Discussiones Mathematicae
Probability and Statistics 37 (2017) 157–163
doi:10.7151/dmps.1189

ON NONNEGATIVE MINIMUM BIASED ESTIMATION IN THE LINEAR REGRESSION MODELS

Mariusz Grządziel

Department of Mathematics
Wrocław University of Environmental and Life Sciences
Grunwaldzka 53, 50–357 Wrocław, Poland

e-mail: mariusz.grzadziel@upwr.edu.pl

Abstract

The problem of the nonnegative estimation of the parametric function $\gamma(\beta,\sigma) = \beta' H\beta + h\sigma^2$ in the linear regression model $\mathcal{M}\{y, X\beta, \sigma^2 I\}$, where H is a nonnegative definite matrix and h is a nonnegative scalar, attracted attention of many researchers. S. Gnot, G. Trenkler and R. Zmyślony [J. Multivar. Anal. 54 (1995), 113–125] proposed an approach in which γ is estimated by a quadratic form $y'Ay$, where A is a nonnegative definite matrix that satisfies an appropriate optimality criterion associated with minimizing the bias of the estimator.

In the paper, we revisit this approach to estimating γ. In particular, we discuss various methods of computing the matrix A, which in the general case is not given explicitly.

Keywords: linear regression model, nonnegative minimum biased estimators, mean squared error.

2010 Mathematics Subject Classification: 62J05, 62H12.

The paper is dedicated to Professor Roman Zmyślony on His 70-th birthday

1. Introduction

Let us consider the linear regression model $\mathcal{M}\{y, X\beta, \sigma^2 I_n\}$, where y is an n-dimensional normally distributed random vector with

$$E(y) = X\beta, \quad Cov(y) = \sigma^2 I_n,$$

X is a known $n \times p$ matrix of rank p, $0 < p < n$, β is an unknown p-dimensional vector of fixed parameters and I_n stands for the identity matrix of order n.

For a given nonnegative definite $p \times p$ matrix H and a given nonnegative scalar h we are interested in estimation of the following parametric function

$$\gamma(\beta, \sigma) = \beta' H \beta + h\sigma^2. \tag{1}$$

The importance of the problem of estimating γ stems from the fact that the total mean squared error of a linear estimator Ly of a parametric function $K\beta$ is of the form (1):

$$TMSE(Ly; K\beta) = E[(LX\beta + L\epsilon - K\beta)'(LX\beta + L\epsilon - K\beta)] \tag{2}$$

$$= E[((LX - K)\beta + L\epsilon)'((LX - K)\beta + L\epsilon)] \tag{3}$$

$$= \beta' H_0 \beta + h_0 \sigma^2, \tag{4}$$

where $H_0 = (LX - K)'(LX - K)$ and $h_0 = trLL'$. Using estimators of (2) for comparing biased estimators of the parameter vector β is discussed in [9] while the applications to the problem of variable selection are discussed in [16, Chapter 11] and [8].

The problem of estimating γ when H is the identity matrix of order p and $h = 0$ was considered by Brook and Moore [6]. They have shown that the squared length of the least squares coefficient vector is positively biased. This indicates that there is a need to consider alternative approaches for estimating $\gamma(\beta, \sigma) = \beta'\beta$.

Let us note that the so called naive estimator of γ given by

$$\hat{\beta}' H \hat{\beta} + h\hat{\sigma}^2, \tag{5}$$

where

$$\hat{\beta} = (X'X)^{-1}X'y \quad \text{and} \quad \hat{\sigma}^2 = y'My/(n-p) \tag{6}$$

with $M = I_n - X(X'X)^{-1}X'$, has bias $\sigma^2 \operatorname{tr}\{H(X'X)^{-1}\}$. On the other hand, the estimator

$$\hat{\beta}' H \hat{\beta} + \hat{\sigma}^2 \left[h - tr\{H(X'X)^{-1}\}\right] \tag{7}$$

is unbiased for γ and has uniform minimum variance as a function of the minimal sufficient and complete statistics $\hat{\beta}$ and $\hat{\sigma}^2$. Let us observe that the estimator (7) is nonnegative if and only if $h \geq \operatorname{tr}\{H(X'X)^{-1}\}$. Thus if $h < \operatorname{tr}\{H(X'X)^{-1}\}$, the problem arises: how to estimate γ by a quadratic estimator $y'Ay$ with nonnegative definite A; compare [10, p. 114].

