ISSN 1066-5307, Mathematical Methods of Statistics, 2009, Vol. 18, No. 4, pp. 375–396. © Allerton Press, Inc., 2009.

Conservative Confidence Ellipsoids for Linear Model Parameters

A. L. Rukhin[1*]

[1] Statist. Engineering Division, National Inst. of Standards and Technology, USA

Received February 17, 2009; in final form, November 12, 2009

Abstract—This paper studies properties of conservative confidence ellipsoids for parameters of a general linear model. These regions are obtained on the basis of a linear estimator when only a vague knowledge of (heterogeneous) error variances is available. The required optimization problem is formulated and the solution space is described. The relationship of this problem to moments of quadratic forms in Gaussian random variables and to multiple hypergeometric functions is demonstrated. We explore the situation when the least favorable variances are equal. An example of a telephone switching study is considered.

Key words: Binet–Cauchy formula, Dirichlet averages, elementary symmetric functions, elliptic integrals, multilinear forms, quadratic forms in normal vectors, Schur product, zonal polynomials.

2000 Mathematics Subject Classification: primary 62J05, 62F25; secondary 62E17, 62P30, 60E99.

DOI: 10.3103/S1066530709040048

1. INTRODUCTION

Confidence regions for parameters of a linear model are important both in theory and applications of statistics. In the classical linear model setting, with a given matrix of error covariances, the exact $1-\alpha$ confidence ellipsoids are well understood. However if the error covariance matrix is not known up to a scalar factor, it is much more difficult to find a confidence region with a guaranteed coverage probability.

Consider the general linear model, $Y \sim N_n(X\theta, \Sigma)$ with a diagonal matrix Σ. Thus with θ denoting an r-dimensional parameter, $r \le n$, and X a design matrix of size $n \times r$ of rank r,

$$Y = X\theta + \epsilon. \tag{1}$$

The vector ϵ formed by independent Gaussian errors ϵ_j, $j = 1, \ldots, n$, has zero mean and unknown variances σ_j^2. In the particular, common mean case, when $r = 1$, $X = e$ is the n-dimensional vector of ones. The classical least squares estimate $\hat{\theta}$, determined from the equation, $X^T\Sigma^{-1}X\hat{\theta} = X^T\Sigma^{-1}Y$, clearly depends on Σ which typically is (at least partially) unknown.

We work with linear unbiased estimators of the form $\delta = WY$, where W is a $r \times n$ matrix. Then $WX = I$, and $\mathrm{Var}(\delta) = W\Sigma W^T$ is the covariance matrix of δ. Here and further I denotes the identity matrix whose size is clear from the context. An alternative form is $\delta = (X^TQX)^{-1}X^TQY$ with a $n \times n$ matrix Q, where Q is interpreted as a guess about Σ^{-1}. Thus, $W = (X^TQX)^{-1}X^TQ$. If $\Sigma = Q^{-1}$, then $\mathrm{Var}(\delta) = (X^TQX)^{-1}$. However, even if $\hat{Q}^{-1}$ is an unbiased estimator of Σ, the commonly used matrix statistics $(X^T\hat{Q}X)^{-1}$ typically underestimates $\mathrm{Var}(\delta)$.

The new estimator $\widehat{\mathrm{Var}(\delta)}$ of the covariance matrix of δ suggested by the author [15, 17] is designed to mitigate this bias and to assure conservative nature of the corresponding confidence region. Put

$$\widehat{\mathrm{Var}(\delta)} = [Y^T(I - XW)^TQ(I - XW)Y](X^TQX)^{-1}. \tag{2}$$

[*] E-mail: andrew.rukhin@nist.gov

A confidence ellipsoid for θ based on δ is given by

$$(\delta-\theta)^T\widehat{\mathrm{Var}(\delta)}^{-1}(\delta-\theta)\le t^2. \tag{3}$$

Actually in (2) one can have a more general quadratic form in the residuals $(I-XW)Y$, but the present choice minimizes the expected volume of (3) when Q^{-1} is the true error covariance matrix [17]. The appropriate choice of the matrix Q is an important issue in using our method. In many problems of meta-analysis one has to combine results of m independent but heterogeneous studies, where each study produces an unbiased estimate of its linear function(s) of θ of uncertain accuracy. Then $X^T=(X_1,\dots,X_m)^T$, with X_i representing a known $n_i\times r$ matrix, $i=1,\dots,m$, $n=n_1+\cdots+n_m$. Under a similar partitioning of $\epsilon^T=(\epsilon_1,\dots,\epsilon_m)^T$, the n_i-dimensional normal vector ϵ_i^T has independent components with common variance σ_i^2, which can be estimated on the basis of available (but commonly scarce) data. Then Q can be formed by diagonal matrix blocks q_iI, $i=1,\dots,m$, with q_i^{-1} being the estimate of σ_i^2 in the ith study.

Under this scenario, Shalaevsky [18] obtained an approximation to the coverage probability of confidence ellipsoids for θ when $\min n_i\to\infty$. This condition is rarely if ever met in applications. Fuller and Rao [6] investigated a two-stage estimation procedure when $m\to\infty$. Carroll and Ruppert [5] suggested several parametric models for heteroscedasticity. Wu [20] devised some versions of jackknife variance estimators and compared them to the bootstrap method. The asymptotic behavior of these methods was elucidated in [13] and [19].

While these authors are mainly interested in point estimation of the parameters, our approach is motivated by construction of confidence regions. In a similar vein, Adkins and Hill [1] explored the coverage probability and the volume of a percentile bootstrap confidence ellipsoid centered at the Stein estimator.

Our goal is to determine the coverage probability of (3) for any Σ at least for large t. In Section 3 we discuss the asymptotic behavior of this probability which matches that of the probability when t^2 is a multiple of $F_{r,n-r}(\alpha)$, the αth percentile of F-distribution with r and $n-r$ degrees of freedom. Section 2 contains needed results about the moments of quadratic forms in Gaussian random variables relating them to multiple hypergeometric functions known as Dirichlet averages [4] and to zonal polynomials [13]. The least favorable variances leading to the worst coverage probability depend only on the matrices X and Q, and can be found as a solution to the optimization problem formulated and studied in Section 4. The situation when these variances are equal is also described there. The optimization problem can be viewed as maximization of a function of r multilinear forms in reciprocals of the variances for a given value of their product. Section 5 gives some examples. All proofs are collected in the Appendix.

2. MOMENTS OF QUADRATIC FORMS IN NORMAL VARIABLES

This section contains basic mathematical results needed to derive conservative confidence ellipsoids.

2.1. Asymptotics and F-Approximation

Let S_r be the unit sphere in r-dimensional space parametrized by $\omega_1,\dots,\omega_r$, $\sum\omega_i^2=1$. Then with $d\omega$ denoting the normalized uniform distribution over S_r, define for $\lambda_i>0$, $i=1\dots,r$, and real q

$$R_q(\lambda_1,\dots,\lambda_r)=\int\limits_{S_r}\Big[\sum_i\lambda_i\omega_i^2\Big]^q\,d\omega. \tag{4}$$

The following result is proven in [17].

Theorem 2.1. *Let $Z_1,\dots,Z_n$, $n\ge r$, be independent standard normal variables, and $\lambda_{r+1},\dots,\lambda_n$ be fixed positive numbers. As $\lambda_1,\dots,\lambda_r\to 0$,*

$$\lim\frac{P\left(\sum_{i=1}^r\lambda_iZ_i^2>\sum_{k=r+1}^n\lambda_kZ_k^2\right)}{R_{(n-r)/2}(\lambda_1,\dots,\lambda_r)}=\frac{\Gamma(n/2)}{\sqrt{\lambda_{r+1}\cdots\lambda_n}\,\Gamma(r/2)\Gamma((n-r+2)/2)},$$

where the function $R_{(n-r)/2}(\lambda_1,\dots,\lambda_r)$ is defined by (4).

When $\lambda_i = t^{-2}$, $i = 1, \dots, r$, $R_{(n-r)/2} = t^{r-n}$, $\lambda_k = 1$, $k = r+1, \dots, n$, this formula agrees with the known result for the tail probabilities of $F_{r,n-r}$, an F random variable with r and $n-r$ degrees of freedom. Indeed as $t \to \infty$,

$$P\left(F_{r,n-r} > \frac{(n-r)t^2}{r}\right) \sim \frac{\Gamma(n/2)}{\Gamma(r/2)\Gamma((n-r+2)/2)t^{n-r}}.$$

According to Theorem 2.1 for $t \to \infty$,

$$P\left(\sum_{i=1}^{r} \lambda_i Z_i^2 > t^2 \sum_{k=r+1}^{n} \lambda_k Z_k^2\right) \sim P\left(\frac{(\lambda_{r+1}\cdots\lambda_n)^{1/(n-r)}}{[R_{(n-r)/2}(\lambda_1,\dots,\lambda_r)]^{2/(n-r)}} F_{r,n-r} > \frac{(n-r)t^2}{r}\right). \quad (5)$$

Thus, in the tail probability approximation for the ratio of two quadratic forms, $\sum_{i=1}^{r} \lambda_i Z_i^2$ and $\sum_{k=r+1}^{n} \lambda_k Z_k^2$, the λ_k's in the denominator are to be replaced by their geometric mean, while the λ_i's in the numerator are replaced by their spherical average $[R_{(n-r)/2}(\lambda_1,\dots,\lambda_r)]^{2/(n-r)}$.

The formulas in Section 7.2 [17] show that the probabilities (5) can be expressed via sums of integrals,

$$\int\limits_{S_r} \Big(\sum_i \lambda_i t^{-2}\omega_i^2\Big)^{(n-r)/2} \prod_k \Big(\sum_i \lambda_i t^{-2}\omega_i^2 + \lambda_k\Big)^{-\nu_k - 1/2} d\omega.$$

Since each of these integrals is bounded by $R_{(n-r)/2}(\lambda_1 t^{-2},\dots,\lambda_r t^{-2}) \prod_k \lambda_k^{-\nu_k-1/2}$, one has for any fixed t,

$$P\left(\frac{\sum_{i=1}^{r}\lambda_i Z_i^2}{\sum_{k=r+1}^{n}\lambda_k Z_k^2} > t^2\right) \le \frac{R_{(n-r)/2}(\lambda_1 t^{-2},\dots,\lambda_r t^{-2})\Gamma(n/2)}{\sqrt{\lambda_{r+1}\cdots\lambda_n}\,\Gamma(r/2)\Gamma((n-r+2)/2)}.$$

According to the Okamoto inequality [11, p. 303],

$$\begin{aligned}
P\left(\frac{\sum_{i=1}^{r}\lambda_i Z_i^2}{\sum_{k=r+1}^{n}\lambda_k Z_k^2} > t^2\right) &\le P\left(\frac{\sum_{i=1}^{r}\lambda_i Z_i^2}{(\lambda_{r+1}\cdots\lambda_n)^{1/(n-r)}\sum_{k=r+1}^{n} Z_k^2} > t^2\right) \\
&= P\left(\frac{r\sum_i \lambda_i\omega_i^2}{(n-r)(\lambda_{r+1}\cdots\lambda_n)^{1/(n-r)}} F_{r,n-r} > t^2\right) \\
&\sim \frac{R_{(n-r)/2}(\lambda_1,\dots,\lambda_r)\Gamma(n/2)}{t^{n-r}\sqrt{\lambda_{r+1}\cdots\lambda_n}\,\Gamma(r/2)\Gamma((n-r+2)/2)}
\end{aligned}$$

as $t \to \infty$. Thus, our F-approximation for the ratio of the quadratic forms is conservative by giving an upper bound on the probability of large values of this ratio.

