cemmap

centre for microdata methods and practice

A simple four-moment approximation to the distribution of a positive definite quadratic form, with applications to testing

Grant Hillier
Raymond O'Brien

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP02/22

Economic and Social Research Council

A simple four-moment approximation to the distribution of a positive definite quadratic form, with applications to testing.*

Grant Hillier and Raymond O'Brien
University of Southampton

Earliest version: October 2017
This version: December 2021

Abstract

The exact distribution of a quadratic form in n standard normal variables, Q, say, (or, equivalently, a linear combination of independent chi-squared variates) is, except in special cases, quite complicated. This has led to many proposals for approximating the distribution by a more tractable form. These approximations typically exploit the fact that the cumulants of the distribution are quite simple, and include both saddlepoint methods, and methods that replace the actual statistic with a statistic with the same low-order cumulants (or moments). In this paper we propose an approximation of this type that matches the first four moments of the distribution. Its advantage over other methods is that it is extremely easy to implement, and, as we shall show, it is almost as accurate as the best of the other proposed methods (which matches the first eight cumulants). Using the same approach, we also suggest an approximation to the distribution of the analogue of a regression $t-statistic$ in cases where the numerator is standard normal, but the denominator is $\sqrt{Q}$, with Q an independent quadratic form (but not chisquared). This is also shown to work extremely well. The approach has applications in many disciplines, from statistics and econometrics through to theoretical physics.

*Address correspondence to Grant Hillier; email grant.hllr@gmail.com

1 Introduction

We are concerned with statistics of the form

$$Q = \sum_{i=1}^{m} a_i \chi^2(n_i) = y' D y, y \sim N(0, I_n) \tag{1}$$

where $n = \sum_{i=1}^{m} n_i$ and

$$D = \text{diag}\{a_i I_{n_i}\}, i = 1, .., m, \tag{2}$$

with $a_i > 0$ for $i = 1, .., m$. The $\chi^2(n_i)$ denote independent chi-square random variables, the $i-th$ having degrees of freedom n_i. Examples occur frequently in statistics, in many different contexts, and there is an extensive literature on their properties.[1]

The cumulants of Q are quite simple, and are well-known to be

$$\kappa_r = 2^{r-1}(r-1)! p_r, \tag{3}$$

where $p_r = \sum_{i=1}^{m} n_i a_i^r = tr[D^r]$ is the $r-th$ power-sum symmetric function of the the elements of D. However, the exact density and distribution function are quite complicated, and this has motivated an extensive literature on approximations. For a recent survey and many references see Bodenham and Adams (2016) (B&A hereafter).

Before proceeding we note that if $n_i = 1$ for all i the distribution of Q is invariant under permutations of the a_i. This property will be useful in limiting the size of the domain used for some of the evaluation procedures discussed later.

Our first purpose in this paper will be to suggest a simple, accurate, approximation to the distribution of Q that is new. We then provide some evidence on the performance of the suggested approximation, showing that it works extremely well. Next, we discuss an application of the approximation to the related problem of approximating the distribution of certain Student-t-like test statistics, these having the form

$$T = \frac{Z}{\sqrt{Q}},$$

[1] In case the problem demands - as it often does - that some of the a_i are positive, some negative, then Q will be the difference between two statistics of the type we deal with here. Thus, the results are applicable to more general situations than it might at first appear.

with $Z \sim N(0,1)$ and Q independent of Z and defined as in equation (1). This application was suggested by the unpublished work reported in Hansen (2017). Following some correspondence between us, a revised version of Hansen (2017), Hansen (2021), suggests applying our approach to the problem of approximating the distribution of the White t-ratio. Like ourselves, Hansen (2021) finds that this works extremely well. Finally, we extend the Student-t approach to analogues of $F-$statistics of the form $F = Q_1/Q_2$, with both numerator and denominator having the form (1).

We will make extensive use of hypergeometric functions, of both scalar and matrix argument. The reader is referred to Muirhead (1982) for the matrix-argument case. In the case of scalar argument these are defined by the formula

$$ {}_pF_q(b_1, ..., b_p; c_1, ..., c_q; x) = \sum_{i=0}^{\infty} \left[\frac{\Pi_{j=1}^{p}(b_j)_i}{\Pi_{k=1}^{q}(c_k)_i} \right] \frac{x^i}{i!}, \tag{4} $$

where $(c)_i = c(c+1)...(c+i-1)$ is the usual Pochammer symbol. The series converges for all x if $p \leq q$, for $|x| < 1$ if $p = q+1$, and for no x if $p > q+1$. For the most part $p = q = 1$ in our applications.

2 Exact and approximate densities

We first consider the exact density of Q. The derivation given is essentially due to James (1964). In the density of y transform to $x = D^{\frac{1}{2}}y$, so that $Q = x'x$ and $x \sim N(0, D)$,

$$ pdf(x) = (2\pi)^{-\frac{n}{2}} |D|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} x' D^{-1} x \right\}. \tag{5} $$

Transform to $(q = x'x, v = x(x'x)^{-\frac{1}{2}})$. The Jacobian is

$$ (dx) = \frac{1}{2} q^{\frac{n}{2}-1} dq (v'dv), $$

where the last term denotes the (un-normalized) invariant measure on the surface $S_n = \{v \in R^n : v'v = 1\}$, the unit sphere in R^n. To obtain the density of q we integrate out v, but first introduce a tuning parameter $\alpha > 0$, and write the exponent as

$$ -\frac{1}{2} q v' D^{-1} v = -\frac{1}{2} q\alpha + \frac{1}{2} q\alpha v' [I_n - (\alpha D)^{-1}] v $$

Integrating out v gives the expression

$$
\begin{aligned}
pdf_Q(q) &= \frac{|D|^{-\frac{1}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} q^{\frac{n}{2}-1} \exp\{-\frac{1}{2}\alpha q\} \\
&\quad \times {}_1F_1\left(\frac{1}{2}, \frac{n}{2}; \frac{1}{2}\alpha q[I_n - (\alpha D)^{-1}]\right). \qquad (6)
\end{aligned}
$$

The constant α can be chosen to accelerate convergence of the series expansion of the hypergeometric function.[2]

In this case the hypergeometric function is a matrix-argument hypergeometric function (Muirhead (1982), Ch. 7), with a series expansion in terms of zonal polynomials, but in this special case the series involves only the top-order zonal polynomials $C_j(\cdot)$. If α is chosen so that $\alpha D > I$ (elementwise), the density and cdf have representations as infinite discrete mixtures of chi-square densities and cdfs, respectively.[3] That is,

$$
pdf_Q(q) = \sum_{j=0}^{\infty} b_j g_{n+2j}(\alpha q)\alpha dq, \qquad (7)
$$

$$
\Pr\{Q < z\} = \sum_{j=0}^{\infty} b_j G_{n+2j}(\alpha z), \qquad (8)
$$

[2]There are several steps to this result, albeit familiar ones in the multivariate literature. The first is to interpret integration over S_n as over the orthogonal group $O(n)$, with v the first column of $H \in O(n)$. This permits invoking the fundamental identity for zonal polynomials $C_\kappa(\cdot)$, where κ is a partition of k,

$$
\int_{O(n)} C_\kappa(AH'BH)(dH) = C_\kappa(A)C_\kappa(B)/C_\kappa(I_n),
$$

see James (1960). Then one applies a special case of the identity

$$
\frac{C_\kappa(I_m)}{C_\kappa(I_n)} = \frac{(\frac{m}{2})_\kappa}{(\frac{n}{2})_\kappa},
$$

which follows from equation (38) in Constantine (1964). In our case κ is the top partition (k), and $m = 1$, producing equation (6).

