



University of Zurich
Zurich Open Repository and Archive

Winterthurerstr. 190
CH-8057 Zurich
<http://www.zora.uzh.ch>

Year: 2009

Evaluating the density of ratios of noncentral quadratic forms in normal variables

Broda, S; Paoletta, M S

Broda, S; Paoletta, M S (2009). Evaluating the density of ratios of noncentral quadratic forms in normal variables. *Computational Statistics & Data Analysis*, 53(4):1264-1270.

Postprint available at:
<http://www.zora.uzh.ch>

Posted at the Zurich Open Repository and Archive, University of Zurich.
<http://www.zora.uzh.ch>

Originally published at:
Computational Statistics & Data Analysis 2009, 53(4):1264-1270.

Evaluating the density of ratios of noncentral quadratic forms in normal variables

Abstract

Two computable expressions for the exact density of a ratio of quadratic forms in Gaussian random vectors are derived, one of which is restricted to special cases of the problem. Ratios of this type are ubiquitous in econometrics, but their density, unlike the corresponding cumulative distribution function, has not received much attention to date. The new algorithms complement those available for the latter. The included performance study demonstrates the accuracy of the two algorithms, both absolute and relative to each other, and allows general recommendations on their use to be made.

Evaluating the Density of Ratios of Noncentral Quadratic Forms in Normal Variables

S. Broda ^{a,*}, M.S. Paoletta ^a

^a*Swiss Banking Institute, University of Zurich*

Abstract

Two computable expressions for the exact density of a ratio of quadratic forms in Gaussian random vectors are derived, one of which is restricted to special cases of the problem. Ratios of this type are ubiquitous in econometrics, but their density, unlike the corresponding cumulative distribution function, has not received much attention to date. The new algorithms complement those available for the latter. The included performance study demonstrates the accuracy of the two algorithms, both absolute and relative to each other, and allows general recommendations on their use to be made.

Key words: Quadratic Form, Characteristic Function

1 Introduction

Ratios of quadratic forms in Gaussian random vectors arise in a vast number of contexts; most prominently, they appear in testing problems in linear models, but also in the estimation of autoregressive models with exogenous variables. Consequently, there exists a sizeable literature on the calculation of their moments (see, e.g., Paoletta, 2003) and, starting with the seminal paper of Imhof (1961), their cumulative distribution functions (c.d.f.). While some results have been obtained for the density of a quadratic form (Lu, 2006), it appears that to date, no efficient algorithm for the computation of the exact density of a ratio of two such forms has been devised; previous work is either limited to special cases (Kamanu, 2006), or approximate in nature, e.g., the saddlepoint approximations of Lieberman (1994), Marsh (1998) and Butler and Paoletta (2008), or the inverted-beta procedure of De Juan and Arroyo (2008). The quality of these approximations notwithstanding, it is useful to have available an exact expression, not least for establishing the accuracy of an approximation.

* Corresponding author. Present address: Swiss Banking Institute, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

Email address: broda@isb.uzh.ch (S. Broda).

In this note, a computable expression is given for the general case of a ratio of noncentral quadratic forms in Gaussian random vectors, along with a computationally simpler expression for the practically relevant special case of two central quadratic forms. The plan is as follows. After discussing a few preliminaries, two different proofs for our main result are derived in Sections 3 and 4, respectively. Section 5 discusses the aforementioned special case. We conclude with a brief performance comparison.

2 Notation and Setup

For symmetric matrices \mathbf{A}^* and \mathbf{B}^* , consider the ratio of quadratic forms

$$R = \frac{\mathbf{Z}'\mathbf{A}^*\mathbf{Z}}{\mathbf{Z}'\mathbf{B}^*\mathbf{Z}},$$

where $\mathbf{Z} \sim N_n(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} > 0$, and \mathbf{B}^* is positive semidefinite with at least one eigenvalue strictly greater than zero. This last requirement ensures that the denominator of R is almost surely positive. Note that without any loss of generality, we may assume that $\boldsymbol{\Sigma} = \mathbf{I}$, because

$$R = \frac{\mathbf{Z}'\mathbf{A}^*\mathbf{Z}}{\mathbf{Z}'\mathbf{B}^*\mathbf{Z}} =_d \frac{\mathbf{X}'\mathbf{A}\mathbf{X}}{\mathbf{X}'\mathbf{B}\mathbf{X}},$$

where $\mathbf{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{A}^*\boldsymbol{\Sigma}^{\frac{1}{2}}$, $\mathbf{B} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{B}^*\boldsymbol{\Sigma}^{\frac{1}{2}}$, $\mathbf{X} = \boldsymbol{\Sigma}^{-1/2}\mathbf{Z} \sim N(\boldsymbol{\mu}, \mathbf{I})$, and $\boldsymbol{\mu} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu}^*$.

Next, construct the spectral decomposition

$$\mathbf{A} - r\mathbf{B} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}', \tag{1}$$

where $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$, the λ_i are the eigenvalues of $\mathbf{A} - r\mathbf{B}$, some of which, depending on \mathbf{A} and \mathbf{B} , might be zero, and $\mathbf{p}_i = (p_{1i}, \dots, p_{ni})'$ is the (column) eigenvector of $\mathbf{A} - r\mathbf{B}$ corresponding to λ_i .