2.2. Dirichlet Averages

We discuss now the properties of functions (4) whose role is important in the study of such tail probabilities for the ratio of two quadratic forms. The function $R_q(\lambda_1,\dots,\lambda_r)$ is a special case of the so-called Dirichlet averages,

$$R_q(b,z) = \int \left(\sum_1^r z_k u_k\right)^q d\mu_b(u),$$

where integration is over a unit simplex in R^r, $u_1,\dots,u_r \ge 0$, $\sum u_i = 1$, and μ_b is the Dirichlet distribution with parameter b [4]. Indeed, in our notation,

$$R_q(\lambda_1,\dots,\lambda_r) = R_q(b,(\lambda_1,\dots,\lambda_r))$$

with $b = (\frac{1}{2},\dots,\frac{1}{2})$.

If λ_i are eigenvalues of an $r \times r$ positive definite symmetric matrix Λ (notation: $\lambda_i = \lambda_i(\Lambda)$), then with $Z^T = (Z_1, \ldots, Z_r)$ for $q > -r/2$,

$$R_q(\lambda_1, \ldots, \lambda_r) = \frac{\Gamma(r/2)}{2^q \Gamma(q + r/2)} E(Z^T \Lambda Z)^q = \frac{E(Z^T \Lambda Z)^q}{E(Z^T Z)^q}. \tag{6}$$

Thus, the Dirichlet averages can be used to express moments of quadratic forms in normal variables. In (4) some of the λ's (say, s of them, $s = 0, 1, \ldots, r-1$) can vanish. For positive q,

$$R_q(0, \ldots, 0, \lambda_{s+1}, \ldots, \lambda_r) = \frac{\Gamma(q + (r-s)/2)\Gamma(r/2)}{\Gamma(q + r/2)\Gamma((r-s)/2)} R_q(\lambda_{s+1}, \ldots, \lambda_r). \tag{7}$$

This formula shows that (6) holds for positive semidefinite $r \times r$ matrices Λ.

Two basic equations,

$$\sum_k \frac{\partial}{\partial \lambda_k} R_q = q R_{q-1} \tag{8}$$

and

$$\sum_k \lambda_k \frac{\partial}{\partial \lambda_k} R_q = q R_q,$$

follow from the definition.

According to Euler's transformation (Theorem 6.8-3 in [4]),

$$R_{(n-r)/2}\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_r}\right) = \left(\prod_i \sqrt{\lambda_i}\right) R_{-n/2}(\lambda_1, \ldots, \lambda_r). \tag{9}$$

This fact allows to relate moments of quadratic forms in normal variables defined by a positive definite matrix and by its inverse [16].

For $0 < q < r/2$,

$$R_{-q}(\lambda_1, \ldots, \lambda_r) = \frac{1}{B(q, r/2 - q)} \int_0^1 \frac{u^{q-1}(1-u)^{r/2-q-1}\, du}{\prod_i \sqrt{1 - u + u\lambda_i}},$$

$$R_{-r/2}(\lambda_1, \ldots, \lambda_r) = \frac{1}{\prod_i \sqrt{\lambda_i}},$$

and the formulas for $R_{-n/2}(\lambda_1, \ldots, \lambda_r)$, $n > r$, can be obtained from these by repeated use of (8).

It follows from [4] that for $n \geq r/2$, $r \geq 1$, $\lambda_1, \ldots, \lambda_r > 0$,

$$R_{-n/2}(\lambda_1, \ldots, \lambda_r) \geq \left(\prod_i \lambda_i\right)^{-n/(2r)} \geq \left(\frac{\sum_i \lambda_i}{r}\right)^{-n/2} \tag{10}$$

with equality attained in either of these inequalities if and only if $\lambda_i \equiv \lambda$.

When $r = 2$, and n is a semi-integer, functions $R_n(\lambda_1, \lambda_2)$ are essentially complete elliptic integrals of the second and the first kind,

$$R_{-1/2}(\lambda_1, \lambda_2) = \frac{2}{\pi} \int_0^{\pi/2} \frac{d\phi}{\sqrt{\lambda_1 \cos^2\phi + \lambda_2 \sin^2\phi}} = \frac{2}{\pi\sqrt{\lambda_1}}\, \mathbf{K}\left(\sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_1}}\right),$$

$$R_{1/2}(\lambda_1, \lambda_2) = \frac{2}{\pi} \int_0^{\pi/2} \sqrt{\lambda_1 \cos^2\phi + \lambda_2 \sin^2\phi}\, d\phi = \frac{2\sqrt{\lambda_1}}{\pi}\, \mathbf{E}\left(\sqrt{\frac{\lambda_1 - \lambda_2}{\lambda_1}}\right),$$

when $\lambda_1 \geq \lambda_2$.

For arbitrary positive integer q, one can use *zonal* polynomials $C_\kappa(\Lambda)$ to get

$$R_q(\lambda_1,\dots,\lambda_r) = \int (e_r^T O\Lambda O^T e_r)^q \, dO = \int [\mathrm{tr}(e_r e_r^T O\Lambda O^T)]^q \, dO$$
$$= \sum_\kappa \int C_\kappa(e_r e_r^T O\Lambda O^T)\, dO = \sum_\kappa \frac{C_\kappa(e_r e_r^T)C_\kappa(\Lambda)}{C_\kappa(I)},$$

where κ denotes partitions of the number q into r parts, and integration is over the group of all $r\times r$ orthogonal matrices O. Here and further e_i will denote unit basis vectors whose dimension is clear from context. Above one can take $e_r = (0,\dots,0,1)^T$.

Since $C_\kappa(e_r e_r^T) = 0$, unless $\kappa = (q)$ is the partition $q = q + 0 + \cdots + 0$, in which case $C_\kappa(e_r e_r^T) = 1$,

$$R_q(\lambda_1,\dots,\lambda_r) = \frac{C_{(q)}(\Lambda)}{C_{(q)}(I)} = \frac{C_{(q)}(\Lambda)\Gamma(q+1/2)\Gamma(r/2)}{\sqrt{\pi}\,\Gamma(q+r/2)}.$$

Thus,

$$R_q(\lambda_1,\dots,\lambda_r) = \frac{\Gamma(q+1/2)\Gamma(r/2)}{\sqrt{\pi}\,\Gamma(q+r/2)} \sum_\kappa c_\kappa M_\kappa(\Lambda)$$

with positive coefficients c_κ (which can be found from recurrence formula (14), Section 7.4 in [12]). The monomial symmetric functions, $M_\kappa(\Lambda) = \sum \prod_{k=1}^s \lambda_{i_k}^{\nu_k}$, are defined by the partition κ with s nonzero parts $\nu_1,\dots,\nu_s$, and the summation is over the distinct permutations $i_1,\dots,i_s$ of s different integers out of $1,\dots,r$.

It follows that for a positive integer q,

$$R_q(\lambda_1,\dots,\lambda_r) \le \frac{\Gamma(q+1/2)\Gamma(r/2)}{\sqrt{\pi}\,\Gamma(q+r/2)} \left(\sum \lambda_i\right)^q, \tag{11}$$

which is true for any positive (possibly non-integer) q, $q \ge 1$. When $0 < q < 1$, $R_q(\lambda_1,\dots,\lambda_r) \le [(\sum \lambda_i)/r]^q$ (formula (2.14) in [3]).

Integration by parts shows that for all pairwise different λ_k,

$$\frac{1}{2}\left[\frac{\partial R_q}{\partial \lambda_k} - \frac{\partial R_q}{\partial \lambda_\ell}\right] = (\lambda_k - \lambda_\ell)\frac{\partial^2 R_q}{\partial \lambda_k \partial \lambda_\ell} = q(q-1)(\lambda_k - \lambda_\ell)\int_{S_r} \omega_k^2 \omega_\ell^2 [\sum \lambda_i \omega_i^2]^{q-2}\, d\omega,$$

so that when $q(q-1) \ge 0$,

$$(\lambda_k - \lambda_\ell)\left[\frac{\partial R_q}{\partial \lambda_k} - \frac{\partial R_q}{\partial \lambda_\ell}\right] \ge 0.$$

Then $R_q(\lambda_1,\dots,\lambda_r)$ is a Schur-convex function (Schur-concave when $q(q-1) < 0$). According to the Minkowski inequality for $q > 1$, $R_q^{1/q}(\lambda_1,\dots,\lambda_r)$ is a convex function.

3. CONSERVATIVE CONFIDENCE REGIONS: AN OPTIMIZATION PROBLEM

In this section we use the following result proven in [17]. Recall that $F_{r,n-r}$ denotes a F random variable with r and $n-r$ degrees of freedom.

Theorem 3.1. *Let $\delta = WY$ be a linear unbiased estimator of θ in* (1). *With $\widehat{\mathrm{Var}(\delta)}$ defined by* (2) *and*

$$\mu_i = \lambda_i\Big((WQ^{-1}W^T)^{-1/2}(X^T\Sigma^{-1}X)^{-1}(WQ^{-1}W^T)^{-1/2}\Big), \qquad i = 1,\dots,r,$$

one has

$$\lim_{t\to\infty} t^{n-r} P_\Sigma\Big((\delta-\theta)^T \widehat{\mathrm{Var}(\delta)}^{-1}(\delta-\theta) > t^2\Big)$$

$$= \frac{R_{(n-r)/2}(\mu_1,\dots,\mu_r)\Gamma(n/2)}{\det(X^T\Sigma^{-1}XWQ^{-1}W^T)^{1/2}\det(\Sigma Q)^{1/2}\Gamma(r/2)\Gamma((n-r+2)/2)}$$

$$= \lim_{t\to\infty} t^{n-r} P\left(\frac{rF_{r,n-r}}{n-r} > t^2\left[\frac{\det(X^T\Sigma^{-1}XWQ^{-1}W^T)\det(\Sigma Q)}{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}\right]^{1/(n-r)}\right). \qquad (12)$$

According to (12), our confidence ellipsoid (3) has the approximate confidence coefficient $1-\alpha$ for fixed Σ, when

$$t^2 = \frac{rF_{r,n-r}(\alpha)}{n-r}\left[\frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(X^T\Sigma^{-1}XWQ^{-1}W^T)\det(\Sigma Q)}\right]^{1/(n-r)}.$$

As Σ is unknown, a conservative procedure corresponds to

$$t_0^2 = \frac{rF_{r,n-r}(\alpha)}{(n-r)[\det(WQ^{-1}W^T)\det(Q)]^{1/(n-r)}}\sup_\Sigma\left[\frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(X^T\Sigma^{-1}X)\det(\Sigma)}\right]^{1/(n-r)}. \qquad (13)$$

Determination of t_0^2 is the central problem of this work. It is tantamount to solving the optimization problem in (13) with regard to all positive definite matrices Σ. However, there are situations where it is reasonable to restrict the error covariance matrices Σ to a smaller set.