[3]Non-negativity of the coefficients need not be imposed, as long as the resulting probabilities are non-negative. In the present case it is evident from the derivation given that this holds for all α.

where $g_v(\cdot)$ and $G_v(\cdot)$ denote the density and cumulative distribution functions, respectively, of a $\chi^2(v)$ variate, and the coefficients

$$b_j = |\alpha D|^{-\frac{1}{2}} \frac{(\frac{1}{2})_j}{j!} C_j \left(I_n - (\alpha D)^{-1} \right) \tag{9}$$

are non-negative and sum to unity. The b_j therefore can be interpreted as probabilities associated with a discrete random variable J (i.e., $b_j = \Pr\{J = j\}$). The distribution of Q thus belongs to the same family as the non-central chi-square distribution, which also has this type of mixture representation, in that case with Poisson weights $b_j = e^{-\frac{1}{2}\lambda}(\lambda/2)^j/j!$.

It is the presence of the top-order zonal polynomials $C_j(\cdot)$ in the distribution that makes it difficult to work with, and interpret, although there are simple recursions for generating these polynomials (see below). However, there is clearly an incentive to approximate, and many approximations for the distribution are available in the literature. The simplest is probably that due to Fisher: treat Q as a multiple of a chi-squared variate, $Q = \tau\chi^2(v)$, choosing τ and v so that the first two cumulants of the exact and approximating distributions agree (also sometimes attributed to Satterthwaite (1946) and Welch (1947)). A generalization of this matching three moments was suggested by Solomon and Stephens (1977). Another simple approximation that matches the first three cumulants is due to Imhof (1961), and Buckley and Eagleson's (1988) application of Hall (1984). Closer to the approach that we suggest is that of Lindsay, Pilla, and Basak (2000), who suggest approximations based on linear combinations of Gamma variates, but these are quite complicated to implement. Finally, one can apply saddlepoint techniques to obtain approximations to the distribution of Q, see Wood, Booth, and Butler (1993) and Butler (2007). Here we suggest an approximation based on a linear combination of two chi-square variates, with parameters chosen to match the first four cumulants of Q. It turns out that this is extremely easy to implement, and our evaluations suggest that it is, for the most part, extremely accurate.

2.1 Note on the top-order zonal polynomials

For an $n \times n$ symmetric matrix A, define the normalized top-order zonal polynomials

$$d_j(A) = \frac{(\frac{1}{2})_j}{j!} C_j(A). \tag{10}$$

The function

$$D(t) = |I_n - tA|^{-\frac{1}{2}} = \sum_{j=0}^{\infty} t^j d_j(A) \tag{11}$$

is a generating function for the d_j, and may easily be used to obtain the recursive relation (omitting the argument matrix)

$$d_j = \frac{1}{2j} \sum_{r=0}^{j} p_r d_{j-r}, d_0 = 1. \tag{12}$$

(see Hillier, Kan, and Wang (2009)). The p_r are again the power-sums $p_r = tr[A^r]$. This recursion has in effect been known since von Neumann (1941), Pitman and Robbins (1949), and Ruben (1962). See also James (1964). It is the basis of the algorithm by Sheil and O'Muircheartaigh, (1977) (hereafter S&O'M). Thus, the coefficients in the mixture representation of the cdf of Q can be generated recursively, and fairly simply, from the power sums p_r, and the eigenvalues of D are not needed. Note, though, that the recursion requires a computation time of $O(j^2)$ for each j, so is not ideal as one moves further into the series (8). A much more efficient recursion based on the elementary symmetric functions is given in Hillier, Kan, and Wang (2009). This is computationally significantly less demanding, because it has length at most n rather than j. In the "exact" calculations used later to assess the accuracy of the appproximations we always use this efficient recursion, rather than that based on power sums.

Remark 1 *It is clear that the density and all properties of Q depend on the $2m$ underlying parameters of the problem, $(a_1, ..., a_m)$ and $(n_1, ..., n_m)$ only through the power-sums $p_r = \sum_{i=1}^{m} n_i a_i^r, r \geq 1$. In view of the results in Hillier, Kan, and Wang (2009), or, equivalently, the Newton-Girard identities, the coefficients in (8) can be generated recursively from the elementary symmetric functions, e_r say, which vanish for $r > n$. In studying the behaviour of any approximation to the exact density, therefore, one obviously needs to explore a parameter space consisting of different configurations of the e_r, $r = 1, .., n$ (not the infinite sequence of power sums).*

2.2 Exact density when $m = 2$

In the case $m = 2$, i.e., $Q = a_1\chi^2_{n_1} + a_2\chi^2_{n_2}$, the exact density is tractable by elementary methods, and is given by (see Appendix A):

$$pdf_Q(q) = \frac{q^{\frac{n}{2}-1}\exp\left\{-\frac{1}{2}qa_1^{-1}\right\}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})a_1^{\frac{n_1}{2}}a_2^{\frac{n_2}{2}}}{}_1F_1\left(\frac{n_2}{2},\frac{n}{2};\frac{1}{2}q\left(a_1^{-1}-a_2^{-1}\right)\right), \tag{13}$$

where $n = n_1 + n_2$. Note that if $a_1 = a_2 = a$, $Q = a\chi^2_n$. Putting $\phi = a_1^{-1}, \psi = a_1/a_2$, the density becomes

$$pdf_Q(q) = \frac{\phi^{\frac{n}{2}}\psi^{\frac{n_2}{2}}\exp\left\{-\frac{1}{2}\phi q\right\}q^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}{}_1F_1\left(\frac{n_2}{2},\frac{n}{2};\frac{1}{2}q\phi\left(1-\psi\right)\right). \tag{14}$$

Note that, in this notation, $Q = \left(\psi\chi^2_{n_1} + \chi^2_{n_2}\right)/\phi\psi$. This is evidently a special case of the exact result above. Here, though, the hypergeometric function has scalar argument, and the degrees of freedom parameters differ. The corresponding cdf is

$$\Pr\{Q < z\} = \psi^{\frac{n_2}{2}}\sum_{j=0}^{\infty}\frac{(\frac{n_2}{2})_j}{j!}(1-\psi)^j G_{n+2j}(\phi z), \tag{15}$$

again an infinite discrete mixture of chi-square cdfs when $0 < \psi < 1$, the weights now being given by

$$b_j = \frac{(\frac{n_2}{2})_j}{j!}\psi^{\frac{n_2}{2}}(1-\psi)^j \tag{16}$$

(and ϕ replacing α^{-1}). It is this distribution that we will suggest using as an approximation to the distribution of Q.

2.2.1 Truncation error control when $m = 2$

The suggested approximations to the *pdf* and *cdf* of Q are evidently infinite series, rather than elementary functions, and the series must be truncated at some point for numerical work. However, by a straightforward adaptation of a result of Rider (1962) it can be shown that the error committed when the series is truncated after l terms is bounded above by $\Pr\{B\left(\frac{n_2}{2}, l\right) > \psi\}G_{n+2l}(\phi z)$, where $B(\frac{n_2}{2}, l)$ is a variate with a $Beta(\frac{n_2}{2}, l)$ distribution. It is therefore possible to control the truncation error very easily. We now briefly explain this argument.

