On Two Nonequivalent Measures of Complexity

DONALD E. RAMIREZ

Abstract —A careful distinction is made between the two definitions of complexity introduced by van Emden for data compression. The paper shows that the two concepts are nonequivalent. A compromise criterion is introduced. Inherently different properties of these two concepts of complexity are proved. These differences are crucial to the choice of a performance criterion in model building. The Maklad–Nichols decision rule is explicitly derived in the case of a multivariate linear regression model.

I. Introduction

A FUNDAMENTAL difficulty in statistical analysis is the selection of an appropriate model and the determination of the dimension of the best model. Model selection is a common problem when a statistical model contains many parameters. Presently in the literature of control theory and econometrics, there is considerable interest in choosing models based on the parsimony of the parameters. The selection of a parsimonious model requires the aid of a model selection criterion.

Akaike [1], in his seminal paper, introduced a new entropy based information criterion called Akaike information criterion AIC for the identification and comparison of statistical models among a class of competing models. For a model M_k with k parameters, the unknown parameters are estimated by their maximum likelihood estimators. A measure of fit is the likelihood function $L(\hat{\theta}(k))$. The AIC value is

$$\mathrm{AIC}(M_k) = -2\ln L(\hat{\epsilon}(k)) + 2k \qquad (1)$$

in which the first term is the measure of fit and the second term is a penalty term. The AIC rule is to choose the model with the smallest AIC value.

One problem with this method is that the minimum may not be unique. Another problem can occur if the available models do not contain an adequate model.

In a recent paper by Maklad and Nichols [2] in these Transactions, a model structure discrimination criterion was presented that was designed to cope with these problems. The criterion is based on minimizing the total complexity in a statistical model using both the complexity of the parameter estimators and the complexity of the estimated residuals. The measure of complexity used is the one that was introduced by M. H. van Emden [3].

Van Emden's complexity is a measure of the difference between a whole and the noninteracting composition of its components. For $X' = (X_1, \cdots, X_n)$ a vector of random variables, with covariance matrix V, the complexity of the matrix V is defined in terms of the positive eigenvalues $\lambda_1, \cdots, \lambda_n$ of V by the rule:

$$\varphi_1(V) = \frac{n}{2}\ln(\mathrm{trace}(V)/n) - \frac{1}{2}\ln(\det(V)) = -\frac{1}{2}\ln\left(\prod_{i=1}^{n}(\lambda_i/\bar{\lambda})\right), \qquad (2)$$

with $\bar{\lambda}$ the average of the λ_i's. So $\varphi_1(V) \geqslant 0$ with $\varphi_1(V) = 0$ only when all $\lambda_i = \bar{\lambda}$.

Van Emden also uses another formula for the complexity of V. This formula was incorrectly asserted to coincide, apart from a constant factor, with the first two terms of the Taylor series expansion of $\varphi_1(\cdot)$. It is given by the rule:

$$\varphi(V) = \frac{1}{n}\|V\|^2 - \left(\frac{\mathrm{trace}(V)}{n}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(\lambda_i - \bar{\lambda})^2 \qquad (3)$$

with $\|V\|^2 = \mathrm{trace}(VV')$, the square of the Frobenius norm associated with V. Thus $\varphi(V) \geqslant 0$ with $\varphi(V) = 0$ only when all $\lambda_i = \bar{\lambda}$.

This paper compares the definitions for complexity. Their properties are discussed and the two concepts of complexity are shown to be inherently different and, indeed, nonequivalent. The Maklad–Nichols criterion is derived in the case of a multivariate regression model and a detailed example is explored.

II. Nonequivalence of Complexities

We now compare these two formulations of model complexity and we show that they are nonequivalent. A new criterion $\varphi_2(\cdot)$ is introduced and is shown to be equivalent to $\varphi_1(\cdot)$ and similar in form to $\varphi(\cdot)$.

Theorem: For a positive-definite matrix V, the complexity $\varphi_1(\cdot)$ is scale-invariant with $\varphi_1(cV) = \varphi_1(V)$, $c > 0$. The complexity $\varphi(\cdot)$ is translation-invariant with $\varphi(V + cI) = \varphi(V)$, $V + cI$ positive-definite.