The most important case is described in Section 1. For simplicity assume that for $i=1,\dots,m$,

$$Y_i = X_0\theta + \epsilon_i,$$

where X_0 is a $p\times r$ matrix of rank r, $r\le p$, and $\epsilon_i \sim N_p(0,\sigma_i^2 I)$. Then with $Y^T=(Y_1^T,\dots,Y_m^T)$, (1) holds for $X^T=(X_0^T,\dots,X_0^T)$, $n=pm$. It is natural to take Q having the same form as Σ, i.e., let $\hat{Q}=Q$ be formed by diagonal matrix blocks q_iI, $i=1,\dots,m$, where

$$q_i^{-1} = \hat\sigma_i^2 = \frac{1}{p-r}(Y_i - X_0\delta_i)^T(Y_i-X_0\delta_i)$$

and

$$\delta_i = (X_0^TX_0)^{-1}X_0^TY_i.$$

Then $X^T\Sigma^{-1}X = (\sum\sigma_i^{-2})X_0^TX_0$, and since $(WQ^{-1}W^T)^{-1} = X^TQX = (\sum q_i)X_0^TX_0$, all eigenvalues μ_i in (12) are equal, $\mu_i \equiv (\sum q_i)/(\sum\sigma_i^{-2})$. It follows that for block scalar matrices Σ,

$$\sup_\Sigma \frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(X^T\Sigma^{-1}X)\det(\Sigma)} = \max_{\sigma_1^2,\dots,\sigma_m^2}\frac{(\sum q_i)^{n-r}}{(\sum\sigma_i^{-2})^n(\prod\sigma_i^2)^p\det(X_0^TX_0)}$$
$$= \frac{(\sum q_i)^{n-r}}{m^n\det(X_0^TX_0)},$$

so that with $\omega_i = q_i/(\sum_j q_j)$, $i=1,\dots,m$,

$$t_0^2 = \frac{rF_{r,n-r}(\alpha)}{(n-r)m^{n/(n-r)}\prod\omega_i^{p/(n-r)}}. \qquad (14)$$

Notice that $\delta = \sum\omega_i\delta_i$, and

$$\widehat{\mathrm{Var}(\delta)} = \left[\sum_i \omega_i(Y_i-X_0\delta)^T(Y_i-X_0\delta)\right](X_0^TX_0)^{-1}.$$

Thus the ellipsoid (3) has the form

$$(\delta-\theta)^T(X_0^TX_0)(\delta-\theta) \le t_0^2\sum_i\omega_i(Y_i-X_0\delta)^T(Y_i-X_0\delta),$$

where t_0^2 is given by (14).

When $r = 1$,

$$R^2_{(n-r)/2}(\mu) = \mu^{n-1} = \left(\frac{1}{(WQ^{-1}W^T)(X^T\Sigma^{-1}X)}\right)^{n-1}.$$

For a given value of $\det(\Sigma) = \prod_j \sigma_j^2$, the minimum of $X^T\Sigma^{-1}X = \sum_j X_{j1}^2\sigma_j^{-2}$ is attained when $\sigma_j^2 \propto X_{j1}^2$, so that

$$\inf_{\Sigma}(X^T\Sigma^{-1}X)^n \det(\Sigma) = n^n \prod_1^n X_{j1}^2.$$

Thus, if now $\omega_i = q_i X_{i1}^2/(\sum q_j X_{j1}^2)$, then

$$t_0^2 = \frac{F_{1,n-1}(\alpha)}{(n-1)n^{n/(n-1)}\prod \omega_i^{n/(n-1)}},$$

which coincides with (14). Note that $\delta = \sum \omega_i Y_i/X_{i1}$, and

$$\widehat{\mathrm{Var}(\delta)} = \sum \omega_i (Y_i/X_{i1} - \delta)^2.$$

With T_{n-1} denoting a t-distributed random variable with $n-1$ degrees of freedom, we have [15]

$$\sup_{\Sigma} \lim_{t\to\infty} t^{n-1} P_\Sigma\Big((\delta-\theta)^T \widehat{\mathrm{Var}(\delta)}^{-1}(\delta-\theta) \ge t^2\Big) = \lim_{t\to\infty} t^{n-1} P\left(\frac{T_{n-1}^2}{n-1} > t^2\left[\prod_1^n (n\omega_i)\right]^{1/(n-1)}\right),$$

so that $\delta \pm t_{n-1}(\alpha/2)\prod(n\omega_i)^{-1/(2(n-1))}\sqrt{\widehat{\mathrm{Var}(\delta)}/(n-1)}$ is an approximate $(1-\alpha)$ confidence interval for β.

The same optimization problem (minimization of $\sum_j B_j^2\sigma_j^{-2}$ for the given value $\prod_j \sigma_j^2$) occurs for larger values of r, which are considered in the next section. Before doing that we reduce (13) to a canonical form. Notice that $WQ^{-1}W^T = X^TQX$, so that

$$\lambda_i\big((WQ^{-1}W^T)^{-1}(X^T\Sigma^{-1}X)^{-1}\big) = \lambda_i\big((X^TQX)(X^T\Sigma^{-1}X)^{-1}\big).$$

Let $A = Q^{1/2}X(X^TQX)^{-1/2}$, so that $A^TA = I$. By replacing the original design matrix X by A, one obtains

$$X^T\Sigma^{-1}X = (X^TQX)^{1/2}[A^T(Q^{1/2}\Sigma Q^{1/2})^{-1}A](X^TQX)^{1/2}$$

and

$$\det(X^T\Sigma^{-1}X) = \det(X^TQX)\det(A^T(Q\Sigma)^{-1}A).$$

Since Q is diagonal,

$$\begin{aligned}\sup_{\Sigma}&\left[\frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(X^T\Sigma^{-1}X)\det(\Sigma)}\right]^{1/(n-r)}\\ &= \sup_{\Sigma}\frac{R^2_{(n-r)/2}\big(\lambda_1((A^T\Sigma^{-1}A)^{-1}),\dots,\lambda_r((A^T\Sigma^{-1}A)^{-1})\big)}{\det(A^T\Sigma^{-1}A)\det(\Sigma)}\frac{\det(Q)}{\det(X^TQX)}.\end{aligned}$$

4. LEAST FAVORABLE VARIANCES AND LOWER BOUNDS

According to (13),

$$t_0^2 = \frac{rF_{r,n-r}(\alpha)}{n-r}G(A), \tag{15}$$

where

$$G(A) = \sup_{\Sigma} \left[\frac{R^2_{(n-r)/2}\big(\lambda_1((A^T\Sigma^{-1}A)^{-1}), \ldots, \lambda_r((A^T\Sigma^{-1}A)^{-1})\big)}{\det(A^T\Sigma^{-1}A)\det(\Sigma)} \right]^{1/(n-r)}.$$

In general, determination of $G(A)$, which can be interpreted as an adjustment factor to the percentile of an F-distribution, involves a solution of an optimization problem, where Σ is an arbitrary positive definite matrix.

An alternative representation for $G(A)$ is obtained from (9)

$$G(A) = \sup_{\Sigma} \left[\frac{R^2_{-n/2}(\lambda_1, \ldots, \lambda_r)}{\det(\Sigma)} \right]^{1/(n-r)}, \tag{16}$$

$\lambda_i = \lambda_i(A^T\Sigma^{-1}A)$, $i = 1, \ldots, r$.

By putting $\Sigma = I$, we get $G(A) \geq 1$, but a stronger inequality follows from (10),

$$[G(A)]^{n-r} \geq \sup_{\Sigma} \left[\det(\Sigma) \left(\frac{\operatorname{tr}(A^T\Sigma^{-1}A)}{r} \right)^n \right]^{-1}.$$

The minimum of the function, $\operatorname{tr}(A^T\Sigma^{-1}A) = \operatorname{tr}(AA^T\Sigma^{-1}) = \sum_{ij} a_{ji}^2 \sigma_j^{-2}$, for the given value $\det(\Sigma)$ is attained when $\sigma_j^2 \propto \sum_i a_{ji}^2$, with a_{ji}, $j = 1, \ldots, n$, $i = 1, \ldots, r$, denoting the elements of the matrix A. It follows that

$$[G(A)]^{n-r} \geq \left(\frac{r}{n} \right)^n \prod_j \frac{1}{\sum_i a_{ji}^2}. \tag{17}$$

This bound is attained when all λ's are equal, in particular, if $r = 1$, but in general, there is no explicit formula for the maximizing matrix in (16) or for $G(A)$. However, Theorem 4.1 proven in the Appendix shows that any stationary point $\Sigma = \operatorname{diag}(\sigma_1^2, \cdots, \sigma_n^2)$ has the form

$$\Sigma = \operatorname{diag}(AFA^T), \tag{18}$$

where F is a $r \times r$ matrix polynomial in $A^T\Sigma^{-1}A$ of degree $r - 1$ whose scalar coefficients depend only on the eigenvalues $\lambda_1, \ldots, \lambda_r$ of $A^T\Sigma^{-1}A$.

It follows that the maximizing vector, $\operatorname{diag}(\Sigma)$, is in the subspace spanned by the vectors with coordinates $a_{\cdot i}a_{\cdot k}, 1 \leq i \leq k \leq r$. This subspace $\mathbb{K}$ has the dimension at most $r(r+1)/2$, and its orthogonal complement is $\{x\colon A^T(\operatorname{diag}\ x)A = 0\}$. In practice, $r(r+1)/2 \ll n$, so that reduction to the solution space is substantial.

Also

$$\operatorname{tr}(\Sigma) = \sum_j \sigma_j^2 = \operatorname{tr}(AFA^T) = \operatorname{tr}(FA^TA) = \operatorname{tr}(F),$$

and the vector $(\sigma_1^2, \ldots, \sigma_n^2)$ is majorized by the vector of the eigenvalues of the matrix AFA^T, formed by $n - r$ zeros and r eigenvalues of the matrix F, which are shown in Lemma 2 to coincide with the gradient $-\nabla \log R^2_{-n/2}$.

For a matrix B denote by

$$B \begin{pmatrix} k_1 & \cdots & k_r \\ j_1 & \cdots & j_r \end{pmatrix}$$

its minor corresponding to the rows indexed by $k_1, \cdots, k_r$ and columns indexed by $j_1, \cdots, j_r$.

Here is the formulation of our main result.

Theorem 4.1. *The matrix* $\Sigma = \operatorname{diag}(\sigma_1^2, \dots, \sigma_n^2)$ *which maximizes* (16) *admits representation* (18). *The vector* $(\sigma_1^2, \dots, \sigma_n^2)^T$ *belongs to the positive orthant of the defined above subspace* $\mathbb{K}$ *of dimension at most* $r(r+1)/2$. *The lower bound* (17) *is valid, and the estimate* (39) *holds for any* $q = 1, \dots, r$. *For* $q = r-1$ *it has the form*

$$G(A) \geq \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^{2/(n-r)} \frac{1}{(n-r+1)^{(n-r+1)/(n-r)}}$$
$$\times \max_{1 \leq j_1 < \dots < j_{r-1} \leq n} \frac{AA^T \begin{pmatrix} j_1 & \cdots & j_{r-1} \\ j_1 & \cdots & j_{r-1} \end{pmatrix}}{\left[\prod_{j \neq j_1, \dots, j_{r-1}} AA^T \begin{pmatrix} j_1 & \cdots & j_{r-1} & j \\ j_1 & \cdots & j_{r-1} & j \end{pmatrix}\right]^{1/(n-r)}}. \tag{19}$$

The upper bound (40) *and* (19) *show that* $G(A)$ *is finite if and only if all* $r \times r$ *submatrices of* A *are non-singular.*

The proof of Theorem 4.1 shows that the optimization problem can be split in two parts. As $F = ODO^T$, the first consists in minimization of $\prod_j e_j^T AODO^T A^T e_j$ over all orthogonal matrices O for a fixed diagonal matrix $D = D(\lambda_1, \dots, \lambda_r) = -\operatorname{diag}(\partial/\partial\lambda_1, \dots, \partial/\partial\lambda_r) \log R_{-n/2}^2$. The second subproblem is that of maximization in λ_i of $R_{-n/2}^2(\lambda_1, \dots, \lambda_r)[\min_O \prod_j e_j^T AODO^T A^T e_j]^{-1}$. Thus,

$$[G(A)]^{n-r} = \sup_D \frac{R_{-n/2}^2(\lambda_1, \dots, \lambda_r)}{\min_O \prod_j e_j^T AODO^T A^T e_j}. \tag{20}$$

As an example, consider the minimization problem in Σ of the function $\det(\Sigma)\det(A^T\Sigma^{-1}A)^{n/r}$ (i.e., replace $R_{-n/2}$ in (16) by the first lower bound in (10)). According to Theorem 4.1, any stationary point has the form

$$\Sigma = \frac{n}{r} \operatorname{diag}(A(A^T\Sigma^{-1}A)^{-1}A^T)$$

as in this situation $F = nr^{-1}\det(A^T\Sigma^{-1}A)(A^T\Sigma^{-1}A)^{-1}$.