Setting $\psi = (1+t)^{-1}$, and $a = n_2/2$, we are interested in the partial sum of the first l terms,

$$S_l(a,t) = \sum_{j=0}^{l-1} \frac{(a)_j}{j!} t^j (1+t)^{-(a+j)}, \tag{17}$$

since this provides a bound on the error incurred by terminating the series expansion for $\Pr\{Q < z\}$. The error is bounded above by $(1 - S_l(a,t))G_{n+2l}(\phi z)$. Now, for $n_2 = 2r$ even we have $a = n_2/2 = r$, so

$$b_j(r,t) = \frac{(r+j-1)!}{j!(r-1)!} t^j (1+t)^{-(r+j)}, \tag{18}$$

the negative Binomial distribution. In this case, Rider (1962) showed that the partial sum S_l is related to the incomplete Beta function:

$$S_l(r,t) = \sum_{j=0}^{l-1} \frac{(r)_j}{j!} t^j (1+t)^{-(r+j)} = \Pr\{B(r,l) < 1/(1+t)\}.$$

This continues to hold for any real $r > 0$, integer or not. So we have, for $a > 0$,

$$S_l(a,t) = \psi^a \sum_{j=0}^{l-1} \frac{(a)_j}{j!} (1-\psi)^j = \Pr\{B\,(a,l) < \psi\}. \tag{19}$$

Thus, the truncation error after l terms is bounded above by $\Pr\{B\,(a,l) > \psi\} G_{n+2l}(\phi z)$.

To see this more clearly, let $\psi = (1+t)^{-1}$, and for $a > 0$,

$$1 - S_l(a,t) = \sum_{j=l}^{\infty} \frac{(a)_j}{j!} t^j (1+t)^{-(a+j)}. \tag{20}$$

Differentiating,

$$\begin{aligned} -S_l'(a,t) &= \sum_{j=l}^{\infty} \frac{(a)_j}{j!} \left[j t^{j-1} (1+t)^{-(a+j)} - (a+j) t^j (1+t)^{-(a+j+1)} \right] \\ &= \frac{(a)_l}{(l-1)!} t^{l-1} (1+t)^{-(a+l)} + \sum_{j=l}^{\infty} \frac{[(a)_{j+1} - (a+j)(a)_j]}{j!} t^j (1+t)^{-(a+j+1)} \end{aligned} \tag{21}$$

But the terms $[(a)_{j+1} - (a+j)(a)_j]$ vanish for all j, and this does not depend on a being an integer. Hence, as in Rider, for any $a > 0$,

$$-S_l'(a,t) = \frac{(a)_l}{(l-1)!} t^{l-1}(1+t)^{-(a+l)}. \tag{22}$$

The remaining steps follow those in Rider exactly (just substitute $\psi = (1+t)^{-1}$ at the end).

3 Suggested Approximation

The suggestion is to use the variate $\tilde{Q} = a_1\chi^2_{n_1} + a_2\chi^2_{n_2} = (\psi\chi^2_{n_1} + \chi^2_{n_2})/\phi\psi$ as an approximation to Q, choosing the four free parameters (n_1, n_2, ϕ, ψ) so that the first four cumulants of $\tilde{Q}$ agree with those of Q. The approximation to the cdf will then have the form (15), with (n_1, n_2, ϕ, ψ) replaced by the cumulant-matching values. Without loss of generality we can assume that $a_1 < a_2$, so that $0 < \psi < 1$, and $\phi > 0$. Although the approximation to the cdf is, like the exact expression, an infinite series, the evaluation of that series is much simpler than for the exact expression, and only the first four power sums p_r are needed. Surprisingly, the parameter values that achieve equality of the first four cumulants are easily obtained.

3.1 Matching Cumulants

To simplify notation, put $n_1 = k, n_2 = l$, and $\phi\psi = c$. In this notation the cumulants of $\tilde{Q} = (\psi\chi^2_k + \chi^2_l)/c$ are given by

$$\hat{\kappa}_r = 2^{r-1}(r-1)!\,(\psi^r k + l)\,/c^r. \tag{23}$$

We want to choose the parameters (l, k, ψ, c) so that the first four cumulants of $\tilde{Q}$ agree with those of Q itself, i.e., $\hat{\kappa}_r = \kappa_r$ for $r = 1, 2, 3, 4$. This produces the four equations:

$$k\psi + l = cp_1, k\psi^2 + l = c^2 p_2 \tag{24}$$
$$k\psi^3 + l = c^3 p_3, k\psi^4 + l = c^4 p_4, \tag{25}$$

to be solved for (ψ, k, l) and c. We assume that the weights a_i are not all equal (see Remark 3 below).

Eliminating l from the first two equations gives

$$k\psi(\psi - 1) = c(cp_2 - p_1), \tag{26}$$

and from the second and third gives

$$k\psi^2(\psi - 1) = c^2(cp_3 - p_2).$$

Thus, taking the ratio,

$$\psi = \frac{c(cp_3 - p_2)}{(cp_2 - p_1)}. \tag{27}$$

Repeating this sequence, but beginning at the second equation, gives the pair

$$\begin{aligned} k\psi^2(\psi - 1) &= c^2(cp_3 - p_2), \\ k\psi^3(\psi - 1) &= c^3(cp_4 - p_3), \end{aligned}$$

from which

$$\psi = \frac{c(cp_4 - p_3)}{(cp_3 - p_2)}.$$

Equating the two expressions for ψ gives a quadratic equation in c alone:

$$g(c) = (cp_2 - p_1)(cp_4 - p_3) - (cp_3 - p_2)^2 = 0, \tag{28}$$

or

$$g(c) = c^2 - c\tau + \delta = 0, \tag{29}$$

with

$$\tau = \frac{p_1p_4 - p_2p_3}{p_2p_4 - p_3^2}, \delta = \frac{p_1p_3 - p_2^2}{p_2p_4 - p_3^2}. \tag{30}$$

We show in Appendix B that the smaller root of $g(c) = 0$, c_1 say, is the appropriate choice for c, i.e.,

$$c_1 = \frac{1}{2}\left[\tau - \sqrt{\tau^2 - 4\delta}\right]. \tag{31}$$

This value determines ψ, equation (26) then provides k, and $l = cp_1 - k\psi$. The complete solution is thus, in terms of c_1:

$$\psi = \frac{c_1(c_1p_3 - p_2)}{c_1p_2 - p_1}; \phi = \frac{(c_1p_2 - p_1)}{(c_1p_3 - p_2)}, \tag{32}$$

$$k = \frac{c_1(p_1 - c_1p_2)}{\psi(1 - \psi)}; l = c_1p_1 - k\psi. \tag{33}$$

Defining $v = k + l = c_1 p_1 + k(1 - \psi)$, the approximating cdf has the form

$$\Pr\{Q < z\} \simeq \psi^{\frac{l}{2}} \sum_{j=0}^{\infty} \frac{(\frac{l}{2})_j}{j!} (1 - \psi)^j G_{v+2j}(\phi z), \tag{34}$$

with (k, l, ψ, ϕ) as given above.

