GEORGE S. DONATOS - GEORGE C. MICHAILIDIS

AN ALTERNATIVE APPROACH TO THE KALMAN'S SCHEME OF IDENTIFICATION IN THE PRESENCE OF MULTICOLLINEARITY

PLAN

ABSTRACT

In this paper we present an alternative approach to the problem of multicollinearity stemming from Kalman's system theoretic approach on identification from noisy data, and link it to the most frequently used in practice, when dealing with multicollinear data, procedure of ridge regression. Especially we describe Kalman's scheme of identification and discuss its use in the presence of multicollinearity and its association with ridge regression.

Consider the general linear regression model

$$y = X\beta + \varepsilon \tag{1}$$

where y is a nx1 vector of observations, X a nxc design matrix, β a cx1 vector of parameters to be estimated and ε a nx1 vector of disturbances with the following properties: $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I_n$.

The LS estimator is defined as

$$\hat{\beta} = (X'X)^{-1} X'y \tag{2}$$

If the columns of X are highly correlated, $X'X$ becomes "nearly singular"; i.e. $X'X$ has a wide eigenvalue spectrum. The consequences of such a situation are that the LS estimates of β, are "unstable" in the sense that a small perturbation of the design matrix X will produce large changes in $\hat{\beta}$, and in many cases present incorrect signs.

Different approaches are suggested in the literature to remedy the problem of multicollinearity[1]. However, most of them either require extensive prior information regarding the β vector and larger samples that are very unlikely to be available to the data analyst or require him to reduce his informational demands by considering only a subspace of the parameter space. The only technique that deals with the problem of multicollinearity by considering the original data set is ridge regression.

The technique of ridge regression, first proposed by Hoerl and Kennard (1970) which has become a very popular tool among data analysts faced with a high degree of multicollinearity in their data. By using a ridge estimator, one hopes to both stabilize one's estimates (lower the condition number of the design matrix) and improve upon the squared error loss (mean square error (MSE)) of the least squares (LS) estimator[2]. Both the stabilization of the estimates of β and the dominance in MSE of the LS estimator gave rise to algorithms in computing the biasing parameter k.

The first approach focuses on the numerical properties of the ridge regression estimator by attempting to bring the system closer to an orthogonal one, while the second approach emphasizes the statistical properties of the ridge estimator and is closely linked to the James-Stein (1961) estimator. The first approach offered the ridge trace algorithm suggested by Hoerl and Kennard (1970), which consists of plotting the elements of the vector $\hat{\beta}(k)$ against

[1] See Maddala (1978) Chapter 10.
[2] See Appendix.

k to help choosing k in a "stable region". This algorithm has been extensively used in empirical work[3]. The second approach provided several ridge algorithms that can be classified in the following categories: i) algorithms that minimize the MSE function, such as those proposed by Goldstein and Smith (1974), Hoerl, Kennard and Baldwin (1975), Obenchain (1975), and Hoerl and Kennard (1976). ii) algorithms that have a Bayesian interpretation by assuming a prior normal distribution centered at zero for the unknown parameter vector β, such as those suggested by Lindley and Smith (1972) with a hierarchical Bayesian procedure, Lawless and Wang (1976) and Dempster, Schatzoff and Wermuth (1977) following an empirical Bayesian procedure. iii) algorithms that minimize the residual sum of squares subject to a constraint on the length of the coefficient vector β, such as the one proposed by McDonald and Galarneau (1975) and Gunst and Mason (1977).

Both approaches give rise to stochastic estimators, when k is also a function of the random vector y and not only of the nonstochastic design matrix X. To see this, notice that the ridge trace depends on y as pointed out by Coniffe and Stone (1973) who observed that the "stable region" of observed $\hat{\beta}_i(k)$ is stochastic itself. Moreover, regarding the second class of ridge estimators notice that the range of k values for which the ridge estimator dominates the LS estimator in MSE depends on the unknown β as well as σ^2, the k resulting from Bayesian procedures is the ratio of the sample variance to the variance of the prior distribution and the k stemming from constrained least squares is subject to a somehow arbitrary constraint. In order to make these algorithms operational it is necessary to use estimates of β and σ^2, which in turn depend on y and thus make the resulting estimators stochastic. A major drawback of stochastic estimators is that it is impossible to determine analytically their statistical properties, and therefore one must resort to Monte Carlo studies[4].