It follows that the vector of diagonal elements of D is

$$\left(\frac{\partial}{\partial\lambda_1}, \dots, \frac{\partial}{\partial\lambda_r}\right) \log\det(A^T\Sigma^{-1}A) = \left(\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_r}\right),$$

and

$$[G(A)]^{n-r} \geq \left(\frac{r}{n}\right)^n \sup_\Sigma \frac{[\det(A^T\Sigma^{-1}A)]^{-n/r}}{\det\left(\operatorname{diag}(A(A^T\Sigma^{-1}A)^{-1}A^T)\right)}$$
$$= \left(\frac{r}{n}\right)^n \sup_{\lambda_1, \dots, \lambda_r} \frac{(\lambda_1 \cdots \lambda_r)^{-n/r}}{\min_O \det\left(\operatorname{diag}(AODO^TA^T)\right)}.$$

It is of interest to investigate the situation when the least favorable variances are equal, in which case $G(A)$ can be found from (17). The next result deals with a slightly more general case when $A^T\Sigma^{-1}A$ is a scalar matrix. Denote

$$C = \frac{n}{r} \operatorname{diag}(AA^T)$$

and let $B \odot B$ be the Schur (Hadamard) square of B. The role of the Schur product in statistical applications, in particular, in variance estimation is discussed by Horn and Horn [9].

Theorem 4.2. *Assume that for a stationary point* $\mathrm{diag}(\Sigma)$ *of* (16), $A^T\Sigma^{-1}A = I$. *Then*

$$A^T C^{-1} A = I,$$

i.e.,

$$A^T[\mathrm{diag}(AA^T)]^{-1}A = \frac{n}{r}I, \tag{21}$$

and $\Sigma = C$. *Conversely, if* (21) *holds, the diagonal matrix* C *leads to a stationary point. This point corresponds to a maximum if and only if the matrix*

$$I + \frac{n-r}{n(r+2)}ee^T - \frac{n(n+2)}{r(r+2)}(BO \odot BO)(BO \odot BO)^T \tag{22}$$

is positive semidefinite for any $r \times r$ *orthogonal matrix* O *and* $B = C^{-1/2}A$.

5. TWO-DIMENSIONAL PARAMETER, $r = 2$

5.1. Examples

When $r = 2$, to evaluate $G(A)$ in (16) one needs $R_{-n/2}(\lambda_1, \lambda_2)$, which satisfies the equation

$$R_{-n/2}(\lambda_1, \lambda_2) = \frac{R_{(n-2)/2}(\lambda_1, \lambda_2)}{(\lambda_1\lambda_2)^{(n-1)/2}}. \tag{23}$$

It is more practical to calculate the ratio

$$H_m(\lambda_1, \lambda_2) = \frac{R_{m+1}(\lambda_1, \lambda_2)}{R_m(\lambda_1, \lambda_2)} = \frac{\lambda_1\lambda_2}{H_{-m-1}(\lambda_1, \lambda_2)}$$

by using the recurrent formula (identity (5.9-25) in [4])

$$H_m(\lambda_1, \lambda_2) = \frac{(2m+1)(\lambda_1 + \lambda_2)}{2(m+1)} - \frac{m\lambda_1\lambda_2}{(m+1)H_{m-1}(\lambda_1, \lambda_2)}.$$

If $n/2$ is an integer, the initial value is $H_0(\lambda_1, \lambda_2) = (\lambda_1 + \lambda_2)/2$; when $n/2$ is a semi-integer, one has to use $H_{-1/2}(\lambda_1, \lambda_2) = R_{1/2}(\lambda_1, \lambda_2)/R_{-1/2}(\lambda_1, \lambda_2)$ via the formulas for elliptic integrals given in Section 2.2. The formulas

$$R_{m+1}(\lambda_1, \lambda_2) = \prod_{0 \le k \le m} H_{m-k}(\lambda_1, \lambda_2), \qquad m \quad \text{integer},$$

$$R_{m+1}(\lambda_1, \lambda_2) = \prod_{0 \le k < m} H_{m-k}(\lambda_1, \lambda_2) R_{-1/2}(\lambda_1, \lambda_2), \qquad m \quad \text{semi-integer},$$

give a computationally efficient way to calculate R_m even for large m.

The function $H_{n/2-1} = H_{n/2-1}(\lambda_1, \lambda_2)$ provides the values of negative log-derivatives of the function $R^2_{-n/2}(\lambda_1, \lambda_2)$ denoted by d_1 and d_2. Indeed, because of (8), $\lambda_1 d_1 + \lambda_2 d_2 = n$ and $d_1 + d_2 = nR_{-n/2-1}/R_{-n/2} = n(\lambda_1\lambda_2)^{-1}H_{n/2-1}$. Lemma 2 shows that the diagonal matrix of eigenvalues of $A^T\Sigma^{-1}A$ is $D = \mathrm{diag}(d_1, d_2)$.

According to Theorem 4.1, instead of maximizing a function of n variables, it suffices to minimize for fixed λ_1, λ_2, $\rho = (d_1 - d_2)/(d_1 + d_2)$, $\tan\phi_j = a_{j2}/a_{j1}$, $|\phi_j| \le \pi/2$, $j = 1, \ldots, n$,

$$\begin{aligned}\prod_j e_j^T AODO^T A^T e_j &= \prod_j (a_{j1}^2 + a_{j2}^2)\big(d_1\cos^2(\phi_j + \phi) + d_2\sin^2(\phi_j + \phi)\big) \\ &= 2^{-n}(d_1 + d_2)^n \prod_j (a_{j1}^2 + a_{j2}^2) \prod_j \big(1 + \rho\cos 2(\phi_j + \phi)\big)\end{aligned}$$

over 2×2 orthogonal matrices $O = [\cos\phi, \sin\phi; -\sin\phi, \cos\phi]$, and then to optimize a function of two variables λ_1 and λ_2. The first task is an example of a classical problem of finding extrema of trigonometric polynomials (see Section VI, [14]). If $\hat\phi = \hat\phi(\lambda_1, \lambda_2) = \arg\min_\phi \prod_j (1 + \rho\cos 2(\phi_j + \phi))$, then the least favorable variances are $\sigma_j^2 \propto 1 + \rho \cos 2(\phi_j + \hat\phi)$.

Without a generality loss we can assume that $\lambda_1 = x \le 1 = \lambda_2$. Then from (20) and (23)

$$G(A) = \frac{2^{n/(n-2)}}{n^{n/(n-2)} \prod_j (a_{j1}^2 + a_{j2}^2)^{1/(n-2)}} \times \max_{x:\, 0\le x\le 1} \frac{x^{1/(n-2)}[R_{n/2-1}(x,1)]^{2/(n-2)}}{[H_{n/2-1}(x,1)]^{n/(n-2)}[\prod_j (1+\rho\cos 2(\phi_j+\hat\phi))]^{1/(n-2)}}. \tag{24}$$

When $x \to 0$, we have $\rho \to 1$ and $x/(1-\rho) \to (n-1)/2$. One can show that

$$\frac{\prod_j \left(1 + \rho\cos 2(\phi_j + \hat\phi)\right)}{1-\rho} \to 2^{n-1} \min_k \prod_{j:\, j\ne k} \sin^2(\phi_j - \phi_k),$$

which is positive provided that all angles ϕ_j are different. By (7),

$$R_{n/2-1}(x,1) \to \frac{\Gamma((n-1)/2)}{\sqrt{\pi}\Gamma(n/2)}$$

and

$$H_{n/2-1}(x,1) \to \frac{n-1}{n}.$$

Therefore, taking the limit when $x \to 0$ in (24), one gets

$$\begin{aligned} G(A) &\ge \frac{[\Gamma((n-1)/2)]^{2/(n-2)}}{(n-1)^{(n-1)/(n-2)}\pi^{1/(n-2)}[\Gamma(n/2)]^{2/(n-2)}} \times \frac{1}{[\prod_j (a_{j1}^2 + a_{j2}^2)]^{1/(n-2)} \min_k [\prod_{j:\, j\ne k} \sin^2(\phi_j - \phi_k)]^{1/(n-2)}} \\ &= \frac{[\Gamma((n-1)/2)]^{2/(n-2)}}{(n-1)^{(n-1)/(n-2)}\pi^{1/(n-2)}[\Gamma(n/2)]^{2/(n-2)}} \max_k \frac{(a_{k1}^2 + a_{k2}^2)}{\prod_{j\ne k} |a_{k1}a_{j2} - a_{k2}a_{j1}|^{2/(n-2)}}, \end{aligned} \tag{25}$$

which coincides with (19) and which is a sharp bound.

For example, when $n = 3$, $-\pi/2 < \alpha < \pi/2$,

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\alpha & \cos(\alpha+\pi/3) & \cos(\alpha-\pi/3) \\ \sin\alpha & \sin(\alpha+\pi/3) & \sin(\alpha-\pi/3) \end{bmatrix}^T.$$

Then $\phi_1 = \alpha$, $\phi_2 = \alpha + \pi/3$, $\phi_3 = \alpha - \pi/3$, and for a fixed ρ,

$$\min_\phi \prod_j \left(1 + \rho\cos 2(\phi_j + \phi)\right) = 1 - \frac{3\rho^2}{4} - \frac{\rho^3}{4}.$$

Since $\rho = 1 + x - 2x/H_{1/2}(x,1)$, the function $xR_{1/2}^2(x,1)H_{1/2}^{-3}(x,1)(1 - 3\rho^2/4 - \rho^3/4)^{-1}$ of x, $0 \le x \le 1$, is monotonically increasing and its maximum is attained when $x = 1$, i.e., when $\Sigma = I$. The same fact can be obtained from Theorem 4.2, since $\operatorname{diag}(AA^T) = 2/3e$, $C = I$, and the matrix (22) is positive definite.

