Statistica Sinica **20** (2010), 000-000

CONFIDENCE REGIONS FOR PARAMETERS OF LINEAR MODELS

Andrew L. Rukhin

National Institute of Standards and Technology
and University of Maryland at Baltimore County

Abstract: A method is suggested for constructing a conservative confidence region for the parameters of a linear model on the basis of a linear estimator. In meta-analytical applications, when the results of independent but heterogeneous studies are to be combined, this region can be employed with little to no knowledge of error variances. The formulas for the smallest volume and the corresponding critical constant are derived. The method is compared to several resampling schemes by Monte Carlo simulation, and particular cases of one or two parameters are examined.

Key words and phrases: Dirichlet averages, general linear model, jackknife variance estimators, meta-analysis, quadratic forms in normal vectors, weighted bootstrap.

1. Introduction

The estimation of parameters in a linear model is one of the important problems of statistics. In its classical setting, with a given matrix of error covariances, the solution is well known. However if the error covariance matrix cannot be assumed known or proportional to a given matrix, the problem becomes much more challenging.

Consider the general linear model, $Y \sim N_p(X\beta, \Sigma)$ with a diagonal matrix Σ. Thus if the parameter β is r-dimensional, $r \le p$, and X is a $p \times r$ design matrix of rank r,

$$Y = X\beta + \epsilon. \tag{1.1}$$

The vector ϵ is formed by independent errors ϵ_j, $j = 1, \ldots, p$, with zero mean and unknown variances σ_j^2. In the meta-analysis context, there are p independent but heterogeneous studies, with each study producing an unbiased estimate of its linear function of β. The accuracy of this estimator may not be given. The classical least squares estimate $\hat{\beta}$, determined from the equation, $X^T\Sigma^{-1}X\hat{\beta} = X^T\Sigma^{-1}Y$, clearly depends on Σ which typically is (at least partially) unknown.

Carroll and Ruppert (1982) investigate different parametric models for heteroscedasticity. Fuller and Rao (1978) study a two-stage estimation procedure assuming that observations compose several groups with constant variance within each group. The residuals derived from ordinary least squares are employed to estimate the covariance matrix, which is then used in the generalized least squares. Wu (1986) devised a class of jackknife variance estimators and compared them to the bootstrap method. The asymptotic behavior of these methods is elucidated in Shao and Tu (1995, Chap. 7) and Basu and Chatterjee (2002). While these authors are mainly interested in distributional properties, our approach is motivated by the desire to construct confidence regions for the parameter β.

We look at linear unbiased estimators of the form $\delta = WY$, with an $r \times p$ matrix W. Then, with I denoting the identity matrix (whose size is usually clear from the context), $WX = I$, and $\mathrm{Var}\,(\delta) = W\Sigma W^T$ is the covariance matrix of δ. A natural form of the estimator is $\delta = (X^TQX)^{-1}X^TQY$ with a $p \times p$ diagonal matrix Q, where Q is thought of as an approximation to Σ^{-1}. Thus, we take $W = (X^TQX)^{-1}X^TQ$. If $Q = \Sigma^{-1}$, then $\mathrm{Var}\,(\delta) = (X^TQX)^{-1}$. However, the commonly used estimator of $\mathrm{Var}\,(\delta)$, $(X^TQX)^{-1}$, typically underestimates this matrix.

To adjust for this bias and to derive a conservative confidence ellipsoid, a new estimator $\widehat{\mathrm{Var}\,(\delta)}$ of $\mathrm{Var}\,(\delta)$ is suggested here. It is a scalar multiple of $(X^TQX)^{-1}$,

$$\widehat{\mathrm{Var}\,(\delta)} = [Y^T(I - XW)^TS(I - XW)Y](X^TQX)^{-1}. \tag{1.2}$$

The non-negative definite $p \times p$ matrix S defining the quadratic form in the residuals $Y - XWY$ in (1.2) allows many choices, like $S = Q$, or S corresponding to the jackknife variance estimator discussed in Section 6. When $r = 1$, one can take S leading to the statistic suggested by Horn, Horn, and Duncan (1975).

A confidence ellipsoid for β based on δ is given by

$$(\delta - \beta)^T\widehat{\mathrm{Var}\,(\delta)}^{-1}(\delta - \beta) \le t^2. \tag{1.3}$$

Our goal is to determine the coverage probability of this ellipsoid for any Σ at least for large t.

In the common mean case, when $r = 1$ and X is the p-dimensional vector of ones, this problem was considered by Rukhin (2007). The maximal non-coverage probability as a function of t equals 1 for sufficiently small t, but for large t it coincides with the tail probability of the t_{p-1}-distribution. Then the maximum is attained when all σ's are equal, and one can determine the adjustment factor G (which depends on t, Q and S), $G \ge 1$, such that

$$\sup_{\Sigma} P\Big((\delta - \beta)^T\widehat{\mathrm{Var}\,(\delta)}^{-1}(\delta - \beta) > t^2G\Big) = P\Big(|T_{p-1}| > t\Big),$$

where T_{p-1} denotes a t-distributed random variable with $p-1$ degrees of freedom.

In Section 2 the asymptotic behavior of the coverage probability for any r is shown to be that of the probability when t^2 is a multiple of a percentile of the F-distribution with r and $p-r$ degrees of freedom. Section 3 contains needed results about the moments of indefinite quadratic forms in Gaussian random variables, relating them to Dirichlet averages (Carlson (1977)). Section 5 discusses the volume of the confidence ellipsoid and suggests an optimal choice of S and the corresponding value of t^2 in (1.3). Sections 4 and 6 give some examples and simulation results. All proofs are collected in the Appendix.

Motivation for the problem comes from interlaboratory studies where the data is influenced by systematic, laboratory-specific errors (the so-called type B uncertainties), and where error variances cannot be reliably estimated. Indeed, the possibility of "unrealistic" uncertainties which do not take into account all possible sources of errors, is widely recognized (as well as the fact that in such a situation the classical least squares procedures cannot be employed.) Paule and Mandel (1970) describe a study in which several laboratories performed measurements via different techniques of gold vapor pressure as a function of the absolute temperature T in the (individual for each laboratory) range from $1,300$ to $2,100$K. According to the heat law, the logarithm of pressure $\log P$ is a linear function of $1/T$. The data on P was collected from 38 runs of ten laboratories with the number of observations at each run varying from 5 to 31. After removal of obvious outliers and the results of one dubious laboratory, there are a total of 375 different temperature points. A natural assumption is that the error variance depends only on the run within each individual laboratory (and not on the temperature value). This study then fits the model (1.1) with $p = 375$, under the additional condition that there are only 38 different values of σ_j^2 corresponding to each run. The matrix X is formed by the reciprocals of T employed by each laboratory; the diagonal matrix Q is composed of reciprocals of the sample variances obtained for each laboratory.

2. Conservative Confidence Regions

Let S_r be the unit sphere in r-dimensional space parametrized by $\omega_1, \ldots, \omega_r$ with $\sum \omega_i^2 = 1$. Define for $\lambda_i > 0$, $i = 1, \ldots, r$, and real q,

$$H_q(\lambda_1, \ldots, \lambda_r) = \int_{S_r} \left[\sum_i \lambda_i \omega_i^2 \right]^q d\omega, \tag{2.1}$$

where $d\omega$ denotes the normalized uniform distribution over S_r. Denote by $F_{r,p-r}$ a F random variable with r and $p-r$ degrees of freedom, and by $F_{r,p-r}(\alpha)$ the critical point of its distribution, let $\lambda_i = \lambda_i(\Lambda)$ denote the eigenvalues of a matrix Λ.