1.1. Notation

For a given positive integer k we will denote by $\mathbb{R}^k$, $\mathbb{S}^k$, $\mathbb{S}^k_+$ and $\mathbb{S}^k_{++}$ the k-dimensional Euclidean vector space, the Euclidean space of symmetric matrices of order k, the convex cone of symmetric nonnegative definite matrices of order k and the convex cone of symmetric positive definite matrices of order k, respectively.

For $A \in \mathbb{S}^k$ we will denote its positive part by A_+ and its negative part by A_-. Let us recall that these matrices can be defined by

$$A_+ = \sum_{i=1}^{k} \max(a_i, 0) u_i u_i', \quad A_- = -\sum_{i=1}^{k} \min(a_i, 0) u_i u_i',$$

where $a_1, a_2, \ldots, a_k$ are the eigenvalues of A and $u_1, \ldots, u_k$ are the corresponding orthonormal eigenvectors.

2. Minimum biased quadratic estimation of γ

Gnot, Trenkler and Zmyślony [10, p. 115] have defined the nonnegative minimum biased estimator of γ as the unique solution of a certain convex optimization problem.

Definition 1 [10]**.** The nonnegative minimum biased estimator of the parametric function γ is given by

$$\hat{\beta}' C_H \hat{\beta} + c_H \hat{\sigma}^2, \tag{8}$$

where the pair $< C_H, c_H >$ is the unique solution of the optimization problem

$$\text{minimize} \quad \operatorname{tr}\left\{(H - C)^2\right\} + \left[\operatorname{tr}\left\{C(X'X)^{-1}\right\} + c - h\right]^2 \tag{9}$$

$$\text{subject to} \quad C \in \mathbb{S}^p_+ \quad \text{and} \quad c \geq 0. \tag{10}$$

Let us note that if $h \geq \operatorname{tr}\{H(X'X)^{-1}\}$ then $C_H = H$, $c_H = h - \operatorname{tr}\{H(X'X)^{-1}\}$ and (8) is the uniformly minimum variance unbiased estimator of γ. It can be shown that if $h < \operatorname{tr}\{H(X'X)^{-1}\}$ then $c_H = 0$ and

$$C_H = \left[H - \lambda(X'X)^{-1}\right]_+,$$

where λ is the unique root of the equation $f(x) = 0$ with f defined by

$$f(\lambda) = \operatorname{tr}\left\{\left[H - \lambda(X'X)^{-1}\right]_+ (X'X)^{-1}\right\} - h - \lambda, \tag{11}$$

compare [8, p. 68] and [10, p. 119]. The properties of the function f are given in the following lemma.

Lemma 2. (a) *The function $f : \mathbb{R} \mapsto \mathbb{R}$ is continuous and strictly decreasing.*

(b) *If $h < \mathrm{tr}\{H(X'X)^{-1}\}$ then $f(0) > 0$ and $f\left(\mathrm{tr}\{H(X'X)^{-1}\}\right) < 0$.*

Proof. The continuity of f is a consequence of the fact that the orthogonal projection of $A \in \mathbb{S}^n$ onto the convex cone $\mathbb{S}^n_+$ is equal to A_+; compare [14, Theorem 2.1] and [12, Proposition 3.1.3]. Its strictly decreasing monotonicity follows from the fact that the function

$$\mathbb{R} \ni \delta \mapsto \mathrm{tr}\left\{\left[H - \delta(X'X)^{-1}\right]_- (X'X)^{-1}\right\}$$

is nondecreasing; compare [10, Lemma 2.2]. The last inequality in the part (b) also follows from it. This completes the proof. ■

The above lemma suggests using the bisection method, the Brent's method [5, Chapter 4] or the Ridders' method [15, 146–150] for finding the unique root of the function f. The implementations of these methods can be found e.g. in the *R* package *pracma* [4].

The unique root of f can be also found using the following result due to Gnot, Trenkler and Zmyślony [10].

Theorem 3 ([10], Theorem 2.2). *Let us assume that $h < \mathrm{tr}\{H(X'X)^{-1}\}$ and let the sequence (δ_n) be defined by*

$$\delta_0 = [\mathrm{tr}\{H(X'X)^{-1}\} - h]/\tau, \tag{12}$$

$$\delta_n = [\mathrm{tr}\{[H - \delta_{n-1}(X'X)^{-1}]_-(X'X)^{-1}\}]/\tau + \delta_0, \quad n = 1, 2, \ldots, \tag{13}$$

where $\tau = 1 + \mathrm{tr}\{(X'X)^{-2}\}$. Then (δ_n) converges to the unique root of f.