Algorithms for computing the c.d.f. of R typically rely on the following equality: define $\mathbf{W} := \mathbf{P}'\mathbf{X} \sim N(\boldsymbol{\nu}, \mathbf{I}_n)$, where $\boldsymbol{\nu} := \mathbf{P}'\boldsymbol{\mu} = (\nu_1, \dots, \nu_n)'$. Then the c.d.f. of R can be written

$$\begin{aligned} \Pr(R \leq r) &= \Pr(\mathbf{X}'\mathbf{A}\mathbf{X} \leq r\mathbf{X}'\mathbf{B}\mathbf{X}) = \Pr(\mathbf{X}'(\mathbf{A} - r\mathbf{B})\mathbf{X} \leq 0) \\ &= \Pr(\mathbf{X}'\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'\mathbf{X} \leq 0) = \Pr(\mathbf{W}'\boldsymbol{\Lambda}\mathbf{W} \leq 0) = F_S(0), \end{aligned} \tag{2}$$

where $S := \sum_{i=1}^n \lambda_i W_i^2$ and $W_i^2 \stackrel{\text{ind}}{\sim} \chi^2(1, \theta_i)$, i.e., as the c.d.f. of a constructed random variable S at zero. The distribution of S is that of a weighted sum of noncentral χ^2 random variables, each with one degree of freedom and noncentrality parameter $\theta_i = \nu_i^2$, $i = 1, \dots, n$. Efficient algorithms exist for evaluating the c.d.f. of S , and, thus, that of R . For evaluating the density, however, relationship (2) is of no use, so that different techniques must be employed.

3 The General Case

This section covers the most general setting, in which possibly $\boldsymbol{\mu} \neq \mathbf{0}$, and the λ_i are permitted to be multiple eigenvalues. The algorithm we present here is based on the following result from Geary (1944). Let N and D be continuous random variables with joint characteristic function $\varphi_{N,D}$ and such that $\Pr(D > 0) = 1$ and $\mathbb{E}[D] < \infty$. Geary (1944) showed that the density of $R = N/D$ can be written as

$$f_R(r) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \left[\frac{\partial \varphi_{N,D}(s,t)}{\partial t} \right]_{t=-rs} ds, \quad (3)$$

where $i^2 = -1$. We set $N = \mathbf{X}'\mathbf{A}\mathbf{X}$ and $D = \mathbf{X}'\mathbf{B}\mathbf{X}$. Because the joint moment generating function $\mathbb{M}(s,t)$, say, of N and D exists, we have that $\varphi(s,t) = \mathbb{M}(is,it)$, and we can rewrite (3) as

$$f_R(r) = \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} [\mathbb{M}^*(is)] ds, \quad (4)$$

where $\mathbb{M}^*(s) := [\partial \mathbb{M}(s,t)/\partial t]_{t=-rs}$ is given in Butler and Paoletta (2008) as

$$\mathbb{M}^*(s) = \left[\prod_{i=1}^n (1 - 2s\lambda_i)^{-1/2} \right] \exp \left\{ s \sum_{i=1}^n \frac{\lambda_i \nu_i^2}{1 - 2s\lambda_i} \right\} \left[\operatorname{tr} \mathbf{D}^{-1}\mathbf{H} + \boldsymbol{\nu}'\mathbf{D}^{-1}\mathbf{H}\mathbf{D}^{-1}\boldsymbol{\nu} \right],$$

$\mathbf{D} = \mathbf{I} - 2s\boldsymbol{\Lambda}$, $\mathbf{H} = \mathbf{P}'\mathbf{B}\mathbf{P}$, and we have exploited the fact that $\operatorname{Re}[\mathbb{M}^*(is)]$ is an even function of s . For use with software packages not supporting complex arithmetic, the following result will be proven.

Proposition 1 *Let $R = \mathbf{X}'\mathbf{A}\mathbf{X}/\mathbf{X}'\mathbf{B}\mathbf{X}$, where $\mathbf{X} \sim \mathbf{N}(\boldsymbol{\mu}, \mathbf{I})$ and $\mathbf{B} \geq 0$ with at least one strictly positive eigenvalue. Then the density of R is*

$$f_R(r) = \frac{1}{\pi} \int_0^{\infty} \frac{\rho(u) \cos \beta(u) - u\delta(u) \sin \beta(u)}{2\gamma(u)} du, \quad (5)$$

where

$$\begin{aligned} \beta(u) &= \frac{1}{2} \sum_{i=1}^n \arctan a_i + \frac{\theta_i a_i}{c_i}, & \gamma(u) &= \exp \left\{ \frac{1}{2} \sum_{i=1}^n \frac{\theta_i b_i}{c_i} + \frac{1}{4} \ln c_i \right\}, \\ \rho(u) &= \operatorname{tr} \mathbf{H}\mathbf{F}^{-1} + \boldsymbol{\nu}'\mathbf{F}^{-1}(\mathbf{H} - u^2\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Lambda})\mathbf{F}^{-1}\boldsymbol{\nu}, & \delta(u) &= \operatorname{tr} \mathbf{H}\boldsymbol{\Lambda}\mathbf{F}^{-1} + 2\boldsymbol{\nu}'\mathbf{F}^{-1}\mathbf{H}\boldsymbol{\Lambda}\mathbf{F}^{-1}\boldsymbol{\nu}, \end{aligned}$$