Theorem 2.1. *Let $\delta = WY$ be a linear unbiased estimator of β in* (1.1). *With $\widehat{\mathrm{Var}\,(\delta)}$ defined by* (1.2) *and*

$$\mu_i = \lambda_i\Big((X^TQX)^{1/2}(X^T\Sigma^{-1}X)^{-1}(X^TQX)^{1/2}\Big), \quad i = 1, \ldots, r,$$

one has

$$\begin{aligned}
&\lim_{t\to\infty} t^{p-r} P_\Sigma\Big((\delta-\beta)^T \widehat{\mathrm{Var}\,(\delta)}^{-1}(\delta-\beta) > t^2\Big)\\
&\quad= \frac{H_{(p-r)/2}(\mu_1, \ldots, \mu_r)\Gamma(p/2)}{\det(X^T\Sigma^{-1}XWS^{-1}W^T)^{1/2}\det(\Sigma S)^{1/2}\Gamma(r/2)\Gamma((p-r+2)/2)}\\
&\quad= \lim_{t\to\infty} t^{p-r} P\left(\frac{rF_{r,p-r}}{p-r} > t^2\left[\frac{\det(X^T\Sigma^{-1}XWS^{-1}W^T)\det(\Sigma S)}{H^2_{(p-r)/2}(\mu_1, \ldots, \mu_r)}\right]^{1/(p-r)}\right). \quad (2.2)
\end{aligned}$$

According to (2.2), the ellipsoid (1.3) has approximate confidence $1-\alpha$ for fixed Σ, if

$$t^2 = t^2(\Sigma) = \frac{rF_{r,p-r}(\alpha)}{p-r}\left[\frac{H^2_{(p-r)/2}(\mu_1, \ldots, \mu_r)}{\det(X^T\Sigma^{-1}XWS^{-1}W^T)\det(\Sigma S)}\right]^{1/(p-r)}.$$

As Σ is unknown, a conservative procedure corresponds to

$$t_0^2 = \frac{rF_{r,p-r}(\alpha)G}{(p-r)}\left[\frac{\det(WQ^{-1}W^T)\det(Q)}{\det(WS^{-1}W^T)\det(S)}\right]^{1/(p-r)}, \quad (2.3)$$

where the adjustment factor,

$$\begin{aligned}
G &= \sup_\Sigma \left[\frac{H^2_{(p-r)/2}(\mu_1, \ldots, \mu_r)}{\det(WQ^{-1}W^T)\det(Q)\det(X^T\Sigma^{-1}X)\det(\Sigma)}\right]^{1/(p-r)}\\
&= \sup_\Sigma \left[\frac{H^2_{(p-r)/2}(\lambda_1((A^T\Sigma^{-1}A)^{-1}), \ldots, \lambda_r((A^T\Sigma^{-1}A)^{-1}))}{\det(A^T\Sigma^{-1}A)\det(\Sigma)}\right]^{1/(p-r)} \quad (2.4)
\end{aligned}$$

does not depend on S, but only on X and Q through $A = Q^{1/2}X(X^TQX)^{-1/2}$, so that $A^TA = I$.

To see that, replace Σ by $\tilde{\Sigma} = Q\Sigma$ in (2.4). Since $\mu_i = \lambda_i((A^T\tilde{\Sigma}^{-1}A)^{-1})$ and $\det(A^T\tilde{\Sigma}^{-1}A) = \det(X^T\Sigma^{-1}X)/\det(X^TQX) = \det(X^T\Sigma^{-1}X)\times\det(WQ^{-1}W^T)$, it follows that $t_0^2 = \max_\Sigma t^2(\Sigma)$. Of course $G \geq 1$, which is seen from (2.4) when $\Sigma = I$.

If $r = 1$,

$$H_{(p-r)/2}(\mu) = \mu^{(p-1)/2} = \left(\frac{X^TQX}{X^T\Sigma^{-1}X}\right)^{(p-1)/2}.$$

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

$$\sup_{\Sigma} \frac{X^TQX}{(X^T\Sigma^{-1}X)^{p/(p-1)}\det(\Sigma)^{1/(p-1)}} = \frac{X^TQX}{p^{p/(p-1)}(\prod_1^p X_{j1}^2)^{1/(p-1)}}$$

and, according to (2.4), $G^{p-1} = (X^TQX)^{p-1}/[p^p \det(WQ^{-1}W^T)\det(Q)\prod_1^p X_{j1}^2]$. Thus for $r = 1$,

$$t_0^2 = \frac{F_{1,p-1}(\alpha)(X^TQX)}{(p-1)p^{p/(p-1)}[WS^{-1}W^T\det(S)]^{1/(p-1)}(\prod_1^p X_{j1}^2)^{1/(p-1)}}.$$

3. Moments of Quadratic Forms in Normal Variables: Asymptotics and Special Functions

This section contains the needed results for the distribution of quadratic forms in normal variables.

Theorem 3.1. *Let $Z_1, \ldots, Z_p$ be independent standard normal variables, and $\lambda_{r+1}, \ldots, \lambda_p$ be fixed positive numbers. As $\lambda_1, \ldots, \lambda_r \to 0$,*

$$\lim \frac{P(\sum_{i=1}^r \lambda_i Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2)}{H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)} = \frac{\Gamma(p/2)}{\sqrt{\lambda_{r+1}\cdots\lambda_p}\,\Gamma(r/2)\Gamma((p-r+2)/2)}.$$

When $\lambda_i = t^{-2}$, $i = 1, \ldots, r$, $\lambda_k = 1$, $k = r+1, \ldots, p$, $H_{(p-r)/2} = t^{r-p}$, and the formula in Theorem 3.1 agrees with the well-known result for the tail probabilities of $F_{r,p-r}$, a F random variable with r and $p-r$ degrees of freedom. Indeed as $t \to \infty$,

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

The proof of the Theorem 3.1 shows that

$$\begin{aligned} P\left(\frac{\sum_{i=1}^r \lambda_i Z_i^2}{\sum_{k=r+1}^p \lambda_k Z_k^2} > t^2\right) &\le P\left(\frac{[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)}\sum_{i=1}^r Z_i^2}{(\lambda_{r+1}\cdots\lambda_p)^{1/(p-r)}\sum_{k=r+1}^p Z_k^2} > t^2\right) \\ &= P\left(\frac{r[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)}}{(p-r)(\lambda_{r+1}\cdots\lambda_p)^{1/(p-r)}}F_{r,p-r} > t^2\right). \end{aligned}$$

In other words, when approximating the tail probability for the ratio of two quadratic forms, the λ_i's in the numerator are to be replaced by their *spherical* average $[H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)]^{2/(p-r)}$, and the λ_k's in the denominator by their geometric mean.

