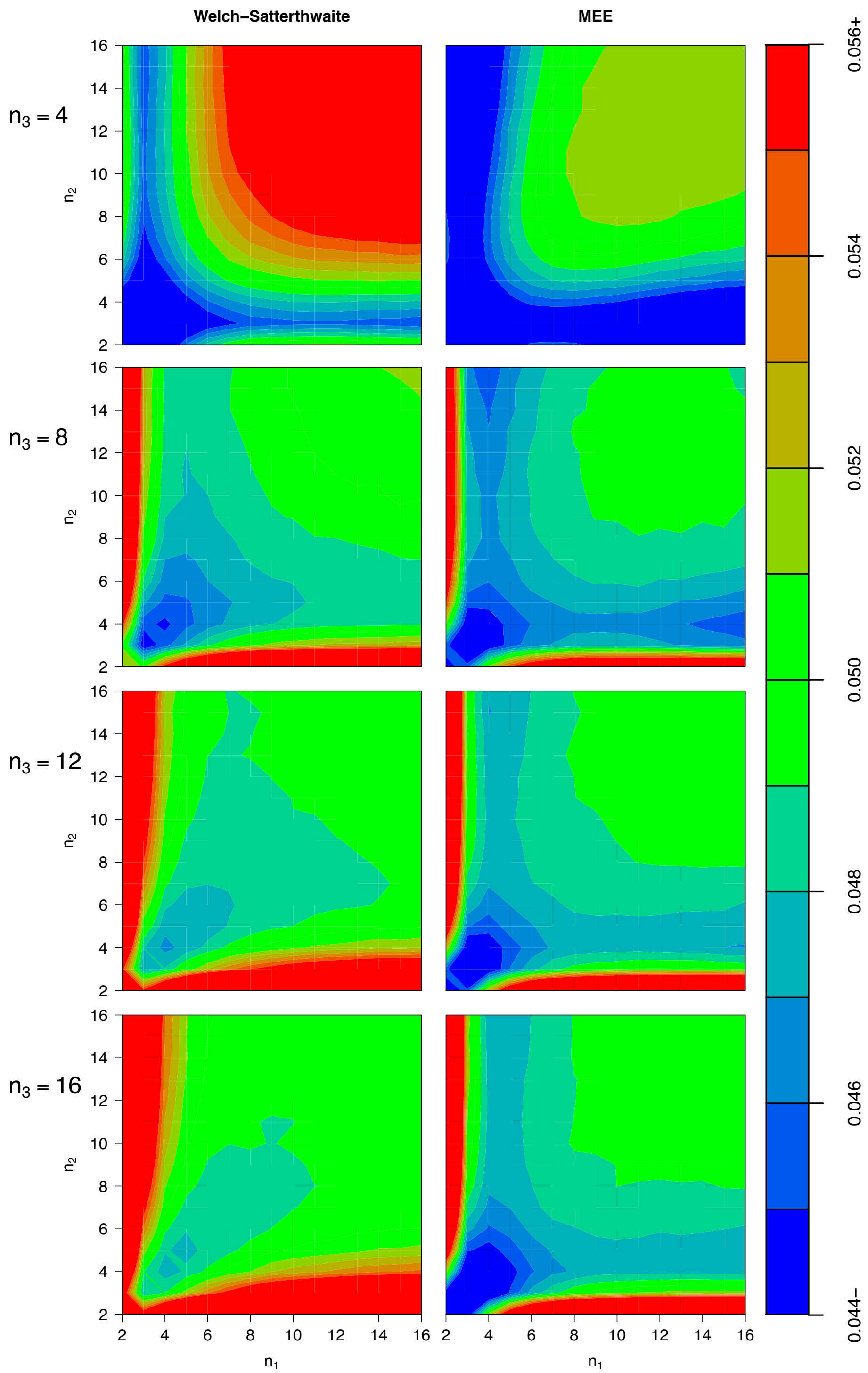

FIGURE 8. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 2$, $\sigma_2^2 = 2$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

