Working Paper 92-08
February 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

A NOTE ON THE MULTIVARIATE BOX-COX TRANSFORMATION TO NORMALITY

Santiago Velilla*

Abstract

We study some aspects of the multivariate Box-Cox transformation to normality which have received little attention in the literature.

Key words
Asymptotic relative efficiency (ARE); Concentrated log-likelihood; Multivariate Box-Cox transformation.

* Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. 28903-Getafe (Madrid), Spain.

A NOTE ON THE MULTIVARIATE BOX-COX TRANSFORMATION TO NORMALITY

Santiago Velilla

Departamento de Estadística, Universidad Carlos III de Madrid, 28903-Getafe (Madrid), Spain.

Abstract: We study some aspects of the multivariate Box-Cox transformation to normality which have received little attention in the literature.

Key words: Asymptotic relative efficiency (ARE); Concentrated log-likelihood; Multivariate Box-Cox transformation.

1. INTRODUCTION

Let X be a random variable which takes values denoted by x. If the distribution of X is not normal, it is sometimes convenient to consider transformations that help to normalize the observed data. When x>0, a useful family of transformations is the family of Box and Cox (1964):

$$x^{(\lambda)}=\begin{cases}\dfrac{x^{\lambda}-1}{\lambda}, & \lambda\neq 0;\\ \log x, & \lambda=0.\end{cases} \qquad (1)$$

Note that the transformation in (1) is indexed by the scalar parameter λ.

Consider now a p-variate random vector $\mathbf{X}=(X_1, \ldots, X_p)'$ such that all its components take positive values. When the distribution of $\mathbf{X}$ is not multivariate normal, Andrews et al. (1971) have given the following generalization of (1). We have a p-vector $\Lambda=(\lambda_1, \ldots, \lambda_p)'$ of transformation parameters, one for each dimension, such that the model

$$X^{(\Lambda)}=(X_1^{(\lambda_1)}, \ldots, X_p^{(\lambda_p)})' \sim N_p(\mu,\Sigma) \qquad (2)$$

holds. We will write the parameters in model (2) in the form $\Theta=(\Lambda,\mu,\Sigma)$, where $\mu=(\mu_1, \ldots,\mu_p)'$ and $\Sigma=(\sigma_{ij})_{pxp}$.

With few exceptions (see, for example, section 5.3 in the book by Gnanadesikan (1977) or section 4.2 in the book by Seber (1984)), the multivariate Box-Cox transformation has received little attention in the literature. Computational and inferential procedures remain to be explored. In this note we derive, in section 2, a result which is shown to be useful for computational purposes. In section 3, we propose a general methodology for making inferences about the parameter Λ. Finally, in section 4, we study some efficiency properties of the MLE estimator $\hat{\Lambda}$ of Λ in model (2).

2. A COMPACT EXPRESSION FOR THE CONCENTRATED LOG-LIKELIHOOD FOR Λ

Let $\mathcal{X}=(x_{ij})=(x_1, \ldots,x_p)$ a nxp data matrix from model (2) and assume that the rows of the transformed data matrix $\mathcal{X}^{(\Lambda)}$, namely, $\mathbf{x}_i^{(\Lambda)}=(x_{i1}^{(\lambda_1)}, \ldots,x_{ip}^{(\lambda_p)})'$, i=1, ...,n, are i.i.d. $N_p(\mu,\Sigma)$. If $L(\Theta)=L(\Lambda,\mu,\Sigma)$ represents the associated log-likelihood, standard normal theory shows that the concentrated log-likelihood $L_{max}(\Lambda)$ for the transformation parameter is (up to an additive constant),

$$L_{max}(\Lambda)=-(n/2)\log[\,|\mathcal{X}^{(\Lambda)'}(I_n-1_n1_n'/n)\mathcal{X}^{(\Lambda)}|\,]+\log[J_\Lambda(\mathcal{X})], \qquad (3)$$

where $J_\Lambda(\mathcal{X})=\prod_{j=1}^{p} J_{\lambda_j}$, $J_{\lambda_j}=J_{\lambda_j}(x_j)=\prod_{i=1}^{n}|dx_{ij}^{(\lambda_j)}/dx_{ij}|=(\prod_{i=1}^{n}x_{ij})^{\lambda_j-1}$, is the jacobian of the transformation. Define, for each column j of the data matrix $\mathcal{X}$, j=1, ...,p, the n vector of normalizad variables