λ_i are the eigenvalues of $\mathbf{A} - r\mathbf{B}$, \mathbf{p}_i the corresponding eigenvectors, $a_i = \lambda_i u$, $b_i = a_i^2$, $c_i = 1 + b_i$, $\theta_i = \nu_i^2 = (\mathbf{p}_i'\boldsymbol{\mu})^2$, $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, $\mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$, $\mathbf{H} = \mathbf{P}'\mathbf{B}\mathbf{P}$, and $\mathbf{F} = \mathbf{I} + u^2\boldsymbol{\Lambda}^2$.

Proof. Let

$$\psi(s) := \left[\prod_{i=1}^n (1 - 2s\lambda_i)^{-1/2} \right] \exp \left\{ s \sum_{i=1}^n \frac{\lambda_i \nu_i^2}{1 - 2s\lambda_i} \right\}$$

and

$$\Upsilon(s) := \left[\operatorname{tr} \mathbf{D}^{-1}\mathbf{H} + \boldsymbol{\nu}'\mathbf{D}^{-1}\mathbf{H}\mathbf{D}^{-1}\boldsymbol{\nu} \right].$$

Then

$$\begin{aligned}
\operatorname{Re} \mathbb{M}^*(is) &= \operatorname{Re} [\psi(is) \Upsilon(is)] \\
&= \operatorname{Re} \psi(is) \operatorname{Re} \Upsilon(is) - \operatorname{Im} \psi(is) \operatorname{Im} \Upsilon(is) \\
&= |\psi(is)| [\cos \arg \psi(is) \operatorname{Re} \Upsilon(is) - \sin \arg \psi(is) \operatorname{Im} \Upsilon(is)].
\end{aligned}$$

As shown in Imhof (1961),

$$|\psi(is)| = [\gamma(2s)]^{-1} \quad \text{and} \quad \arg \psi(is) = \beta(2s),$$

where $\gamma(u)$ and $\beta(u)$ are as in (5). Denoting the elements of \mathbf{H} as (h_{jk}) ,

$$\begin{aligned}
\Upsilon(is) &= \sum_{j=1}^n \frac{h_{jj}}{1 - 2is\lambda_j} + \sum_{j=1}^n \sum_{k=1}^n \frac{\nu_j \nu_k h_{jk}}{(1 - 2is\lambda_j)(1 - 2is\lambda_k)} \\
&= \sum_{j=1}^n \frac{h_{jj}(1 + 2is\lambda_j)}{1 + 4s^2\lambda_j^2} + \sum_{j=1}^n \sum_{k=1}^n \frac{\nu_j \nu_k h_{jk}(1 + 2is\lambda_j)(1 + 2is\lambda_k)}{(1 + 4s^2\lambda_j^2)(1 + 4s^2\lambda_k^2)}
\end{aligned}$$

so that

$$\operatorname{Re} \Upsilon(is) = \sum_{j=1}^n \frac{h_{jj}}{1 + 4s^2\lambda_j^2} + \sum_{j=1}^n \sum_{k=1}^n \frac{\nu_j \nu_k h_{jk}(1 - 4s^2\lambda_j\lambda_k)}{(1 + 4s^2\lambda_j^2)(1 + 4s^2\lambda_k^2)} = \rho(2s)$$

where

$$\rho(u) := \operatorname{tr} \mathbf{H} \mathbf{F}^{-1} + \boldsymbol{\nu}' \mathbf{F}^{-1} (\mathbf{H} - u^2 \boldsymbol{\Lambda} \mathbf{H} \boldsymbol{\Lambda}) \mathbf{F}^{-1} \boldsymbol{\nu}$$

and $\mathbf{F} := \mathbf{I} + u^2 \boldsymbol{\Lambda}^2$.

Similarly,

$$\operatorname{Im} \Upsilon(is) = \sum_{j=1}^n \frac{2sh_{jj}\lambda_j}{1 + 4s^2\lambda_j^2} + \sum_{j=1}^n \sum_{k=1}^n \frac{2s\nu_j\nu_k h_{jk}(\lambda_j + \lambda_k)}{(1 + 4s^2\lambda_j^2)(1 + 4s^2\lambda_k^2)} = 2s\delta(2s),$$

where

$$\delta(u) := \operatorname{tr} \mathbf{H} \boldsymbol{\Lambda} \mathbf{F}^{-1} + 2\boldsymbol{\nu}' \mathbf{F}^{-1} \mathbf{H} \boldsymbol{\Lambda} \mathbf{F}^{-1} \boldsymbol{\nu}.$$

Thus

$$\operatorname{Re} \mathbb{M}^*(is) = \frac{\rho(2s) \cos \beta(2s) - 2s\delta(2s) \sin \beta(2s)}{\gamma(2s)},$$

which, upon changing variables to $u = 2s$, yields (5). ■

For use with canned routines for numerical integration, we have found it convenient to map the integral onto a finite interval via the substitution $u = (1 - t)/t$, replacing the integrand at $t = 0$ with its limit, which is zero. The only remaining potential caveat is that integrands of the type appearing in (5) can be highly oscillatory for certain problems, thus rendering numerical integration difficult. In our experiments however, this behavior only occurred for very small values of n in very extreme parameter constellations, which are unlikely to be encountered in real-world applications such as those considered in Section 6.