Manuscript received September 6, 1987; revised April 15, 1988 and March 18, 1989.

D. E. Ramirez is with the Department of Mathematics, University of Virginia, Charlottesville, VA 22903-0908.

IEEE Log Number 8928402.

0018-9472/89/0900-1073$01.00 ©1989 IEEE

Proof: Let $\lambda_1,\cdots,\lambda_n$ be the eigenvalues of V. Then

$$\begin{aligned}\varphi_1(cV) &= -\frac{1}{2}\ln\left(\prod_{i=1}^{n} c\lambda_i/c\bar{\lambda}\right)\\ &= -\frac{1}{2}\ln\left(\prod_{i=1}^{n}\lambda_i/\bar{\lambda}\right)\\ &= \varphi_1(V).\end{aligned} \tag{4}$$

Also

$$\begin{aligned}\varphi(cI+V) &= \frac{1}{n}\sum_{i=1}\left((\lambda_i+c)-(\bar{\lambda}+c)\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})^2\\ &= \varphi(V).\end{aligned} \tag{5}$$

Thus given a positive-definite matrix U with eigenvalues $\lambda_1,\cdots,\lambda_n$, let $V=U+cI$ with $c>-\min\{\lambda_1,\cdots,\lambda_n\}$. Then $\varphi(U)=\varphi(V)$. But $\varphi_1(V)=-1/2\ln(\prod_{i=1}^{n}\lambda_i+c/\bar{\lambda}+c)$ has

$$\varphi_1(V)\to 0 \quad \text{as } c\to\infty,$$

and (6)

$$\varphi_1(V)\to\infty \quad \text{as } c\to -\min\{\lambda_1\cdots\lambda_n\}.$$

Similarly, given U, now let $V=cU$ with $c>0$. Then $\varphi_1(U)=\varphi_1(V)$. But $\varphi(V)=1/n\sum_{i=1}^{n}(c\lambda_i-c\bar{\lambda})^2$ has

$$\varphi(V)\to 0 \quad \text{as } c\to 0,$$

and (7)

$$\varphi(V)\to\infty \quad \text{as } c\to\infty.$$

Thus we have established the following result.

Theorem: The complexities $\varphi_1(\cdot)$ and $\varphi(\cdot)$ are nonequivalent.

The proof, given in van Emden [3, p. 62], that the two complexities, $\varphi_1(\cdot)$ and $\varphi(\cdot)$, are second order equivalent, uses an incorrect sign in the expansion of the power series of $\ln(x)$. The correction expansion is given in the following.

Theorem:

$$\varphi_1(V) \doteq \frac{1}{4}\sum_{i=1}^{n}\left(\frac{\lambda_i-\bar{\lambda}}{\bar{\lambda}}\right)^2. \tag{8}$$

Proof: Let $\lambda_1,\cdots,\lambda_n$ be the eigenvalues of V. Using the approximation

$$\ln(x) \doteq (x-1)-\frac{1}{2}(x-1)^2 \tag{9}$$

we have

$$\begin{aligned}\varphi_1(V) &= -\frac{1}{2}\ln\left(\prod_{i=1}^{n}\lambda_i/\bar{\lambda}\right)\\ &\doteq -\frac{1}{2}\sum_{i=1}^{n}\left[(\lambda_i/\bar{\lambda}-1)-\frac{1}{2}(\lambda_i/\bar{\lambda}-1)^2\right]\\ &= -\frac{1}{4}\bar{\lambda}^2\sum_{i=1}^{n}\left(2\bar{\lambda}(\lambda_i-\bar{\lambda})-(\lambda_i-\bar{\lambda})^2\right)\\ &= \frac{1}{4}\bar{\lambda}^2\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})(\lambda_i-3\bar{\lambda})\\ &= \frac{1}{4}\bar{\lambda}^2\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})(\lambda_i-\bar{\lambda}-2\bar{\lambda})\\ &= \frac{1}{4}\bar{\lambda}^2\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})^2-\frac{2\bar{\lambda}}{4\bar{\lambda}^2}\sum_{i=1}^{n}(\lambda_i-\bar{\lambda})\\ &= \frac{1}{4}\sum_{i=1}^{n}(\lambda_i-\bar{\lambda}/\bar{\lambda})^2.\end{aligned} \tag{10}$$