$Z_j^{(\lambda_j)} = (z_{ij}^{(\lambda_j)})$, $z_{ij}^{(\lambda_j)} = J_{\lambda_j}^{-1/n} x_{ij}^{(\lambda_j)}$, $i=1, \ldots, n$. We have the following result.

LEMMA 2.1. Let $\mathbf{Z}^{(\Lambda)} = (Z_1^{(\lambda_1)}, \ldots, Z_p^{(\lambda_p)})$ be the nxp matrix of normalized variables. The concentrated log-likelihood for Λ is (up to an additive constant)

$$L_{max}(\Lambda) = -\frac{n}{2} \log \left[\left| \mathbf{Z}^{(\Lambda)'} (I_n - 1_n 1_n'/n) \mathbf{Z}^{(\Lambda)} \right| \right]. \qquad (4)$$

PROOF. Define $D = \mathrm{diag}(J_{\lambda_1}^{1/n}, \ldots, J_{\lambda_p}^{1/n})$ so that $-(n/2)\log[J_\Lambda^{-2/n}(\mathcal{X})] = -(n/2)\log[|D|^{-2}]$. Taking account that $\mathbf{Z}^{(\Lambda)} = \mathcal{X}^{(\Lambda)} D^{-1}$ and recalling expression (3), the result (4) follows. ■

For computational purposes, expression (4) shows that the MLE estimator $\hat{\Lambda}$ is obtained by minimizing in Λ the determinant

$$\left| \mathbf{Z}^{(\Lambda)'} (I_n - 1_n 1_n'/n) \mathbf{Z}^{(\Lambda)} \right|, \qquad (5)$$

which depends on the nxp matrix $\mathbf{Z}^{(\Lambda)}$. It is easily seen that expression (5) generalizes the minimization problem associated with the determination of the MLE of λ in the scalar case of model (1). Recall that the (i,j) element of the determinant in (5) is of the form $\sum_{k=1}^{n} (z_{ki}^{(\lambda_i)} - \bar{z}_i^{(\lambda_i)})(z_{kj}^{(\lambda_j)} - \bar{z}_j^{(\lambda_j)})$, where $\bar{z}_i^{(\lambda_i)} = (1/n) \sum_{k=1}^{n} z_{ki}^{(\lambda_i)}$ and $\bar{z}_j^{(\lambda_j)} = (1/n) \sum_{k=1}^{n} z_{kj}^{(\lambda_j)}$, $i,j=1, \ldots, p$. In principle, minimization of (5) can be carried out by using the appropriate numerical subroutine. However, it is common to encounter computing overflow problems in the determination of the jacobian $J_\Lambda(\mathcal{X})$, particularly when n is large. We can use lemma 2.1 above to derive the following invariance property of $\hat{\Lambda}$ which is useful for overcoming these inconveniences.

THEOREM 2.1. Let $S=\mathrm{diag}(s_1, \ldots, s_p)$, where $s_j>0$, $j=1, \ldots, p$, and consider the scaled data matrix $\tilde{\mathcal{X}}=\mathcal{X}S^{-1}$. Write, in obvious notation, the corresponding matrix $\tilde{Z}^{(\Lambda)}=(\tilde{Z}_1^{(\lambda_1)}, \ldots, \tilde{Z}_p^{(\lambda_p)})$. We have,

$$|\tilde{Z}^{(\Lambda)'}(I_n-1_n1_n'/n)\tilde{Z}^{(\Lambda)}|=|S|^{-2}|Z^{(\Lambda)'}(I_n-1_n1_n'/n)Z^{(\Lambda)}|. \tag{6}$$