However, all the existing solutions to the problem of multicollinearity take for granted the existence of only one linear relationship in the data, i.e. the one between y and the X's. For noise[5] free data this approach is correct since the design matrix X is of full column rank. If, on the other hand, the variables in X are also subject to noise contamination the assumption of existence of a single linear relationship within the data may be incorrect and in this case the problem will not be the one of multicollinearity and how to get stable estimates, but rather how to detect the number of linear relationships admitted by the data and how to reexamine the modelling process. Following this line of inquiry we

[3] See for example Watson and White (1976), Mahajan, Jain and Bergier (1977), Kvalseth (1979), Lee (1980), Erickson (1981).

[4] Most of the references provided above in the identification of ridge algorithms contain extensive Monte Carlo simulations.

[5] We interpret noise in a broad sense, i.e. not only as round-off errors and aggregation errors, but even as any causal or random factor affecting the realisation of the variables in the design matrix X which can not be modelled.

need to identify whether there exists a single or more than one linear relationships in the data set under examination. The system theory literature provides us with an exact mathematical result to deal with our problem, although the motivation for this result comes from a different line of research and is presented next.

2. KALMAN'S APPROACH TO IDENTIFICATION

Kalman's approach to identification from real data influenced by system theory deals with the following issue. Consider a (sample or population) covariance matrix Σ for c variables $x_1, x_2, \ldots, x_c$. In case these variables are noise free[6] they are linked by m linear relations if and only if there exists a matrix B of rank m such that

$$B\Sigma = 0 \tag{3}$$

Notice that in order to solve the indeterminancy in (3) the matrix B has to be normalized in one of the following ways: either make any arbitrary element (usually the first one) of every column of B equal to 1 or make the length of each column of B equal to 1.

In case each variable x_i, $i = 1, \ldots, c$ has some unknown amount of noise in it we can write

$$x_i = \hat{x}_i + v_i, \quad i = 1, \ldots, c \tag{4}$$

where hat denotes the exact values and v the noise. It is assumed that the noise variables are zero-mean, independent of one another as well as of the $\hat{x}_i$'s. It is also assument that the linear relations link the noise free part of the variables. The latter assumption gives

$$B\hat{\Sigma} = 0 \tag{5}$$

while the former together with (4) give

$$\Sigma = \hat{\Sigma} + V \tag{6}$$

[6] As usually assumed in the linear regression setting for the variables in the design matrix X.

where $V > 0$ [7] and diagonal, and $\Sigma > V > 0$. As Kalman (1982a) observed V has to be diagonal in order to be consistent with the hypothesis that noise variables should not be modelable.

Given $\Sigma > 0$, since the x_i's are well defined we have a well posed mathematical problem, namely: given Σ find $B, \hat{\Sigma}$ and V such that $B\hat{\Sigma} = 0$, $\hat{\Sigma} > 0$ and symmetric and $V > 0$ and diagonal. Kalman (1982a) shows the existence of such a solution (p. 175), although it might not be unique. However, this problem admits something unique, an invariant, namely the maximum rank of B. Using the following definition we get:

Definition: The corank of a square nxn matrix A is defined as

$$corank(A) = n - rank(A) \tag{7}$$

Then,

$$maxrank(B) = maxcorank(\hat{\Sigma}) = maxcorank(\Sigma) = m \tag{8}$$

Using (8) the problem of studying multicollinearity reduces to the study of the maximally singular matrix Σ. Kalman (1982b) proves that $m = 1$ (p. 149), namely the data admit only one linear relationship, if and only if Σ^{-1} is inverse positive, which means that all entries of Σ^{-1} are strictly positive or this can be achieved by a suitable choice of sign changes in the definition of the variables $x_1, x_2, ..., x_c$. We give next a slightly different proof of this Theorem.

Theorem: $m = 1$, if and only if, Σ^{-1} has strictly positive elements, possibly after sign changes of rows and corresponding columns.