The simplified version of $\varphi_1(V)$ is thus given by

$$\begin{aligned}\varphi_2(V) &= \frac{n}{4}\varphi(V)/(\text{trace}(V)/n)^2\\ &= \frac{n}{4}\cdot\frac{1}{n}\sum_{i=1}^{n}(\lambda_i-\bar{\lambda}/\bar{\lambda})^2.\end{aligned} \tag{11}$$

The complexity of $\varphi_2(V)$ measures the relative variation in the eigenvalues while $\varphi(V)$ measures the absolute variation in the eigenvalues. It has the desirable property of being scale-invariant, and $\varphi_2(V)\geqslant 0$ with $\varphi_2(V)=0$ only when all $\lambda_i=\bar{\lambda}$.

Theorem: The complexity of $\varphi_2(\cdot)$ is scale-invariant.

In the next section, we will join together random variables of different dimensions and types. In this case, it is important to include the additional terms in (11).

III. Decision Rule

Maklad and Nichols [2] proposed using minimum complexity as a performance criterion in model selection. Suppose the parameters are estimated by $\hat{\theta}$ and the residuals by $\hat{\epsilon}|\hat{\theta}$. They propose the rules:

$$\varphi_1(\hat{\theta})+\varphi_1(\hat{\epsilon}|\hat{\theta}) \tag{12}$$

and

$$\varphi(\hat{\theta})+\varphi(\hat{\epsilon}|\hat{\theta}). \tag{13}$$

For computational simplicity, they used (13) in their time series example. However because the measure of the parameter complexity should be scale-invariant, this author has found $\varphi_2(\cdot)$ to be an attractive alternative with

$$\varphi_2(\hat{\epsilon})+\varphi_2(\hat{\epsilon}|\hat{\theta}). \tag{14}$$

The procedure (14) includes the appropriate terms for joining random variables of different dimensions. It has the desirable properties of (1) being scale-invariant, (2) being information-based, (3) being geometric in formulation, and (4) being second-order equivalent to $\varphi_1(\cdot)$.

As Maklad and Nichols [2] have shown, these criteria measure both the accuracy of the estimated parameters and the whiteness of the model residuals.

IV. Kronecker Products

To apply the Maklad–Nichols model selection criterion to multivariate regression models, we first need to derive the relationships between complexity and Kronecker products of the covariance matrices that occur in multivariate analysis. Two important properties that Kronecker products possess are given next and can be found in Searle [4] or Magnus and Neudecker [5].

Lemma: For square matrices U and V with ranks a and b respectively

$$\operatorname{trace}(U\otimes V) = \operatorname{trace}(U)\operatorname{trace}(V),$$
$$\det(U\otimes V) = (\det(U))^b(\det(V))^a. \tag{15}$$

For conformable matrices,

$$(U\otimes V)' = U'\otimes V',$$
$$(A\otimes B)(U\otimes V) = AU\otimes BV. \tag{16}$$

$\varphi_1(\cdot)$ has the following property that is useful with multivariate models.

Theorem: If U and V are positive-definite matrices with ranks a and b respectively, then

$$\varphi_1(U\otimes V) = b\varphi_1(U) + a\varphi_1(V). \tag{17}$$

Proof:

$$\begin{aligned}\varphi_1(U\otimes V) &= \frac{ab}{2}\ln\left[\frac{\operatorname{trace}(U\otimes V)}{ab}\right] - \frac{1}{2}\ln(\det(U\otimes V))\\ &= \frac{ab}{2}\ln[\operatorname{trace}(U)/a\operatorname{trace}(V)/b]\\ &\quad - \frac{1}{2}\ln\left[(\det(U))^b(\det(V))^a\right]\\ &= b\left[\frac{a}{2}\ln(\operatorname{trace}(U)/a) - \frac{1}{2}\ln(\det(U))\right]\\ &\quad + a\left[\frac{b}{2}\ln(\operatorname{trace}(V)/b) - \frac{1}{2}\ln(\det(U))\right]\\ &= b\varphi_1(U) + a\varphi_1(V).\end{aligned} \tag{18}$$

$\varphi(\cdot)$ satisfies the following property that is useful when working with multivariate models.

