ON THE USE OF GENERAL POSITIVE QUADRATIC FORMS IN SIMULTANEOUS INFERENCE

by

Yosef Hochberg

Highway Safety Research Center
and
Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1030

September 1975

ON THE USE OF GENERAL POSITIVE QUADRATIC FORMS IN SIMULTANEOUS INFERENCE

by

Yosef Hochberg

Highway Safety Research Center
and
Department of Biostatistics
University of North Carolina at Chapel Hill

1. Introduction and Summary. The matrix in the quadratic form used to project simultaneous confidence intervals for linear functions of the parameter vector in Scheffé's S-method is conventionally chosen to be the inverse of the dispersion matrix of the allied estimators. (See Scheffé, 1959, Ch. 3 and Miller, 1966, Ch. 2).

The first attempt to introduce optimality criteria for choosing the 'best' quadratic form seems to be that of Hoel (1951) where the case of confidence bands for a simple straight line regression is discussed. Hoel (1951) introduced a weight function for the values of the independent variable and proposed the criterion of minimum weighted average distance between the bands. It turns out that the Working-Hotelling quadratic form (a special case of the S-method in straight line regression) is not necessarily optimal. The optimal choice depends on the specific distribution attached to the 'independent' variable.

Bohrer (1973) proved that the S-method is optimal (for practical confidence levels) under the criterion of minimum average length of confidence intervals for linear functions with uniform weight in some elliptical region. Besides this result, it seems that little has been recorded on features of the S-method, (see discussion following Wynn and Bloomfield, 1971).

The first goal of this work is to discuss an additional optimality feature of the S-method. This is given in Section 2.

Another point discussed here is the use of general quadratic forms as a convenient tool for discriminating among various linear parametric functions (by the allocation of lengths of confidence intervals) in a way different than that obtained by using the S-method.

In a linear regression setup several writers (see, e.g., Wynn and Bloomfield, 1971, and their references) have sharpened the usual bands obtained from the S-method by restricting the bands to finite intervals. We feel that problems might arise where an investigator has main interest in a region which is away from the vector of means of the independent variables but will still like to have bands over the entire space. This situation is discussed in Section 3 where elliptical bands are constructed with minimum length of confidence interval set at a prespecified point.

Similar techniques can be used in any ANOVA setup. General p.d. quadratic forms can be used to construct discriminatory simultaneous confidence intervals on different types of estimable parametric functions.

2. On optimality of the Working-Hotelling-Scheffé method. (The S-method).

Let $\underset{\sim}{\theta} = (\theta_1,\ldots,\theta_k)' \epsilon R^k$ be the unknown vector of parameters, let $\hat{\underset{\sim}{\theta}} = (\hat{\theta}_1,\ldots,\hat{\theta}_k)' \epsilon R^k$ be the corresponding estimators with the distributional property $\hat{\underset{\sim}{\theta}} \sim N(\underset{\sim}{\theta}, \sigma^2 \underset{\sim}{V})$, where $\underset{\sim}{V}$ is a positive definite (p.d.) and completely known, σ^2 is unknown but an unbiased estimator s_ν^2 is available such that s_ν^2 is independent of $\hat{\underset{\sim}{\theta}}$ and $\nu s_\nu^2/\sigma^2$ is distributed as a central chi-square variate with ν d.f.

In its utmost generality, the projection technique for setting simultaneous confidence intervals on all $\underset{\sim}{\ell}'\underset{\sim}{\theta}$, $\underset{\sim}{\ell} \epsilon R^k$ starts with a general convex confidence set $\mathcal{D}$, say, for $\underset{\sim}{\theta}$, then projecting on $\underset{\sim}{\ell}$ the corresponding pair of tangents to $\mathcal{D}$. The S-method chooses for $\mathcal{D}$ the set $\mathcal{D}^* = \{\underset{\sim}{\theta}: (\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})'\underset{\sim}{V}^{-1}(\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}) \leq s_\nu d\}$ where d is determined by the confidence coefficient (See Scheffé, 1959, Ch. 3). Among all sets of the same confidence coefficient, $\mathcal{D}^*$ has minimum volume. This can be verified by simple arguments. However, we are looking for an optimality characteristic of the S-method in terms of the projected confidence intervals for the $\underset{\sim}{\ell}'\underset{\sim}{\theta}$. We will now discuss such an optimality feature of $\mathcal{D}^*$ in the class of all ellipsoids. First we give the simultaneous estimation procedure based on a general p.d. quadratic form in $\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta}$. Let $\underset{\sim}{M}$: kxk be any p.d. matrix and put $\underset{\sim}{t} = (\hat{\underset{\sim}{\theta}}-\underset{\sim}{\theta})/s_\nu$, then we have