PROOF. Write $\tilde{\mathcal{X}}=(\tilde{x}_1, \ldots, \tilde{x}_p)$ so that $\tilde{x}_j=x_j/s_j$. Let $q_j=-s_j^{-1}J_{\lambda_j}^{-1/n}(x_j)s_j^{(\lambda_j)}$, $j=1, \ldots, p$. Using $J_{\lambda_j}(\tilde{x}_j)=(\prod_{i=1}^{n}\tilde{x}_{ij})^{\lambda_j-1}=s_j^{n(1-\lambda_j)}J_{\lambda_j}(x_j)$, it is not difficult to show that $\tilde{z}_{ij}^{(\lambda_j)}=J_{\lambda_j}^{-1/n}(\tilde{x}_j)\tilde{x}_{ij}^{(\lambda_j)}=s_j^{-1}z_{ij}^{(\lambda_j)}+q_j$. Therefore, we have

$$\tilde{Z}^{(\Lambda)}=Z^{(\Lambda)}S^{-1}+1_nQ', \tag{7}$$

where Q is the p×1 vector $Q=(q_1, \ldots, q_p)'$. Expression (6) follows directly from (7) by observing that $(I_n-1_n1_n'/n)1_n=0$. ■

As a consecuence of (6), the determination of $\hat{\Lambda}$ is not affected by scaling appropriately the data matrix $\mathcal{X}$ with a collection of constants $\{s_j\}$ not depending on Λ. For example, in practical applications with positive data, a convenient choice of the jth scale factor is $s_j=\max_{1\le i\le n} x_{ij}$, $j=1, \ldots, p$, so that the modified data matrix $\tilde{\mathcal{X}}$ has all its entries between 0 and 1.

3. INFERENCE ON THE TRANSFORMATION PARAMETER Λ

Under model (2), we can obtain, on the basis of asymptotic considerations, that the set of all Λ such that

$$2[L_{max}(\hat{\Lambda})-L_{max}(\Lambda)]\le\chi^2_{p,\alpha}, \tag{8}$$

where $\chi^2_{p,\alpha}$ is the upper 100α% point of a chi-squared distribution with p degrees of freedom, is an approximate (1-α)x100% confidence region for Λ. This is the methodology proposed by Andrews et al. (1971) for making inferences about Λ (see also Gnanadesikan (1977, sec. 5.3)). In this section we propose an alternative approach based on the expression (4) for $L_{max}(\Lambda)$.

Under suitable regularity conditions, we have that

$$\hat{\Lambda} \sim AN_p[\Lambda, (1/n)J(\Lambda,\Lambda)], \quad (9)$$

where $J(\Lambda,\Lambda)$ is the corresponding block for the parameter Λ in the inverse of the information matrix $I(\Theta)=-E[\partial/\partial\theta(\partial L(\Theta)/\partial\Theta)']$ under model (2). A general method for obtaining an estimate for $(1/n)J(\Lambda,\Lambda)$, as exposed, for example in Seber and Wild (1989, p. 38), is to compute

$$-\left[\frac{\partial}{\partial\Lambda}\left(\frac{\partial}{\partial\Lambda} L_{max}(\Lambda)\right)'\Bigg|_{\Lambda=\hat{\Lambda}}\right]^{-1}. \quad (10)$$

The main problem in dealing with (10) lies in finding the expression for the second partial derivatives $\partial^2 L_{max}(\Lambda)/\partial\lambda_r\partial\lambda_s$, r,s=1, ...,p, at $\Lambda=\hat{\Lambda}$. These are obtained as follows.