Theorem: If U and V are positive-definite matrices with ranks a and b respectively, then

$$\varphi(U\otimes V) = \frac{1}{2}\|U\|^2\frac{1}{b}\|V\|^2 - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2. \tag{19}$$

Proof:

$$\begin{aligned}\varphi(U\otimes V) &= \frac{1}{ab}\|U\otimes V\|^2 - (\operatorname{trace}(U\otimes V)/ab)^2\\ &= \frac{1}{ab}\operatorname{trace}((U\otimes V)(U\otimes V)')\\ &\quad - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2\\ &= \frac{1}{ab}\operatorname{trace}((U\otimes V)(U'\otimes V'))\\ &\quad - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2\\ &= \frac{1}{ab}\operatorname{trace}(UU'\otimes VV')\\ &\quad - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2\\ &= \operatorname{trace}(UU')/a\operatorname{trace}(VV')/b\\ &\quad - \left(\frac{\operatorname{trace}(U)}{a}\right)^2\left(\frac{\operatorname{trace}(V)}{b}\right)^2\\ &= \frac{1}{a}\|U\|^2\frac{1}{b}\|V\|^2\\ &\quad - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2.\end{aligned} \tag{20}$$

$\varphi_2(\cdot)$ satisfies the following condition.

Theorem: If U and V are positive-definite matrices with ranks a and b respectively, then

$$\varphi_2(U\otimes V) = \frac{ab}{4}\left(\frac{1}{a}\|U\|^2\frac{1}{b}\|V\|^2 - (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2/ (\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)\right)^2 \tag{21}$$

Proof:

$$\begin{aligned}\varphi_2(U\otimes V) &= ab/4\varphi(U\otimes V)/(\operatorname{trace}(U\otimes V)/ab)^2\\ &= \frac{ab}{4}\left(\frac{1}{a}\|U\|^2\frac{1}{b}\|V\|^2 - \left(\frac{\operatorname{trace}(U)}{a}\right)^2\left(\frac{\operatorname{trace}(V)}{b}\right)^2\right.\\ &\qquad \left./(\operatorname{trace}(U)/a)^2(\operatorname{trace}(V)/b)^2\right).\end{aligned} \tag{22}$$

Thus the Kronecker product of V with an $n\times n$ identity matrix I_n has

$$\varphi_1(V\otimes I_n) = n\varphi_1(V), \tag{23}$$
$$\varphi(V\otimes I_n) = \varphi(V) \tag{24}$$

and

$$\varphi_2(V\otimes I_n) = n\varphi_2(V). \tag{25}$$

V. Multivariate Regression

The Maklad–Nichols decision rule will be explicitly derived in the case of multivariate regression models. Using standard notation, for example from Arnold [6], the model is

$$\underset{(n\times p)}{Y} = \underset{(n\times q)}{X}\ \underset{(q\times p)}{B} + \underset{(n\times p)}{E} \tag{26}$$

with

$$\hat{E} \sim MVN\left(0, \underset{n\times n}{I}, \underset{p\times p}{\Sigma}\right). \tag{27}$$

The parameters are estimated by

$$\hat{B} = (X'X)^{-1}X'Y \tag{28}$$

with

$$\hat{B} \sim MVN\left(B, \underset{q\times q}{(X'X)^{-1}}, \underset{p\times p}{\Sigma}\right). \tag{29}$$

The estimators $\hat{B}$ and $\hat{E}$ as vectors have covariance structures respectively

$$\operatorname{cov}(\hat{B}^v) = (X'X)^{-1}\otimes\Sigma \tag{30}$$

and

$$\operatorname{cov}(\hat{E}^v) = I\otimes\Sigma. \tag{31}$$

Theorem: The decision rule of Maklad and Nichols based on (12) for multivariate regression analysis is

$$\begin{aligned}\varphi_1(\hat{B}, \hat{E}|\hat{B}) = &\frac{pq}{2}\ln\left(\operatorname{trace}(X'X)^{-1}/q\right)\\ &+ p(n+q)/2\ln\left(\operatorname{trace}(\Sigma)/p\right)\\ &-\frac{p}{2}\ln\left(\det(X'X)^{-1}\right)\\ &-(n+q)/2\ln\left(\det(\Sigma)\right).\end{aligned} \tag{32}$$