Gnot, Trenkler and Zmyślony [10] have proved that in the case when $HX'X = X'XH$ the function f is piecewise linear and the unique root of the function f can be expressed in an explicit form.

Theorem 4 ([10], Theorem 2.1). *If $HX'X = X'XH$ and $h < \mathrm{tr}\{H(X'X)^{-1}\}$, then there exist $x_1, \ldots, x_q$ such that $0 < x_1 < x_2 < \cdots < x_q < \mathrm{tr}\{H(X'X)^{-1}\}$ and the restriction of f to $[x_i, x_{i+1}]$, $i = 1, \ldots, q$, with $x_0 = 0$ and $x_{q+1} = \mathrm{tr}\{H(X'X)^{-1}\}$, is an affine function.*

Proof. There exists an orthogonal matrix ($P = [v_1|\cdots|v_p]$) of order p such that

$$X'X = P\,\mathrm{diag}(\lambda_1, \ldots, \lambda_p)P' = \sum_{i=1}^{p} \lambda_i v_i v_i',$$

$$H = P\,\mathrm{diag}(\eta_1, \ldots, \eta_p)P' = \sum_{i=1}^{p} \eta_i v_i v_i',$$

where $\lambda_1, \dots, \lambda_p$ and $\eta_1, \dots, \eta_p$ are the eigenvalues of $X'X$ and H, respectively, satisfying the condition

$$\lambda_1\eta_1 \geq \lambda_2\eta_2 \geq \cdots \geq \lambda_p\eta_p. \tag{14}$$

Let $\xi_1 < \xi_2 \cdots < \xi_s$, $s \leq p$, be an ordered sequence consisting of all distinct elements of the set $\{\lambda_1\eta_1, \lambda_2\eta_2, \dots, \lambda_p\eta_p\}$. It can be seen that $q = s$ and $x_i = \xi_i$, $i = 1, 2, \dots, q$, satisfy the desired conditions. ■

Remark 5. If $HX'X = X'XH$ and $h < \operatorname{tr}\{H(X'X)^{-1}\}$, then the unique root of f is given by

$$t_m = \frac{\sum_{i=1}^m \eta_i/\lambda_i - h}{1 + \sum_{i=1}^m 1/\lambda_i^2},$$

where m is the largest positive integer not exceeding p such that

$$\eta_m - \frac{t_m}{\lambda_m} > 0;$$

compare [10, pp. 116–117] and [17, p. 31].

Let us also recall that if $h < \operatorname{tr} H(X'X)^{-1}$ and the matrices H and $X'X$ commute, the nonnegative minimum biased estimator of γ is admissible in the class of all nonnegative quadratic estimators with respect to the mean squared error risk function. This result is also due to Gnot, Trenkler and Zmyślony; compare [10, Theorem 3.1].

3. Future research issues

The problem of finding the unique root of the function f defined by (11) is challenging when the matrices H and $X'X$ do not commute. In this case it is necessary to use the iterative procedure described in Theorem 3 or to apply a root-finding procedure such as the bisection method. In both cases one has to compute the appropriate eigenvectors and eigenvalues the matrix $H - \lambda(X'X)^{-1}$ so as to compute $[H - \lambda(X'X)^{-1}]_+$ in each iteration step.

In [8] and [11] the modified *regula falsi* (also known as the *Illinois method*, compare [13, p. 95]) was used for finding the approximate value of the unique root of f. This method, known to be "extraordinarily effective" in many practical situations, proved to be good also in this case.

Professor Roman Zmyślony proposed the following conjecture concerning the function f.

Conjecture 6. *The function f defined by* (11) *is convex.*

Remark 7. It is easy to show that the function f is convex if $HX'X = X'XH$. Attempts to prove or disprove this conjecture when this condition is not satisfied were not successful so far to the best knowledge of the author.

Remark 8. The conjectured convexity of f would suggest using the Ridders' method [15, pp. 146–150] for finding the unique root of f in the case when $HX'X \neq X'XH$.