Proof: i) In order (5) to hold, $\hat{\Sigma}\beta = 0$ must hold, or equivalently by (6) $(I - \Sigma^{-1} V)\beta = 0$. It follows then that β is an eigenvector of $\Sigma^{-1} V$ with corresponding eigenvalue equal to one. By Frobenius Theorem (Gantmacher (1959), p. 47) for nonegative matrices, we get that β is strictly positive, since ex hypothesis Σ^{-1} is stricity positive and V is a covariance matrix. So, β lies in the cone of Σ^{-1} and so does by (6) the null space of $\hat{\Sigma}$. Hence, all elements of this null space lie in the strictly positive or negative orthnat. Consequently, this null space can be at most one dimensional, and thus $m = 1$.

ii) If not all elements of Σ^{-1} have compatible signs, then there exist two columns i and j that they do not lie in the same orthant. So, in the cone of these two columns there is a vector β with a zero element. But, by Hakim et al (1976, p. 14, Corollaire 2.4) this implies that $m \geq 2$ which contradicts the hypothesis $m = 1$.

[7] The symbol " >0 " means a positive definite square matrix.

We prove next some sufficient conditions for Σ^{-1} to be inverse positive in the general linear regression setting. Define the multivariable $Z = [y \mid X]$ and Σ be the covariance matrix of Z.

Lemma 1: Any nonsingular covariance matrix Σ with all nonpositive cross covariance terms is inverse positive.

Proof: Corollary 2 (page 85) in Varga (1962) states that: if A is a real, symmetric and nonsingular nxn irreducible matrix, where $a_{i,j}) \leq 0$ for all $i \neq j$, the $A^{-1} > 0$ (inverse positive) if and only if A is positive definite. It can be easily then seen that since Σ is a real covariance matrix it is automatically symmetric, nonsingular and positive definite. Moreover, ex hypothesis all its off-diagonal terms are nonpositive. Thus, the result follows trivially from Varga's Corollary.

Lemma 2: If the columns of X are orthogonal, then Σ is inverse positive.

Proof: Because all the columns of X are orthogonal the submatrix Σ_x is diagonal with positive elements. Moreover, all the cross covariances of y with the columns of X can be made negative with an appropriate variable sign change. As long as Σ is positive definite, using Lemma 1 we get the result.

Notice that Lemma 2 gives a formal justification for using linear regression (only one linear relationship between y and X) when the columns of X are orthogonal (as assumed in the linear regression context but rarely satisfied in practice).

It remains to show the connection between ridge regression and the invariant corank of Σ, $m(\Sigma)$. Consider the mapping

$$k \rightarrow m(\Sigma + k I_c) \tag{9}$$

for Σ being a constant, symmetric, positive definite covariance matrix and k varying on $(0, \infty)$. Since $m(I_c) = c$ (the identity matrix admits c linear relations), it is obvious that

$$lim_{k \to \infty} m(\Sigma + kI_c) = c + 1 \tag{10}$$

and hence intuitively clear that $m(\Sigma + k I_c)$ is a monotonic function of k. Adding k to the diagonal elements of $X'X$ in ridge regression is like adding noise in Σ_x and hence in Σ. Then if Σ is near the boundary where m changes from 1 to 2 a small value of k is enough to bring the change about. But if $m(\Sigma + k I_c) = 2$ we should abandon the prejudice of looking for a single linear relationship in the data and should either search for a simultaneous equations model where we would model the relationship between our x variables, or drop some of the x's. Moreover, this approach provides a way of checking the admissibility of k obtained by a ridge algorithm. Let k^* be the k that makes $m.(\Sigma + k I_c) = 2$. Then, the optimal k in any ridge algorithm should lie in the interval $(0, k^*)$ in order to be meaningful. The problem arises because all the existing ridge algorithms are derived in order to remedy the statistical problems of the LS estimator and they do not take into consideration the nature of the linear relations embedded in the data.

3. APPLICATIONS

We illustrate our point regarding multicollinearity and ridge regression with two examples from the literature. In the two applications we discusse a heuristic explanation of the merits of Kalman's approach to identification when dealing with multicollinear data and when applying the technique of ridge regression by using both the Σ^{-1} and the AR matrices. The first application appears in Neter, Wasserman and Kutner (1985)[8]. The data covariance matrix Σ is given next, with an appropriate sign change for X_1 along with Σ^{-1}, which is inverse positive.