Write $M(\Lambda)=\mathbf{Z}^{(\Lambda)'}(I_n-1_n1_n'/n)\mathbf{Z}^{(\Lambda)}$ and define, for each j=1, ...,p, $W_j^{(\lambda_j)}=\partial Z_j^{(\lambda_j)}/\partial\lambda_j$ and $U_j^{(\lambda_j)}=\partial W_j^{(\lambda_j)}/\partial\lambda_j=\partial^2 Z_j^{(\lambda_j)}/\partial\lambda_j^2$. For a general nxp matrix $H=(h_1, \ldots, h_p)$, we will write $H_j(u)$ for the nxp matrix obtained from H by replacing its jth column by the nx1 vector u. We will also write $H_{jk}(u,v)$ for the nxp matrix obtained replacing its jth and kth column by, respectively, u and v. If we put Z_j, W_j and U_j for the corresponding functions $Z_j^{(\lambda_j)}$, $W_j^{(\lambda_j)}$ and $U_j^{(\lambda_j)}$ evaluated in the jth coordinate of the MLE estimator $\hat{\Lambda}$, we have the following result.

THEOREM 3.1. Let $\mathbf{Z}=\mathbf{Z}^{(\hat{\Lambda})}=(Z_1, \ldots, Z_p)$ and $A=(I_n-1_n1_n'/n)$. We have:

$$a)\ \partial L_{max}(\Lambda)/\partial\lambda_r\Big|_{\Lambda=\hat{\Lambda}} = -n|M(\hat{\Lambda})|^{-1}|\mathbf{Z}'_r(W_r)A\mathbf{Z}|;$$

$$b)\ \partial^2 L_{max}(\Lambda)/\partial\lambda_r^2\Big|_{\Lambda=\hat{\Lambda}} = -n|M(\hat{\Lambda})|^{-1}[|\mathbf{Z}'_r(W_r)A\mathbf{Z}_r(W_r)| + |\mathbf{Z}'A\mathbf{Z}_r(U_r)|],$$

for r=1, ...,p. We also have:

$$c)\ \partial^2 L_{max}(\Lambda)/\partial\lambda_r\partial\lambda_s\Big|_{\Lambda=\hat{\Lambda}} = -n|M(\hat{\Lambda})|^{-1}[|\mathbf{Z}'_{rs}(W_r,W_s)A\mathbf{Z}| + |\mathbf{Z}'_r(W_r)A\mathbf{Z}_s(W_s)|],$$

for r,s=1, ...,p ($r\neq s$).

The proof of this theorem is based on the following lemma.

LEMMA 3.1. Let C, D, and B be three nxp matrices and let E be a symmetric matrix of nxn. If e_i represents the ith canonical vector of $\mathbb{R}^n$, we have, for i=1, ...,n, j=1, ...,p, the following differentiation formulae:

$$a)\ \partial|C'ED|/\partial c_{ij} = |C'_j(e_i)ED|;\quad b)\ \partial|C'ED|/\partial d_{ij} = |C'ED_j(e_i)|;$$

$$c)\ \partial|B'EB|/\partial b_{ij} = 2|B'_j(e_i)EB|.$$

PROOF. See appendix.

PROOF OF THEOREM 3.1. a) By the chain rule and part c) of lemma 3.1 above, we have

$$\partial L_{max}(\Lambda)/\partial\lambda_r\Big|_{\Lambda=\hat{\Lambda}} = -n|M(\hat{\Lambda})|^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}|\mathbf{Z}'_j(e_i)A\mathbf{Z}|(\partial z_{ij}^{(\lambda_j)}/\partial\lambda_r\Big|_{\Lambda=\hat{\Lambda}})$$

$$= -n|M(\hat{\Lambda})|^{-1}\sum_{i=1}^{n}|\mathbf{Z}'_r(e_i)A\mathbf{Z}|w_{ir}$$

$$= -n|M(\hat{\Lambda})|^{-1}\sum_{i=1}^{n}|\mathbf{Z}'_r(w_{ir}e_i)A\mathbf{Z}|$$

$$= -n|M(\hat{\Lambda})|^{-1}|\mathbf{Z}'_r(W_r)A\mathbf{Z}|.$$

Parts b) and c) are obtained in a similar way by using parts a) and b) of lemma 3.1 and by recalling the fact that $\partial|M(\Lambda)|/\partial\Lambda\Big|_{\Lambda=\hat{\Lambda}}=0$. ■

For practical purposes, we get that the set of all Λ-values such that

$$(\Lambda-\hat{\Lambda})'H(\hat{\Lambda})(\Lambda-\hat{\Lambda})\leq\chi^2_{p,\alpha}, \qquad (11)$$

where $H(\hat{\Lambda})$ is the inverse of the matrix in (10), is an approximate $(1-\alpha)\times100\%$ confidence ellipsoid for the transformation parameter Λ. A general expression for the functions $W_j^{(\lambda_j)}$ and $U_j^{(\lambda_j)}$ can be found in Atkinson and Lawrance (1989).