4 Distinct Eigenvalues

This section gives an alternative derivation for the density when all the nonzero eigenvalues are simple, i.e., each λ_i has multiplicity one. Our starting point is the result in (2): provided an expression for the c.d.f. is available, straightforward differentiation will yield an expression for the density. We base our algorithm on the expression for the c.d.f. given in Imhof (1961). Kamanu (2006) applied a similar idea, but considered only the central case ($\boldsymbol{\mu} = \mathbf{0}$), for which an even more efficient algorithm can be constructed, as will be demonstrated in Section 5 below.

Applied to the case at hand, Imhof's result is

$$F_R(r) = \Pr[\mathbf{W}'\boldsymbol{\Lambda}\mathbf{W} \leq 0] = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \beta(u)}{u\gamma(u)} du, \quad (6)$$

where $\beta(u)$ and $\gamma(u)$ are as in (5) above.

An expression for the density is obtained by differentiating (6) with respect to r , for which we will require derivatives of λ_i . Using a result from Magnus (1985), if λ_i is a simple eigenvalue of $\mathbf{A} - r\mathbf{B}$, then

$$\begin{aligned} \dot{\lambda}_i &:= \frac{d}{dr} \lambda_i(r) = -\mathbf{p}'_i \mathbf{B} \mathbf{p}_i = -h_{ii} \quad \text{and} \\ \dot{\mathbf{p}}_i &:= \left[\frac{d}{dr} p_{1i}(r), \dots, \frac{d}{dr} p_{ni}(r) \right]' = -\mathbf{P} \mathbf{Q}_i^+ \mathbf{P}' \mathbf{B} \mathbf{p}_i = -\mathbf{P} \mathbf{Q}_i^+ \mathbf{h}_i, \end{aligned} \quad (7)$$

where $\mathbf{Q}_i = (\lambda_i \mathbf{I}_n - \boldsymbol{\Lambda})$, \mathbf{h}_i is the i th column of $\mathbf{H} = \mathbf{P}' \mathbf{B} \mathbf{P}$, and a superscript $+$ on a matrix denotes its Moore-Penrose inverse. Writing, as before, $a_i = \lambda_i u$, $b_i = a_i^2$, $c_i = 1 + b_i$, and $\theta_i = \nu_i^2 = (\mathbf{p}'_i \boldsymbol{\mu})^2$, the derivative of θ_i is

$$\dot{\theta}_i := \frac{d}{dr} \theta_i(r) = 2\mathbf{p}'_i \boldsymbol{\mu} \boldsymbol{\mu}' \dot{\mathbf{p}}_i = -2\nu_i \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_j h_{ij}}{\lambda_i - \lambda_j},$$

and so

$$f_R(r) = \frac{1}{\pi} \int_0^\infty \left[\frac{\gamma_r(u) \sin \beta(u)}{u(\gamma(u))^2} - \frac{\beta_r(u) \cos \beta(u)}{u\gamma(u)} \right] du, \quad (8)$$

where

$$\begin{aligned} \beta_r(u) &:= \frac{d}{dr} \beta(u) = \frac{1}{2} \sum_{i=1}^n \frac{u \dot{\lambda}_i}{c_i} + \frac{\dot{\theta}_i a_i}{c_i} + \frac{\theta_i u \dot{\lambda}_i}{c_i} - \frac{2\theta_i b_i u \dot{\lambda}_i}{c_i^2} \\ &= -\frac{u}{2} \sum_{i=1}^n \left[\frac{h_{ii}}{1 + u^2 \lambda_i^2} + \frac{\nu_i^2 h_{ii} (1 - u^2 \lambda_i^2)}{(1 + u^2 \lambda_i^2)^2} - \frac{\dot{\theta}_i \lambda_i}{1 + u^2 \lambda_i^2} \right] \\ &= -\frac{u}{2} \left[\sum_{i=1}^n \frac{h_{ii}}{1 + u^2 \lambda_i^2} + \sum_{i=1}^n \sum_{j=1}^n \frac{\nu_i \nu_j h_{ij} (1 - u^2 \lambda_i \lambda_j)}{(1 + u^2 \lambda_i^2)(1 + u^2 \lambda_j^2)} \right] =: -\frac{u}{2} \rho(u), \end{aligned}$$