$$\Sigma = \begin{pmatrix} 25.2331 & -24.2923 & -8.3867 & -21.6295 \\ -24.2923 & 27.4012 & 1.6164 & 23.4704 \\ -8.3867 & 1.6164 & 13.3017 & 2.6527 \\ -21.6295 & 23.4704 & 2.6527 & 26.0731 \end{pmatrix}$$

$$\Sigma^{-1} = \begin{pmatrix} 31.7148 & 26.4194 & 16.6189 & 0.8367 \\ 26.4194 & 22.1687 & 13.8535 & 0.5515 \\ 16.6189 & 13.8535 & 8.7858 & 0.4220 \\ 0.8367 & 0.5515 & 0.4220 & 0.1931 \end{pmatrix}$$

The columns of the two matrices correspond to the variables tricepts skinfold thickness, thigh circumferences, midarm circumferences (expanatory variables) and body fat (dependent variable). Normalizing Σ^{-1}, or in other words compute all direct and reverse regressions we get

$$AR = \begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ 0.8330 & 0.8391 & 0.8336 & 0.6591 \\ 0.5240 & 0.5244 & 0.5287 & 0.5044 \\ 0.0264 & 0.0209 & 0.0254 & 0.2308 \end{pmatrix}$$

[8] Table 11.2, p. 385.

which indicates that the solution space is a very thin tetraedron, and with some noise it will become a triedron and thus one parameter will not be identifiable simply by a linear model. Adding $k = .0978$ along the diagonal elements of Σ_x we get

$$\Sigma^{-1} = \begin{pmatrix} 4.4898 & 3.6802 & 2.3316 & 0.1745 \\ 3.6802 & 3.1746 & 1.9205 & -0.0001 \\ 2.3316 & 1.9205 & 1.2876 & 0.0744 \\ 0.1745 & -0.0001 & 0.0744 & 0.1757 \end{pmatrix}$$

which is not inverse positive. However, the ridge solution suggested in Neter, Wasserman and Kutner (1985) based on the examination of the ridge trace and the variance inflation factor gives a value for $k = .02$[9], which lies within the interval (0, 0.0978) and therefore is an admissible biasing parameter. With $k = .02$ the condition number of the inverse of $X'X$ improves by 56 percent (from 1085.78 to 483.11) and there is a small improvement in the MSE as well. In this application we are in an "optimal situation". The data admit just one linear relationship (or in other words we face the classical problem of multicollinearity) and the ridge solution improves the stability of the estimated coefficients and lies at the same time within the admissible region.

In the next application the data covariance matrix is not inverse positive and we show how to practically detect additional linear relationships[10]. As previously stated, the normalized columns of Σ^{-1} give all the regression vectors (AR) of Σ. It can be shown that the i^{th} AR is the LS estimate of the coefficients when the i^{th} variable is assumed noisy and all others noise free, or to put it differently the i^{th} AR is the normalized eigenvector corresponding to the zero eigenvalue of

$$V = \Sigma - D \qquad (11)$$

where $D = diag\,(0, \ldots, d^i, \ldots,)$ and d^i is such that V is singular and positive semidefinite. Kalman (1982a) shows that d^i is the reciprocal of the i^{th} diagonal element of Σ^{-1} and can be interpreted as an upper bound on the noise variance of the i^{th} variable.

When Σ^{-1} is inverse positive, the upper and lower bounds of every coefficient will have the same sign, and the solution space will be bounded, otherwise it is not. However, we can use this information in detecting the additional

[9] We also computed the Lawless and Wang (1976) and the Hoerl, Kennard and Baldwin (1975) algorithms which give values for k within the admissible range.

[10] Remember that they do not exist precise mathematical results for detecting the exact number of linear relationships in an arbitrary data covariance matrix.

linear relations existing in the data. This can be best demonstrated using an application given in Tinter (1946).

The data covariance matrix Σ is given next

$$\Sigma = \begin{pmatrix} 1.2126 & 0.5362 & 0.0876 & -0.0727 & 0.5320 \\ 0.5362 & 0.5576 & 0.2071 & 0.0750 & 0.2200 \\ 0.0876 & 0.2071 & 0.1064 & 0.0545 & 0.0303 \\ -0.0727 & 0.0750 & 0.0545 & 0.0500 & -0.0385 \\ 0.5320 & 0.2200 & 0.0303 & -0.0385 & 0.2407 \end{pmatrix} x\ 10^3$$

The variables are prices paid to farmers, national income, agricultural production, a time trend and prices paid to farmers respectively. The upper bounds on the noise variances are

$$(27.7682 \quad 32.8249 \quad 10.3431 \quad 0 \quad 6.2635)$$

where d^4 is set to zero because the third variable is a time trend and is assumed to be noise free.