EXAMPLE 1. A bivariate sample $(X_i,Y_i)'$ of size n=50 is generated through a bivariate lognormal model

$$(\log X,\log Y)'\sim N_2[\begin{pmatrix}0\\0\end{pmatrix};\begin{pmatrix}1 & .5\\ .5 & 1\end{pmatrix}].$$

By minimizing the determinant (5), we get $\hat{\Lambda}=(-.020,-.25)'$, with estimated variance-covariance matrix

$$\begin{pmatrix}.0197 & .0001\\ .0001 & .0181\end{pmatrix}.$$

Figure 1 shows the approximate 95% joint confidence ellipsoid (11) with boundary given by $\chi^2_{2,\alpha}=5.991$. ■

Figure 1

The advantage of this approach, in relation to the confidence region (8), is that approximate confidence regions can be explicitly computed. This is important because, when p≥2, the usual graphical estimation of the transformation parameter, which is extremely useful in the scalar case, is no longer feasible.

4. EFFICIENCY CONSIDERATIONS

Recall that model (2) implies the p marginal models

$$X_j^{(\lambda_j)} \sim N(\mu_j, \sigma_{jj}), \qquad (10)$$

j=1, ...,p. Model (2) and the family of models (10) are equivalent when the components of the random vector $\mathbf{X}=(X_1, \ldots, X_p)'$ are independent. Andrews et al. (1971) compare the results of fitting by maximum likelihood the model (2) and each of the p marginal models (10) separately and pose the general question if there is something to gain in using the model (2) in relation to the collection of p marginal models (10). A possible answer can be found in the theory what follows.

Define $\hat{\Lambda}_M=(\hat{\lambda}_{1M}, \ldots,\hat{\lambda}_{pM})'$, where $\hat{\lambda}_{jM}$ stands for the MLE estimator of λ_j under the jth marginal model. Given that (9) holds, if we can prove that

$$\hat{\Lambda}_M \sim AN_p[\Lambda,(1/n)J_M(\Lambda,\Lambda)],$$

asymptotic efficiency considerations indicate that the choice between $\hat{\Lambda}$ and $\hat{\Lambda}_M$ depends on the relation between the matrices $J(\Lambda,\Lambda)$ and $J_M(\Lambda,\Lambda)$. In fact, we have the following theorem.

THEOREM 4.1. Assuming appropriate regularity conditions, we have

a) $\hat{\Lambda}_M \sim AN_p[\Lambda,(1/n)J_M(\Lambda,\Lambda)]$;

b) $J_M(\Lambda,\Lambda) \geq J(\Lambda,\Lambda)$, in the sense that the p×p matrix $J_M(\Lambda,\Lambda)-J(\Lambda,\Lambda)$ is non negative definite.

PROOF. To simplify the notation, we will make the proof only in the case p=2. The ideas in the case of general p are similar. Note that, when p=2, the set of parameters under model (2) can be written as $\Theta=(\Lambda',\Phi')'$, $\Phi=(\Phi_1',\Phi_2)'$, where $\Phi_1=(\mu_1,\mu_2,\sigma_{11},\sigma_{22})'$ and $\Phi_2=\sigma_{12}$. Recall that, relabelling, $\Theta=(\Lambda,\Phi)=(\theta_1',\theta_2',\sigma_{12})'$, where $\theta_j=(\lambda_j,\mu_j,\sigma_{jj})'$, $j=1,2$.