If $n = 4$, the function of x in (24) can reach its maximum either at $x = 1$, or at $x = 0$, or in the middle of the unit interval. For example, if for $0 < \alpha < \pi/2$,

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 & 1 \\ \sin\alpha & \cos\alpha & 1 & 0 \end{bmatrix}^T,$$

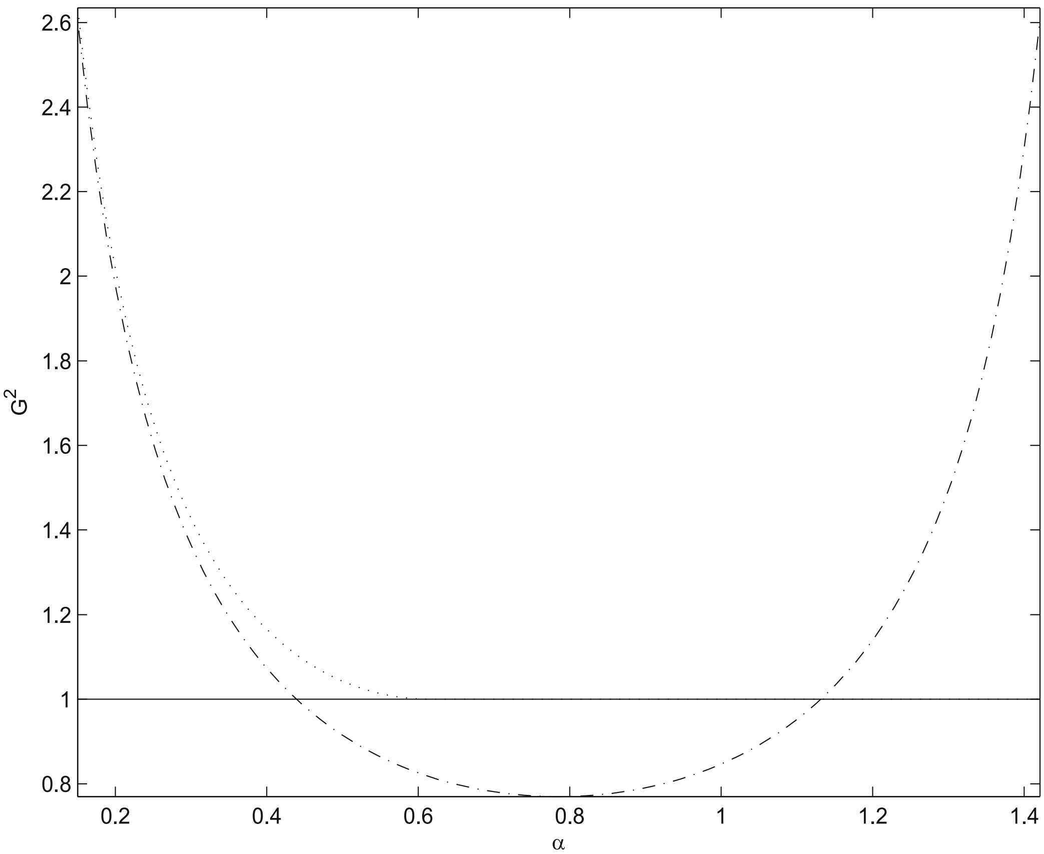

Fig. 1. The graphs of $G^2(A)$ as a function of α and of (25) (dash-dotted line) when $n = 4$, $r = 2$; $G(A)$ is given by the dotted line up to $\arcsin(1/\sqrt{3}) = 0.6155\ldots$, becomes equal to the solid line, which is the value of (17), up to the point $\pi/2 - \arcsin(4/(3\sqrt{3}))/2 = 1.1315\ldots$, after which it equals to the value given in (25).

so that $C = I$, Theorem 4.2 shows that the (global) maximum is attained at $x = 1$ when and only when $\arcsin(1/\sqrt{3}) \le \alpha \le \pi/2 - \arcsin(4/(3\sqrt{3}))/2$. Otherwise $x = 1$ is a saddle point. When $0 < \alpha < \arcsin(\sqrt{2/3})/2$, the global maximum is attained in the middle of the unit interval; if $\pi/2 - \arcsin(4/(3\sqrt{3}))/2 < \alpha < \pi/2$, the maximum is at the boundary $x = 0$. The lower bound (19) for $G(A)$ in this situation is $4/(3\sqrt{3}\sin 2\alpha)$ (see Fig. 1).

When $n \ge 5$, the maximum typically is attained at $x = 1$, so that (19) reduces to an equality. For example, if

$$a_{j1} = a_{j1}^0 = \sqrt{\frac{2}{n}} \cos \frac{\pi(j-1)}{n}, \qquad a_{j2} = a_{j2}^0 = \sqrt{\frac{2}{n}} \sin \frac{\pi(j-1)}{n}, \quad j = 1, \ldots, n,$$

so that for the matrix A_0 formed by these entries, $\mathrm{diag}(A_0 A_0^T) = 2n^{-1}e$, and $\Sigma = I$ is a stationary point. Theorem 4.2 can be used to show that the global maximum can be attained there only if $n \le 6$, and it is a saddle point for $n \ge 7$. Before demonstrating this fact and the extremal nature of the matrix A_0, we can conclude that for large n this method recommends prescribing the variance zero to the study $\jmath$ which is the most similar to other studies in the sense that it minimizes $\prod_{k:\, k \ne \jmath} |a_{k1}a_{\jmath 2} - a_{k2}a_{\jmath 1}|$.

To show that a_{j1}^0, a_{j2}^0 minimize the bound (25) over all design matrices A, notice that for a fixed k, $\prod_{j:\, j \ne k} \sin^2(\phi_j - \phi_k)$ is a Schur-concave function of $\phi_2 - \phi_1, \ldots, \phi_n - \phi_{n-1}, \phi_1 - \phi_n$. It follows that

$$\prod_j (a_{j1}^2 + a_{j2}^2) \min_k \prod_{j:\, j \ne k} \sin^2(\phi_j - \phi_k) \le \left(\frac{2}{n}\right)^n \prod_{j=1}^{n-1} \sin^2 \frac{\pi j}{n} = \left(\frac{1}{2n}\right)^{n-2}$$

(see formula (1.393) in [7]). Thus for any A

$$G(A) \geq \frac{2n[\Gamma((n-1)/2)]^{2/(n-2)}}{\pi^{1/(n-2)}[\Gamma(n/2)]^{2/(n-2)}(n-1)^{(n-1)/(n-2)}}.$$

This inequality confirms that for $n \geq 7$, $G(A) > 1$, and $\Sigma = I$ cannot be the point of maximum.

5.2. Application

As an example we consider the telephone switching data set presented in [10]. This data provides penetrations of new switching systems in five states taken over a period of twelve years. The square root of the penetration measure given by the vector $Y^T = (Y_1, \ldots, Y_5)^T$ seems to be fitted well by a linear function of time. See Fig. 2.

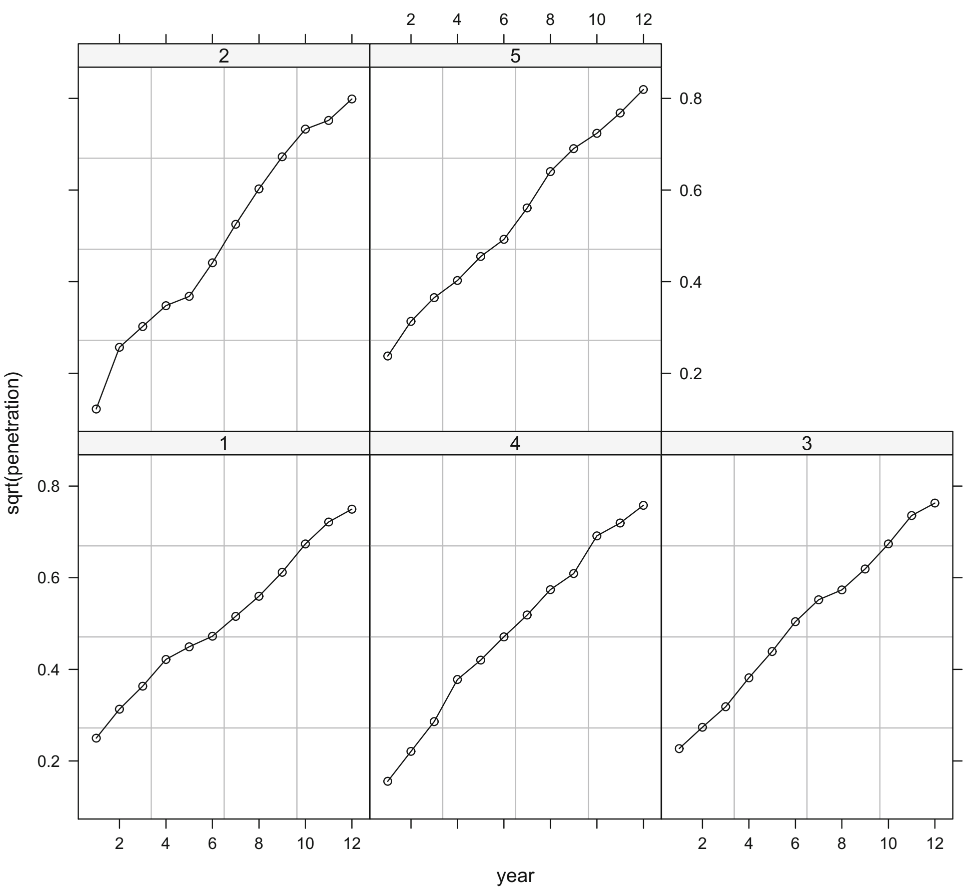

Fig. 2. The telephone switching data.

We used the replicated linear model with $r = 2$, $m = 5$, $p = 12$ and $n = 5 \times 12 = 60$ assuming as in derivation of (14) that the error variances for each state are equal. They can be estimated by the within state residual variances, so that one can take as q_i the reciprocals of these variances or rather the unbiased estimates of σ_i^{-2}. The first column of the matrix X_0 is formed by the vector of ones, the second is formed by time (in years), and in these units the linear estimator $\delta = (X_0^T X_0)^{-1} X_0^T \sum \omega_i Y_i$, equals to $(0.184, 0.050)^T$.

Calculations after (14) give $t_0^2 = 0.125$. Fig. 3 depicts the 95% conservative confidence ellipsoid for θ along with the traditional (smaller) confidence ellipsoid

$$(\delta - \theta)^T (X^T Q X)(\delta - \theta) \leq \chi_2^2(0.95).$$

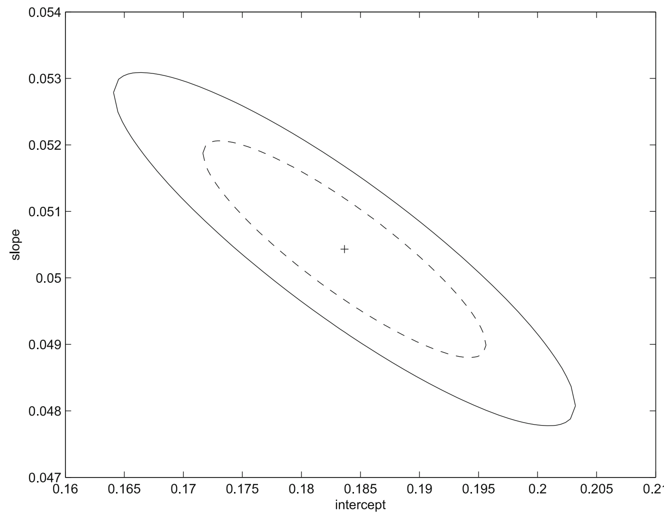

Fig. 3. Confidence ellipsoids for the parameters in the telephone switching study.

6. APPENDIX

6.1. Elementary Symmetric Polynomials: Lemma 1

Let for $1 \le k_i \le n$,

$$A_{k_1 \ldots k_r} = A \begin{pmatrix} k_1 & \cdots & k_r \\ 1 & \cdots & r \end{pmatrix}. \tag{26}$$

The Binet–Cauchy formula (Section 0.8.7, [8]) shows that

$$\det(A^T \Sigma^{-1} A) = \frac{1}{r!} \sum_{k_1,\ldots,k_r} A^2_{k_1\ldots k_r} \sigma_{k_1}^{-2} \cdots \sigma_{k_r}^{-2} = \sum_{1 \le k_1 < \ldots < k_r \le n} A^2_{k_1\ldots k_r} \sigma_{k_1}^{-2} \cdots \sigma_{k_r}^{-2},$$

which is a multilinear form in $\sigma_1^{-2}, \ldots, \sigma_n^{-2}$, as $A_{k_1\ldots k_r} = 0$, unless all indices $k_1, \ldots, k_r$ are different. The following lemma provides an extension of this fact to any elementary symmetric polynomial.