The AR matrix is given by

$$AR = \begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ -0.3451 & -2.4516 & -3.6422 & -0.0200 \\ 0.2994 & 3.1601 & 8.9674 & 0.0837 \\ 0.1770 & 1.6003 & -3.2736 & -0.4019 \\ -1.9044 & -0.1101 & -0.5323 & -2.3279 \end{pmatrix}$$

The third row is the only one presenting a sign descrepancy. If we manage to find a diagonal noise covariance matrix such that the (4, 2) (or the (4, 1)) element could be made zero, then this would reveal an additional linear relation among the first three and the last variable. Consider the following diagonal matrix $G = diag\ (0, 0, 0, 6.19)$. Notice that $det\ (\Sigma_{24} - D) = 0$, where Σ_{24} is the cofactor of the (4, 2) element of Σ. Moreover notice that $g_4 = 6.19$ is less than $d_5 = 6.26$, and hence it provides a feasible solution. Computing the AR matrix of the $\Sigma - S'$ matrix, where $S' = diag\,(0, 0, 0, 0, 6.19)$ yields

$$AR = \begin{pmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 \\ -0.0246 & -0.5081 & -0.1989 & -0.0200 \\ 0.0868 & 0.7012 & 0.5224 & 0.0837 \\ -0.3936 & 0.0000 & -0.5437 & -0.4019 \\ -2.3218 & -1.8827 & -2.2392 & -2.3279 \end{pmatrix}$$

which does not present any sign discrepancy among its rows and reveals a linear relation between the first three and the last variables. We can now use this information in deciding how to model or more generally how to approach the question we study with data set in hand.

4. CONCLUDING REMARKS

When we have noisy data (like in most empirical applications) the uncertainty in the data will be inherited by the model. Most of the times it is covered behind some ad hoc modeling mechanism but it is not eliminated. The situation becomes more difficult in the presence of multicollinearity, since the degree of the uncertainty increases by the presence of underlying linear relationships in the design matrix X (see the second example). A widely used procedure to deal with this problem is ridge regression with interesting numerical and statistical features. However, as we showed with the example above it solves problems up to a point, but once the data with the addition of some noise admit more than one linear relationships it is pure prejudice to continue using the linear model and try to improve the solution by the use of ridge regression. The real solution lies in the better understanding of our data, and the questions we expect to answer with the particular data set in hand. Ridge regression offers in many situations a great potential but it is not a panacea and certainly can not substitute the modelling process.

APPENDIX

To correct the problem of the ill-conditioning of $X'X$, Hoerl and Kennard (1970) proposed the ridge estimator

$$\hat{\beta}(k) = (X'X + kI_d)^{-1} X'y, \quad k > 0 \tag{12}$$

Adding the constant k to the diagonal elements of $X'X$ before inverting, accounts to increasing each eigenvalue of $X'X$ by k, since the ridge estimator can be rewritten as:

$$\hat{\beta}(k) = (C'(\Lambda + kI_d)C)^{-1} X'y, \qquad C'X'XC = \Lambda, \quad C'C = I_c \tag{13}$$

where $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_d)$, λ_i, $i = 1, ...c$ being the eigenvalues of $X'X$, and C be the matrix of the corresponding orthonormal eigenvectors.

To see that the ridge estimator is more stable than $\hat{\beta}$, we note that the condition number of the matrix being inverted in (11) is decreased. The condition number of a matrix is a measure of ill–conditioning and is defined as

$$CN(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \tag{14}$$

where $\lambda_{max}(A)$ and $\lambda_{min}(A)$ denote the largest and smallest eigenvalues of the matrix A respectively. Large values of $CN(A)$ imply that A is ill–conditioned. Since,

$$\frac{\lambda_1 + k}{\lambda_c + k} \langle \frac{\lambda_1}{\lambda_c} \tag{15}$$

for $k > 0$, the ridge estimator in (11) is relieving the ill–conditioning problem of $X'X$.

The other important property of the ridge estimator is the "ridge existence theorem". This theorem asserts that for a fixed parameter vector β_o, there exists a value k depending on β_o, such that the MSE of $\hat{\beta}(k)$ is smaller than the MSE of $\hat{\beta}$ (see Theorem 4.3, p. 62, Hoerl and Kennard (1970)).