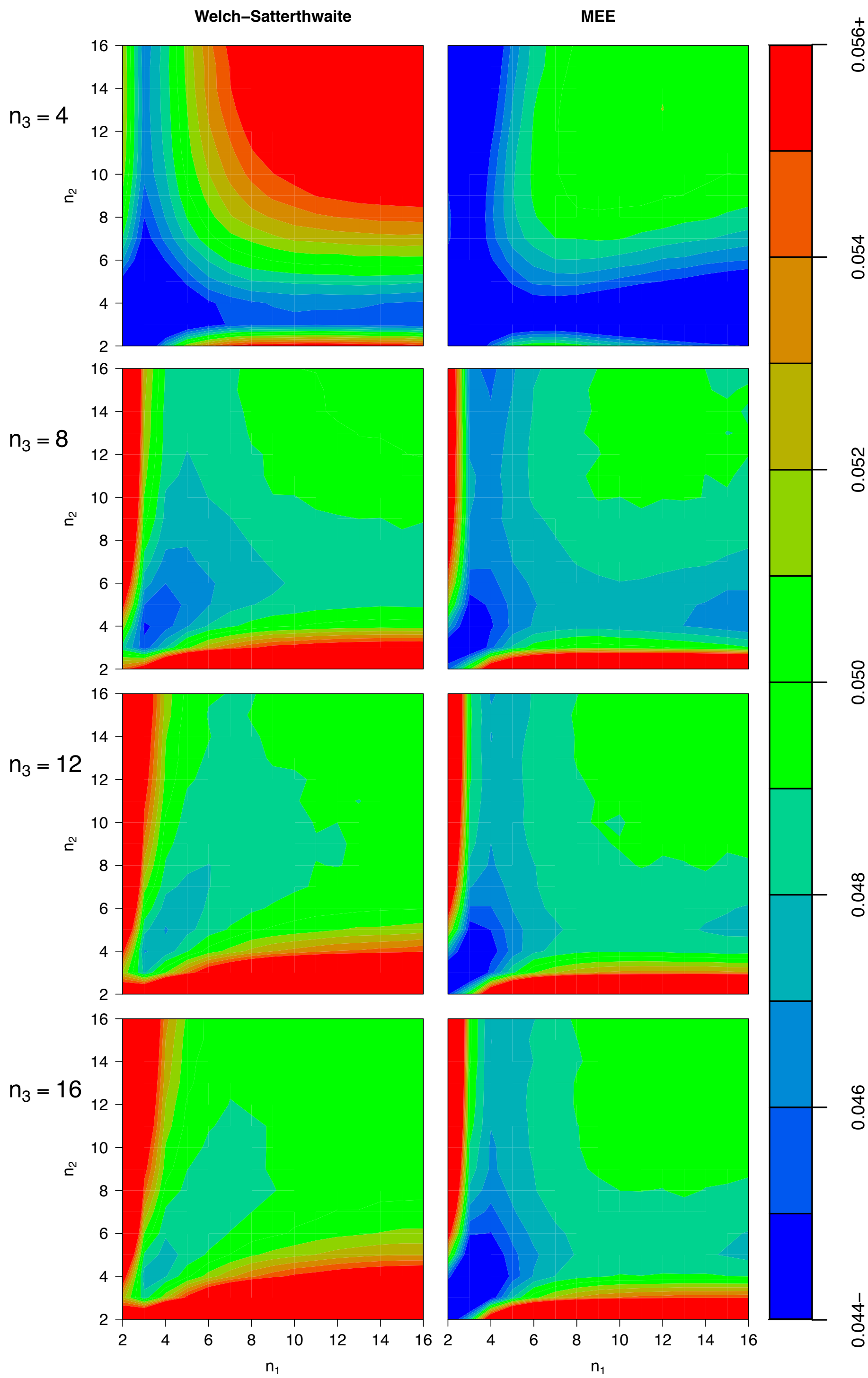

FIGURE 9. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 2$, $\sigma_2^2 = 3$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