and the last equality follows from the partial fraction decomposition

$$\begin{aligned}
-\sum_{i=1}^n \frac{\dot{\theta}_i \lambda_i}{1+u^2 \lambda_i^2} &= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j \lambda_i h_{ij}}{(1+u^2 \lambda_i^2)(\lambda_i - \lambda_j)} \\
&= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j \lambda_j h_{ij}}{(1+u^2 \lambda_i^2)(\lambda_i - \lambda_j)} + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j h_{ij} (1 - u^2 \lambda_i \lambda_j)}{(1+u^2 \lambda_i^2)(1+u^2 \lambda_j^2)} \\
\Leftrightarrow -\sum_{i=1}^n \frac{\dot{\theta}_i \lambda_i}{1+u^2 \lambda_i^2} &= \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j h_{ij} (1 - u^2 \lambda_i \lambda_j)}{(1+u^2 \lambda_i^2)(1+u^2 \lambda_j^2)}.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\gamma_r(u) &:= \frac{d}{dr} \gamma(u) = \gamma(u) \left(\frac{1}{2} \sum_{i=1}^n \frac{\dot{\theta}_i b_i}{c_i} + \frac{2\theta_i a_i u \dot{\lambda}_i}{c_i} - \frac{2\theta_i a_i b_i u \dot{\lambda}_i}{c_i^2} + \frac{a_i u \dot{\lambda}_i}{c_i} \right) \\
&= -\frac{\gamma(u) u^2}{2} \sum_{i=1}^n \left[\frac{\lambda_i h_{ii}}{1+u^2 \lambda_i^2} + \frac{2\nu_i^2 \lambda_i h_{ii}}{(1+u^2 \lambda_i^2)^2} - \frac{\dot{\theta}_i \lambda_i^2}{1+u^2 \lambda_i^2} \right] \\
&= -\frac{\gamma(u) u^2}{2} \left[\sum_{i=1}^n \frac{\lambda_i h_{ii}}{1+u^2 \lambda_i^2} + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{2\nu_i \nu_j \lambda_i h_{ij}}{(1+u^2 \lambda_i^2)(1+u^2 \lambda_j^2)} \right] \\
&=: -\frac{\gamma(u) u^2}{2} \delta(u),
\end{aligned}$$

and the last equality follows from the partial fraction decomposition

$$\begin{aligned}
-\frac{\dot{\theta}_i \lambda_i^2}{1+u^2 \lambda_i^2} &= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j \lambda_i^2 h_{ij}}{(1+u^2 \lambda_i^2)(\lambda_i - \lambda_j)} \\
&= 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j \lambda_j^2 h_{ij}}{(1+u^2 \lambda_i^2)(\lambda_i - \lambda_j)} + 2 \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{\nu_i \nu_j (\lambda_i + \lambda_j) h_{ij}}{(1+u^2 \lambda_i^2)(1+u^2 \lambda_j^2)} \\
\Leftrightarrow -\frac{\dot{\theta}_i \lambda_i^2}{1+u^2 \lambda_i^2} &= \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n \frac{2\nu_i \nu_j \lambda_i h_{ij}}{(1+u^2 \lambda_i^2)(1+u^2 \lambda_j^2)}.
\end{aligned}$$

Substituting $\beta_r(u)$ and $\gamma_r(u)$ in (8) gives (5).

5 Central Case, Distinct Eigenvalues

In this section we discuss the special case where $\boldsymbol{\mu} = \mathbf{0}$ and all the λ_i have multiplicity one. The reason for treating this case separately, apart from its practical relevance, is that a more efficient algorithm for computing the c.d.f. (and thus, by differentiation, the density) of R is available in this setting. The result is usually attributed to Pan (1964), though a

similar expression appears in Grad and Solomon (1955); see also Farebrother (1984). A proof can be found in Durbin and Watson (1971). We give the expression in a slightly different form, which is more suited for our purposes.

Let the eigenvalues λ_i be sorted in descending fashion, and denote by v the number of nonnegative eigenvalues. Then

$$F_R(r) = 1 + \frac{1}{\pi} \sum_{j=1}^{\lceil v/2 \rceil} (-1)^j \int_{-1}^1 \exp \{g_j(t)\} \frac{dt}{(1-t^2)^{\frac{1}{2}}} \quad (9)$$

where

$$g_j(t) = \begin{cases} -\frac{1}{2} \sum_{\substack{i=1 \\ i \neq 2j, 2j-1}}^n \ln \left| 1 - 2\lambda_i [\lambda_{2j}(1-t) + \lambda_{2j-1}(1+t)]^{-1} \right|, & \text{if } j \leq v/2 \\ -\frac{1}{2} \sum_{\substack{i=1 \\ i \neq 2j-1}}^n \ln \left| 1 - 2\lambda_i (\lambda_{2j-1}(1-t))^{-1} \right|, & \text{if } j = (v+1)/2. \end{cases}$$

Each integral in (9) can be approximated by Gauss-Chebyshev quadrature as

$$\frac{1}{\pi} \int_{-1}^1 \exp \{g_j(t)\} \frac{dt}{(1-t^2)^{\frac{1}{2}}} \approx \frac{1}{N} \sum_{k=1}^N \exp \{g_j(y_k^{(N)})\}, \quad (10)$$

where $y_k^{(N)} = \cos\left(\frac{(2k-1)\pi}{2N}\right)$. A value of $N = 12$ is usually sufficient; see the discussion in Section 6. If $v > n/2$, it is computationally more efficient to evaluate $1 - F_{-R}(-r)$.