$$\Pr\{|\underset{\sim}{x}'\underset{\sim}{t}| \leq \xi_\alpha^{\frac{1}{2}} (\underset{\sim}{x}'\underset{\sim}{M}^{-1}\underset{\sim}{x})^{\frac{1}{2}}, \forall\ \underset{\sim}{x} \epsilon \underset{\sim}{R}^k\} = 1 - \alpha \qquad (2.1)$$

where ξ_α is the $(1-\alpha)$-th quantile of a random variable distributed as $\nu \sum_{j=1}^{k} \lambda_j \chi_1^2/\chi_\nu^2 = Q$, where χ_ν^2 denotes a central chi-square variate of ν d.f., the λ_j's are the k non-zero roots of $\underset{\sim}{V}\underset{\sim}{M}$ and all the chi-square variates in the expression given for Q are independent. This is easily proved using the well known result on the distribution of a general p.d. quadratic form in normal variates (see for example Box, 1954) and Scheffé's projection argument. Note that only when k = 2 both λ_1 and λ_2 must be positive. For k > 2 we may have some negative roots. These roots will, however, satisfy the conditions $\Sigma\lambda_i > 0$ and $\Sigma(1/\lambda_i) > 0$, see Graybill (1969), Ch. 12. The problem of obtaining ξ_α or estimating it is discussed later.

For any given $\underset{\sim}{x}$ we define LCI = $2(\xi \underset{\sim}{x}'\underset{\sim}{M}^{-1}\underset{\sim}{x})^{\frac{1}{2}}$. This is the length

between the upper and lower $1-\alpha$ confidence bounds on $\underset{\sim}{x}'\underset{\sim}{t}$ in (2.1). Our problem is to choose a 'good' $\underset{\sim}{M}$. Rather than trying to minimize the weighted mean of the LCI we will look for an $\underset{\sim}{M}$ that minimizes the weighted mean of the squared LCI's. This approach is more convenient than Hoel's approach, since on introducing a distribution for the values of $\underset{\sim}{x}\in R^k$ we find that the weighted mean squared LCI depends only on the mean vector and the dispersion matrix of $\underset{\sim}{x}$ while the mean weighted LCI depends on the specific distribution assumed. Thus, assume that a weight function (a distribution) is introduced for the values of $\underset{\sim}{x}$ in R^k such that the mean vector of $\underset{\sim}{x}$ is $\underset{\sim}{u}$ and the dispersion matrix is $\underset{\sim}{\Sigma}$. The expected squared LCI is then seen from (2.1) to be

$$4\xi_\alpha[\text{Trace}(\underset{\sim}{M}^{-1}\underset{\sim}{\Sigma}) + \underset{\sim}{u}'\underset{\sim}{M}^{-1}\underset{\sim}{u}]. \tag{2.2}$$

Explicit minimization of (2.2) with respect to a real p.d. matrix $\underset{\sim}{M}$ for any given $\underset{\sim}{\mu}$ and $\underset{\sim}{\Sigma}$ seems difficult. (Note that ξ_α is a rather complicated function of $\underset{\sim}{M}$). Clearly the S-method (i.e., $\underset{\sim}{M} = \underset{\sim}{V}^{-1}$) is not expected to be optimal for all $\underset{\sim}{\Sigma}$ or for $\underset{\sim}{u} \neq \underset{\sim}{0}$. If we take $\underset{\sim}{u} = \underset{\sim}{0}$ and $\underset{\sim}{\Sigma} = \underset{\sim}{V}^{-1}$ (this is quite reasonable in a linear regression setup where the independent variables have been centered around their means) we get for (2.2) the simplified expression

$$4\xi_\alpha \sum_{j=1}^{k} (1/\lambda_j) \tag{2.3}$$

where the λ_j's are the roots of $\underset{\sim}{V}\underset{\sim}{M}$.