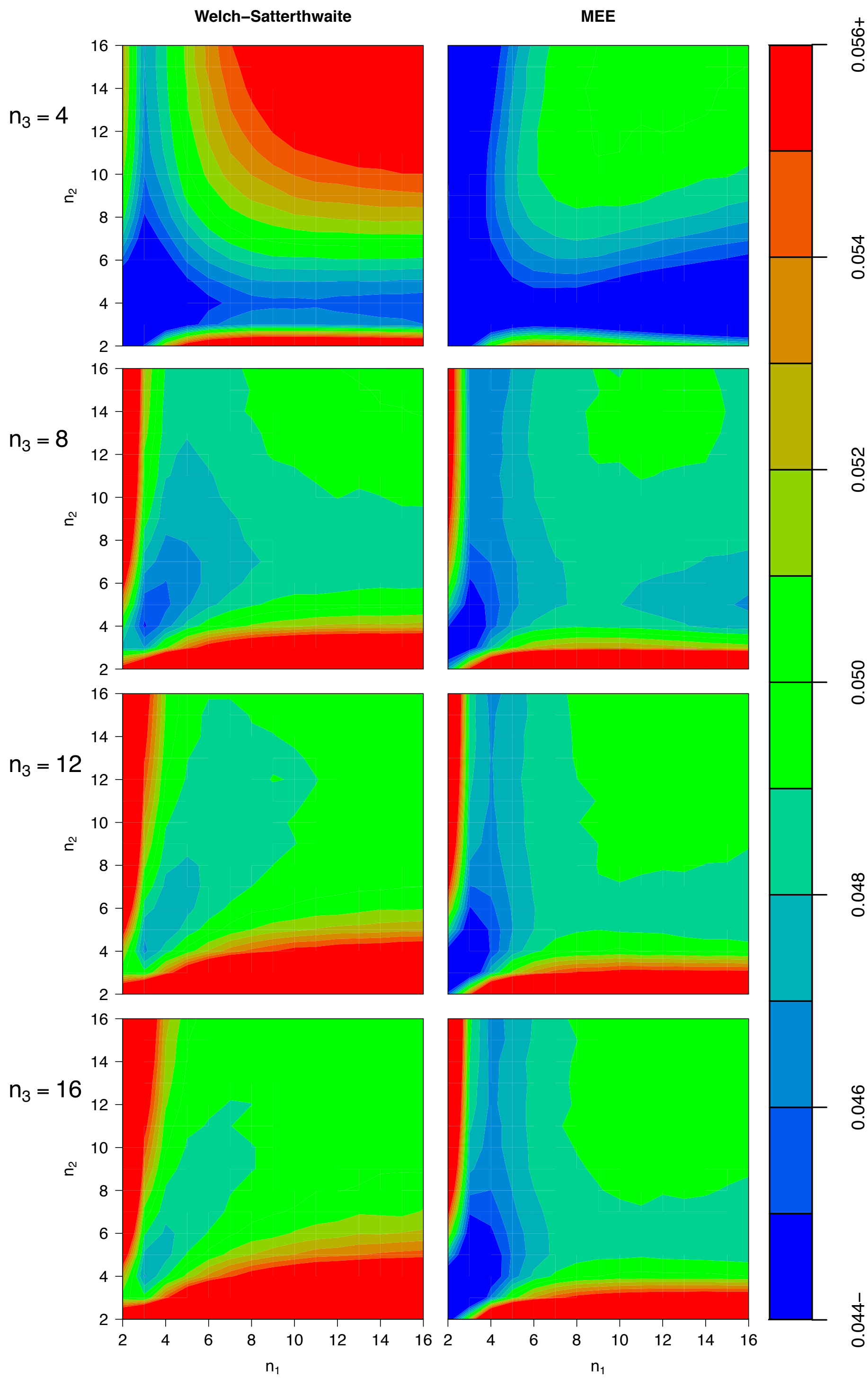

FIGURE 10. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 2$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