Lemma 1. *If $\mathbb{S}_\ell(\lambda_1, \ldots, \lambda_r), \ell = 0, 1, \ldots, r$, is the ℓth elementary symmetric polynomial in the eigenvalues of the matrix $A^T \Sigma^{-1} A$, then*

$$\mathbb{S}_\ell(\lambda_1, \ldots, \lambda_r) = \frac{1}{r!} \binom{r}{\ell} \sum_{k_1,\ldots,k_r} A^2_{k_1\ldots k_r} \sigma_{k_{r-\ell+1}}^{-2} \cdots \sigma_{k_r}^{-2}.$$

Proof. Since $A^T A = I$, the characteristic polynomial of the matrix $A^T \Sigma^{-1} A$ has the form

$$\begin{aligned}
\phi_{A^T \Sigma^{-1} A}(\lambda) &= \det(A^T \Sigma^{-1} A - \lambda I) = \det(A^T (\Sigma^{-1} - \lambda I) A) \\
&= \frac{1}{r!} \sum_{k_1,\ldots,k_r} A^2_{k_1\ldots k_r} (\sigma_{k_1}^{-2} - \lambda) \cdots (\sigma_{k_r}^{-2} - \lambda) \\
&= \frac{1}{r!} \sum_{k_1,\ldots,k_r} A^2_{k_1\ldots k_r} \sum_{\ell=0}^{r} (-1)^{r-\ell} \lambda^{r-\ell} \mathbb{S}_\ell(\sigma_{k_1}^{-2}, \ldots, \sigma_{k_r}^{-2}) \\
&= \frac{1}{r!} \sum_{\ell=0}^{r} (-1)^{r-\ell} \lambda^{r-\ell} \binom{r}{\ell} \sum_{k_1,\ldots,k_r} A^2_{k_1\ldots k_r} \sigma_{k_{r-\ell+1}}^{-2} \cdots \sigma_{k_r}^{-2}.
\end{aligned}$$

It follows that

$$\sum_{\ell=0}^{r}(-1)^{r-\ell}\lambda^{r-\ell}\mathbb{S}_\ell(\lambda_1,\dots,\lambda_r)=\prod_1^r(\lambda_i-\lambda)=\phi_{A^T\Sigma^{-1}A}(\lambda)$$
$$=\frac{1}{r!}\sum_{\ell=0}^{r}(-1)^{r-\ell}\lambda^{r-\ell}\binom{r}{\ell}\sum_{k_1,\dots,k_r}A^2_{k_1\dots k_r}\sigma^{-2}_{k_{r-\ell+1}}\cdots\sigma^{-2}_{k_r},$$

and this proves Lemma 1. □

By equating the trace of ℓth compound matrix of $A^T\Sigma^{-1}A$ [8], which is known to be an elementary symmetric function $\mathbb{S}_\ell$, to its expression in Lemma 1, we see that for any fixed indices $j_1,\dots,j_\ell$, $1\le\ell\le r$,

$$\frac{1}{(r-\ell)!}\sum_{k_{\ell+1},\dots,k_r}A^2_{j_1\cdots j_\ell k_{\ell+1}\dots k_r}=AA^T\begin{pmatrix}j_1&\cdots&j_\ell\\j_1&\cdots&j_\ell\end{pmatrix}.\tag{27}$$

In terms of the normalized elementary symmetric functions of $\lambda_1,\dots,\lambda_r$,

$$\mathbb{E}_\ell(\lambda_1,\dots,\lambda_r)=\left[\binom{r}{\ell}\right]^{-1}\mathbb{S}_\ell(\lambda_1,\dots,\lambda_r)=\left[\binom{r}{\ell}\right]^{-1}\sum_{1\le i_1<\dots<i_\ell\le r}\lambda_{i_1}\cdots\lambda_{i_\ell},$$

the identity of Lemma 1 means that for $\ell=0,1\dots,r$,

$$\mathbb{E}_\ell(\lambda_1,\dots,\lambda_r)=\frac{1}{r!}\sum_{k_1,\dots,k_r}A^2_{k_1\dots k_r}\sigma^{-2}_{k_{r-\ell+1}}\cdots\sigma^{-2}_{k_r}.\tag{28}$$

Because of Lemma 1, the gradient of $\det(A^T\Sigma^{-1}A)$,

$$\nabla\det(A^T\Sigma^{-1}A)=\left(\frac{\partial}{\partial\sigma_1^{-2}},\dots,\frac{\partial}{\partial\sigma_n^{-2}}\right)\det(A^T\Sigma^{-1}A),$$

as the function of $\sigma_\jmath^{-2}$, $\jmath=1,\dots,n$, has coordinates

$$[\nabla\det(A^T\Sigma^{-1}A)]_\jmath=[\nabla\mathbb{E}_r]_\jmath=\frac{r}{r!}\sum_{k_2,\dots,k_r}A^2_{\jmath k_2\dots k_r}\sigma^{-2}_{k_2}\cdots\sigma^{-2}_{k_r}.$$

More generally, by differentiating (28) with respect to $\sigma_\jmath^{-2}$, one obtains

$$\frac{\partial}{\partial\sigma_\jmath^{-2}}\mathbb{E}_\ell(\lambda_1,\dots,\lambda_r)=\frac{\ell}{r!}\sum_{k_2,\dots,k_r}A^2_{\jmath k_2\dots k_r}\sigma^{-2}_{k_{r-\ell+2}}\cdots\sigma^{-2}_{k_r},\tag{29}$$

so that for $\ell=1,\dots,r$,

$$\sum_\jmath\frac{\partial}{\partial\sigma_\jmath^{-2}}\mathbb{E}_\ell(\lambda_1,\dots,\lambda_r)=\ell\mathbb{E}_{\ell-1}(\lambda_1,\dots,\lambda_r).$$

6.2. Derivatives of Eigenvalues: Lemma 2

Let M be the $r\times r$ matrix with elements

$$M_{k\ell}=\frac{\partial}{\partial\lambda_k}\mathbb{E}_\ell(\lambda_1,\dots,\lambda_r)=\left[\binom{r}{\ell}\right]^{-1}\sum_{1\le k_1<\dots<k_{\ell-1}\le r,\,k_i\ne k}\lambda_{k_1}\cdots\lambda_{k_{\ell-1}},\tag{30}$$

and denote by E the $n\times r$ matrix whose $(\jmath,\ell)$th element is given by (29). Then with the matrix L formed by elements $\frac{\partial}{\partial\sigma_\jmath^{-2}}\lambda_i$, $\jmath=1,\dots,n$, $i=1,\dots,r$,

$$E=LM.\tag{31}$$

If

$$\Delta_{sj_2\dots j_r} = A\begin{pmatrix} j_2 & \cdots & j_s & j_{s+1} & \cdots & j_r \\ 1 & \cdots & s-1 & s+1 & \cdots & r \end{pmatrix}$$

is the cofactor of the element $a_{\jmath s}$ in the matrix A, then $A_{\jmath j_2\dots j_r} = \sum_s (-1)^{\jmath+s} a_{\jmath s}\Delta_{sj_2\dots j_r}$. Thus, one gets from (29) a useful representation for the elements of the matrix E:

$$\frac{\partial \mathbb{E}_\ell}{\partial \sigma_\jmath^{-2}} = \frac{\ell}{r!}\sum_{j_2,\dots,j_r,s,u}(-1)^{s+u} a_{\jmath s}a_{\jmath u}\Delta_{sj_2\dots j_r}\Delta_{uj_2\dots j_r}\sigma_{j_{r-\ell+2}}^{-2}\cdots\sigma_{j_r}^{-2} = e_\jmath^T A H_\ell A^T e_\jmath. \tag{32}$$

Here the $r\times r$ matrix H_ℓ, $\ell = 1,\dots,r$, has the elements

$$\frac{\ell(-1)^{s+u}}{r!}\sum_{j_2,\dots,j_r}\Delta_{sj_2\dots j_r}\Delta_{uj_2\dots j_r}\sigma_{j_{r-\ell+2}}^{-2}\cdots\sigma_{j_r}^{-2}, \qquad u,s = 1,\dots,r.$$

Clearly, H_r is the (classical) adjoint of the matrix $A^T\Sigma^{-1}A$. For $\ell = 1,\dots,r-1$,

$$(\ell+1)H_\ell = \sum_\jmath \frac{\partial H_{\ell+1}}{\partial\sigma_\jmath^{-2}}.$$

Induction on this recurrent formula proves that if $A^T\Sigma^{-1}A$ is a nonsingular matrix, then

$$H_\ell = \left[\binom{r}{\ell}\right]^{-1}\sum_{k=\ell}^{r}(-1)^{k-\ell}\binom{r}{k}\mathbb{E}_k(\lambda_1,\dots,\lambda_r)(A^T\Sigma^{-1}A)^{\ell-k-1}.$$

The Cayley–Hamilton formula according to which

$$\sum_{k=0}^{r}(-1)^{r-k}\binom{r}{k}\mathbb{E}_k(\lambda_1,\dots,\lambda_r)(A^T\Sigma^{-1}A)^{r-k} = 0,$$

shows that in the general case for $\ell = 1,\dots,r$,

$$H_\ell = \left[\binom{r}{\ell}\right]^{-1}\sum_{k=0}^{\ell-1}(-1)^k\binom{r}{\ell-k-1}\mathbb{E}_{\ell-k-1}(\lambda_1,\dots,\lambda_r)(A^T\Sigma^{-1}A)^k. \tag{33}$$

In particular, $H_1 = r^{-1}I$, $H_2 = 2(\mathrm{tr}(A^T\Sigma^{-1}A)I - A^T\Sigma^{-1}A)/[r(r-1)]$, and all matrices H_ℓ commute.

Now (31), (32), and (33) imply the first statement of the following lemma.

Lemma 2. *Assume that all $\lambda_i, i = 1,\dots,r$, are different, so that the matrix M in* (30) *is nonsingular. With $M_{i\ell}^-$ denoting the elements of the matrix M^{-1}, one has*

$$\frac{\partial\lambda_i}{\partial\sigma_\jmath^{-2}} = e_\jmath^T A\Big(\sum_\ell M_{i\ell}^- H_\ell\Big)A^T e_\jmath, \tag{34}$$

where the matrices H_ℓ, $\ell = 1,\dots,r$, satisfy (33). *If $A^T\Sigma^{-1}A = O\Lambda O^T$ with the diagonal matrix $\Lambda = \mathrm{diag}(\lambda_1,\dots,\lambda_r)$ and an orthogonal $r\times r$ matrix O, then*

$$\frac{\partial\lambda_i}{\partial\sigma_\jmath^{-2}} = (e_\jmath^T A O e_i)^2. \tag{35}$$

To prove the second statement of this lemma notice that (33) implies that

$$\binom{r}{\ell}\lambda_s(H_\ell) = \sum_{k=0}^{\ell-1}(-1)^k\lambda_s^k\mathbb{S}_{\ell-k-1}(\lambda_1,\dots,\lambda_r)$$

$$= \sum_{k=0}^{\ell-1} (-1)^k \lambda_s^k \Big[\lambda_s \mathbb{S}_{\ell-k-2}(\lambda_1, \dots, \lambda_{s-1}, \lambda_{s+1}, \dots, \lambda_r) \\ + \mathbb{S}_{\ell-k-1}(\lambda_1, \dots, \lambda_{s-1}, \lambda_{s+1}, \dots, \lambda_r) \Big] \\ = \mathbb{S}_{\ell-1}(\lambda_1, \dots, \lambda_{s-1}, \lambda_{s+1}, \dots, \lambda_r) = \binom{r}{\ell} M_{\ell s}.$$

Thus, assuming without generality loss that all λ's are different, we see that

$$\sum_\ell M_{i\ell}^- H_\ell = O\Big(\sum_\ell M_{i\ell}^- \operatorname{diag}(\lambda_1(H_\ell) \dots, \lambda_r(H_\ell)) \Big) O^T \\ = O\Big(\sum_\ell M_{i\ell}^- \operatorname{diag}(M_{\ell 1}, \dots, M_{\ell r}) \Big) O^T = O e_i M M^{-1} e_i^T O^T = O e_i e_i^T O^T,$$

which because of (34) proves (35).