Straightforward calculation then shows that the density of R can be computed by inserting $\dot{g}_j(y_k^{(N)})$ after the summation sign in (10), where

$$\dot{g}_j(t) := \begin{cases} \sum_{\substack{i=1 \\ i \neq 2j, 2j-1}}^n \frac{\dot{\lambda}_i - [\dot{\lambda}_{2j}(1-t) + \dot{\lambda}_{2j-1}(1+t)] [\lambda_{2j}(1-t) + \lambda_{2j-1}(1+t)]^{-1} \lambda_i}{\lambda_{2j}(1-t) + \lambda_{2j-1}(1+t) - 2\lambda_i}, & \text{if } j \leq v/2 \\ \sum_{\substack{i=1 \\ i \neq 2j-1}}^n \frac{\dot{\lambda}_i - [\dot{\lambda}_{2j-1}] [\lambda_{2j-1}]^{-1} \lambda_i}{\lambda_{2j-1}(1-t) - 2\lambda_i}, & \text{if } j = (v+1)/2. \end{cases}$$

6 Performance Comparison and Conclusions

The algorithms described above have been implemented in the Matlab programming language by the authors (the programs are available upon request), using the built-in routine `quadl` for the evaluation of the integral in (5), mapped onto the interval $(0, 1]$ via the substitution $u = (1-t)/t$. In order to facilitate a comparison of the accuracy and speed of the different algorithms, a setting must be chosen in which both procedures are applicable, viz., that of Section 5. The null and alternative distributions of most tests in linear models

fall into this category, and the corresponding tail probabilities are commonly computed by means of Imhof’s procedure; a recent example is Bivand et al. (2008). We consider two examples: i) the density corresponding to the upper bound distribution of the Durbin-Watson statistic, as used in Farebrother (1980) for comparing Imhof’s and Pan’s procedures for the distribution function, and ii) the density of the least squares estimator in an AR(1) model with regressors, as required for the mode-adjusted estimator of Broda et al. (2007).

6.1 Lower Bound of the Durbin-Watson Statistic

Farebrother considered the lower bound distribution of the Durbin-Watson statistic in a linear regression with 5 regressors, corresponding to $\mathbf{B} = \mathbf{I}_{T-5}$ and $\mathbf{A} = \text{diag}(a_1, \dots, a_{T-5})$, where

$$a_i = 2 - 2 \cos[(T - i)\pi/T], \quad i \in \{1, \dots, T - 5\},$$

and T is the sample size. While this is certainly not the most natural application for evaluating the density, our interest here is primarily in the performance of the algorithms *per se*, and replicating the setup of Farebrother (1980) allows us to compare our results to those for the distribution function obtained therein. Our findings are largely similar.

Table 1 contains results on the accuracy and speed (in seconds, using Matlab on a 2.8 GHz Intel Pentium 4) of algorithms (5) (‘Geary’) and (10) (‘Pan’). For the former, the tolerance in Matlab’s integration routine was set to $10^{-\eta}$, for $\eta \in \{3, 5, 7\}$. For the latter, the number of terms in the sum was set to $N \in \{3, 6, 12, 24\}$. The ‘exact’ values have been obtained from (5) with $\eta = 12$, and the measure of accuracy reported is minus the base-ten logarithm of the absolute relative error, which corresponds roughly to the number of correct significant digits. We evaluate the density at the lower 5% quantile of the distribution, for different sample sizes ranging between 10 and 200. The results can be summarized as follows. Like for the c.d.f., the ‘Pan’ algorithm with $N = 12$ terms in the sum achieves at least 10-digit accuracy for values of $T \leq 70$, and deteriorates for large values of T . The ‘Geary’ algorithm with $\eta = 7$ achieves 10-digit accuracy for all cases considered, but, expectedly, is considerably slower.

6.2 Least Squares Estimator for an AR(1)

As our second example, we consider the least squares estimator for an AR(1) model with exogenous regressors. The matrices \mathbf{A} and \mathbf{B} entering in (1) are defined in Broda et al. (2007); they depend on the sample size T , the regressor matrix \mathbf{X} , and the autoregressive parameter, α . For this study we choose a regressor matrix consisting of a column of ones and a column of increasing natural numbers from 1 to T , corresponding to an intercept and a linear time trend. The autoregressive parameter is fixed at unity, and as before, we evaluate the density at the lower 5% quantile, which corresponds to the critical value of a unit root test based on the least squares estimator.