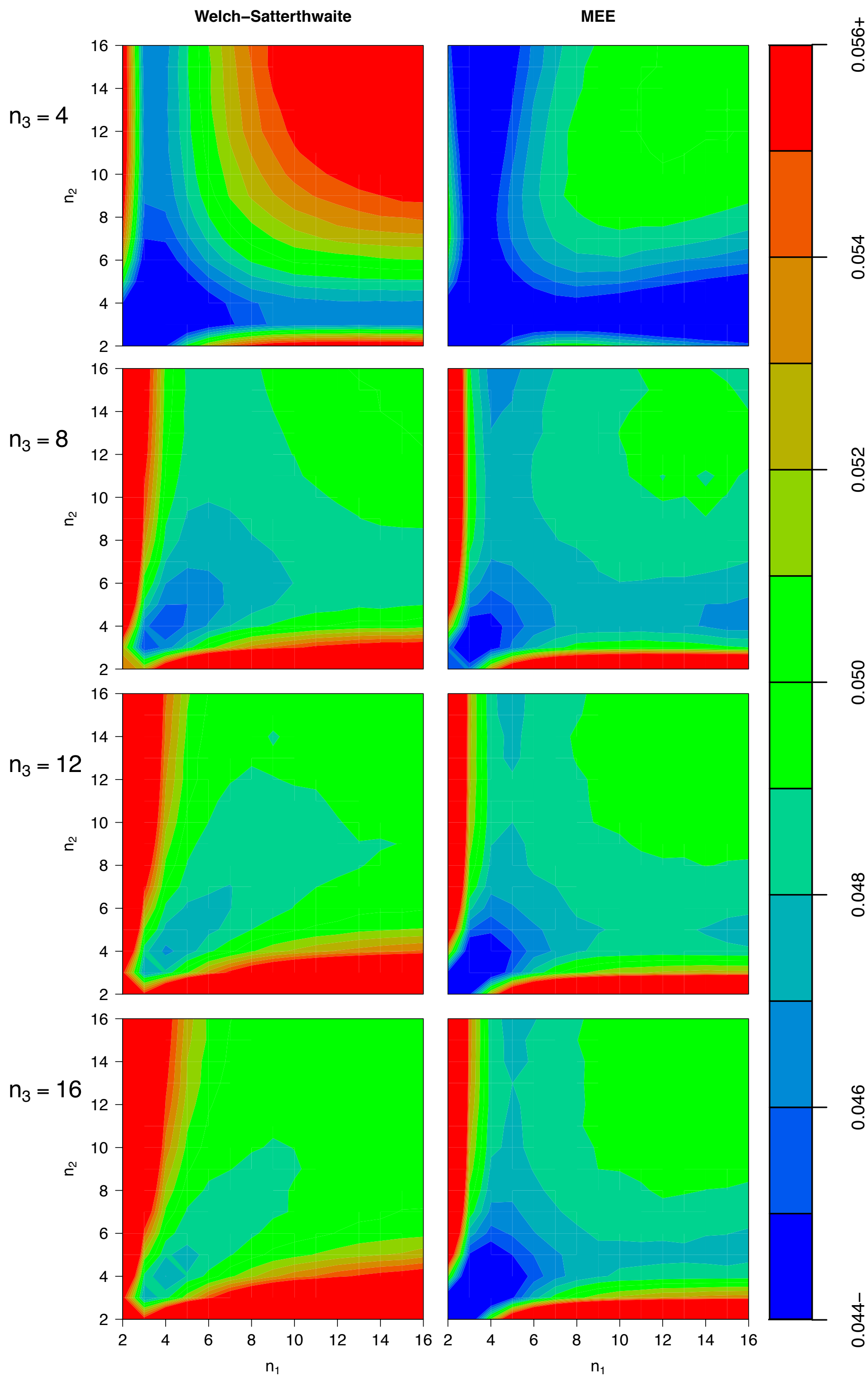

FIGURE 11. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 3$, $\sigma_2^2 = 3$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