Remark 2 *Of course, in the case $m = 2$ the approximation should be exact, and it is straightforward, if tedious, to confirm that that is the case.*

Remark 3 *It is immediate from the argument in Appendix B that ψ, ϕ, and k are positive when the a_i are not equal. The positivity of l can also be established with a little extra algebra; see Appendix B. Hence, the cumulants $\hat{\kappa}_r$ implied by these values are all positive, as they should be.*

Remark 4 *Note that this solution process fails if the weights a_i are all equal to a, say, since then $p_r = na^r$. However, in that case the system is satisfied by $k = n, l = 0$, and $\psi = ca$, so that $\tilde{Q} = a\chi_n^2 = Q$, and no approximation is involved. It may also be unstable if the a_i are close to being equal, and in this case an alternative approximation should be sought.*

Remark 5 *The values for (k, l) and hence $v = k + l$ obtained here will not be integers, so strictly speaking the distributions appearing in the mixture representation (34) are those of Gamma variates, rather than Chi-square, but we ignore this distinction since it is irrelevant.*

Remark 6 *Note also that, in the approximation (34) it is not merely a matter of replacing the (complicated) coefficients b_j in the exact expression by simpler terms. Both the parameter v of the chi-square cdfs involved, and its argument ϕz, are derived from the cumulant-matching equations.*

3.2 Evaluation Summary: Quadratic form

Comparisons of numerical accuracy and resource requirements of various exact procedures (Imhof (1961), Sheil and O'Muircheartaigh (1977)), and several moment-matching approximations for this problem, have been reviewed recently in B&A. Bodenham and Adams conclude that the four-term approximation suggested by Lindsay et al. (2000) (hereafter LPB), which we call G4, is generally superior to other methods, although the difficulty calculating the components needed to implement it is significant. Rather than comparing our method with several of the others in this category, we shall therefore evaluate our method against G4 and the "exact" methods available.[4]We also include in the comparison three saddlepoint-based methods discussed by Wood (1993). These have a normal base, and two implementations of a chi-square base.

The LPB method we call G4 uses a mixture of 4 gamma variates with the same shape parameter, different scale parameters, and, with appropriate mixture probabilities, matches the first 8 moments of Q. B&A provide a greatly simplified exposition of G4, clarifying the calculation of the scale parameters μ_j. In their evaluations LPB reported, among other things, results for 4 configurations of the a_i from Solomon & Stevens (1977), and 14 (different) configurations from Wood (1989).[5] In these configurations the n_i are all treated as unity, but when one counts only distinct a_i, of the 14 Wood cases, 6 have 2 distinct a_i, for which $\tilde{Q}$ is exact, 6 have 3 distinct, and 2 have 4 different a_i.

Our first evaluation exercise uses, for $n = 4$, the same 18 configurations of the a_i, and also the same 10 quantiles, as those used by LPB, i.e., those corresponding to the probabilities

$$P = \{0.01, 0.025, 0.50, 0.10, 0.25, 0.75, 0.90, 0.95, 0.975, 0.990\}$$

There are thus 180 points of comparison. Note that this means we are evaluating the ability of the various methods to approximate the *entire distribution*, not just the tails. We give analogous results focussing on the upper tail of the distribution seperately. The exact quantiles were computed using the S&O'M exact method, but rather than using the Ruben recursions for the weights in the cdf, this is implemented by using the HKW efficient recursion. The results of this exercise are given in the following Table:

Table 1a: Comparison of the three methods over the entire distribution

[4]The methods which can in principle be iterated to converge to an arbitrary precision are Davies' (1980) implementation of Imhof's exact procedure, and Ruben's (1962) series implemented by Sheil and O'Muircheartaigh (1977), or in a faster but perhaps slightly riskier version by Farebrother (1984a).

[5]B&A object to the partiality of such investigations of the parameter space. They instead used a uniform random search over the a_i.

	MAE	AAE	4F%	3F%	2F%	1%	2.5%
$\tilde{\mathrm{Q}}$	0.0022	0.00035	48	72	100	100	100
G4	0.0212	0.00066	52	86	97.2	97.8	100
S*	0.014	0.0017	15	49	89.4	96.7	100

In Table 1a, MAE = maximum absolute error, AAE = average absolute error, rF% = percentage of cases accurate to r significant digits. The last two columns give the percentage of cases in which the absolute error was less than 1%, or less than 2.5%. The results given for saddlepoint methods, S*, are those of the best of the three saddlepoint methods (usually one of the methods with a chi-square base). Evidently all methods are quite accurate, but our simple approach ($\tilde{\mathrm{Q}}$) dominates the other methods on almost all criteria. It must be said, however, that these results favour $\tilde{\mathrm{Q}}$, since 6 of the 18 points used have $m = 2$, when $\tilde{Q}$ is exact.

Focussing next on the upper tail, the analogous results are given based on the quantiles corresponding to probabilities 0.90(0.005)0.995, and the same set of coefficients. The next table gives the results of this exercise (omitting the last two columns).

Table 1t: Performance of the approximations in the upper tail

	MAE	AAE	4F%	3F%	2F%
$\tilde{Q}$	0.0005	0.00009	54	99.6	100
$G4$	0.0009	0.00009	67	97.1	100
S^*	0.0045	0.00112	12	49.0	100

The values labelled S^* are for the best of the three saddlepoint methods, but these are clearly dominated by the other two methods. Again, the four-moment approach using $\tilde{Q}$ is very nearly as good as $G4$, and dominates it on some criteria.

Next, we conducted a similar exercise in which the weights a_i are varied, with $n = m = 3, 4$, and for each method we search for the configuration with the worst-case error. In this case the quantiles used were those of the 15 probabilities

$$P = \{0.01, 0.025, 0.05, 0.1(0.1)0.9, 0.975, 0.99\},$$

where $a(b)c$ means "from a by increments of b to c". The a_i were, without loss of generality, normalised so that $\Sigma_{i=1}^{n} a_i = n$, and we searched over the grid $0.02(0.02)n/2$ for each a_i. The results were as given in the following Table (omitting the case $n = 2$ where $\tilde{Q}$ is exact):[6]

Table 2a: Comparison of methods; worst cases over a grid of coefficients, $n = 3, 4$.

	MAE	AAE	4F%	3F%	2F%	1%	2.5%
$\tilde{Q}, n = 3$	0.0075	0.00080	37	68	96.7	100	100
$G4, n = 3$	0.0241	0.00043	56	81	98.7	99.7	100
$\tilde{Q}, n = 4$	0.0116	0.00099	22	56	96.9	99.8	100
$G4, n = 4$	0.0288	0.00038	48	81	99.2	99.9	100
S^*	0.0111	0.00193	25	51	84.4	97.3	100

Overall, bearing in mind that the results in the table refer to the worst cases, both $\tilde{Q}$ and $G4$ perform well, both having errors of 1% or below in over 99% of cases. In terms of MAE $\tilde{Q}$ is superior, but in terms of AAE $G4$ is better, but the differences are small in both cases. The analogue of Table 2a, but now focussing on the upper tail of the distribution (and omitting the last two columns again), yields the results given in the following table:

Table 2t: Comparison of methods; worst cases over a grid of coefficients, $n = 3, 4$; upper tail of distribution

	MAE	AAE	4F%	3F%	2F%
$\tilde{Q}, n = 3$	0.00058	0.000061	69	99.6	100
$G4, n = 3$	0.00019	0.000014	93	100	100
$\tilde{Q}, n = 4$	0.00079	0.000093	53	98.3	100
$G4, n = 4$	0.00034	0.000021	89	100	100
S^{**}	0.00405	0.00792	17	52.4	100

[6] Because of the symmetry of Q as a function of the coefficients, there were 84,000 points of comparison for $n = 4$. For $n = 5$ this would be 24 times as large.

Here, S^{**} is the analogue of S^* in Table 1t. Again, $\tilde{Q}$ and $G4$ dominate the saddlepoint methods, and both are accurate to at least three figures in at least 98% of cases. This is certainly more precision than is typically needed for testing purposes.

Both evaluation exercises so far have been cases with all $n_i = 1$. We next seek to explore the influence of the n_i on the approximations when $n_i > 1$ for some i. To do so we examined four cases with $m = 3$ corresponding to the a_i-configurations that produced the worst positive and negative errors for $\tilde{Q}$ and $G4$ when all $n_i = 1$. The implicated configurations are given in the second column of Table 3 below.[7]

For these chosen configurations we then search over a grid of values for the n_i, $n_i \in \{1, 2, 3, 4, 5, 10, 20, 30, 50\}$ (i.e., $9^3 = 729$ triplets), and for 10 P-values, 5 in the lower tail $\{0.01, 0.025, 0.05, 0.1, 0.25\}$, and the corresponding 5 in the upper tail. The results are summarized in Table 3 below. The last 2 columns give the percentage of instances in which each method has better than 3-digit accuracy.