Proof: Using (17),

$$\begin{aligned}\varphi_1(\hat{B}^v) &= \varphi_1\left(\underset{q\times q}{(X'X)^{-1}}\otimes\underset{p\times p}{\Sigma}\right)\\ &= p\varphi_1\left((X'X)^{-1}\right) + q\varphi_1(\Sigma)\\ &= \frac{pq}{2}\ln\left(\operatorname{trace}(X'X)^{-1}/q\right)\\ &\quad -\frac{p}{2}\ln\left(\det(X'X)^{-1}\right)\\ &\quad +\frac{qp}{2}\ln\left(\operatorname{trace}(\Sigma)/p\right) - \frac{q}{2}\ln\left(\det(\Sigma)\right)\end{aligned} \tag{33}$$

and

$$\begin{aligned}\varphi_1(\hat{E}^v) &= \varphi_1\left(\underset{n\times n}{I}\otimes\underset{p\times p}{\Sigma}\right)\\ &= p\varphi_1(I) + n\varphi_1(\Sigma)\\ &= \frac{np}{2}\ln\left(\operatorname{trace}(\Sigma)/p\right) - \frac{n}{2}\ln\left(\det(\Sigma)\right).\end{aligned} \tag{34}$$

Now combine (33) and (34).

In the case of multiple regression with $p=1$, the covariance matrix of the residuals is diagonal of the form $\sigma^2 I$ and has complexity equal to zero. In this case the decision rule is simply the complexity of the covariance of the estimated parameters, $\varphi_1(\hat{B}) = \varphi_1((X'X)^{-1})$.

This last theorem corrects an error in Bozdogan [7, formula (6.2)] in the derivation of $\varphi_1(\cdot)$ for multivariate regression analysis models. Bozdogan has dubbed the decision rule ICOMP (information complexity).

As an example of the usefulness of the Maklad–Nichols decision rule, we will apply the procedure to the multivariate two-way analysis of variance (Manova) example given in Johnson and Wichern [7, p. 255]. In this example three responses (tear resistance, glass, and opacity) have been observed on 20 samples of plastic film in a two-way fixed-effects model. The two factors are the rate of extrusion (α) and amount of an additive (β). Each factor is measured at two levels (low and high) and the experiment is repeated five times for each of the four cases.

The data matrix Y is 20×3 and the experimental design matrix X is 20×4 with:

$$\underset{(20\times 3)}{Y} = \begin{bmatrix} 6.5 & 9.5 & 4.4\\ 6.2 & 9.9 & 6.4\\ 5.8 & 9.6 & 3.0\\ 6.5 & 9.6 & 4.1\\ 6.5 & 9.2 & 0.8\\ 6.7 & 9.1 & 2.8\\ 6.6 & 9.3 & 4.1\\ 7.2 & 8.3 & 3.8\\ 7.1 & 8.4 & 1.6\\ 6.8 & 8.5 & 3.4\\ 6.9 & 9.1 & 5.7\\ 7.2 & 10.0 & 2.0\\ 6.9 & 9.9 & 3.9\\ 6.1 & 9.5 & 1.9\\ 6.3 & 9.4 & 5.7\\ 7.1 & 9.2 & 8.4\\ 7.0 & 8.8 & 5.2\\ 7.2 & 9.7 & 6.9\\ 7.5 & 10.1 & 2.7\\ 7.6 & 9.2 & 1.9 \end{bmatrix}$$

$$\underset{(20\times 4)}{X} = \begin{bmatrix} 1 & -1 & -1 & 1\\ 1 & -1 & -1 & 1\\ 1 & -1 & -1 & 1\\ 1 & -1 & -1 & 1\\ 1 & -1 & -1 & 1\\ 1 & 1 & -1 & -1\\ 1 & 1 & -1 & -1\\ 1 & 1 & -1 & -1\\ 1 & 1 & -1 & -1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & 1 & -1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{35}$$

The coefficient matrix with γ the interaction term is 4×3 with:

$$\underset{(4\times 3)}{B} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}. \tag{36}$$

The full model or cell-mean model solves (26) with X and B given by (35) and (36), respectively. The additive model reduces X and B by deleting the last column and row, respectively.