The function H_q is a particular case of the so-called Dirichlet averages (Carlson (1977, Chap. 5)), i.e.,

$$H_q(\lambda_1, \dots, \lambda_r) = \int \left[\sum_1^r \lambda_k u_k \right]^q d\mu_b(u),$$

where integration is over a unit simplex in R^r and μ_b is the Dirichlet distribution with the parameter $b = (1/2, \dots, 1/2)$. If a $r \times r$ positive definite symmetric matrix Λ has eigenvalues λ_i, then with $Z \sim N_r(0, I)$ and $q > -r/2$,

$$H_q(\lambda_1, \dots, \lambda_r) = \frac{\Gamma(r/2)}{2^q \Gamma(q + r/2)} E(Z^T \Lambda Z)^q. \tag{3.1}$$

Meng (2005) gives a number of useful formulas for such moments and discusses their statistical applications.

4. Example: $r = 2$

To evaluate G when $r = 2$, one needs $H_{(p-2)/2}(\lambda_1^{-1}, \lambda_2^{-1})$ with $\lambda_i = \lambda_i(A^T \Sigma^{-1} A)$, $i = 1, 2$. By symmetry and homogeneity of this function, one gets

$$H_{(p-2)/2}(\lambda_1^{-1}, \lambda_2^{-1}) = \frac{H_{(p-2)/2}(\lambda_1, \lambda_2)}{(\lambda_1 \lambda_2)^{(p-2)/2}}.$$

It is more practical to calculate the ratio $R_m(\lambda_1, \lambda_2) = H_{m+1}(\lambda_1, \lambda_2)/H_m(\lambda_1, \lambda_2)$, as it satisfies the recurrence formula

$$R_{m+1}(\lambda_1, \lambda_2) = \frac{(2m+1)(\lambda_1 + \lambda_2)}{2(m+1)} - \frac{m \lambda_1 \lambda_2}{(m+1) R_m(\lambda_1, \lambda_2)},$$

(Carlson (1977, p.101)).

According to Lemma 1 in the Appendix, to find G one has to minimize for fixed h_1, h_2, $\rho = (h_1 - h_2)/(h_1 + h_2)$, with $\tan \phi_j = a_{j2}/a_{j1}$, $j = 1, \dots, p$,

$$\begin{aligned}
\prod_j e_j^T AODO^T A^T e_j &= \prod_j \left(a_{j1}^2 + a_{j2}^2 \right) \left(h_1 \cos^2(\phi_j - \phi) + h_2 \sin^2(\phi_j - \phi) \right) \\
&= 2^{-p} (h_1 + h_2)^p \prod_j \left(a_{j1}^2 + a_{j2}^2 \right) \prod_j \left(1 + \rho \cos 2(\phi_j - \phi) \right) \\
&= \det(\Sigma)
\end{aligned}$$

over 2×2 orthogonal matrices $O = [\cos \phi, -\sin \phi; \sin \phi, \cos \phi]$. This is an example of the classical problem of finding the minimum of a trigonometric polynomial. Its statistical interpretation is maximum likelihood estimation of the rotation parameter ϕ, $\arg\min_\phi \sum \log(1+\rho\cos(\phi_j-\phi))$, in the family of densities $(2\pi)^{-1}\sqrt{1-\rho^2}[1-\rho\cos 2(\cdot-\phi)]^{-1}$. These densities are popular as models for wind directions, where they arise as a distribution of the polar angle in a bivariate normal vector whose coordinates have variances σ_1^2 and σ_2^2. Then $\rho=(\sigma_2^2-\sigma_1^2)/(\sigma_1^2+\sigma_2^2)$.

For large p, one can estimate the "true" density $f(\phi)$ of angles ϕ_j on the unit circle, $0\le\phi<2\pi$, and then approximate this minimum by

$$\exp\left\{p\min_\psi\int_0^{2\pi}\log(1+\rho\cos(\phi-\psi))f(\phi)d\phi\right\}.$$

Assuming that $\lambda_1=x\le 1=\lambda_2$,

$$\begin{aligned}G&=\frac{2^{p/(p-2)}}{p^{p/(p-2)}\prod_j(a_{j1}^2+a_{j2}^2)^{1/(p-2)}}\\&\quad\times\max_{x:0\le x\le 1}\frac{x^{1/(p-2)}[H_{(p-2)/2}(x,1)]^{2/(p-2)}}{[R_{p/2-1}(x,1)]^{p/(p-2)}[\min_\phi\prod_j(1+\rho\cos 2(\phi_j-\phi))]^{1/(p-2)}}.\end{aligned}\tag{4.1}$$

When $x\to 0$, $\rho\to 1$,

$$\begin{gathered}\frac{x}{1-\rho}\to\frac{p-1}{2},\\ \frac{\min_\phi\prod_j(1+\rho\cos 2(\phi_j-\phi))}{1-\rho}\to 2^{p-1}\min_k\prod_{j:j\ne k}\sin^2(\phi_j-\phi_k),\end{gathered}$$

which is positive provided all angles ϕ_j are different. Thus

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

Numerous examples show that when $p\ge 5$, this bound is typically attained, and this is the case for the interlaboratory studies by Paule and Mandel (1970) of gold vapor pressure.

Assuming there are 38 different values of σ_j^2 corresponding to the different laboratory runs, the experimental data allows one to estimate the sample variance

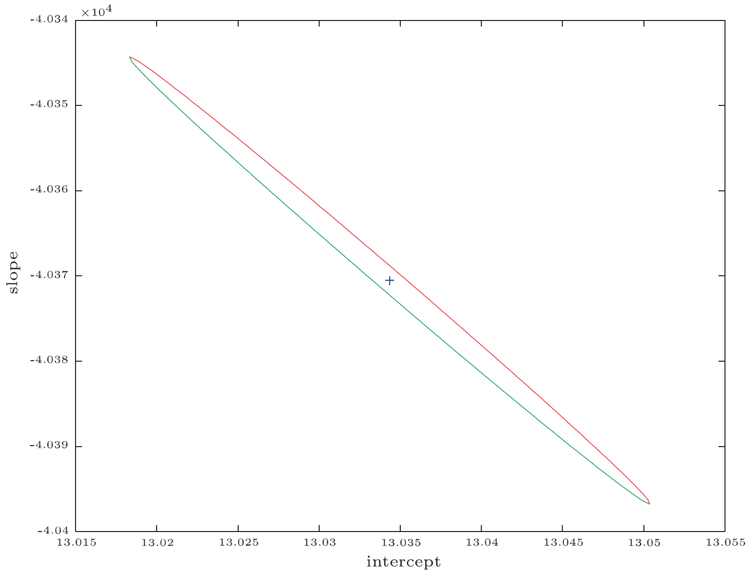

Figure 1. Confidence ellipsoid for β in the gold vapor pressure study (the value of the estimator δ is marked by a '+').

for each run, and reciprocals of these variances give the diagonal matrix Q. The matrix A is then determined from the design matrix X formed by pairs $(1, 1/T_j)$. We used $p = 375$ different temperature points T_j given in Table 2 and employed $1/T10^4$ in K^{-1} units. Thus, this procedure coincides with the one suggested by Fuller and Rao (1978). Numerical determination of $G(A)$, employing the scheme above for $\alpha = 0.05$, gives $\phi_{opt} = 2.19$, $\lambda_2 = 1$, $\lambda_1 = 0$, $G(A) = 9.23$, and the formula (5.4) from the next section provides the value $t_0^2 = 0.1495$.