Table 3: Comparisons over a grid of n_i values; $m = 3$, worst-case weights.

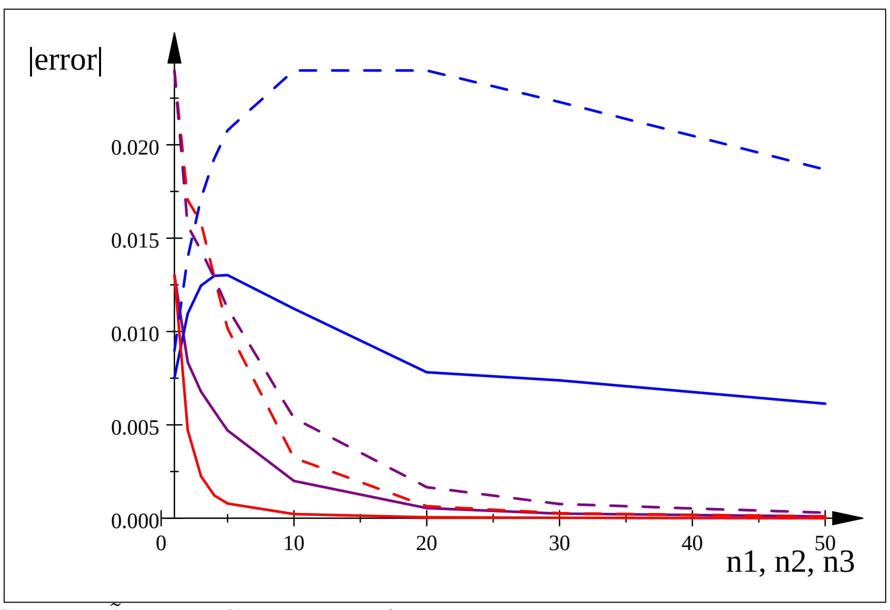
Case	weights	$\tilde{Q}$ max	$G4$ max	$\tilde{Q}$ mean	$G4$ mean	$\tilde{Q}$%	$G4$%
1	0.105, 0.990, 1.905	0.013022	0.023986	0.000281	0.000786	88.09	76.5
2	0.145, 0.985, 1.870	0.010048	0.018989	0.000237	0.000761	88.5	74.98
3	0.225, 0.040, 2.735	0.008631	0.074395	0.000143	0.003871	95	50
4	0.140, 0.145, 2.715	0.000005	0.047503	0.000000	0.004430	100	48

Again we find that $\tilde{Q}$ dominates $G4$ on all criteria, in this case by a fairly wide margin: the former is correct to 3 digits in all but 12% of cases, but $G4$ strays in some 23%.

Remark 7 *The grid search just described effectively entails searching over a large array of values for the elementary symmmetric functions e_r upon which the distribution depends. Thus, whilst the search uses $m = 3$, it is in fact quite a comprehensive evaluation of the methods in terms of the quantities that really matter. See Remark 1 above.*

[7] It should be noted that the elementary symmetric functions e_r for these four base cases are (3,2.256,.267), (3,2.1899,0.198), (3,0.7338,0.0246), and (3,0.7941,0.055) respectively. Thus, the e_r for cases 1 and 2 are very different from those for cases 3 and 4.

The results of this exercise can also be summarized by viewing the trajectories of the largest error as the n_i vary. We took the four cases used above (the worst positive and negative errors for each of $\tilde{Q}$ and $G4$), and did a grid search over the 9 n_i values given above, and the same 10 quantiles. We then graphed, for each i, and each value of n_i, the largest error found in a search over the values of the other two n_j, and over the ten quantiles. Figure 1 illustrates the results of this exercise for Case 1. The solid lines relate to $\tilde{Q}$, the dashed lines to $G4$. For example, in Fig. 1, starting from $\tilde{Q}'s$ worst case when $n_1 = n_2 = n_3 = 1$ (with weights $\{0.105, 0.990, 1.905\}$), the solid blue line shows the maximum error of $\tilde{Q}$ over the grid as n_2 varies from zero to 50. The purple and red lines show the same behaviour for increasing n_1 and n_3, respectively. Thus, the three solid lines reveal that increasing n_2, the multiplicity of the smallest weight, is least effective at reducing the error. The dashed lines are the analogous trajectories for $G4$, and similar patterns are revealed.



Case 1, $\tilde{Q}$ solid, G4 dashes. At $n_i = 10$ the curves are in the order $n_3 < n_1 < n_2$.

Figure 1 shows that, when evaluated by these "worst case" metrics, $\tilde{Q}$ is again convincingly superior to $G4$. The analogous plots for Cases 2 - 4 are available from the authors.[8] The equivalent analysis for $m = 4$ is problematic, because

[8]The trajectories are L-shaped or unimodal, but the mode may be at some distance

Case 1 (worst positive error for $\tilde{Q}$, all $n_i = 1$) has two equal weights, and thus can be included in the set of $m = 3$ cases. The same applies to Case 3 (worst positive error for $G4$). Unsurprisingly, though, all trajectories exhibit the same pattern: large n_i helps, but much more rapidly when attached to the larger weights.

In summary, the several types of evidence described in this section clearly indicate that the approximation based on $\tilde{Q}$ outperforms that based on $G4$, and it also outperforms various saddlepoint methods that have been proposed. Since $\tilde{Q}$ is also significantly simpler to implement than the other methods considered, its use can be confidently recommended.

4 Application: Student-t-like Tests

In recent work on the finite sample properties of a number of tests that are routinely used in applied work, Hansen (2021) points out that, in a variety of contexts, the test statistics in question have the form[9]

$$T = \frac{Z}{\sqrt{Q}}, \tag{35}$$

where $Z \sim N(0,1)$, and Z and Q independent, as with a standard $t-$statistic, but instead of having $Q \sim \chi^2(v)$, we have

$$Q = \sum_{i=1}^{m} a_i \chi^2(n_i), \tag{36}$$

a positive linear combination of independent χ^2 variates. He presents simulation evidence which shows that assuming a Student-t distribution for T - the usual approach - can produce tests that are badly over-sized (with corresponding distortions of the coverage levels of confidence sets).

Using essentially the infinite-mixture representation for $pdf_Q(q)$ in equation (7), Hansen shows that the distribution of T is a mixture of Student-t

from the origin. The most delayed decline is Case 3, weights $\{0.225, 0.040, 2.735\}$, n_2 trajectory, for $G4$. The peak is at $n_2 = 30$, but the subsequent decline is slow. This illustrates that increasing the frequency for the smallest weight is not quickly effective on its own.

[9] In practice the distribution assumptions made here may be asymptotic, rather than finite-sample. Nevertheless, the densities involved, and the need for approximations to them, remain the same.

distributions of the same form. We prove this more simply below. However, there are a number of computational issues associated with the implementation of this exact result, so there is an incentive to replace the distribution of Q by an approximation. Using simulation studies, Hansen (2021) explores the accuracy of several different approximations, including ours, and concludes that our approach is superior. In the next section we examine the accuracy of this approximation using exact computations, as we have done earlier for $\tilde{Q}$. These too support our approach.

