SINET: Ethiop. J. Sci., 21(2):273–277, 1998
 ISSN: 0379–2897

Short communication

A GENERALIZED DETERMINISTICALLY SHRUNKEN ESTIMATOR

Eshetu Wencheko

Department of Statistics, Faculty of Science,
Addis Ababa University, PO Box 1176, Addis Ababa, Ethiopia

ABSTRACT: In this paper we introduce a shrinkage-strategy by means of which the variance inflation factors of the ordinary least squares estimator of the regression coefficient vector obtained under one type of data ill-conditioning, namely multicollinearity, can be controlled. This is attained through a transformation of the least squares estimator. In doing so the transformed estimator is a linear biased estimator, the so-called deterministic shrunken estimator. The estimator thus obtained is an improvement of the least squares estimator as it guarantees reduction in magnitudes of the variance inflation factors. The result could be useful in econometric applications, especially in forecasting.

Key words/phrases: Multicollinearity, scalar risk, shrunken estimator, variance inflation factors

INTRODUCTION

Consider the multiple linear regression model given by

$$y = X\beta + \epsilon \quad E(\epsilon) = 0 \quad Cov(\epsilon) = \sigma^2 I_n, \tag{1}$$

where $y = (y_1, \ldots, y_n)'$ is a vector of observations, $\beta = (\beta_1, \ldots, \beta_p)'$ is an unknown fixed vector of regression coefficients, $X = (x_{ij})$ is a nxp non-stochastic regressor matrix of full column rank, $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ is a random vector of error terms with a common unknown variance $\sigma^2 > 0$.

It is known that, although unbiased, the ordinary least squares estimator (LS) of the coefficient vector β given by $b = (X'X)^{-1}X'y$ is not reliable in many regards if two or more regressors are strongly dependent on one another.

Alternatives to the LS provide the so called class of generalized ridge estimators (Obenchain, 1975; 1978) which consists of biased linear estimators of β. Members of this class of estimators enjoy superiority in performance over the LS on sub-spaces of the domain of definition of the regression parameters. Such dominance can be expressed by way of comparison of risks; usually using the squared scalar or matrix-valued mean square error. A disadvantage, however, is that the theoretical comparison-statements depend on unknown biasing quantities as well as β and/or σ^2.

Another criterion for comparing linear biased estimators of β with the LS is by way of component-wise comparison of magnitudes of the variance inflation factors (VIF). The usefulness of VIF's was advocated, among others, by Marquardt (1970), Marquardt and Snee (1975), Montgomery and Askin (1981) and Trenkler (1981).

Variance inflation factors are the diagonal elements of the covariance matrix of a linear estimator of the coefficient vector. On this issue Marquardt (1970) states: "A rule of thumb for choosing the amount of bias to allow with ill conditioned data, ..., is that the maximum variance inflation factor usually should be larger than 1 but not certainly as large as 10".

The aim of this paper is to use an affine transformation of the LS computed for multi-collinear data such that the resulting shrunken estimator can be shown to be superior to the LS in terms of VIF-comparison. The next section is devoted to such discussion.

THE DETERMINISTICALLY SHRUNKEN ESTIMATOR

The class of generalized ridge estimators following Obenchain is given as

$$M: = \{ Ly: L = Z\,(X'X)^{-1}X'\,, \quad Z: = P\Delta P' \} \tag{2}$$

where $\Delta: = diag(\delta_1, ..., \delta_p),\ \delta_j \in [0,\ 1],$

and the matrix P results from the spectral decomposition $X'X = P\Lambda P'$. Further, Λ is the diagonal matrix of eigenvalues $\lambda_j > 0$, $j = 1, 2, \ldots, p$ of $X'X$ with the corresponding matrix P of ortho-normal eigenvectors and $Z = XP$.

A member of this class that we consider in this paper is a generalization of the shrunken estimator (Mayer and Willke, 1973; Stein, 1960), which shall henceforth be referred to as the generalized deterministic shrunken estimator.

This estimator has the algebraic form

$$b(C) = Cb,\ C = \mathrm{diag}(c_1, \ldots, c_p),\ c_j \in (0, 1], \tag{3}$$

implying that it belongs to M with $\Delta = C$.