We note that (2.3) is invariant to $\Sigma\lambda_i$ if the λ_i are changed proportionally, thus we may arbitrarily fix $\Sigma\lambda_i$. For $k = 2$ we now let $\lambda_1 + \lambda_2 = 1$ and prove that (2.3) is minimized when $\lambda_1 = \lambda_2 = \frac{1}{2}$. It is not

difficult to verify that this is implied by the following result. Let X, Z, and Y be i.i.d. random variables where X and Y are distributed as χ_1^2 variates, then we have:

Lemma 2.1. For any $a > 0$ the quantity

$$\Pr\left\{\frac{\lambda X + (1-\lambda)Y}{Z} \leq a\lambda(1-\lambda)\right\},$$

is maximized when $\lambda = \frac{1}{2}$.

Proof. The problem is automatically reduced to proving that the quantity

$$f(\lambda,a) = \Pr\{\lambda X + (1-\lambda)Y \leq a\lambda(1-\lambda)\}$$

is minimized by $\lambda = \frac{1}{2}$ for all values of a. On letting $h(\cdot)$ denote the probability density function of a χ_1^2 variable we have

$$f' = \frac{df(\lambda,a)}{d\lambda} = \int_{y=0}^{a\lambda} h[(1-\lambda)(a - \tfrac{y}{\lambda})]h(y)(\tfrac{y}{\lambda^2} - a)dy.$$

When substituting for $h(\cdot)$ and $\lambda = \frac{1}{2}$ we get

$$f'(\lambda = \tfrac{1}{2}) = \int_{y=0}^{a/2} [(\tfrac{A}{2} - y)y]^{-\frac{1}{2}}(4y - a) = 0.$$

The second derivative, f", is negative at $\lambda = \frac{1}{2}$ since

$$f''(\lambda = \tfrac{1}{2}) = \int_{0}^{a/2} h(y)[(4y - a)^2 h'(\tfrac{a}{2} - y) - 16yh(\tfrac{a}{2} - y)]dy$$

and $h'(x)$, the derivative of $h(x)$ is negative for all values of $x > 0$.

This demonstrates that $\lambda = \frac{1}{2}$ is at least a local maximum. It seems difficult to show analytically that this is the global maximum.

However, numerical integration (see Table 1) supports the statement that at $\lambda = 1/2$ the functions $f(\lambda,a)$ indeed obtains a unique global maximum for any $a > 0$.

3. Simultaneous confidence bands of regression surfaces with minimum LCI at an arbitrary point.

The Working-Hotelling-Scheffé bands on regression surfaces have minimum LCI at the vector of means of the independent variables. In some experimental situations it might be appropriate to have simultaneous confidence bands the LCI of which is minimized at a different point. This state of affairs arises, for example, when the bands are constructed mainly for prediction purposes and the range of future most likely values for the independent variables is removed from their current sample means. We now consider regression surfaces with intercepts and suppose that the independent variables have been centered around their means. We let $\underset{\sim}{M}^{-1} = (m_{ij}^{(-1)})_{i,j=1,\ldots,k}$, $\underset{\sim}{V} = (v_{ij})_{i,j=1,\ldots,k}$ and look for a p.d. matrix $\underset{\sim}{M}$ such that the LCI (obtained from (2.1)) at a point $\underset{\sim}{x} + \underset{\sim}{\mu} = (1, x_2+\mu_2, \ldots, x_k+\mu_k)$ is proportional to the LCI obtained by the S-method at the point $\underset{\sim}{x}$, for all $\underset{\sim}{x}$. Thus, we must satisfy the identity

$$(\underset{\sim}{x}-\underset{\sim}{\mu})'\underset{\sim}{V}(\underset{\sim}{x}-\underset{\sim}{\mu}) = \underset{\sim}{x}'\underset{\sim}{M}^{-1}\underset{\sim}{x}, \quad \text{for all } \underset{\sim}{x} = (1,x_2,\ldots,x_k). \tag{3.1}$$

This gives

$$\left.\begin{aligned} m_{ij}^{(-1)} &= v_{ij}, \quad i,j = 2,\ldots,k \\ m_{1i}^{(-1)} &= m_{i1}^{(-1)} = -\sum_{j=2}^{k} \mu_j v_{ij}, \quad i = 2,\ldots,k \\ m_{11}^{(-1)} &= v_{11} + \sum_{i=2}^{k}\sum_{j=2}^{k} v_{ij}\mu_i\mu_j \,. \end{aligned}\right\} \tag{3.2}$$