a) Let $L_j(\theta_j)$ be the log-likelihood of θ_j under the jth marginal model and $\hat{\theta}_{jM}$ the corresponding MLE estimator of θ_j. Assuming regularity conditions, we can write

$$n^{1/2}\begin{pmatrix}\hat{\theta}_{1M}-\theta_1\\ \hat{\theta}_{2M}-\theta_2\end{pmatrix}=\begin{pmatrix}K(\theta_1) & 0_{3x3}\\ 0_{3x3} & K(\theta_2)\end{pmatrix}(1/n^{1/2})\begin{pmatrix}\partial L_1(\theta_1)/\partial\theta_1\\ \partial L_2(\theta_2)/\partial\theta_2\end{pmatrix}+o_p(1), \qquad (11)$$

where $K(\theta_1)$ and $K(\theta_2)$ converge in probability to the inverses of the corresponding information matrices $I(\theta_1)$ and $I(\theta_2)$. Therefore, the left hand side of (11) is asymptotically normal with mean zero and variance-covariance matrix given by

$$\begin{pmatrix}I(\theta_1)^{-1} & I(\theta_1)^{-1}M(\theta_1,\theta_2)I(\theta_2)^{-1}\\ I(\theta_2)^{-1}M(\theta_2,\theta_1)I(\theta_1)^{-1} & I(\theta_2)^{-1}\end{pmatrix}, \qquad (12)$$

where $M(\theta_1,\theta_2)=E[(\partial L_1(\theta_1/\partial\theta_1)(\partial L_2(\theta_2/\partial\theta_2)']$, and $M(\theta_2,\theta_1)=M(\theta_1,\theta_2)'$. This implies that $n^{1/2}(\hat{\Lambda}_M-\Lambda)$ is also asymptotically normal with mean zero and variance-covariance matrix $J_M(\Lambda,\Lambda)$ given by the corresponding submatrix of the matrix (12) above.

b) From a standard expansion of the likelihood equation for model (2), we can obtain

$$n^{1/2}(\hat{\sigma}_{12}-\sigma_{12})=K(\theta_1,\theta_2,\sigma_{12})(1/n^{1/2})(\partial L/\partial\Theta)+o_p(1), \qquad (13)$$

where $K(\theta_1,\theta_2,\sigma_{12})$ is a 1x7 matrix which converges in probability to the corresponding submatrix of the inverse of the information matrix under model (2). By adjoining expansions (11) and (13) we get that $n^{1/2}[(\hat{\theta}_{1M}-\theta_1)',(\hat{\theta}_{2M}-\theta_2)',\hat{\sigma}_{12}-\sigma_{12}]'$ is asymptotically normal with mean zero and variance-covariance matrix $V(\Lambda,\Phi)$, say. By the asymptotic optimality of the MLE estimator of (Λ,Φ) under model (2), we get that

$V(\Lambda,\Phi)\geq I(\Lambda,\Phi)^{-1}$ and the same is true for the submatrices which correspond to Λ . Therefore, $J_M(\Lambda,\Lambda)\geq J(\Lambda,\Lambda)$. ■

As a conclusion, the estimates obtained by fitting each of the marginal models separately are less efficient than the joint MLE $\hat{\Lambda}$. As the following examples show, the loss of efficiency might be severe and , therefore, $\hat{\Lambda}$ is preferable to $\hat{\Lambda}_M$.

EXAMPLE 2. Consider the case p=2 and $\Lambda=(0,0)'$. After lengthy algebra, it can be shown that

$$J_M(\Lambda,\Lambda)=D^{1/2}\begin{pmatrix}1 \; ; & \rho^3\\ \rho^3; & 1\end{pmatrix}D^{1/2}, \tag{14}$$

where $D=\text{diag}(2/(3\sigma_{11}),2/(3\sigma_{22}))$, and ρ is the correlation coefficient between logX and logY. We also obtain

$$J(\Lambda,\Lambda)=D^{1/2}\begin{pmatrix}p_{11}(\rho); & p_{12}(\rho)\\ p_{12}(\rho); & p_{22}(\rho)\end{pmatrix}D^{1/2}, \tag{15}$$