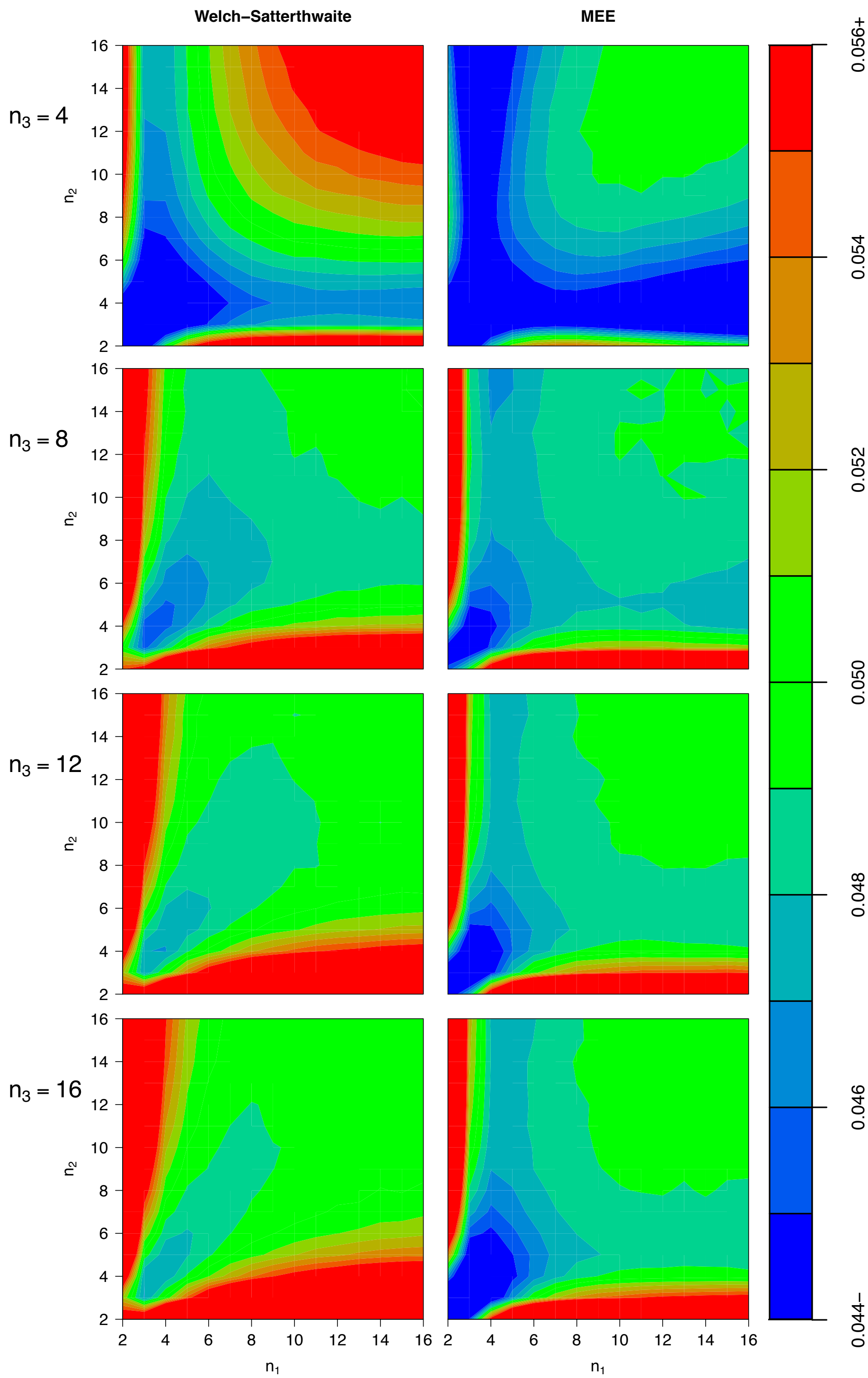

FIGURE 12. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 3$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.