When k = 2, for example, we have

$$\underset{\sim}{V} = \begin{bmatrix} 1/n & 0 \\ 0 & 1/\Sigma x_i^2 \end{bmatrix} \quad ; \quad \underset{\sim}{M}^{-1} = \begin{bmatrix} 1/n+\mu^2/\Sigma x_i^2 & -\mu/\Sigma x_i^2 \\ -\mu/\Sigma x_i^2 & 1/\Sigma x_i^2 \end{bmatrix}$$

which gives

$$\underset{\sim\sim}{VM} = \begin{bmatrix} 1 & \mu \\ n\mu/\Sigma x_i^2 & 1+n\mu^2/\Sigma x_i^2 \end{bmatrix} .$$

The roots of $\underset{\sim\sim}{VM}$ are thus determined from the equation

$$(1-\lambda)(1-\lambda+a) - a = 0, \quad a = n\mu^2/\Sigma x_i^2 . \tag{3.3}$$

As noted earlier, when k = 2 both roots are positive and the implementation of the method is easy by using the tabulated values of ξ_α given in Table 2. The values of ξ_α depend on ν,α and the roots λ_1,λ_2, however, one may fix $\lambda_1 + \lambda_2 = 2$, say, as we did in our tables. This way the number of parameters is three: ν, α, and λ, $0 < \lambda < 1$. The use of the method and a comparison with the S-method is demonstrated by the following numerical example in a simple linear regression setup

$$E(Y_i) = \alpha + \beta X_i, \quad i = 1,\ldots,n.$$

Example. (From Draper and Smith (1966), 1.2) (μ chosen arbitrarily). Using lower case letters x, x_i, etc. for deviations from the means, we have n = 25, $\bar{X}$ = 52.6; $\bar{Y}$ = 9.424; s_ν^2 = .7926, Σx_i^2 = 7154.42, $\hat{\beta}$ = -0.079829, μ = -17.2. Using Scheffé's method we have the bands

$$\bar{Y} + \hat{\beta}x \pm s_\nu\{2F_{2,23}^{(\alpha)}[.04 + .00014x^2]\}^{1/2}.$$

Using the above described method which sets the minimum LCI at μ = -17.2

we get here

$$M^{-1} = \begin{bmatrix} .08135 & .00240 \\ .00240 & .00014 \end{bmatrix};$$

$$a = 1.03377;\ \lambda_1 = 2.65751;\ \lambda_2 = 0.37629.$$

So the bands are

$$\bar{Y} + \hat{\beta}x \pm s_\nu\{\xi_\alpha(0.08135 - 0.00480x + 0.00014x^2)\}^{1/2}. \qquad (3.4)$$

To obtain ξ_α from our table one should have used the entries $\nu = 23$, $\lambda = 2\ \mathrm{Min}(\lambda_1,\lambda_2)/(\lambda_1 + \lambda_2) = 0.24806$, and an appropriate α. The value obtained from the table must then be multiplied by $(\lambda_1+\lambda_2)/2$. These entries are not provided by our tables and thus either interpolation or a simple well known approximation, that we now consider, must be used. That approximation is given by

$$\hat{\xi}_\alpha = ghF^{(\alpha)}_{h,\nu}$$

where $F^{(\alpha)}_{h,\nu}$ is the $(1-\alpha)$-th quantile of a central F distribution with d.f. h and ν and where

$$g = \frac{\Sigma\lambda_j^2}{\Sigma\lambda_j} \quad \text{and} \quad h = \frac{(\Sigma\lambda_j)^2}{(\Sigma\lambda_j)^2}.$$

See for example Box (1954). This approximation is based on equating the first two moments of Q (see definition following (2.1)) with that of a $g\nu x_h^2/x_\nu^2$ variable and solving for g and h. It can be used conveniently for any $k > 2$ for which exact tabulation is not practical

For our example and $\alpha = 0.10$ we have: $g = 2.37456$, $h = 1.27762$,

and $\hat{\xi}_{.10} = 8.49991$. The exact value (see Section 4 for details on computations) is $\xi_{.10} = 8.31868 = (3.0338)(5.484)/2$, where the value 5.484 was computed as the entries of Table 2 for $\lambda = 2\,\mathrm{Min}(\lambda_1,\lambda_2)/(\lambda_1 + \lambda_2) = 0.248$ and $\nu = 23$. Both bands, i.e., the S-method bands and those based on (3.4) with the exact evaluation of $\xi_{.10}$, are given in Figure 1.