4.1 Exact and approximate distribution function

Since $Z \sim N(0,1)$ and is independent of Q, the conditional distribution of T given $Q = q$ is $N(0, q^{-1})$. The unconditional distribution can be derived from this in the obvious way (see Appendix B), but there is an easier way, as follows. The conditional distribution is evidently symmetric about the origin, and, conditionally,

$$\Pr\{T^2 < z|Q = q\} = \Pr\{\chi^2(1) < qz\} = G_1(qz), \tag{37}$$

so that, unconditionally, from the mixture representation of $pdf_Q(q)$ in equation (7),

$$\Pr\{T^2 < z\} = E_Q\left[G_1(qz)\right] = \sum_{j=0}^{\infty} b_j E_{x\sim\chi^2(n+2j)}\left[G_1(xz/\delta)\right]. \tag{38}$$

The cdf of the χ_v^2 distribution is denoted here and elsewhere by $G_v(z) = \Pr\{\chi_v^2 \leq z\}$. But, directly from the definition of an F_{v_1,v_2} variate, a straightforward conditioning argument yields:

$$\Pr\{F_{v_1,v_2} < z\} = E_{x\sim\chi^2(v_2)}\left[G_{v_1}\left(\frac{xv_1z}{v_2}\right)\right]. \tag{39}$$

Thus, we immediately obtain an expression for the (exact) cdf of T^2 :

$$\Pr\{T^2 < z\} = E_Q\left[G_1(qz)\right] = \sum_{j=0}^{\infty} b_j \Pr\left\{F_{1,n+2j} < (n+2j)z/\alpha\right\}. \tag{40}$$

This mixture representation of the cdf can be used with the actual b_j in equation (9), but of course is subject to the same computational issues.

Instead, we suggest using the approximation for the distribution of Q, with parameters chosen as described earlier. That is, with

$$b_j = (\frac{l}{2})_j \psi^{\frac{l}{2}} (1-\psi)^j / j!, \tag{41}$$

and the terms $\Pr\{F_{1,n+2j} < (n+2j)z/\alpha\}$ replaced by $\Pr\{F_{1,v+2j} < (v+2j)z\phi\}$. That is, we suggest using the approximation

$$\Pr\{T^2 < z\} \simeq \psi^{\frac{l}{2}} \sum_{j=0}^{\infty} \frac{(\frac{l}{2})_j}{j!} (1-\psi)^j \Pr\{F_{1,v+2j} < (v+2j)z\phi\}, \tag{42}$$

with the values (k, l, ψ, ϕ) those given earlier for approximating the cdf of Q. We denote a variate with this cdf by $\tilde{T}^2$.

Remark 8 *Again, the value of $v = k+l$ produced by equating cumulants will not be an integer, so the random variables $F_{1,v+2j}$ are not strictly $F-$distributed, but again this distinction is immaterial.*

Remark 9 *Theorem 2 in Hansen (2021) is essentially this result, but Hansen's proof is somewhat different.*

4.2 Evaluation Summary: T-test

The cdf of $\tilde{T}^2$, the approximation to T^2, and of T^2 itself, are evaluated using the HKW approach, and Imhof's exact method, adapted to the case where the "conditional" distributions in the mixture representation are F-distributions rather than Chi-squared. These are accurate to 10^{-6} or better. Since there is no competing approximation in this case, the performance of $\tilde{T}^2$ can only be evaluated relative to the exact distribution.[10]

To begin with we use the same 18 configurations of the weights a_i as above, choosing, as before, the a_i at which the maximum error of $\tilde{Q}$ occurs when all $n_i = 1$. We then explore the same grid of the n_i, and obtain the results summarized in the following table:

[10] Hansen (2021) also considers the three-moment approximation suggested by Buckley and Eagleson's (1988) application of Hall (1984). However, our approach is found to dominate this method as well.

Table 4: Summary of performance of $\tilde{T}^2$; $m = 3, 4$.

	MAE	n_i	AAE	$3F\%$
$m = 3$	0.004838	$\{1, 2, 1\}$	< 0.000025	$> 99\%$
$m = 4$	0.006208	$\{1, 1, 3, 1\}$	< 0.000027	$> 99\%$

In the case $m = 3$ all trajectories decline monotonically, the largest error at $n_i = 50$ being 0.000147 for, predictably, the small weight 0.04. For $m = 4$ all trajectories decline monotonically beyond $n_i = 3$, the largest error at $n_i = 50$ being 0.000360 for the weight of 0.08. The largest error at $n_i = 10$ is 0.000114.

One can instead use the weights giving the worst errors for $\tilde{T}^2$ (rather than those worst for $\tilde{Q}$) when all $n_i = 1$, then explore the grid of n_i values. It happens that the errors for $\tilde{T}^2$ all have the same sign. The implied configurations and maximum errors are as given in the following table:

Table 5: Cases giving the worst errors for $\tilde{T}^2$ when all $n_i = 1$.

m	error	a_1	a_2	a_3	a_4	Prob.
3	0.003902	1.01	0.08	1.91		0.975
4	0.005544	1.26	0.08	0.08	2.58	0.975
4*	0.005521	1.26	0.07	0.09	2.58	0.975

Note that for $m = 4$ the worst case has $a_2 = a_3$ (so actually $m = 3$). In the last row of the Table we separate these, with very little impact on the maximum error, so we report the trajectories for n_i with the weights in the last row of the table. The maximum error rises as the $n_i = 1$ "constraint" is relaxed, but for $m = 3$ only to 0.005565 at $\{n_1, n_2, n_3\} = \{1, 3, 1\}$ and $P = 0.975$, and for $m = 4$ to 0.006519 at $\{1, 3, 1, 1\}$ and the same P. Errors decline monotonically with increasing n_i beyond these points. The largest error for $n_i = 50$ is 0.000483 (at $\{2, 50, 1, 1\}, P = 0.975$).

In short, $\tilde{T}^2$ provides an excellent approximation to the distribution of T^2 - even more accurate than $\tilde{Q}$ provides for Q. This conclusion is strongly reinforced by the simulation results given in Hansen (2021). It is worth noting that Hansen (2021) also explores the computation cost of several approximation methods, and these results too are favourable to our approach.[11]

[11] The consequences of using a suitably rescaled $F(1, n)$ variate to approximate the actual

5 Generalization: F-Type Test statistics

A related problem of the same type is to consider the density of a ratio of two independent quadratic forms, i.e., the generalized version of an $F-$statistic:

$$F = \frac{y_1' D_1 y_1}{y_2' D_2 y_2} = \frac{Q_1}{Q_2}, y_i \sim N(0, I_{n_i}), i = 1, 2; y_1 \perp y_2. \tag{43}$$

Using the results given earlier, the exact density of $F = Q_1/Q_2$ is readily obtained, and is:

$$pdf_F(f) = \sum_{j,k=0}^{\infty} b_{1j} b_{2k} (\phi_1/\phi_2)^{\frac{n_1}{2}} \left\{ \frac{\Gamma(j+k+\frac{n_1+n_2}{2})}{\Gamma(j+\frac{n_1}{2})\Gamma(k+\frac{n_2}{2})} f^{\frac{n_1}{2}+j-1} (1+\phi_1 f/\phi_2)^{-(\frac{n}{2}+j+k)} \right\}, \tag{44}$$

with

$$b_{ij} = |\phi_i D_i|^{-\frac{1}{2}} \frac{(\frac{1}{2})_j}{j!} C_j(I_{n_i} - (\phi_i D_i)^{-1}), i = 1, 2. \tag{45}$$