The formula (1.2) leads to an approximate 95%-confidence ellipsoid for β based on a F-distribution with $r = 2$ and $p - r = 373$ degrees of freedom. This ellipsoid portrayed in Figure 1 provides useful information about the joint nature of the slope and the intercept in this study. Monte Carlo study indicates that the coverage probability of this ellipsoid is about 0.977 when the errors have the covariance matrix Q^{-1}. For the least favorable σ_j^2, the coverage probability is 0.987.

5. Volume of the Confidence Set

The volume of the confidence ellipsoid (1.3) is

$$\frac{t^r \pi^{r/2} [Y^T (I - XW)^T S (I - XW) Y]^{r/2}}{\Gamma((r+2)/2) \sqrt{\det(X^T Q X)}}.$$

Under an error covariance matrix Σ_0 its expected value, because of (3.1), has the form

$$\Delta = \frac{t^r \pi^{r/2} E_{\Sigma_0}(Y^T(I-XW)^T S(I-XW)Y)^{r/2}}{\Gamma((r+2)/2)\sqrt{\det(X^TQX)}}$$
$$= \frac{t^r \pi^{r/2} E(Z^T \Sigma_0^{1/2}(I-XW)^T S(I-XW)\Sigma_0^{1/2} Z)^{r/2}}{\Gamma((r+2)/2)\sqrt{\det(X^TQX)}}.$$

The matrix $\Sigma_0^{1/2}(I-XW)^T S(I-XW)\Sigma_0^{1/2}$ has rank $p-r$. Let η_j, $j=1,\ldots,p$ denote its eigenvalues (r of them are zeros.) Then for t_0^2 in (2.3),

$$\Delta = \left[\frac{rF_{r,p-r}(\alpha)G}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2}\Gamma((p+r)/2)}{\Gamma(p/2)\Gamma((r+2)/2)} \det(X^TQX)^{(2r-p)/[2(p-r)]}$$
$$\times \frac{H_{r/2}(\eta_1,\ldots,\eta_p)\det(Q)^{r/[2(p-r)]}}{[\det(S)\det(WS^{-1}W^T)\det(X^TQX)^2]^{r/[2(p-r)]}}. \quad (5.1)$$

With $A = Q^{1/2}X(X^TQX)^{-1/2}$, $\Pi = I - AA^T$ is a projection matrix.

Theorem 5.1. *The expected volume* Δ *of the confidence ellipsoid* (1.3) *under the error covariance matrix* Σ_0 *has the form* (5.1). *With* G *defined by* (2.4)*, its minimal value,*

$$\min_S \Delta = \left[\frac{rF_{r,p-r}(\alpha)G}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2}\Gamma(p/2)}{\Gamma((p-r)/2)\Gamma((r+2)/2)}$$
$$\times \frac{\prod_{k=r+1}^{p}[\lambda_k(\Pi\Sigma_0 Q\Pi)]^{r/[2(p-r)]}}{\sqrt{\det(X^TQX)}}, \quad (5.2)$$

is attained when $Q^{-1/2}SQ^{-1/2}$ *is the generalized, Moore-Penrose inverse of* $Q^{1/2}$ $(I-XW)\Sigma_0(I-XW)^TQ^{1/2}$.

In practice, the matrix Σ_0 is unknown, but Q^{-1} is a suitable surrogate. If $\Sigma_0 = Q^{-1}$, then $S = Q(I-XW)$, which gives the same result in (1.2) as $S = Q$, since $WX = I$. Then (5.2) simplifies to

$$\min_S \Delta = \left[\frac{rF_{r,p-r}(\alpha)G}{p-r}\right]^{r/2} \frac{(2\pi)^{r/2}\Gamma(p/2)}{\sqrt{\det(X^TQX)}\Gamma((p-r)/2)\Gamma((r+2)/2)}. \quad (5.3)$$

According to (2.3), one gets for this S, when $\Sigma_0 = Q^{-1}$,

$$t_0^2 = \frac{rF_{r,p-r}(\alpha)}{p-r}G. \quad (5.4)$$

Thus G with $(G \geq 1)$ can be interpreted as the adjustment factor to the percentile of an F-distribution needed to obtain a conservative $1-\alpha$ confidence region.

When $r=1$, $X_{j1} \equiv 1$, (i.e., in the common mean case), Theorem 5.1 shows that the shortest interval for this common mean based on the weighted means statistic $\delta = \sum \omega_i Y_i$, $\omega_i \geq 0$, $\sum \omega_i = 1$, corresponds to the quadratic form $\sum \omega_i (Y_i - \delta)^2$ in (1.3) (Rukhin (2007)).

6. Further Examples and Simulation Results

We re-examine here the example discussed by Wu (1986), in which $p = 12$, $r = 3$, $X = [X_1, X_2, X_3]$ with X_1 being the 12-dimensional vector of ones, $X_2 = [1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10]^T$, and X_3 formed by the squares of coordinates of X_2. Wu considers two cases, namely that of unequal variances $\sigma_j^2 = x_{j2}/2$, and of equal variances $\sigma_j^2 \equiv 1$. With $Q = \mathrm{diag}(2/X_2)$, $\alpha = 0.05$, calculations give the value $t_0^2 = 6.71$ while, when Q is the identity matrix, $t_0^2 = 16.72$.

We evaluated, by Monte Carlo simulation, the coverage probabilities and average volumes of (1.2) as well as for the confidence ellipsoid based on the estimators $v_{J,8}$, $v_{H(1)}$, v_w, and $v_{J(1)}$ defined in Wu (1986). The retain-eight jackknife variance estimator $v_{J,\rho}$, $\rho = 8$, has the form

$$v_{J,\rho} = \frac{1}{\binom{p-r}{\rho-r+1} \det(XX^T)} \sum_{k_1,\ldots,k_\rho} XX^T(k_1,\ldots,k_\rho) \times (\delta_{k_1,\ldots,k_\rho} - \delta)(\delta_{k_1,\ldots,k_\rho} - \delta)^T,$$

where δ is the least squares estimator, $\delta_{k_1,\ldots,k_\rho}$ is the least squares estimator obtained on the basis of the pairs (x_j, y_j), $j = k_1, \ldots, k_\rho$, and $XX^T(k_1, \ldots, k_\rho)$ denotes the principal minor of the matrix XX^T corresponding to the rows and columns indexed by $k_1, \ldots, k_\rho$.

The delete-one jackknife variance estimator,

$$v_{J(1)} = \sum_j (1 - w_j)(\delta_{(j)} - \delta)(\delta_{(j)} - \delta)^T,$$

is based on the least squares estimators $\delta_{(j)}$ after removal of the j-th observation and on the weights $w_j = \mathrm{diag}(X(X^TX)^{-1}X^T)$ (which cannot exceed one.)