4. Computations. Let X,Y,Z be independent where X and Y each distributed as a χ^2_1 variable and Z distributed as a χ^2_ν/ν. Table 1 was constructed by computing in brute force

$$\Pr\{\lambda X + (1-\lambda)Y \leq a\lambda(1-\lambda)\} = \int_{y=0}^{a\lambda} \Pr\{X \leq (1-\lambda)(a - \tfrac{y}{\lambda})\}h(y)dy$$

where $h(\cdot)$ is the density of a χ^2_1.

The outer integral was computed by the IBM scientific subroutine DQG32 based on Gaussian quadrature. A subroutine for computing the CDF of a χ^2_ν variable was used from the library of the Biostatistics Department of the University of North Carolina at Chapel Hill. The range of integration for the outer integral was split to a number of sub-intervals with attention to the shape of $h(\cdot)$. When $\lambda = .5$ the results of our numerical integrations coincide with the available tables to the third decimal point.

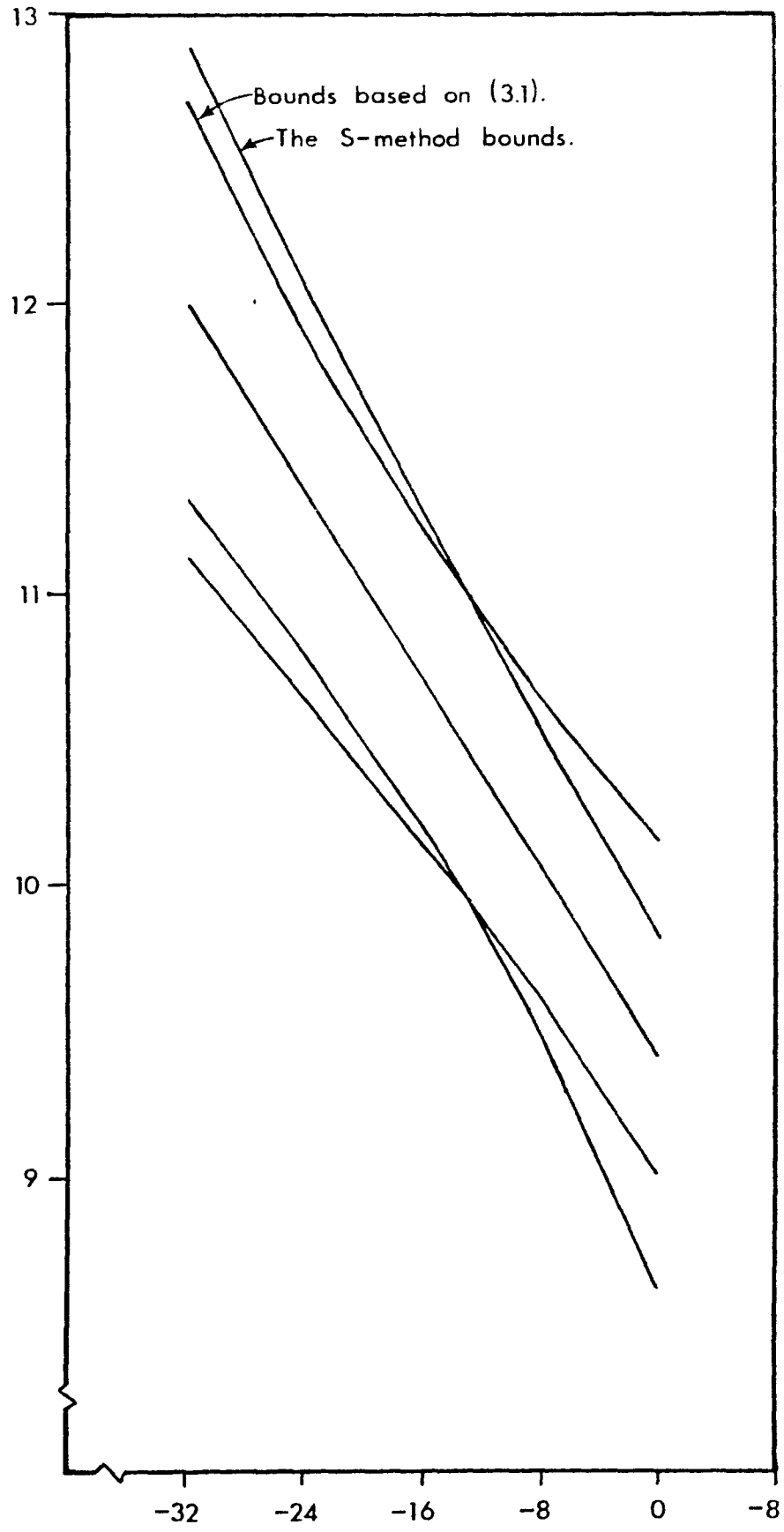
Table 2 involved similar computations. Here we used numerical integrations as above with a simple search procedure that terminated when the gap between the upper and lower bounds became less than --0.01.