The numerical results are reported in Table 2. The performance of the ‘Geary’ algorithm

is very similar to the previous example, as is that of ‘Pan’ for sample sizes smaller than 70. For larger samples, the ‘Pan’ algorithm achieves higher accuracy than in the previous example.

| T | x | Pan | | | | Geary | | |
|-----|-------|-------------------|------------------|------------------|---------------------|-----------------|------------------|-------------------|
| | | $n = 3$ | $n = 6$ | $n = 12$ | $n = 24$ | $\eta = 3$ | $\eta = 5$ | $\eta = 7$ |
| 10 | 2.414 | 2.80 (0.001) | 6.11 (0.001) | 11.75 (0.001) | 11.71 (0.002) | 6.54 (0.011) | 6.54 (0.008) | 6.54 (0.018) |
| 20 | 1.828 | 4.57 (0.003) | 9.17 (0.003) | 15.19 (0.004) | 15.19 (0.006) | 7.58 (0.016) | 7.58 (0.011) | 10.80 (0.029) |
| 30 | 1.739 | 4.39 (0.004) | 8.73 (0.004) | 14.78 (0.007) | 14.89 (0.011) | 6.46 (0.033) | 6.46 (0.024) | 10.84 (0.061) |
| 40 | 1.721 | 4.13 (0.005) | 8.22 (0.007) | 14.26 (0.008) | ∞ (0.032) | 6.54 (0.071) | 8.93 (0.064) | 12.48 (0.100) |
| 50 | 1.721 | 3.71 (0.007) | 7.56 (0.009) | 13.05 (0.011) | 12.78 (0.017) | 8.70 (0.120) | 8.74 (0.132) | 12.66 (0.197) |
| 60 | 1.727 | 3.13 (0.009) | 6.78 (0.011) | 11.91 (0.015) | 11.89 (0.024) | 5.70 (0.088) | 10.66 (0.190) | 13.02 (0.352) |
| 70 | 1.735 | 2.47 (0.011) | 5.90 (0.015) | 10.63 (0.018) | 11.75 (0.029) | 6.16 (0.179) | 9.72 (0.266) | 9.72 (0.403) |
| 80 | 1.743 | 1.81 (0.013) | 4.96 (0.016) | 9.10 (0.021) | 9.91 (0.033) | 5.87 (0.231) | 10.24 (0.347) | 10.24 (0.440) |
| 90 | 1.751 | 1.44 (0.016) | 3.97 (0.019) | 7.92 (0.027) | 8.23 (0.043) | 4.99 (0.211) | 8.03 (0.389) | 11.06 (0.501) |
| 100 | 1.758 | 0.15 (0.018) | 2.93 (0.021) | 8.15 (0.031) | 6.62 (0.046) | 6.57 (0.409) | 12.46 (0.650) | 11.46 (0.792) |
| 150 | 1.788 | -6.28 (0.036) | -2.70 (0.045) | 0.01 (0.059) | 1.41 (0.089) | 6.45 (1.302) | 11.38 (2.228) | 12.58 (3.723) |
| 200 | 1.809 | -12.52 (0.074) | -8.77 (0.075) | -6.22 (0.099) | -4.85 (0.146) | 7.59 (3.179) | 7.08 (4.269) | 12.41 (10.206) |

Table 1

Accuracy, measured as minus the base-ten logarithm of the absolute relative error (and speed, in seconds) of algorithms (5) (‘Geary’) and (10) (‘Pan’) for the density corresponding to the upper bound distribution of the Durbin-Watson statistic at x , for different sample sizes ranging between 10 and 200. A value of ‘ ∞ ’ signifies that the relevant values agree to machine precision.

| T | x | Pan | | | | Geary | | |
|-----|--------|-----------------|-----------------|------------------|------------------|-----------------|------------------|-------------------|
| | | $n = 3$ | $n = 6$ | $n = 12$ | $n = 24$ | $\eta = 3$ | $\eta = 5$ | $\eta = 7$ |
| 10 | -0.323 | 4.01 (0.003) | 9.12 (0.002) | 14.74 (0.003) | 14.80 (0.004) | 7.06 (0.015) | 7.06 (0.011) | 6.85 (0.036) |
| 20 | 0.173 | 2.84 (0.004) | 6.30 (0.004) | 13.25 (0.005) | 14.65 (0.008) | 3.13 (0.015) | 6.64 (0.020) | 7.52 (0.032) |
| 30 | 0.400 | 2.80 (0.006) | 6.26 (0.006) | 13.19 (0.009) | 15.30 (0.015) | 6.13 (0.058) | 8.70 (0.101) | 12.47 (0.135) |
| 40 | 0.530 | 2.78 (0.006) | 6.24 (0.007) | 13.15 (0.010) | 14.99 (0.014) | 6.35 (0.075) | 9.06 (0.114) | 11.52 (0.171) |
| 50 | 0.613 | 2.77 (0.008) | 6.23 (0.009) | 13.13 (0.011) | 15.76 (0.017) | 6.41 (0.099) | 7.81 (0.111) | 9.32 (0.233) |
| 60 | 0.672 | 2.76 (0.009) | 6.22 (0.013) | 13.12 (0.013) | 14.87 (0.028) | 5.85 (0.146) | 8.32 (0.158) | 10.24 (0.335) |
| 70 | 0.715 | 2.76 (0.011) | 6.21 (0.013) | 13.12 (0.017) | 15.88 (0.026) | 5.67 (0.185) | 7.91 (0.228) | 11.97 (0.479) |
| 80 | 0.748 | 2.75 (0.014) | 6.21 (0.016) | 13.11 (0.020) | 15.33 (0.028) | 5.6 (0.260) | 7.62 (0.414) | 11.50 (0.673) |
| 90 | 0.774 | 2.75 (0.016) | 6.20 (0.018) | 13.68 (0.021) | 13.23 (0.029) | 5.57 (0.360) | 8.21 (0.553) | 12.12 (1.160) |
| 100 | 0.795 | 2.74 (0.019) | 6.20 (0.023) | 13.06 (0.025) | 14.12 (0.034) | 5.57 (0.482) | 7.80 (0.949) | 12.09 (1.597) |
| 150 | 0.861 | 2.74 (0.047) | 6.19 (0.047) | 13.10 (0.050) | 15.39 (0.066) | 5.85 (1.315) | 8.84 (3.589) | 8.84 (4.070) |
| 200 | 0.895 | 2.73 (0.095) | 6.19 (0.100) | 13.08 (0.106) | 15.21 (0.125) | 6.77 (3.965) | 10.30 (8.165) | 10.44 (10.829) |