The cdf is thus the doubly-infinite mixture of $F-$distribution functions:

$$\Pr\{F < z\} = \sum_{j,k=0}^{\infty} b_{1j} b_{2k} \Pr\left\{ F_{n_1+2j, n_2+2k} < \frac{\phi_1(n_2+2k)}{\phi_2(n_1+2j)} \right\}. \tag{46}$$

One approach to approximating this would be to use the methods discussed earlier for each component separately to obtain an approximate joint density. This would entail replacing the b_{1j}, b_{2k} by

$$\tilde{b}_{ij} = \psi_i^{\frac{n_{2i}}{2}} \frac{(\frac{n_{2i}}{2})_j}{j!} (1-\psi_i)^j, \tag{47}$$

and using values ψ_i, ϕ_i, n_{1i} and n_{2i} determined as discussed earlier. Preliminary evaluations of this method are encouraging, but we defer a full evaluation to later work.

distribution of T^2 can be explored by similar methods. We find that doing so results in rejection rates that are too large: for nominal sizes of $\{10, 5, 2.5, 1\}\%$ the actual sizes can be as large as $\{21, 14.5, 10, 6\}\%$ when $m = 3$, and as large as $\{19.3, 12.8, 8.3, 4.6\}\%$ when $m = 4$. Moral: don't use the $F(1, n)$ distribution to approximate the distribution of T^2!

6 Concluding Remarks

We have proposed a new, four-moment, approximation to the distribution of a positive definite quadratic form in standard normal variates, or, equivalently, to a linear combination of independent chi-squared variates. The method is simple to implement, and provides an extremely accurate approximation - on most criteria matching the performance of the "best of the rest", the eight-moment approximation suggested by Lindsay et. al. (2000), which is considerably more complicated to implement. We then suggested that the approximation can also be used to approximate the distribution of a $student-t-like$ variate, when the denominator is not a chi-squared variate but a more general quadratic form. This form arises in many applied contexts, and we again find that the proposed approach prides an excellent approximation - certainly more than adequate for practical testing appications.

Finally, we note that results of the type we discuss have wide applicability. For instance, Hansen (2021) shows that they can be used to analyse the properties of the heteroscedasticity-robust t-ratio introduceded by White (1980), and studied by many since. At another extreme, Bausch (2013) shows that the distribution, and approximations to it, have applications to the physics of string vacua.

7 Appendix A: Exact density for the case $m = 2$

Starting from $x_i \sim \chi^2(n_i), i = 1, 2$, we want the density of $q = a_1x_1 + a_2x_2$. Assume, without loss of generality, that $a_2 > a_1 > 0$, and transform to $q_i = a_ix_i$, $i = 1, 2$, to give $q = q_1 + q_2$, and

$$pdf(q_1, q_2) = \frac{q_1^{\frac{n_1}{2}-1} q_2^{\frac{n_2}{2}-1}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})a_1^{\frac{n_1}{2}}a_2^{\frac{n_2}{2}}} \exp\left\{-\frac{1}{2}\left(a_1^{-1}q_1 + a_2^{-1}q_2\right)\right\}. \quad (48)$$

Now transform to $q = q_1 + q_2$ and $b = q_1/q$, leaving

$$pdf(b, q) = \frac{b^{\frac{n_1}{2}-1}(1-b)^{\frac{n_2}{2}-1} q^{\frac{n_1+n_2}{2}-1}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})a_1^{\frac{n_1}{2}}a_2^{\frac{n_2}{2}}} \exp\left\{-\frac{1}{2}q\left(a_1^{-1}b + a_2^{-1}(1-b)\right)\right\}. \quad (49)$$

Integrating out b gives

$$pdf_Q(q) = \frac{q^{\frac{n_1+n_2}{2}-1}\exp\left\{-\frac{1}{2}qa_2^{-1}\right\}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1+n_2}{2})a_1^{\frac{n_1}{2}}a_2^{\frac{n_2}{2}}}\,{}_1F_1\left(\frac{n_1}{2},\frac{n_1+n_2}{2};-\frac{1}{2}q\left(a_1^{-1}-a_2^{-1}\right)\right) \quad (50)$$

$$= \frac{q^{\frac{n_1+n_2}{2}-1}\exp\left\{-\frac{1}{2}qa_1^{-1}\right\}}{2^{\frac{n_1+n_2}{2}}\Gamma(\frac{n_1+n_2}{2})a_1^{\frac{n_1}{2}}a_2^{\frac{n_2}{2}}}\,{}_1F_1\left(\frac{n_2}{2},\frac{n_1+n_2}{2};\frac{1}{2}q\left(a_1^{-1}-a_2^{-1}\right)\right) \quad (51)$$

Putting $\phi = a_1^{-1}$, $\psi = a_1/a_2$, $0 < \psi < 1$, $n = n_1 + n_2$, this becomes

$$pdf_Q(q) = \frac{\phi^{\frac{n}{2}}\psi^{\frac{n_2}{2}}q^{\frac{n}{2}-1}\exp\left\{-\frac{1}{2}q\phi\right\}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\,{}_1F_1\left(\frac{n_2}{2},\frac{n}{2};\frac{1}{2}q\phi\left(1-\psi\right)\right), \quad (52)$$

as reported in the text.

8 Appendix B: the appropriate root of the quadratic

To determine which root of the quadratic $g(c)$ should be used, we seek solutions satisfying $k > 0$ and $0 < \psi < 1$. We make use of a well-known inequality for the power sums p_r, which says that for all j, k, r, a, b, c s.t. $aj + bk = cr$.

$$p_j^a p_k^b \geq p_r^c.$$

(see Reznick, (1983), for instance). With $j = 1, k = 3, a = b = 1, r = c = 2$, we get $p_1p_3 \geq p_2^2$, and with $j = 2, k = 4, a = b = 1, r = 3, c = 2$, we have $p_2p_4 \geq p_3^2$. The inequalities are strict unless all non-zero elements of D are equal.

Now, for $g(\cdot)$ as in equation (28), we have $g(0) = \delta > 0$, $g(p_1/p_2) = -(p_1p_3 - p_2^2)^2/p_2^2 < 0$, and $g(p_2/p_3) = (p_2^2 - p_1p_3)(p_2p_4 - p_3^2)/p_3^2 < 0$, so both roots are real and positive, and the values p_1/p_2 and p_2/p_3 are between the roots c_1 and c_2. For the larger root c_2, $c_2p_3 - p_2$ and $c_2p_2 - p_1$ are both positive, and

$$\begin{aligned} c_2(c_2p_3 - p_2) - (c_2p_2 - p_1) &> \left[p_1(c_2p_3 - p_2) - p_2(c_2p_2 - p_1)\right]/p_2 \\ &= c_2(p_1p_3 - p_2^2)/p_2 \\ &> 0. \end{aligned}$$

Thus, at $c = c_2$, $\psi > 1$. On the other hand, at c_1 we have $c_1 p_2 - p_1$ and $c_1 p_3 - p_2$ both negative, and

$$\begin{aligned} c_1(p_2 - c_1 p_3) - (p_1 - c_1 p_2) &< [p_1(p_2 - c_1 p_3) - p_2(p_1 - c_1 p_2)]/p_2 \\ &= -c_1(p_1 p_3 - p_2^2)/p_2 \\ &< 0, \end{aligned}$$

so that, at $c = c_1$, $0 < \psi < 1$, as required. This is the root that should be used.

To see that $l > 0$ when the a_i are not equal we have

$$l = c_1 p_1 - k\psi = \frac{c_1}{1-\psi}[c_1 p_2 - p_1 \psi].$$

The second term is

$$\begin{aligned} c_1 p_2 - p_1 \psi &= \frac{c_1}{c_1 p_2 - p_1}[p_2(c_1 p_2 - p_1) - p_1(c_1 p_3 - p_2)] \\ &= \frac{c_1^2}{c_1 p_2 - p_1}\left[p_2^2 - p_1 p_3\right] > 0, \end{aligned}$$

since numerator and denominator are both negative.