To compute the nonnegative minimum biased estimator of γ we can also use the fact that the problem (9)–(10) belongs to the class of convex nonlinear semidefinite programs [18] and to the class of cone quadratic programs [1]. The mentioned programs can be solved using such packages as CVXOPT [2] or PENLAB [7]. This approach can be used for finding the solution of a counterpart of the optimization problem (9)–(10) in which the discrepancy function has a more general form: for given $D \in \mathbb{S}^p_+$ and $d > 0$

$$\text{minimize} \quad \operatorname{tr}[(H-C)D]^2 + d\left[\operatorname{tr}\left\{C(X'X)^{-1}\right\} + c - h\right]^2 \tag{15}$$

$$\text{subject to} \quad C \in \mathbb{S}^p_+ \quad \text{and} \quad c \geq 0. \tag{16}$$

Let us also note that the problem of minimizing the Bayesian risk of a quadratic estimator of γ having the form $\hat{\beta}'G\hat{\beta} + g\hat{\sigma}^2$ with given $G \in \mathbb{S}^p_+$ and $g \geq 0$ considered in [3, p. 189] can also be formulated as a cone quadratic program or a convex nonlinear semidefinite program.

The strategies for variable selection in the linear regression model basing on nonnegative minimum biased estimation of the parametric function γ appear to be very promising; compare [8, 11]. It may be thus expected that the approaches to estimating γ proposed by Professor Roman Zmyślony and his co-workers will find many applications in statistics and data science.

References

[1] M. Andersen, J. Dahl, Z. Liu and L. Vandenberghe, *Interior-point methods for large-scale cone programming*, in: Optimization for Machine Learning, Sra, Nowozin and Wright (Ed(s)), (MIT Press, Cambridge, Massachusetts, 2012) 55–83.

[2] M. Andersen, J. Dahl, and L. Vandenberghe, CVXOPT: A Python package for convex optimization (Python package, version 1.1.9. Available at cvxopt.org, 2016).

[3] J. Bojarski and R. Zmyślony, *Nonnegative Bayesian estimators for nonnegative parametric functions in linear models*, Discuss. Math., Algebra Stoch. Methods **18** (1998) 187–193.

[4] H.W. Borchers, pracma: Practical numerical math functions (R package, version 2.07. Available at CRAN.R-project.org/package=pracma, 2017).

[5] R. Brent, Algorithms for minimization without derivatives (Prentice Hall, Englewood Cliffs, NJ, 1973).

[6] R. Brook and T. Moore, *On the expected length of the least squares coefficient vector*, J. Econometrics **12** (1980) 245–246.

[7] J. Fiala, M. Kočvara and M. Stingl, PENLAB: A MATLAB solver for nonlinear semidefinite optimization (arXiv preprint:1311.5240, 2013).

[8] S. Gnot, H. Knautz and G. Trenkler, *Using nonnegative biased quadratic estimation for variable selection in the linear regression models*, in: Proceeding of the International Conference on Linear Statistical Inference LINSTAT '93, Caliński and Kala (Ed(s)), (New York: Springer, 1994) 65–71.

[9] S. Gnot and G. Trenkler, *Nonnegative quadratic estimation of the mean squared errors of minimax estimators in the linear regression model*, Acta Appl. Math. **43** (1996) 71–80.

[10] S. Gnot, G. Trenkler and R. Zmyślony, *Nonnegative minimum biased quadratic estimation in the linear regression model*, J. Multivar. Anal. **54** (1995) 113–125.

[11] M. Grządziel, *A simulation study of a new variable selection strategy based on minimum biased quadratic estimation*, Discuss. Math., Algebra Stoch. Methods **15** (1995) 325–334.

[12] J.B. Hiriart-Urruty and C. Lemarchéal, Fundamentals of Convex Analysis (Springer Verlag, Heidelberg, 2001).

[13] J. Monahan, Numerical methods of statistics (Cambridge University Press, 2011).

[14] J. Malick and H. Sendov, *Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices*, Set-Valued Anal. **14** (2006) 273–293.

[15] J. Kiusalaas, Numerical methods in engineering with Python (Second Edition, Cambridge University Press, 2010).

[16] A. Sen and M.S. Srivastava, Regression Analysis. Theory, Methods and Applications (Springer Verlag, New York- Heidelberg, 1990).

[17] Y. Sukestiyarno, Strategien zur Wahl der Ridge-Parameter auf der Basis nichtnegativer quadratischer Schätzer (Verlag fur Wissenschaft und Forschung, Berlin, 1999).

[18] H. Yamashita and H. Yabe, *A survey of numerical methods for nonlinear semidefinite programming*, J. Oper. Res. Soc. Jpn. **58** (2015) 24–60.

Received 27 September 2017
Accepted 18 November 2017