For the full model, the estimates are

$$\hat{B} = \begin{pmatrix} 6.78 & 9.31 & 3.93 \\ 0.29 & -0.26 & 0.14 \\ 0.19 & 0.17 & 0.49 \\ 0.00 & 0.16 & 0.44 \end{pmatrix},$$

$$\hat{\Sigma} = \begin{pmatrix} 0.09 & 0.00 & -0.15 \\ 0.00 & 0.13 & -0.03 \\ -0.15 & -0.03 & 3.25 \end{pmatrix} \tag{37}$$

and for the additive model the estimates are

$$\hat{B} = \begin{pmatrix} 6.67 & 9.31 & 3.93 \\ 0.29 & -0.26 & 0.14 \\ 0.19 & 0.17 & 0.49 \end{pmatrix},$$

$$\hat{\Sigma} = \begin{pmatrix} 0.09 & 0.00 & -0.15 \\ 0.00 & 0.16 & 0.05 \\ -0.15 & 0.05 & 3.44 \end{pmatrix}. \tag{38}$$

A typical model selection procedure to determine the structure of the Manova model would compute the AIC's values for all the different models. In this example, the best AIC values, using (1), occur with the full model (AIC $=139.0$) and with the additive model (AIC $=138.0$). Since these values are nearly equal, we can use the Maklad–Nichols rule to break the tie. From the last Theorem and formula (32), using complexity as a performance criterion, we have that $\varphi_1(\hat{B}, \hat{E}|\hat{B})$ is 45.6 for the full model and 43.0 for the additive model; and so the additive model is chosen. Finally, we note that the Maklad–Nichols decision rule favors a balanced orthogonal model since in that case $\varphi_1((X'X)^{-1})$ is zero.

VI. Conclusion

The two concepts of complexity given by van Emden have been shown to be nonequivalent and some of their properties that are useful in model selection have been proved. The Maklad–Nichols decision rule for model selection has been investigated in the case of multivariate regression models and the criterion explicitly derived.

Acknowledgment

The author is grateful to Clark Gaylord who modified the author's generalized linear model program Genlin to compute the model selection criteria.

References

[1] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. on Infor. Theory*, B. N. Petrov and F. Csaki, Eds. Budapest, Hungary: Publishing House of Hungarian Academy of Science, pp. 267–281, 1973.

[2] M. S. Maklad and S. T. Nichols, "A new approach to model structure discrimination," *IEEE, Trans. Syst. Man Cybern.*, vol. SMC-10, no. 2, pp. 78–84, Feb. 1980.

[3] M. H. van Emden, "An analysis of complexity," *Mathematical Centre Tracts*, vol. 35. Amsterdam: Mathematical Centrum, 1971.

[4] S. Searle, *Matrix Algebra Useful for Statistics*. New York: Wiley, 1982.

[5] J. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley, 1988.

[6] S. Arnold, *The Theory of Linear Models and Multivariate Analysis*. New York: Wiley, 1981.

[7] H. Bozdogan, "ICOMP: A new model-selection criterion," *Classification and Related Methods and Data Analysis: Proc. First Conf. Inter. Fed. Classification Societies*, Hans H. Bock, Ed. Amsterdam: North-Holland, Spring 1988.

[8] R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*. Englewood Cliffs, N.J.: Prentice Hall, 1988.

Donald E. Ramirez was born in New Orleans, LA on May 21, 1943. He received the B.Sc. and Ph.D. degrees in mathematics from Tulane University in 1963 and 1966, respectively.

During 1966–1967, he was an Office of Naval Research Associate at the University of Washington, Seattle. From 1967–1971, he was an Assistant Professor at the University of Virginia, Charlottesville, and, currently, he is an Associate Professor of Mathematics. During 1975–1988, he was an Editor of the Semigroup Forum. He is a co-author of two books, *Topics in Harmonic Analysis* and *Representations of Commutative Semitopological Semigroups*. From 1983–1984, he served as Director of the Statistics Forum at the University of Virginia. He has published over 45 research articles. His research interests have included approximation of Fourier-Stieltjes transforms and weakly almost periodic functions. His most recent work has been in mathematical statistics and in selection criteria in factor analysis modeling.