9 Appendix C: Distribution of T

Since the conditional distribution of T given $Q = q$ is $N(0, q^{-1})$ the conditional density is:

$$pdf_T(u|Q = q) = \frac{\sqrt{q}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u^2 q\right\}. \tag{53}$$

The unconditional density is therefore the expectation[12]

$$pdf_T(u) = (2\pi)^{-\frac{1}{2}} \int_{q>0} \exp\left\{-\frac{1}{2}u^2 q\right\} \sqrt{q}\, pdf_Q(q)dq. \tag{54}$$

[12]Of course, if $Q \sim \chi_n^2$, this variance-mixture of the $N(0, q^{-1})$ density produces the usual t-distribution.

Using the exact density of Q in equation (6) and evaluating the integral gives

$$\begin{aligned} pdf_T(u) &= \frac{|D|^{-\frac{1}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}(2\pi)^{-\frac{1}{2}}\int_{q>0}\exp\left\{-\frac{1}{2}q(\alpha+u^2)\right\}q^{\frac{n+1}{2}-1} \\ &\quad \times {}_1F_1\left(\frac{1}{2},\frac{n}{2};\frac{1}{2}\alpha q[I_n-(\alpha D)^{-1}]\right)dq \\ &= \frac{|D|^{-\frac{1}{2}}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})}(\alpha+u^2)^{-\frac{n+1}{2}}{}_2F_1\left(\frac{1}{2},\frac{n+1}{2};\frac{n}{2};\frac{\alpha[I_n-(\alpha D)^{-1}]}{(\alpha+u^2)}\right) \end{aligned} \tag{55}$$

That is,

$$pdf_T(u) = \alpha^{-\frac{1}{2}}|\alpha D|^{-\frac{1}{2}}\sum_{j=0}^{\infty}\frac{(\frac{1}{2})_j}{j!}C_j(I_n-(\alpha D)^{-1})\left\{\frac{\Gamma(\frac{n+1+2j}{2})}{\sqrt{\pi}\Gamma(\frac{n+2j}{2})}\left(1+\frac{u^2}{\alpha}\right)^{-\frac{n+2j+1}{2}}\right\}, \tag{56}$$

so that the corresponding distribution function is again the mixture

$$\Pr\{T<u\} = \sum_{j=0}^{\infty} b_j F_{n+2j}\left(u\sqrt{(n+2j)/\alpha}\right), \tag{57}$$

where $F_v(z)$ is the cdf of the Student-t distribution, and the b_j are exactly those appearing in the density of Q (see equation (9)). This is the equation given in Theorem 3 of Hansen (2017) (in slightly different notation).

When Q has only two components ($m=2$) the density $pdf_Q(q)$ is replaced by the simpler form

$$pdf_Q(q) = \frac{\phi^{\frac{n}{2}}\psi^{\frac{n_2}{2}}\exp\left\{-\frac{1}{2}q\phi\right\}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}q^{\frac{n}{2}-1}{}_1F_1\left(\frac{n_2}{2},\frac{n}{2};\frac{1}{2}q\phi\left(1-\psi\right)\right), \tag{58}$$

producing the unconditional cdf in the case $m=2$:

$$\Pr\{T<u\} = \psi^{\frac{n_2}{2}}\sum_{j=0}^{\infty}\frac{(\frac{n_2}{2})_j}{j!}(1-\psi)^j F_{n+2j}\left(u\sqrt{(n+2j)\phi}\right). \tag{59}$$

References

Bodenham, D.A., and Adams, N.M. (2016) A comparison of efficient approximations for a weighted sum of chi-square random variables, *Statistical Computation*, 26, 917-928.

Bausch, J. (2013) On the efficient calculation of a linear combination of chi-square random variables with an application in counting string vacua. *Journal of Physics A: Mathematical and Theoretical,* 46, 1-19.

Buckley, M.J., and Eagleson,G.K. (1988) An approximation to the distribution of quadratic forms in normal random variables, *Australian Journal of Statistics*, 150-159.

Butler, R.W. (2007), *Saddlepoint Approximations with Applications.* Cambridge: Cambridge Univ. Press

Constantine, A. G., (1964) Some non-central problems in multivariate analysis, *Annals of Mathematical Statistics*, 1270-1285.

Farebrother, R.W. (1984a). Algorithm AS 204: the distribution of a positive linear combination of χ^2 random variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 33(3), 332–339.

Farebrother, R.W.(1984b). Remark AS R53: A Remark on Algorithms AS 106, AS 153 and AS 155: The Distribution of a Linear Combination of χ^2 Random Variables, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 33, No. 3, pp. 366-369.

Hall, P. (1984) Chi-square approximations to the distribution of a sum of independent variables, *Annals of Probability,* 11,4, 1028-1036.

Hansen, B. (2017) Slides published on University of Wisconsin Website

Hansen, B. (2021) The Exact Distribution of the White t-ratio, manuscript, University of Wisconsin.

Hillier, G., Kan, R., and Wang, X. (2009), Computationally efficient recursions for top-order invariant polynomials with applications, *Econometric Theory*, 25, 211-242.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, 48, 419-426.

James, A. T. (1960) The distribution of the latent roots of the covariance matrix, *The Annals of Mathematical Statistics*, 31, 151 - 158.

James, A. T. (1964), Distributions of matrix variates and latent roots derived from normal samples, *Annals of Mathematical Statistics*, 475-501.

Lindsay, B.G., Pilla, R.S., Basak, P. (2000), Moment-based approximations of distributions using mixtures: theory and applications. *Annals of the Institute of Statistical Mathematics.* 52(2), 215–230.

Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, NY.

Pitman, E.J.G., and Robbins H., (1949), Application of the method of mixtures to quadratic forms in normal variates, *Annals of Mathematical Statistics*, 20, 552-560.

Reznick, B. (1983), Some inequalities for products of power sums, *Pacific Journal of Mathematics*, 104, 2, 443-463.

Ruben, H., (1962), Probability content of regions under spherical normal distribution, IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. *The Annals of Mathematical Statistics*, 33, 2, 542-570.

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

Sheil, J., and O'Muircheartaigh, I. (1977), The distribution of non-negative quadratic forms in normal variables, *Journal of the Royal Statistical Society, Series C*, 26, 92-98.

Solomon, H. and Stephens, M. A. (1977) Distribution of a Sum of Weighted Chi-Square Variables, *Journal of the American Statistical Association*, **72** (360), 881-885.

Titterington, D.M., Smith, A.F.M and Makov, U.E. (1985) *Statistical Analysis of Finite Mixture Distributions.* Wiley, New York.

von Neumann, J., (1941), Distribution of the ratio of the mean square successive difference to the variance, *The Annals of Mathematical Statistics*, 12, 367-395.

Welch, B.L. (1947) The significance of the difference between two means when the population variances are unequal. *Biometrika,* 29, 350-362.

White, H. (1980) A hereoscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48, 817-838.

Wood, A.T.A. (1989) An F approximation to the distribution of a linear combination of chi-squared variables. *Commun. Stat. Simul. Comput.*18(4), 1439–1456.

Wood, A.T.A, Booth, J.G. and Butler, R.W., (1993) Saddlepoint Approximations to the CDF of Some Statistics with Nonnormal Limit Distributions, *Journal of the American Statistical Association* **88**, 422, 680-686.