A related procedure was given by Hinkley (1977),

$$v_{H(1)} = \frac{1}{1 - r/p} \sum_j (1 - w_j)^2 (\delta_{(j)} - \delta)(\delta_{(j)} - \delta)^T.$$

The weighted bootstrap variance estimator,

$$v_w = \frac{E_\star(\det(X^T D^\star X)(\delta^\star - \delta)(\delta^\star - \delta)^T)}{E_\star(\det(X^T D^\star X))},$$

Table 1. The coverage probabilities (cp) of the confidence ellipsoids and their average volumes (av).

	equal variances						unequal variances					
	(1.3)	$v_{J,8}$	$v_{J(1)}$	$v_{H(1)}$	v_w	(6.1)	(1.3)	$v_{J,8}$	$v_{J(1)}$	$v_{H(1)}$	v_w	(6.1)
cp	0.99	0.52	0.46	0.37	0.46	0.51	0.99	0.83	0.81	0.76	0.73	0.89
av	0.25	0.03	0.02	0.02	0.02	0.11	2.59	0.12	0.75	0.59	0.27	0.18

uses $P_j^\star$ copies of $(y_j, x_{j1}, \ldots, x_{jr})$, $j = 1, \ldots, p$, with $E_\star$ referring to the resampling expected value, and $D^\star = (P_1^\star, \ldots, P_p^\star)$.

We also compared these procedures with the estimation method of the error covariance matrix by Horn, Horn, and Duncan (1975) that suggests estimating $\mathrm{Var}(\delta)$, $\delta = WY$, via

$$\widetilde{\mathrm{Var}\,(\delta)} = W\mathrm{diag}\Big(R^{-1}(I - XW)YY^T(I - XW)^T\Big)W^T, \tag{6.1}$$

where $R = \mathrm{diag}(I - XW)$.

Table 1 illustrates the conservative nature of the ellipsoid (1.3). While it is superior to all other procedures in terms of coverage probability, which is near one, its volume exceeds their volumes.

Acknowledgements

This research was supported by NSA grant #H98230-06-1-0068. The author is grateful to two referees for helpful comments. The help of F. Potra with the calculations in Section 4 is also acknowledged.

Appendix

By replacing Y by $Q^{1/2}Y$, X by $Q^{1/2}X$, and Σ by $Q^{1/2}\Sigma Q^{1/2}$, we will assume in the proofs of Theorems 2.1 and 5.1 that $Q = I$.

A.1. Proof of Theorem 2.1

When $Q = I$, $I - XW = X(X^TX)^{-1}X^T = \Pi$ and $XW = I - \Pi$ are projection matrices.

With the vector Z formed by independent standard normal $Z_1, \ldots, Z_p$,

$$\begin{aligned}
&P\left((\delta - \beta)^T\widehat{\mathrm{Var}\,(\delta)}^{-1}(\delta - \beta) > t^2\right)\\
&\quad = P\left((Y - X\beta)^TW^TX^TXW(Y - X\beta) > t^2Y^T(I - XW)^TS(I - XW)Y\right)\\
&\quad = P(Z^TTZ < 0),
\end{aligned}$$

where $T = \Sigma^{1/2}[\Pi S \Pi - t^{-2}(I - \Pi)]\Sigma^{1/2}$, and $\Sigma^{1/2}$ is the diagonal matrix given by the elements $\sigma_i, i = 1, \ldots, p$. Thus T is congruent via $\Sigma^{1/2}$ to the matrix $\Pi S \Pi - t^{-2}(I - \Pi)$, which does not depend on $\sigma_1, \ldots, \sigma_p$.

The linear operator corresponding to $I - \Pi$ leaves the subspace spanned by the columns of the matrix X invariant, while Π annuls this subspace. Thus, by Sylvester's law of inertia, Theorem 4.5.8 in Horn and Johnson (1985), T (as well as $\Sigma^{1/2} T \Sigma^{1/2}$) must have r negative eigenvalues $\lambda_1, \ldots, \lambda_r$ and $p - r$ positive eigenvalues $\lambda_k, k = r + 1, \ldots, p$. When $t \to \infty$, $\lambda_k \sim \lambda_k(\Sigma^{1/2} \Pi S \Pi \Sigma^{1/2})$. According to perturbation theory for symmetric matrices, $\lambda_i \sim -\sum_j \mu_{ij} t^{-2j}, i = 1, \ldots, r$, is an analytic function in t^{-2} for sufficiently large t.

One can identify $\mu_i = \mu_{i1}$ as non-zero eigenvalues of the matrix $\Pi_0 \Sigma^{1/2}(I - \Pi)\Sigma^{1/2}\Pi_0$, where Π_0 is the projection matrix onto the r-dimensional eigenspace corresponding to the zero eigenvalue of $\Sigma^{1/2}\Pi S \Pi \Sigma^{1/2}$, i.e., the projection matrix onto the r-dimensional space spanned by the columns of $\Sigma^{1/2} X \Sigma^{1/2}$, $\Pi_0 = \Sigma^{-1/2} X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1/2}$. Thus

$$\Pi_0 \Sigma^{1/2}(I - \Pi)\Sigma^{1/2}\Pi_0 = \Sigma^{-1/2} X (X^T \Sigma^{-1} X)^{-1} X^T X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1/2}.$$

The non-zero eigenvalues of this matrix are those of the matrix

$$(X^T \Sigma^{-1} X)^{-1} X^T X (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1/2} \Sigma^{-1/2} X = (X^T \Sigma^{-1} X)^{-1} X^T X,$$

so that μ_i are the eigenvalues of the matrix $(X^T \Sigma^{-1} X)^{-1} X^T X$, and they coincide with $\lambda_i((X^T X)^{1/2} (X^T \Sigma^{-1} X)^{-1} (X^T X)^{1/2})$.

Because of Theorem 3.1,

$$P\left(Z^T T Z < 0\right) = P\left(\sum_{i=1}^{r} |\lambda_i| Z_i^2 > \sum_{k=r+1}^{p} \lambda_k Z_k^2\right)$$
$$\sim \frac{H_{(p-r)/2}(\mu_1, \ldots, \mu_r)\Gamma(p/2)}{t^{p-r}\sqrt{\lambda_{r+1}\cdots\lambda_p}\,\Gamma(r/2)\Gamma((p-r+2)/2)}.$$

To find the product $\lambda_{r+1} \cdots \lambda_p$, notice that with $\phi_C(\lambda) = \det(C - \lambda I)$ denoting the characteristic polynomial of a matrix C,

$$\lambda_{r+1} \cdots \lambda_p = \lim_{\lambda \to 0} \frac{|\phi_{\Sigma^{1/2}\Pi S \Pi \Sigma^{1/2}}(\lambda)|}{\lambda^r}.$$

Let V denote a $2r \times 2r$ symmetric matrix of the form,

$$V = \begin{pmatrix} X^T S X & -I_r \\ -I_r & 0_r \end{pmatrix},$$

where 0_r is the $r \times r$ zero matrix. Then $|\det(V)| = 1$,

$$V^{-1} = \begin{pmatrix} 0_r & -I_r \\ -I_r & -X^T S X \end{pmatrix},$$

$$\Sigma^{1/2}\Pi S\Pi\Sigma^{1/2} - \lambda I = \Sigma^{1/2}S\Sigma^{1/2} - \lambda I + \Sigma^{1/2}(W^T, SX)V\begin{pmatrix} W \\ X^TS \end{pmatrix}\Sigma^{1/2}.$$