The derivation of the moments of stochastic shrunken estimators cannot be given in closed forms as these depend on the stochastic expressions that are being used as shrinkage factors. The analytic study of such estimators is also very difficult. For these and other reasons the practical application and utilization of such estimators is not recommendable. Therefore, our concern here is only about the case of the practically useful deterministic shrunken estimator; meaning the case where the shrinkage matrix C above is deterministic.

It is known that the deterministic shrunken estimator has bias vector and covariance matrix

$$B[b(C)] := (C - I_p)\beta \tag{4}$$

$$Cov[b(C)] := \sigma^2 C(X'X)^{-1} C'. \tag{5}$$

The total variance of b(C), that is the sum total of the diagonal elements of its covariance matrix, is

$$V[b(C)] = \sigma^2 tr C(X'X)^{-1} C'. \tag{6}$$

The scalar risk of b(C) is

$$R[b(C)] = V[b(C)] + tr\,\beta(C-I_p)^2\beta'. \tag{7}$$

In what follows the bias vector (4), the covariance matrix (5) and the total variance (6) above, shall be used in developing a strategy to obtain different sets of VIF's of b(C).

The procedure we follow in order to arrive at the elements of C is based on the idea of reducing only those VIF's with values larger than 10 to ones that lie within the interval suggested by Marquardt (1970), leaving the others unchanged. Hence the strategy used to get the shrinkage matrix C relies upon setting

$$trC(X'X)^{-1}C' = dp\ ,\quad d \in [1,\ 10]$$

out of which the elements of C can be chosen such that

$$c_j = \left(\frac{m^{jj}}{d}\right)^{\frac{-1}{2}},\quad j=1,2,\ldots,p$$

where m^{jj} is the j^{th} diagonal element of $(X'X)^{-1}$.

Once we agree on the choice of c_j's we establish a theoretical condition under which the total variance of the LS

$$V(b) = \sigma^2 tr(X'X)^{-1} \tag{8}$$

can be compared with the scalar risk of b(C).

The scalar risk R[b(C)] for the above choice of c_j's is given as

$$\beta'diag\left[\left(\left(\frac{m^{jj}}{d}\right)^{\frac{-1}{2}}-1\right)^2\right]\beta + p\sigma^2. \tag{9}$$

We formulate the theoretical result of the comparison as follows:

ASSERTION ABOUT RISK COMPARISON

The performance in risk comparison of the LS b and generalized deterministic shrunken estimator b(C) with the aforementioned choice of shrinkage matrix C can be expressed as follows:

$$R[b(C)] < V(b) \leftrightarrow \beta' diag\left[((\frac{m^{jj}}{d})^{\frac{-1}{2}} - 1)^2\right]\beta/\sigma^2 < -p + tr(X'X)^{-1}. \quad (10)$$

REFERENCES

1. Marquardt, D.W. (1970). Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation. *Technometrics* **12**:591–612.
2. Marquardt, D.W. and Snee, R.D. (1975). Ridge regression in practice. *The American Statistician* **29**:3–20.
3. Mayer, L.S. and Willke, T.A. (1973). On biased linear estimation in linear models. *Technometrics* **15**:497–508.
4. Montgomery, D.C. and Askin, R.G. (1981). Problems of non-normality and multicollinearity for forecasting methods based on least squares. *American Institute of Industrial Engineers Transactions* **13**:102–114.
5. Obenchain, R.L. (1975). Ridge analysis following a preliminary test of the shrunken hypothesis. *Technometrics* **17**:431–445.
6. Obenchain, R.L. (1978). Good and optimal ridge estimators. *Annals of Statistics* **6**:1111–1121.
7. Stein, C.M. (1960). Multiple regression: Contributions to probability and statistics. In: *Essays in honour of Harold Hotelling*, pp. 424–443, Stanford University Press, Stanford.
8. Trenkler, G. (1981). *Biased Estimators in the Linear Regression Model.* Oelgeschlager, Gunn and Hain, Cambridge, Massachusetts.