6.3. Proof of Theorem 4.1

6.3.1. Proof of (18). In the maximization problem determining $G(A)$, a stationary point $(\sigma_1^2, \dots, \sigma_n^2)$ satisfies the equations for $j = 1, \dots, n$,

$$\frac{\partial}{\partial \sigma_j^{-2}} \Big[2 \log R_{-n/2}(\lambda_1, \dots, \lambda_r) - \log\det(\Sigma) \Big] = \sigma_j^2 - \sum_i d_i \frac{\partial \lambda_i}{\partial \sigma_j^{-2}} = 0,$$

where $d_i = -2 \frac{\partial \log R_{-n/2}}{\partial \lambda_i}$, $i = 1, \dots, r$. Therefore, with $d = (d_1, \dots, d_r)^T$ and the matrix L defined in Section 6.2, $\operatorname{diag}(\Sigma) = Ld$. Lemma 2 implies (18) for the matrix $F = \sum_{i,\ell} d_i M_{i\ell}^- H_\ell = ODO^T$, $D = \operatorname{diag}(d_1, \dots, d_r)$, so that the eigenvalues of F are equal to d.

Thus, if the maximum is attained when $\sigma_j^2 > 0$, the coordinates of any stationary point admit representation (18).

6.3.2. Maximum at the boundary. We show here that if the maximum is attained at the boundary, then the representation (18) still holds. This fact is more conveniently proven by the direct use of (12) rather than (16).

To do so it suffices to look at the case when $\sigma_{j_1} = \dots = \sigma_{j_q} = 0$ for fixed indices $j_1, \dots, j_q$, $1 \le j_1 < \dots < j_q \le n$. Indeed, if for some j, $\sigma_j \to \infty$, then with $\mu_i = 1/\lambda_i(A^T \Sigma^{-1} A)$,

$$\frac{R^2_{(n-r)/2}(\mu_1, \dots, \mu_r)}{\det(A^T \Sigma^{-1} A) \det(\Sigma)} \to 0.$$

Denote by A_1 a sub-matrix of A formed by the rows with these indices, and let A_2 be the complementary $(n-q) \times r$ sub-matrix (composed by the rows whose indices are different from $j_1, \dots, j_q$).

Then with a similar partitioning Σ_1, Σ_2 of Σ,

$$\det(A^T \Sigma^{-1} A) = \det(A_1^T \Sigma_1^{-1} A_1 + A_2^T \Sigma_2^{-1} A_2) \\ = \det(A_2^T \Sigma_2^{-1} A_2) \det(\Sigma_1^{-1}) \det\big(\Sigma_1 + A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T\big).$$

Thus, if $\Sigma_1 \to 0$,

$$\det(\Sigma_1) \det(A^T \Sigma^{-1} A) \sim \det(A_2^T \Sigma_2^{-1} A_2) \det\big(A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T\big).$$

Similarly,

$$(A^T \Sigma^{-1} A)^{-1} = (A_2^T \Sigma_2^{-1} A_2)^{-1} - (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T [\Sigma_1 + (A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T)^{-1}] A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1},$$

and for $\Sigma_1 \to 0$,

$$(A^T \Sigma^{-1} A)^{-1} \sim \Xi = (A_2^T \Sigma_2^{-1} A_2)^{-1} - (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T [A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1} A_1^T]^{-1} A_1 (A_2^T \Sigma_2^{-1} A_2)^{-1}.$$

Since the matrix

$$(A_2^T\Sigma_2^{-1}A_2)^{-1/2}A_1^T[A_1(A_2^T\Sigma_2^{-1}A_2)^{-1}A_1^T]^{-1}A_1(A_2^T\Sigma_2^{-1}A_2)^{-1/2}$$

is idempotent and has the rank equal to q, the rank of Ξ is $r-q$. It follows that $\mu_i \to 1/\xi_i = \lambda_i(\Xi)$, $i = 1,\dots,r-q$, $\mu_i \to 0$, $i = r-q+1,\dots,r$. Therefore the maximum can be attained at the boundary, $\sigma_{j_1} = \dots = \sigma_{j_q} = 0$, only if $q < r$, as otherwise the limit vanishes.

For notational convenience we take now indices $j_1,\dots,j_q$, $1 \le q < r$, to be $1,\dots,q$, and demonstrate first that the elementary symmetric polynomials of $\xi_1,\dots,\xi_{r-q}$ are multilinear forms in $\sigma_{q+1}^{-2},\dots,\sigma_n^{-2}$. To this end observe that because of Lemma 1 when $\ell = 0,1,\dots,r-q$,

$$\lim_{\Sigma_1\to 0}\det(\Sigma_1)\mathbb{S}_{r-\ell}(\lambda_1,\dots,\lambda_r) = \frac{q!}{r!}\binom{r}{\ell}\binom{r-\ell}{\ell}\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+\ell+1}}\cdots\sigma^{-2}_{k_r}$$

$$= \frac{1}{(r-q)!}\binom{r-q}{\ell}\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+\ell+1}}\cdots\sigma^{-2}_{k_r},$$

and the limit is zero if $\ell > r-q$. For $\ell = 0$, this formula gives

$$\det(A_2^T\Sigma_2^{-1}A_2)\det\left(A_1(A_2^T\Sigma_2^{-1}A_2)^{-1}A_1^T\right) = \frac{1}{(r-q)!}\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+1}}\cdots\sigma^{-2}_{k_r}. \tag{36}$$

It follows that for $\ell \le r-q$,

$$\lim_{\Sigma_1\to 0}\frac{\mathbb{S}_{r-\ell}(\lambda_1,\dots,\lambda_r)}{\mathbb{S}_r(\lambda_1,\dots,\lambda_r)} = \binom{r-q}{\ell}\frac{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+\ell+1}}\cdots\sigma^{-2}_{k_r}}{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+1}}\cdots\sigma^{-2}_{k_r}}.$$

Since

$$\frac{\mathbb{S}_{r-\ell}(\lambda_1,\dots,\lambda_r)}{\mathbb{S}_r(\lambda_1,\dots,\lambda_r)} = \mathbb{S}_\ell(\lambda_1^{-1},\dots,\lambda_r^{-1})$$

$$\to \mathbb{S}_\ell(\xi_1^{-1},\dots,\xi_{r-q}^{-1},0,\dots,0) = \mathbb{S}_\ell(\xi_1^{-1},\dots,\xi_{r-q}^{-1}) = \frac{\mathbb{S}_{r-q-\ell}(\xi_1,\dots,\xi_{r-q})}{\mathbb{S}_{r-q}(\xi_1,\dots,\xi_{r-q})},$$

one obtains

$$\frac{\mathbb{S}_{r-q-\ell}(\xi_1,\dots,\xi_{r-q})}{\mathbb{S}_{r-q}(\xi_1,\dots,\xi_{r-q})} = \binom{r-q}{\ell}\frac{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+\ell+1}}\cdots\sigma^{-2}_{k_r}}{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+1}}\cdots\sigma^{-2}_{k_r}}.$$

By putting $\ell = r-q$, it follows that

$$\xi_1\cdots\xi_{r-q} = \mathbb{S}_{r-q}(\xi_1,\dots,\xi_{r-q}) = \frac{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{q+1}}\cdots\sigma^{-2}_{k_r}}{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}}$$

$$= \frac{\det(A_2^T\Sigma_2^{-1}A_2)\det(A_1(A_2^T\Sigma_2^{-1}A_2)^{-1}A_1^T)}{\det(A_2^TA_2)\det(A_1(A_2^TA_2)^{-1}A_1^T)},$$

and, more generally, for $\ell = 0,\dots,r-q$,

$$\mathbb{E}_\ell(\xi_1,\dots,\xi_{r-q}) = \frac{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}\sigma^{-2}_{k_{r-\ell+1}}\cdots\sigma^{-2}_{k_r}}{\sum_{k_{q+1},\dots,k_r} A^2_{1\cdots qk_{q+1}\dots k_r}}, \tag{37}$$

i.e., all elementary symmetric polynomials in $\xi_i = 1/\lambda_i(\Xi)$, $i = 1,\dots,r-q$, are multilinear forms in σ_j^{-2}, $j = q+1,\dots,n$.

Returning to the proof of (18), we use (7) and derived above formulas, to see that when $\Sigma_1 \to 0$,

$$\frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(A^T\Sigma^{-1}A)\det(\Sigma)} \to \frac{R^2_{(n-r)/2}(\xi_1^{-1},\dots,\xi_{r-q}^{-1},0,\dots,0)}{\det(A_2^T\Sigma_2^{-1}A_2)\det(A_1[A_2^T\Sigma_2^{-1}A_2]^{-1}A_1^T)\det(\Sigma_2)}$$

$$= \left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^2 \left[\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)\right]^{-1} \frac{R^2_{(n-r)/2}(\xi_1^{-1},\dots,\xi_{r-q}^{-1})}{\xi_1\cdots\xi_{r-q}\det(\Sigma_2)}.$$

Thus, according to (9),

$$\sup_{\Sigma:\,\Sigma_1=0} \frac{R^2_{(n-r)/2}(\mu_1,\dots,\mu_r)}{\det(A^T\Sigma^{-1/2}A)\det(\Sigma)} = \max_q \left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^2$$
$$\times \left[\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)\right]^{-1} \sup_{\Sigma_2} \frac{R^2_{(q-n)/2}(\xi_1,\dots,\xi_{r-q})}{\det(\Sigma_2)}. \tag{38}$$

By repeating the argument leading to the formulas (32) and (18) now for the maximizing matrix Σ_2, we see that $\mathrm{diag}(\Sigma_2)$ belongs to the space $\mathbb{K}$.

6.3.3. Lower bound. The same argument as used to prove (17), applied now to (38) shows that

$$G(A) \ge \max_q \left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^{2/(n-r)} \left[\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)\right]^{-1/(n-r)}$$
$$\times \sup_{\Sigma_2} \left[\frac{1}{[(\xi_1+\ldots+\xi_{r-q})/(r-q)]^{n-q}\det(\Sigma_2)}\right]^{1/(n-r)}.$$

When indices $1,\dots,q$ are replaced by $j_1,\dots,j_q$, identity (36) with $\ell = 1$ gives

$$\frac{\xi_1+\cdots+\xi_{r-q}}{r-q} = \frac{\sum_{k_{q+1},\dots,k_r} A^2_{j_1\cdots j_q k_{q+1}\dots k_r}\sigma^{-2}_{k_r}}{\sum_{k_{q+1},\dots,k_r} A^2_{j_1\cdots j_q k_{q+1}\dots k_r}}$$
$$= \frac{\sum_{k_{q+1},\dots,k_r} A^2_{j_1\cdots j_q k_{q+1}\dots k_r}\sigma^{-2}_{k_r}}{(r-q)!\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)}.$$

As in (17), the minimum of the linear (in elements of Σ_2^{-1}) function, $\xi_1+\ldots+\xi_{r-q}$, for the fixed value of $\det(\Sigma_2)$ is attained when

$$\sigma_j^2 \propto \sum_{k_{q+1},\cdots,k_{r-1}} A^2_{j_1\cdots j_q k_{q+1}\cdots k_{r-1} j}, \qquad j \ne j_1,\dots,j_q.$$

Thus,

$$\min_{\Sigma_2}\left[\frac{\xi_1+\ldots+\xi_{r-q}}{r-q}\right]^{n-q}\det(\Sigma_2) = \left[\frac{n-q}{(r-q)!\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)}\right]^{n-q}$$
$$\times \prod_{j\ne j_1,\dots,j_q}\sum_{k_{q+1},\cdots,k_{r-1}} A^2_{j_1\cdots j_q k_{q+1}\cdots k_{r-1} j}.$$