Table 2

Same as Table 1, but for the density of the least squares estimator in an AR(1) model with an intercept and a time trend.

6.3 Conclusions

In view of the above results, the following general recommendation can be made: when it is applicable (i.e., if $\boldsymbol{\mu} = \mathbf{0}$ and the eigenvalues λ_i are distinct), the ‘Pan’ algorithm with $N = 12$ is preferred whenever $T \leq 70$; for all other cases, the ‘Geary’ algorithm with $\eta = 7$ should be used. Strictly speaking, of course, these results pertain only to the problems considered here, but we expect them to hold somewhat more generally.

Acknowledgements

Part of this research has been carried out within the National Centre of Competence in Research “Financial Valuation and Risk Management” (NCCR FINRISK), which is a research program supported by the *Swiss National Science Foundation*. The authors are grateful for the valuable and constructive comments from two anonymous referees and the associate editor.

References

- Bivand, R., Müller, W. G., Reder, M., 2008. Power calculations for global and local Moran’s *I*. *Computational Statistics & Data Analysis* Forthcoming. 8
- Broda, S., Carstensen, K., Paoletta, M. S., 2007. Bias-adjusted estimation in the ARX(1) model. *Computational Statistics & Data Analysis* 51, 3355–3367. 8
- Butler, R., Paoletta, M., 2008. Uniform saddlepoint approximations for ratios of quadratic forms. *Bernoulli* 14, 140–154. 1, 3
- De Juan, A., Arroyo, A. S. M., 2008. European incomplete catching-up. *Empirical Economics* Forthcoming. 1
- Durbin, J., Watson, G. S., 1971. Testing for serial correlation in least squares regression. III. *Biometrika* 58 (1), 1–19. 7
- Farebrother, R. W., 1980. Algorithm AS 153. Pan’s procedure for the tail probabilities of the Durbin-Watson statistic. *Applied Statistics* 29, 224–27. 8
- Farebrother, R. W., 1984. The distribution of a linear combination of central χ^2 random variables. A remark on AS 153: Pan’s procedure for the tail probabilities of the Durbin-Watson statistic. *Applied Statistics* 33, 363–69. 7
- Geary, R. C., 1944. Extension of a theorem by Harald Cramer on the frequency distribution of the quotient of two variables. *Journal of the Royal Statistical Society* 107, 56–57. 3
- Grad, A., Solomon, H., 1955. Distribution of quadratic forms and some applications. *Annals of Mathematical Statistics* 26, 464–77. 7
- Imhof, J. P., 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48, 419–26. 1, 4, 5, 8
- Kamanu, T. K. K., 2006. Location-based estimation of the autoregressive coefficient in ARX(1) models. Thesis, University of the Western Cape. 1, 5

- Lieberman, O., 1994. Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. *Journal of the American Statistical Association* 89, 924–28. 1
- Lu, Z.-H., 2006. The numerical evaluation of the probability density function of a quadratic form in normal variables. *Computational Statistics & Data Analysis* 51, 1986–1996. 1
- Magnus, J. R., 1985. On differentiating eigenvalues and eigenvectors. *Econometric Theory* 1, 179–191. 5
- Marsh, P. W. N., 1998. Saddlepoint approximations for noncentral quadratic forms. *Econometric Theory* 14, 539–559. 1
- Pan, J.-J., 1964. Distributions of the noncircular serial correlation coefficients. *Shuxue Jinzhan* 7, 328–37, translated by N. N. Chan for *Selected Translations in Mathematical Statistics and Probability*, Vol. 7 (1968), 281–92. 6, 8
- Paolella, M. S., 2003. Computing moments of ratios of quadratic forms in normal variables. *Computational Statistics & Data Analysis* 42, 313–331. 1