REFERENCES

Coniffe D. and J. Stone (1973), «A Critical View of Ridge Regression», *The Statistician*, 22, 181-187.

Dempster A. P., Schatzoff M., and N. Wermuth (1977), «A Simulation Study of Alternatives to Ordinary Least Squares», *Journal of the American Statistical Association*, 72, 77-106.

Erickson G. M. (1981), «Using Ridge Regression to Estimate Directly Lagged Effects in Marketing», *Journal of the American Statistical Association*, 76, 766-773.

Gantmacher F. R. (1959), «Matrizenrechnung, Teil II», *VEB, Deutscher Verlag der Wissenschaften*, Berlin.

Goldstein M., and A. F. M. Smith (1974), «Ridge Type Estimators for Regression Analysis», *Journal of the Royal Statistical Society*, Series B, 36, 284-291.

Gunst R. F., and R. L. Mason (1977), «Biased Estimation in Regression: An Evaluation Using Mean Squared Error», *Journal of the American Statistical Association*, 72, 616-628.

Hakim M., Lochard E. O., Olivier J. P., and E. Terouanne (1976), «Sur les Traces de Spearman», *Cahiers de Bureau de Recherche Operationelle*, Université Pierre et Marie Curie, Paris.

Hoerl A. E., and R. W. Kennard (1970), «Ridge Regression: Biased Estimation for Nonorthogonal Prblems», *Technometrics*, 12, 56-67.

Hoerl A. E., Kennard R. W., and K. F. Baldwin (1975), «Ridge Regression: Some Simulations», *Communications in Statistics, Theory and Methods*, 4, 105-123.

Hoerl A. E., and R. W. Kennard (1976), «Ridge Regression Iterative Estimation of the Biasing Parameter», *Communications in Statistics, Theory and Methods*, 5, 77-88.

James W., and C. Stein (1961), «Estimation with Quadratic Loss», *Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability*, 1, 361-379.

Kalman R. E. (1982a), «Identification from Real Data», *Current Developments in the Interface: Economics, Econometrics, Mathematics*, eds. M. Hazelwinkel and A. H. G. Rinnooy Kan, D. Reided, Dordrecht, 161-196.

Kalman R. E. (1982b), «System Identification from Noisy Data», *Dynamical Systems II*, eds. A. Bednarek and L. Cesari, Academic Press, New York, 135-164.

Kvalseth T. O. (1979), «Ridge Regression Models of Urban Crime», *Regional Science and Urban Economics*, 9, 247-260.

Lawless J. F., and P. Wang (1976), «A Simulation Study of Ridge and Other Regression Estimators», *Communications in Statistics, Theory and Methods*, 5, 307-323.

Lee J. H. (1980), «Factor Relationship in Postwar Japanese Agriculture: Application of Ridge Regression to the Translog Production Function», *Economic Studies Quarterly*, 31, 33-44.

Lindley D. V., and A. F. M. Smith (1972), «Bayes Estimates for the Linear Model», *Journal of the Royal Statistical Society*, Series B, 34, 1-41.

Maddala G. S. (1977), *Econometrics*, McGraw-Hill, New York, NY.

Mahajan V., Jain A. K., and M. Bergier (1977), «Parameter Estimation in Marketing Models in the Presence of Multicollinearity», *Journal of Marketing Research*, 14, 1586-1591.

McDonald G. C., and D. I. Galarneau (1975), «A Monte Carlo Evaluation of Some Ridge Type Estimators», *Journal of the American Statistical Association*, 70, 407-416.

Neter J., Wasserman W., and M. H. Kutner (1985), *Applied Linear Statistical Models*, 2nd ed., Irwin, Homewood, Il.

Obenchain R. L. (1975), «Ridge Analysis Following a Preliminary Test of the Shrunken Hypothesis», *Technometrics*, 17, 431-441.

Tinter G. (1946), «Multiple Regression for Systems of Equattions», *Econometrica*, 14, 5-36.

Varga R. S. (1962), *Iterative Matrix Analysis*, Prentice Hall, Englewood Cliffs, N. J.

Watson D. E., and K. J. White (1976), «Forecasting the Demand for Money under Changing Term Structure of Interest Rates: An Application of Ridge Regression», 43, 1096-1105.