If λ does not belong to the spectrum of ΣS, one gets, from Theorem 18.1.1 in Harville (1997),

$$\begin{aligned}\phi_{\Sigma^{1/2}\Pi S\Pi\Sigma^{1/2}}(\lambda) &= \det(\Sigma)\det(S - \lambda\Sigma^{-1})\det(V)\\ &\quad\times\det\left(V^{-1} + \begin{pmatrix} W \\ X^TS \end{pmatrix}(S-\lambda\Sigma^{-1})^{-1}(W^T, SX)\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{|\phi_{\Sigma^{1/2}\Pi S\Pi\Sigma^{1/2}}|}{\lambda^r} &= \frac{\det(\Sigma)\det(S-\lambda\Sigma^{-1})}{\lambda^r}\\ &\times\det\begin{pmatrix} W(S-\lambda\Sigma^{-1})^{-1}W^T & W(S-\lambda\Sigma^{-1})^{-1}SX - I_r \\ X^TS(S-\lambda\Sigma^{-1})^{-1}W^T - I_r & X^TS(S-\lambda\Sigma^{-1})^{-1}SX - X^TSX \end{pmatrix}\\ &\to \det(\Sigma)\det(S)\det(WS^{-1}W^T)\det(X^T\Sigma^{-1}X).\end{aligned}$$

A.2. Proof of Theorem 3.1

For a standard normal variable Z (Erdelyi (1953, Chap. 9, 9.3(3)))

$$P(Z^2 > z) = \frac{1}{\pi}\int_1^\infty \frac{e^{-zu/2}\,du}{u\sqrt{u-1}},\quad z > 0.$$

If $Z_1, \ldots, Z_p$ are independent standard normal variables, then for positive coefficients $\lambda_2, \ldots, \lambda_p$,

$$\begin{aligned}P(Z_1^2 > \sum_2^p \lambda_k Z_k^2) &= \frac{1}{\pi}\int_1^\infty \frac{E\left(e^{-(\sum\lambda_k Z_k^2)u/2}\right)\;du}{u\sqrt{(u-1)}}\\ &= \frac{1}{\pi}\int_1^\infty \frac{du}{u\sqrt{(u-1)\prod_k(1+\lambda_k u)}}.\end{aligned}$$

More generally, when $r = 2m+1$ is an odd positive integer,

$$\begin{aligned}P\Big(\sum_{i=1}^r Z_i^2 > z\Big) &= \int_{z/2}^\infty e^{-u}u^{m-1/2}\,du/\Gamma(m+1/2)\\ &= \sum_{j=1}^m \frac{e^{-z/2}z^{j-1/2}}{2^{j-1/2}\Gamma(j+1/2)} + \frac{1}{\pi}\int_1^\infty \frac{e^{-zu/2}\,du}{u\sqrt{u-1}},\end{aligned}$$

so that

$$P\left(\sum_{i=1}^r Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2\right) = \sum_{j=1}^m \frac{E\left(e^{-\sum_k\lambda_k Z_k^2/2}(\sum_k \lambda_k Z_k^2)^{j-1/2}\right)}{2^{j-1/2}\Gamma(j+1/2)}$$

$$+\frac{1}{\pi}\int_1^\infty \frac{E\left(e^{-u(\sum_k \lambda_k Z_k^2)/2}\right)\,du}{u\sqrt{u-1}}.$$

Known formulas for the normal distribution show that for any positive integer $r, r \le p$,

$$\begin{aligned}
&P\left(\sum_{i=1}^r Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2\right)\\
&= \sum_{j=1}^m \frac{1}{2^j\sqrt{\pi}\Gamma(j+1/2)}\int_1^\infty \frac{E\left(e^{-u(\sum_k \lambda_k Z_k^2)/2}(\sum_k \lambda_k Z_k^2)^j\right)\,du}{\sqrt{u-1}}\\
&+\frac{1}{\pi}\int_1^\infty \frac{E\left(e^{-u(\sum_k \lambda_k Z_k^2)/2}\right)\,du}{u\sqrt{u-1}}\\
&= \frac{1}{\sqrt{\pi}}\sum_{j=1}^m \frac{j!}{\Gamma(j+1/2)}\sum_{\nu_{r+1}+\cdots+\nu_p=j}\prod_k \frac{\lambda_k^{\nu_k}\Gamma(\nu_k+1/2)}{\nu_k!\Gamma(1/2)}\\
&\times\int_1^\infty \frac{du}{\sqrt{u-1}\prod_k(1+u\lambda_k)^{\nu_k+1/2}} + \frac{1}{\pi}\int_1^\infty \frac{du}{u\sqrt{(u-1)\prod(1+u\lambda_k)}}. \qquad \text{(A.1)}
\end{aligned}$$

Here and further $\nu_{r+1},\ldots,\nu_p$ are non-negative integers, and we used the formulas,

$$\frac{1}{\sqrt{z}} = \frac{1}{\sqrt{2\pi}}\int_1^\infty \frac{e^{-z(u-1)/2}\,du}{\sqrt{u-1}}$$

with $z = \sum_k \lambda_k Z_k^2$,

$$E\left(e^{-u\lambda_k Z_k^2/2} Z_k^{2\nu}\right) = \frac{2^\nu\Gamma(\nu+1/2)}{(1+u\lambda_k)^{\nu+1/2}\Gamma(1/2)}$$

combined with the multinomial theorem.

Since $Z_s^2 = (\sum_i Z_i^2)\omega_s^2, s = 1,\ldots,r, \sum_i \omega_i^2 = 1$, one gets

$$P(\sum_i \lambda_i Z_i^2 > z) = P(\sum_i Z_i^2 \sum_i \lambda_i\omega_i^2 > z).$$

By replacing λ_k in (A.1), $k = r+1,\ldots,p$, with $\lambda_k/\sum_i \lambda_i\omega_i^2$, we see that

$$\begin{aligned}
&P\left(\sum_{i=1}^r \lambda_i Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2\right)\\
&= \sum_{j=1}^m \frac{j!}{\Gamma(j+1/2)\Gamma(1/2)}\sum_{\nu_{r+1}+\cdots+\nu_p=j}\prod_k \frac{\lambda_k^{\nu_k}\Gamma(\nu_k+1/2)}{\nu_k!\Gamma(1/2)}
\end{aligned}$$

$$\times \int_{S_r} d\omega \int_1^\infty \frac{(\sum_i \lambda_i \omega_i^2)^{(p-r)/2}\, du}{\sqrt{u-1}\prod_k(\sum_i \lambda_i\omega_i^2 + u\lambda_k)^{\nu_k+1/2}}$$
$$+\frac{1}{\pi}\int_{S_r} d\omega \int_1^\infty \frac{(\sum_i \lambda_i\omega_i^2)^{(p-r)/2}\, du}{u\sqrt{(u-1)\prod_k(\sum_i \lambda_i\omega_i^2 + u\lambda_k)}}.$$