where $p_{11}(\rho) = p_{22}(\rho)= [1+([\rho^2(\rho^2+3)]/[3(3-2\rho^2)])]^{-1}$ and $p_{12}(\rho) = \rho^3[1+(1/3)(\rho^4-3\rho^2+6)]^{-1}$. Note that the diagonal entries of the matrix $J_M(\Lambda,\Lambda)$, namely, $2/(3\sigma_{11})$ and $2/(3\sigma_{22})$, agree with the expression of the asymptotic variance for $\hat{\lambda}_{jM}$, in the case of $\lambda_j=0$, which was obtained by Hinkley (1975).

It is easily shown that the matrix $J_M(\Lambda,\Lambda)-J(\Lambda,\Lambda)$ equals to zero when $\rho=0$ and is positive definite for $0<|\rho|<1$. For $|\rho|<1$, the *asymptotic relative efficiency* (ARE) of the MLE $\hat{\Lambda}$ with respect to $\hat{\Lambda}_M$ is given by (Serfling (1980), p. 141),

$$\text{ARE}(\hat{\Lambda},\hat{\Lambda}_M)=[\,|J(\Lambda,\Lambda)|/|J_M(\Lambda,\Lambda)|\,]^{1/2} \tag{16}$$

$$= \frac{3}{[(1+\rho^2+\rho^4)(\rho^4-3\rho^2+9)]^{1/2}} .$$

Table 1 displays the value of the ARE for selected values of ρ. Recall that the ARE is bounded between 1 (ρ=0) and $(3/7)^{1/2}$=.655 which is the limiting value as $|\rho|$ approaches 1. ■

ρ	.00	.25	.33	.50	.75	.90	.95	.975	.99
ARE	1.000	0.978	0.908	0.792	0.738	0.710	0.683	0.668	0.660

Table 1. ARE($\hat{\lambda},\hat{\lambda}_M$) for selected values of ρ.

EXAMPLE 3. Let (X,Y)' be a bivariate random vector. Consider estimation of the scalar parameter λ from two different models:

a) (Joint mormality) $(X^{(\lambda)},Y)' \sim N_2(\mu,\Sigma)$;

b) (Marginal normality) $X^{(\lambda)} \sim N(\mu_1,\sigma_{11})$.

Let $\hat{\lambda}$ be the MLE of λ under model a) and let $\hat{\lambda}_M$ be the MLE of λ under model b). If av[$\hat{\lambda}_M$] and av[$\hat{\lambda}$] are the respective asymptotic variances of $\hat{\lambda}_M$ and $\hat{\lambda}$, it can be shown that, when λ=0,

$$av[\hat{\lambda}_M]=av[\hat{\lambda}][1+ \frac{\rho^2}{3(1-\rho^2)}]. \tag{17}$$

for $|\rho|<1$. From (17), the two asymptotic variances are the same when ρ=0. However, the loss of efficiency can be very large for values of $|\rho|$ close to one. ■

APPENDIX

PROOF OF LEMMA 3.1. We will proof only part a) since parts b) and c) are similar. Write $C=(c_1, \ldots, c_p)$, $D=(d_1, \ldots, d_p)$ and let $Q=(q_{rs})=C'ED$. Therefore $q_{rs}=c_r'Ed_s$. Standard results on matrix differentiation show that $\partial|Q|/\partial q_{rs}=Q_{rs}$, where Q_{rs} is the cofactor of the element q_{rs} (see, for example, Mardia et al. (1979, p.479)). By the chain rule we have

$$\partial|C'ED|/\partial c_{ij}=\sum_{r=1}^{p}\sum_{s=1}^{p}(\partial|Q|/\partial q_{rs})(\partial q_{rs}/\partial c_{ij})=\sum_{s=1}^{p}Q_{js}(\sum_{\beta=1}^{n}e_{i\beta}d_{\beta s})$$

$$=|C_j'(e_i)ED|.$$

■

REFERENCES

ANDREWS, D. F., GNANADESIKAN, R.& WARNER, J.L. (1971). Transformations of multivariate data. *Biometrics*, **27**, 825-840.