The range for the integral over values of z in calculating

$$\int_{z=0}^{\infty} \int_{y=0}^{az/(2-\lambda)\nu} \Pr\{X \leq \frac{az}{\nu\lambda} - \frac{(2-\lambda)y}{\lambda}\}\, dF_1(y)\, dF_\nu(z),$$

where $F_\nu(\cdot)$ is the CDF of a χ^2_ν variate, was taken from zero to 4ν in all entries of the table.

FIGURE 1.



x = deviations from the mean.

REFERENCES

Bohrer, R. (1973). An optimality property of Scheffé bonus. Annals of Statistics, 1, 766-772.

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one way classification. Ann. Math. Statist., 25, 290-302.

Draper, N. and Smith, H. (1966). Applied Regression Analysis. Wiley, New York.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth, Belmont, California.

Hoel, P. B. (1951). Confidence regions for linear regression. Proc. Second Berkeley Symp. on Math. Statist. and Prob. Berkeley: University of California Press, 75-81.

Miller, R. G., Jr. (1966). Simultaneous Statistical Inference. McGraw-Hill, New York.

Scheffé, H. (1959). The Analysis of Variance. Wiley, New York.

Wynn, H. P. and Bloomfield, P. (1971). Simultaneous confidence in regression analysis. J. Roy. Statist. Soc., Ser. B, 33, 202-217.

Table 1: Values of $f(\lambda,a)$

a	λ 0.100	0.200	0.300	0.400	0.500
0.500	0.070	0.093	0.107	0.114	0.117
1.000	0.133	0.177	0.203	0.217	0.221
1.500	0.189	0.251	0.287	0.306	0.313
2.000	0.238	0.317	0.361	0.386	0.393
2.500	0.284	0.376	0.428	0.457	0.465
3.000	0.325	0.428	0.487	0.518	0.528
3.500	0.361	0.475	0.538	0.572	0.583
4.000	0.394	0.517	0.585	0.620	0.632
4.500	0.425	0.555	0.626	0.663	0.675
5.000	0.454	0.589	0.663	0.701	0.713
5.500	0.480	0.621	0.696	0.735	0.747
6.000	0.505	0.650	0.727	0.765	0.777
6.500	0.526	0.675	0.752	0.791	0.803
7.000	0.547	0.698	0.776	0.814	0.826
7.500	0.568	0.721	0.798	0.836	0.847
8.000	0.586	0.741	0.817	0.854	0.865
8.500	0.603	0.759	0.834	0.870	0.881
9.000	0.620	0.775	0.850	0.885	0.895
9.500	0.635	0.791	0.863	0.897	0.907
10.000	0.649	0.805	0.876	0.909	0.918
10.500	0.663	0.818	0.887	0.919	0.928
11.000	0.676	0.830	0.898	0.928	0.936
11.500	0.688	0.841	0.907	0.936	0.944
12.000	0.699	0.850	0.914	0.942	0.950

Table 2: Values of $\xi_\alpha(\lambda,\nu)^*$

α	ν \ λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	10	9.65	9.33	9.07	8.84	8.66	8.47	8.35	8.28	8.22
	15	8.74	8.54	8.27	7.97	7.82	7.66	7.56	7.45	7.40
	20	8.49	8.10	7.81	7.62	7.43	7.23	7.10	7.03	7.00
	30	8.12	7.82	7.51	7.32	7.16	6.96	6.84	6.73	6.67
	40	7.86	7.66	7.38	7.09	6.89	6.79	6.61	6.58	6.50
0.10	10	6.39	6.20	6.08	6.01	5.95	5.91	5.88	5.86	5.85
	15	5.97	5.81	5.69	5.59	5.52	5.47	5.43	5.41	5.40
	20	5.78	5.62	5.47	5.38	5.30	5.25	5.22	5.19	5.18
	30	5.59	5.42	5.28	5.19	5.09	5.05	5.02	5.00	4.98
	40	5.52	5.35	5.21	5.11	5.02	4.96	4.92	4.90	4.89

*Here $k = 2$, $\lambda_1 + \lambda_2 = 2$ and $\lambda = \min(\lambda_1, \lambda_2)$.