With $H_{(p-r)/2}$ defined by (2.1) when $\lambda_i \to 0, i = 1, \ldots, r$,

$$P\left(\sum_{i=1}^r \lambda_i Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2\right) \sim \frac{H_{(p-r)/2}(\lambda_1,\ldots,\lambda_r)}{\sqrt{\lambda_{r+1}\cdots\lambda_p}}\Big[\sum_{j=1}^m \frac{j!}{\Gamma(j+1/2)\Gamma(1/2)}$$
$$\times \sum_{\nu_{r+1}+\cdots+\nu_p=j}\prod_k \frac{\Gamma(\nu_k+1/2)}{\nu_k!\Gamma(1/2)}\int_1^\infty \frac{du}{u^{j+(p-r)/2}\sqrt{u-1}} + \frac{1}{\pi}\int_1^\infty \frac{du}{u^{(p-r)/2}\sqrt{u-1}}\Big]$$
$$= \frac{H_{(p-r)/2}(\lambda_1,\ldots,\lambda_r)}{\sqrt{\lambda_{r+1}\cdots\lambda_p}}\Big[\sum_{j=1}^m \frac{j!\Gamma(j+(p-r)/2)B(1/2, j+(p-r-1)/2)}{\Gamma(j+1/2)\Gamma(1/2)\Gamma((p-r)/2)j!}$$
$$+\frac{1}{\pi}B(1/2,(p-r+1)/2)\Big] = \frac{H_{(p-r)/2}(\lambda_1,\ldots,\lambda_r)\Gamma(p/2)}{\sqrt{\lambda_{r+1}\cdots\lambda_p}\Gamma(r/2)\Gamma((p-r+2)/2)} \tag{A.2}$$

with $B(a,b)$ denoting the beta function.

The situation, when $r = 2m$ is an even integer, is easier to handle since then

$$P\left(\sum_{i=1}^r Z_i^2 > z\right) = e^{-z/2}\sum_{j=0}^{m-1}\frac{z^j}{2^j j!},$$
$$P\left(\sum_{i=1}^r \lambda_i Z_i^2 > z\right) = \sum_{j=0}^{m-1}\frac{z^j}{2^j j!}\int_{S_r}\exp\left\{-\frac{z}{2\sum_i \lambda_i\omega_i^2}\right\}\frac{d\omega}{(\sum_i \lambda_i\omega_i^2)^j}.$$

Thus

$$P\left(\sum_{i=1}^r \lambda_i Z_i^2 > \sum_{k=r+1}^p \lambda_k Z_k^2\right)$$
$$= \sum_{j=0}^{m-1}\frac{1}{2^j j!}\int_{S_r} E\left(\exp\left\{-\frac{\sum_k \lambda_k Z_k^2}{2\sum_i \lambda_i\omega_i^2}\right\}(\sum_k \lambda_k Z_k^2)^j\right)\frac{d\omega}{(\sum \lambda_i\omega_i^2)^j}$$
$$= \sum_{\nu_{r+1}+\cdots+\nu_p \le m-1}\int_{S_r}\prod_k \frac{\lambda_k^{\nu_k}}{2^{\nu_k}\nu_k!}E\left(\exp\left\{-\frac{\lambda_k Z_k^2}{2\sum_i \lambda_i\omega_i^2}\right\}Z_k^{2\nu_k}\right)\frac{d\omega}{(\sum \lambda_i\omega_i^2)^{\nu_k}}$$
$$= \sum_{\nu_{r+1}+\cdots+\nu_p \le m-1}\int_{S_r}\prod_k \frac{\lambda_k^{\nu_k}\Gamma(\nu_k+1/2)}{\Gamma(1/2)\nu_k!}\frac{(\sum_i \lambda_i\omega_i^2)^{(p-r)/2}\, d\omega}{(\sum_i \lambda_i\omega_i^2 + \lambda_k)^{\nu_k+1/2}},$$

so that as $\lambda_i \to 0, i = 1, \ldots, r$,

$$P\left(\sum_{i=1}^{r} \lambda_i Z_i^2 \geq \sum_{k=r+1}^{p} \lambda_k Z_k^2\right) \sim \frac{H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)}{\sqrt{\lambda_{r+1} \cdots \lambda_p}} \sum_{\nu_{r+1}+\cdots+\nu_p \leq m-1} \prod_k \frac{\Gamma(\nu_k + 1/2)}{\Gamma(1/2)\nu_k!}$$
$$= \frac{H_{(p-r)/2}(\lambda_1, \ldots, \lambda_r)\Gamma(p/2)}{\sqrt{\lambda_{r+1} \cdots \lambda_p}\Gamma(r/2)\Gamma((p-r+2)/2)},$$

and (A.2) holds in this case as well.

A.3. Lemma

Lemma 1. *Let for $r = 2$, $\lambda_i = \lambda_i(A^T \Sigma^{-1} A), i = 1, 2,$ where the $p \times 2$ matrix A with elements a_{ji} is such that $A^T A = I$. Then for $\lambda_1 \neq \lambda_2$,*

$$\frac{\partial \lambda_1}{\partial \sigma_j^{-2}} = \frac{1}{\lambda_1 - \lambda_2} e_j^T A \left(-\lambda_2 I + A^T \Sigma^{-1} A\right) A^T e_j,$$
$$\frac{\partial \lambda_2}{\partial \sigma_j^{-2}} = \frac{1}{\lambda_1 - \lambda_2} e_j^T A \left(\lambda_1 I - A^T \Sigma^{-1} A\right) A^T e_j,$$

where $e_j, j = 1, \ldots, p$, denote the basis vectors.

The proof follows by solving simultaneous equations for the derivatives in Lemma 1 obtained by differentiation of the two identities,

$$\operatorname{tr}(A^T \Sigma^{-1} A) = \operatorname{tr}(A A^T \Sigma^{-1}) = \sum_j \frac{a_{j1}^2 + a_{j2}^2}{\sigma_j^2},$$
$$\det(A^T \Sigma^{-1} A) = \sum_{j,k} \frac{(a_{j1} a_{k2} - a_{j2} a_{k1})^2}{2\sigma_j^2 \sigma_k^2}.$$

In Section 4, $G^{p-2} = \sup_\Sigma H^2_{(p-2)/2}(\lambda_1, \lambda_2)/[(\lambda_1 \lambda_2)^{p-1} \det(\Sigma)]$, so that a stationary point $(\sigma_1^2, \ldots, \sigma_p^2)$ satisfies the equation

$$\frac{\partial}{\partial \sigma_j^{-2}} \left[2 \log H_{(p-2)/2}(\lambda_1, \lambda_2) - (p-1)\log(\lambda_1 \lambda_2) - \log\det(\Sigma)\right]$$
$$= \sigma_j^2 - \sum_i h_i \frac{\partial \lambda_i}{\partial \sigma_j^{-2}} = 0,$$

where $h_i = -2\partial \log H_{(p-2)/2}/\partial \lambda_i + (p-1)/\lambda_i, i = 1, 2$. Lemma 1 implies that $\sigma_j^2 = e_j^T A F A^T e_j$, with $(\lambda_1 - \lambda_2)F = (h_2 \lambda_1 - h_1 \lambda_2)I + (h_1 - h_2)A^T \Sigma^{-1} A$. It is easy to check that the vector of eigenvalues of F coincides with h.

If $A^T \Sigma^{-1} A = O \Lambda O^T$ is the spectral decomposition with the diagonal matrix Λ =diag(λ_1, λ_2), and an orthogonal matrix O, then $F = ODO^T$, where O is the same orthogonal matrix and the diagonal matrix D is formed by the eigenvalues of F.