ATKINSON, A.C. & LAWRANCE, A.J. (1989). A comparison of asymptotically equivalent test statistics for regression transformation. *Biometrika* **76**, 223-229.

BOX, G.E.P.& COX, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, Ser. B., **26**, 211-52.

GNANADESIKAN, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. (J. Wiley: New York).

HINKLEY, D.V. (1975). On power transformations to symmetry. *Biometrika*, **62**, 101-111.

MARDIA, K.V., KENT, J.L. & BIBBY, J.M. (1979). *Multivariate Analysis*. (Academic Press: London).

SEBER, G.A.F. (1984). *Multivariate Observations*. (J. Wiley: New York).

SEBER, G.A.F. & WILD, C.J. (1989). *Nonlinear Regression*. (J. Wiley: New York).

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. (J. Wiley: New York).

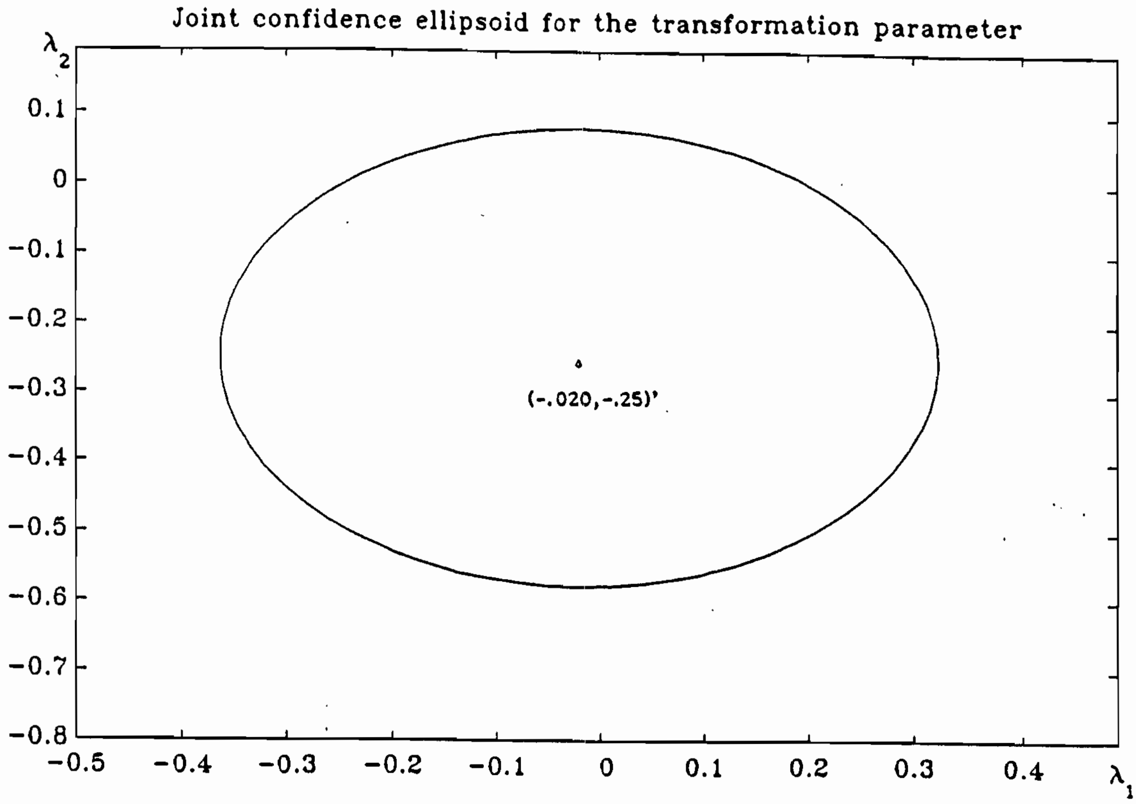

Figure 1. Confidence ellipsoid for $\Lambda=(\lambda_1,\lambda_2)'$.