A lower bound on $G(A)$ obtains

$$G(A) \ge \max_{q=1,\dots,r-1}\left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^{2/(n-r)}\left[\frac{(r-q)!}{n-q}\right]^{(n-q)/(n-r)}$$
$$\times \left[\det(A_2^T A_2)\det(A_1(A_2^T A_2)^{-1}A_1^T)\right]^{(n-q-1)/(n-r)}$$
$$\times \left[\prod_{j\ne j_1,\dots,j_q}\sum_{k_{q+1},\cdots,k_{r-1}} A^2_{j_1\cdots j_q k_{q+1}\cdots k_{r-1} j}\right]^{-1/(n-r)}$$
$$= \max_q \left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^{2/(n-r)} \frac{(r-q)!^{1/(n-r)}}{(n-q)^{(n-q)/(n-r)}}$$
$$\times \max_{j_1,\dots,j_q} \frac{\left[\sum_{k_{q+1},\dots,k_r} A^2_{j_1\cdots j_q k_{q+1}\dots k_r}\right]^{(n-q-1)/(n-r)}}{\left[\prod_{j\ne j_1,\dots,j_q}\sum_{k_{q+1},\dots,k_{r-1}} A^2_{j_1\cdots j_q k_{q+1}\dots k_{r-1} j}\right]^{1/(n-r)}}.$$

The formula (27) implies that

$$G(A) \geq \max_{q=1,\dots,r-1} \left[\frac{\Gamma(r/2)\Gamma((n-q)/2)}{\Gamma(n/2)\Gamma((r-q)/2)}\right]^{2/(n-r)} \left(\frac{r-q}{n-q}\right)^{(n-q)/(n-r)}$$
$$\times \max_{j_1,\dots,j_q} \frac{\left[AA^T\begin{pmatrix} j_1 & \cdots & j_q \\ j_1 & \cdots & j_q \end{pmatrix}\right]^{(n-q-1)/(n-r)}}{\left[\prod_{j\neq j_1,\dots,j_q} AA^T\begin{pmatrix} j_1 & \cdots & j_q & j \\ j_1 & \cdots & j_q & j \end{pmatrix}\right]^{1/(n-r)}}. \tag{39}$$

Note that (17) can be considered as a particular case of (39) when $q = 0$.

When $q = r-1$, (39) gives the exact value of the right-hand side of (38), so that the bound (19) holds. For large n, (19) commonly turns out to be an equality. It also follows from (39) that $G(A)$ can be finite only if all principal $r \times r$ subdeterminants of AA^T are strictly positive, which means that $A^2_{k_1\dots k_r} > 0$ for all $1 \leq k_1 < \cdots < k_r \leq n$.

6.3.4. Upper bound. As another application of (28) we show here that $G(A)$ is finite if

$$\underline{A} = \min_{k_1\neq\cdots\neq k_r} A^2_{k_1\dots k_r} > 0.$$

Indeed, by using (11) and Lemma 1 we see that with

$$\overline{A} = \max_{k_1,\dots,k_{r-1}} \sum_j A^2_{k_1\dots k_{r-1}j}$$

when $n - r \geq 2$,

$$[G(A)]^{n-r} \leq \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^2 \max_\Sigma \frac{(\mathbb{S}_{r-1}(\lambda_1,\dots,\lambda_r))^{n-r}}{\det(A^T\Sigma^{-1}A)^{n-r+1}\det(\Sigma)}$$
$$= \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^2 \max_\Sigma \frac{(\sum_{k_1,\dots,k_r} A^2_{k_1\dots k_r}\sigma^{-2}_{k_1}\cdots\sigma^{-2}_{k_{r-1}}/(r-1)!)^{n-r}}{(\sum_{k_1\dots k_r} A^2_{k_1\dots k_r}\sigma^{-2}_{k_1}\cdots\sigma^{-2}_{k_r}/r!)^{n-r+1}\sigma_1^2\cdots\sigma_n^2}$$
$$\leq \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^2 \frac{\overline{A}^{n-r}}{\underline{A}^{n-r+1}} \max_\Sigma \frac{[\mathbb{S}_{r-1}(\sigma_1^{-2},\dots,\sigma_n^{-2})]^{n-r}\mathbb{S}_n(\sigma_1^{-2},\dots,\sigma_n^{-2})}{[\mathbb{S}_r(\sigma_1^{-2},\dots,\sigma_n^{-2})]^{n-r+1}}$$
$$= \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^2 \frac{\left[\binom{n}{r-1}\overline{A}\,\right]^{n-r}}{\left[\binom{n}{r}\underline{A}\,\right]^{n-r+1}},$$

with the last identity following from the log-concavity in r of normalized elementary symmetric functions $\mathbb{E}_r(\sigma_1^{-2},\dots,\sigma_n^{-2})$ for fixed $\sigma_1^2,\dots,\sigma_n^2$ (Section 1.12, [2]).

Thus,

$$G(A) \leq \left[\frac{\Gamma(r/2)\Gamma((n-r+1)/2)}{\Gamma(n/2)\Gamma(1/2)}\right]^{2/(n-r)} \frac{r}{n-r+1}\left[\binom{n}{r}\right]^{-1/(n-r)} \frac{\overline{A}}{\underline{A}^{(n-r+1)/(n-r)}}. \tag{40}$$

When $n - r = 1$, $G(A) \leq \overline{A}/(2n\underline{A}^2)$.

6.4. Proof of Theorem 4.2

If $A^T\Sigma^{-1}A = I$, then $F = fI$ is a scalar matrix. One has

$$r = \operatorname{tr}(A^T\Sigma^{-1}A) = \operatorname{tr}(\Sigma^{-1}AA^T) = f^{-1}\operatorname{tr}([\operatorname{diag}(AA^T)]^{-1}AA^T) = nf^{-1},$$

so that $f = nr^{-1}$. Theorem 4.1 shows that $\Sigma = C = nr^{-1}\operatorname{diag}(AA^T)$ is a stationary point when and only when (21) holds.

Let $B = C^{-1/2}A$, so that $B^TB = I$ and $\operatorname{diag}(BB^T) = \operatorname{diag}(C^{-1/2}AA^TC^{-1/2}) = rn^{-1}e$. There is a simple relationship between $G(A)$ and $G(B)$,

$$[G(A)]^{n-r} = \max_{\Sigma} \frac{R^2_{-n/2}(\lambda_1(A^T\Sigma^{-1}A), \dots, \lambda_r(A^T\Sigma^{-1}A))}{\det(\Sigma)}$$
$$= \max_{\Sigma_1} \frac{R^2_{-n/2}(\lambda_1(B^T\Sigma_1^{-1}B), \dots, \lambda_r(B^T\Sigma_1^{-1}B))}{\det(C)\det(\Sigma_1)} = \det(C)^{-1}[G(B)]^{n-r},$$

and the maximizer for $G(A)$ is C if and only if the maximizer for $G(B)$ is I. Thus, we can assume that the stationary point is $\Sigma = I$. According to Lemma 2, $L = AO \odot AO$, and $Le = \nabla \operatorname{tr}(A^T\Sigma^{-1}A) = rn^{-1}e$.

The formulas in Section 2.2 show that the gradient of the function $R = -\log R^2_{-n/2}(\lambda_1, \cdots, \lambda_r)$ evaluated at $\lambda_1 = \cdots = \lambda_r = 1$ is $\nabla R = d = nr^{-1}e$, and its Hessian $\nabla^2 R$ is

$$\nabla^2 R = -\frac{n(n+2)}{r(r+2)}I + \frac{n(n-r)}{r^2(r+2)}ee^T.$$

Thus the gradient of the function of $\sigma_1^{-2}, \dots, \sigma_n^{-2}$ in (16) is $\operatorname{diag}(\Sigma) + Ld$, and its Hessian H is

$$H = -\Sigma^2 - L\nabla^2 RL^T - \sum_i d_i \nabla^2 \lambda_i,$$

where $\nabla^2\lambda_i$ is the Hessian of λ_i. One has

$$\sum_i d_i\nabla^2\lambda_i = \frac{n}{r}\nabla^2 \sum_i \lambda_i = \frac{n}{r}\nabla^2 \operatorname{tr}(A^T\Sigma^{-1}A) = 0,$$

so that at $\Sigma = I$,

$$H = -I + L\left[\frac{n(n+2)}{r(r+2)}I - \frac{n(n-r)}{r^2(r+2)}ee^T\right]L^T = -I - \frac{n-r}{n(r+2)}ee^T + \frac{n(n+2)}{r(r+2)}LL^T$$
$$= -I - \frac{n-r}{n(r+2)}ee^T + \frac{n(n+2)}{r(r+2)}(AO \odot AO)(AO \odot AO)^T.$$

For $\Sigma = I$ to be the point of maximum, the matrix above must be negative semidefinite for any orthogonal O, and this happens if and only if (22) holds. □

ACKNOWLEDGMENTS

The author is grateful to Bill Carlson for useful references, and to D. M. Chibisov and A. M. Zubkov for helpful comments.

REFERENCES

1. L. C. Adkins and B. C. Hill, “An Improved Confidence Ellipsoid for the Linear Regression Model”, J. Statist. Comput. Simul. **36**, 9–18 (1990).
2. E. F. Beckenbach and R. Bellman, *Inequalities* (Springer, Berlin, 1961).
3. B. C. Carlson, “Some Inequalities for Hypergeometric Functions”, Proc. Amer. Math. Soc. **17**, 32–39 (1966).
4. B. C. Carlson, *Special Functions of Applied Mathematics* (Academic Press, New York, 1977).
5. R. J. Carroll and D. Ruppert, *Transformation and Weighting in Regression* (Chapman & Hall, New York, 1988).
6. W. A. Fuller and J. N. K. Rao, “Estimation for a Linear Regression with Unknown Diagonal Matrix”, Ann. Statist. **6**, 1149–1158 (1978).
7. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. (Academic Press, San Diego, 2000).
8. R. A. Horn and C. A. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
9. S. A. Horn and R. A. Horn, “Comparison of Estimators of Heteroscedastic Variances in Linear Models”, J. Amer. Statist. Assoc. **70**, 872–879 (1975).

10. E. M. Keramidas and J. C. Lee, “Forecasting Technological Substitutions with Concurrent Short Time Series”, J. Amer. Statist. Assoc. **85**, 625–632 (1990).
11. A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications* (Academic Press, New York, 1979).
12. R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
13. S. D. Peddada and G. Patwardhan, “Jackknife Variance Estimators in Linear Models”, Biometrika **79**, 654–657 (1992).
14. G. Polya and G. Szegö, *Problems and Theorems in Analysis* (Springer, Berlin, 1998).
15. A. L. Rukhin, “Conservative Confidence Intervals Based on Weighted Means Statistics”, Statist. & Probab. Lett. **77**, 853–861 (2007).
16. A. L. Rukhin, “Identities for Moments of Quadratic Forms in Normal Variables”, Statist. & Probab. Lett. **79**, 1004–1007 (2009).
17. A. L. Rukhin, “Confidence Regions for Parameters of Linear Models”, Statist. Sinica **20**, (2010).
18. O. V. Shalaevsky, “Fitting Observations with Unknown Weights by the Method of Least Squares”, in *Proc. Steklov Inst. of Math.* (1968), Vol. 104, pp. 189–214 [in Russian].
19. J. Shao and C. J. F. Wu, “Heteroscedasticity-Robustness of Jackknife Variance Estimators in Linear Models”, Ann. Statist. **15**, 1563–1579 (1987).
20. C. J. F. Wu “Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with Discussion)”, Ann. Statist. **14**, 1261–1350 (1986).