Thus our optimization problem can be separated into two parts. The first consists in minimization of $\prod_{\jmath} e_{\jmath}^T AODO^T A^T e_{\jmath}$ over all orthogonal matrices O for a fixed diagonal matrix $D = D(\lambda_1, \lambda_2)$, and the second in the maximization in λ_1, λ_2 of $H^2_{(p-2)/2}(\lambda_1, \lambda_2)$ $(\lambda_1\lambda_2)^{1-p}$ $[\min_O \prod_{\jmath} e_{\jmath}^T AODO^T A^T e_{\jmath}]^{-1}$. This fact holds for any r.

A.4. Proof of Theorem 5.1

When $Q = I$, $\det(WS^{-1}W^T)\det(X^TX)^2 = \det(X^TS^{-1}X)$. Our goal is to find the matrix $S = U^TLU$ with an orthogonal $p \times p$ U and a diagonal L, say, $\operatorname{diag}(L) = (\ell_1, \ldots, \ell_p)$, $0 \le \ell_1 \le \cdots \le \ell_p$, minimizing the last factor in (5.1), which for $\eta_j = \lambda_j(S\Pi\Sigma_0\Pi)$ can be written as

$$\frac{H_{r/2}(\eta_1, \ldots, \eta_p)}{[\det(S)\det(X^TS^{-1}X)]^{r/[2(p-r)]}}.$$

The function $H_{r/2}(\eta_1, \ldots, \eta_p)$ is Schur convex, so that

$$H_{r/2}(\eta_1, \ldots, \eta_p) \ge H_{r/2}(\ \operatorname{diag}\ (LU\Pi\Sigma_0\Pi U^T)).$$

For the minimizer of the denominator U, the matrix $U\Pi\Sigma_0\Pi U^T$ is diagonal with non-zero elements $\xi_k = \lambda_k(\Pi\Sigma_0\Pi), k = r+1, \ldots, p$.

Let $UX_{i_1\ldots i_r}$ denote the minor of the matrix UX corresponding to the rows indexed by $i_1, \cdots, i_r$, so that by the Cauchy-Binet formula, $\sum UX^2_{i_1\ldots i_r} = r!$ $\det(X^TX)$, and

$$\det(X^TS^{-1}X) = \det(X^TU^TL^{-1}UX) = \frac{1}{r!}\sum_{i_1,\cdots,i_r} UX^2_{i_1\ldots i_r}\ell_{i_1}^{-1}\cdots\ell_{i_r}^{-1}.$$

For the orthogonal matrix U, which maximizes $\det(X^TS^{-1}X)$, one must have $UX^2_{i_1\ldots i_r} = 0$, if $i_1, \cdots, i_r$ is not a permutation of indices $1, \ldots, r$. In other words, the matrix UX has only zeros in the last $p - r$ rows, i.e., the matrix UXX^TU^T is a block-diagonal matrix formed by X^TX in the upper left position, and zero blocks elsewhere. Since $\Pi X = 0$, the matrices XX^T and $\Pi\Sigma_0\Pi$ act on orthogonal subspaces, and an orthogonal matrix U that diagonalizes $\Pi\Sigma_0\Pi$ and reduces XX^T to such a form, exists.

This matrix simultaneously minimizes the denominator and maximizes the numerator,

$$H_{r/2}(\eta_1, \ldots, \eta_p) = H_{r/2}(0, \ldots, 0, \ell_{r+1}\xi_{r+1}, \ldots, \ell_p\xi_p)$$

$$= \frac{[\Gamma(p/2)]^2}{\Gamma((p-r)/2)\Gamma((p+r)/2)} H_{r/2}(\ell_{r+1}\xi_{r+1},\ldots,\ell_p\xi_p),$$

and $\det(X^T S^{-1} X) = \det(X^T X) \prod_{i=1}^r \ell_i^{-1}$. Therefore,

$$\begin{aligned}
&\inf_S \frac{H_{r/2}(\eta_1,\ldots,\eta_p)}{[\det(S)\det(X^T S^{-1}X)]^{r/[2(p-r)]}} \\
&= \min_{\ell_{r+1},\ldots,\ell_p} \frac{H_{r/2}(0,\ldots,0,\ell_{r+1}\xi_{r+1},\ldots,\ell_p\xi_p)}{[\det(X^TX)\prod_{k=r+1}^p \ell_k]^{r/[2(p-r)]}} \\
&= \frac{[\Gamma(p/2)]^2}{\Gamma((p-r)/2)\Gamma((p+r)/2)} \min_{t_1,\ldots,t_{p-r}} \frac{H_{r/2}(t_1,\ldots,t_{p-r})}{(\prod_1^{p-r} t_i)^{r/[2(p-r)]}} \left[\frac{\prod_{r+1}^p \xi_k}{\det(X^TX)}\right]^{r/[2(p-r)]}.
\end{aligned}$$

It follows from Theorem 2, (2.13) in Carlson (1966) that

$$H_{r/2}(t_1,\ldots,t_{p-r}) \geq (\prod_1^{p-r} t_i)^{r/[2(p-r)]},$$

so that the minimal value of the ratio in $t_1,\ldots,t_{p-r}$ is 1, and the minimizing matrix S can be taken to be any positive multiple of the generalized (Moore-Penrose) inverse of $\Pi\Sigma_0\Pi$. Then the eigenvalues η_i are zero (multiplicity r) or one (multiplicity $p-r$), and (5.2) holds.

References

Basu, A. and Chatterjee, S. (2002). Comparison of bootstrap and jackknife estimators in linear regression: second order results. *Statist. Sinica* **12**, 575-598.

Carlson, B. C. (1966). Some inequalities for hypergeometric functions, *Proc. Amer. Math. Soc.* **17**, 32-39.

Carlson, B. C. (1977). *Special Functions of Applied Mathematics.* Academic Press, New York.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Ann. Statist.* **14**, 1261-1350.

Erdelyi, A. (1953). *Higher Transcendental Functions.* McGraw Hill Book Co., New York.

Fuller, W. A. and Rao, J. N. K. (1978). Estimation for a linear regression with unknown diagonal matrix. *Ann. Statist.* **6**, 1149-1158.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective.* Springer, New York.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19**, 285-292.

Horn, R. A. and Johnson, C. A. (1985). *Matrix Analysis.* Cambridge University Press, Cambridge, UK.

Horn, S. A., Horn, R. A. and Duncan, D. B. (1975). Estimating heteroscedastic variance in linear models. *J. Amer. Statist. Assoc.* **70**, 380-385.

Meng, X.-L. (2005). From unit root to Stein's estimator to Fisher's k statistic: if you have a moment I can tell you more. *Statist. Sci.* **20**, 141-162.

Paule, R. C. and Mandel, J. (1970). Analysis of interlaboratory measurements on the vapor pressure of gold. National Bureau of Standards Special Publication, 260-19.

Rukhin, A. L. (2007). Conservative confidence intervals based on weighted means statistics. *Statist. Probab. Lett.* **77**, 1312-1321.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York.

Wu, C. J. F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, (with discussion) *Ann. Statist.* **14**, 1261-1350.

Statistical Engineering Division, National Institute of Standards and Technology, 100 Bureau Dr. Gaithersburg, MD 20899 U.S.A.

E-mail: andrew.rukhin@nist.gov

(Received January 2008; accepted February 2009)