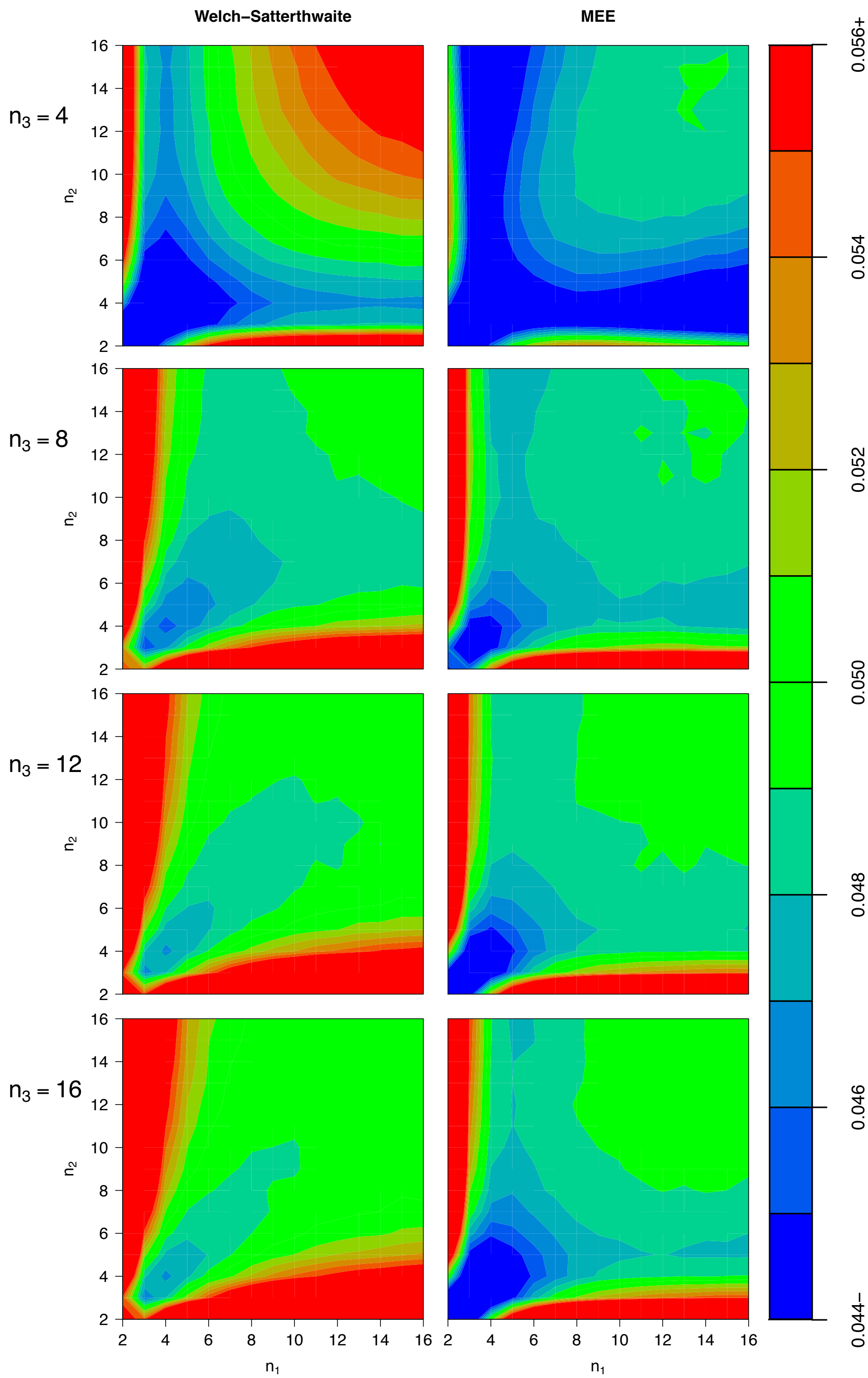

FIGURE 13. Filled contour plots of test level estimates for Welch-Satterthwaite (left column) and MEE (right column) as a function of n_1, n_2, and n_3 at the $\alpha = .05$ level of significance. $\sigma_1^2 = 4$, $\sigma_2^2 = 4$, and $\sigma_3^2 = 1$ for all the plots, while the rows correspond to $n_3 = 4, 8, 12$, and 16, respectively. Each combination is replicated $b = 10^7$ times.