An Unusual Preferences Among Regression Designs

Krystyna Maciąg*

Department of Statistics and Econometrics, Maria Curie-Skłodowska University, Pl. Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland

Abstract: Consider the classical regression design with one explanatory variable taking values $\mathrm{x} = (x_1,\ldots,x_n)'$ and an alternative design based on **Tx**, where $\mathbf{T} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. We reveal an interesting phenomenon that the second design is better, from many reasonable points of view, than the initial one.

Keywords: Linear regression, one explanatory variable, variable transformation, design preference.

1. INTRODUCTION

Regression models with one and two regressors attracted a great attention of many authors, especially in the context of the relation between determination coefficient and the correlation coefficients corresponding to the regressor components. This problem has been arisen by Hamilton [1] and discussed in many papers (cf. [2-10]). It appears that, roughly speaking, the joint information involved in two regressors may be greater than the sum of the information brought by each of them. A paradox of different kind, not referring to regression, but also dealing with statistical information was recently arisen by Kagan and Shepp [11].

Our note reveals one more paradox on the regression background; this time in the context of design preference. It will be shown a rather unexpected fact that the design based on a transformed explanatory variable is better, from many reasonable points of view, than the initial one.

2. UNEXPECTED PREFERENCE AMONG REGRESSION DESIGNS

Consider the simple linear regression model

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1,\ldots,n \quad (n \geq 2),$$

where $y_1,\ldots,y_n$ are realizations of a response variable, $x_1,\ldots,x_n$ are suggested values of an explanatory variable, and not all x_i are the same, $e_1,\ldots,e_n$ are not correlated experimental errors with a common variance σ^2, while α and β are unknown parameters interpreted as the intercept and the regression coefficient of **y** on **x**.

We mention that the variance of the Best Linear Unbiased Estimator (BLUE) of a parametric function $\Psi = c_1\alpha + c_2\beta$, with arbitrary constant coefficients c_1 and c_2, may be presented in the form (see e.g. [12]):

$$\sigma^2 \begin{bmatrix} c_1, & c_2 \end{bmatrix} \mathbf{M}^{-1} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

where $\mathbf{M} = \mathbf{M}(\mathbf{x})$ is the actual information matrix, and namely,

$$\mathbf{M}(\mathrm{x}) = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}.$$

Now consider an alternative design based on the transformed explanatory variable **Tx**, where $\mathbf{T} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. Then

$$\mathbf{M}(\mathbf{Tx}) = \begin{bmatrix} n & 0 \\ 0 & ns_x^2 \end{bmatrix},$$

where $ns_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$. In consequence we get

$$\mathbf{M}^{-1}(\mathbf{x}) = \frac{1}{n^2 s_x^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

and

$$\mathbf{M}^{-1}(\mathbf{Tx}) = \frac{1}{n^2 s_x^2} \begin{bmatrix} ns_x^2 & 0 \\ 0 & n \end{bmatrix}.$$

Since $ns_x^2 \leq \sum x_i^2$ with the strict inequality unless $\bar{x} = 0$, the design induced by **Tx** is at least as good as one induced by **x**, in the sense of minimal variance of the BLUE's of all parameters.

3. FURTHER PREFERENCE CRITERIA

Let us note that (partial) ordering of experimental designs based on the Loewner ordering of the information matrices is very strong and thus applicable rather rarely. For these reasons the statistical literature (cf. for instance [13-15]) suggests some weaker criteria, called also *ϕ-criteria*, based on some scalar functions of **M** or $\mathbf{M}^{-1}$. The most popular of them are: *D*-, *A*- and *E-criterion*. *D-criterion* consists in minimizing the determinant of the inverse information ma-

*Address correspondence to this author at the Department of Statistics and Econometrics, Maria Curie-Skłodowska University, Pl. Marii Curie-Skłodowskiej 5, 20-031 Lublin, Poland;
E-mail: maciag@hektor.umcs.lublin.pl

trix $\mathbf{M}^{-1}$, *A-criterion* - in minimizing its trace, while *E-criterion* - in minimizing its largest eigenvalue. All of these criteria are antitonic (or decreasing) with respect to the Loewner ordering of the information matrices.

It is easy to check that

$$\det(\mathbf{M}^{-1}(\mathbf{Tx})) = \det(\mathbf{M}^{-1}(\mathbf{x})) \quad (1)$$

and

$$\mathrm{tr}(\mathbf{M}^{-1}(\mathbf{Tx})) = \mathrm{tr}(\mathbf{M}^{-1}(\mathbf{x})) - \frac{\bar{x}^2}{ns_x^2} \leq \mathrm{tr}(\mathbf{M}^{-1}(\mathbf{x})), \quad (2)$$

where the equality holds if and only if $\bar{x} = 0$.

Let us denote by λ_1, λ_2 $(\lambda_1 \geq \lambda_2)$ the eigenvalues of $\mathbf{M}^{-1}(\mathbf{x})$ and μ_1, μ_2 $(\mu_1 \geq \mu_2)$ the eigenvalues of $\mathbf{M}^{-1}(\mathbf{Tx})$. We note that (1) and (2) imply the following relations

$$\mu_1 \cdot \mu_2 = \lambda_1 \cdot \lambda_2 \quad \text{and} \quad \mu_1 + \mu_2 \leq \lambda_1 + \lambda_2.$$

This implies immediately that $\lambda_2 \leq \mu_1 \leq \lambda_1$, that is, the largest eigenvalue of $\mathbf{M}^{-1}(\mathbf{Tx})$ is not greater than the largest eigenvalue of $\mathbf{M}^{-1}(\mathbf{x})$.

From above consideration, we can conclude that the regression design corresponding to the transformed explanatory vector **Tx** is preferable than one based on original **x**, in the sense of *A-* and *E-criteria* and equivalent in the sense of *D-criterion* of optimality.

4. NUMERICAL EXAMPLE

Let us consider the linear regression model

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \ldots, n,$$

with the explanatory vector $\mathbf{x} = (1.8, 1.9, 2, 2, 2.1)'$. In this case the information matrix and the inverse information matrix are given by

$$\mathbf{M}(\mathbf{x}) = \begin{bmatrix} 5 & 9.8 \\ 9.8 & 19.26 \end{bmatrix}, \quad \mathbf{M}^{-1}(\mathbf{x}) = \frac{1}{0.26}\begin{bmatrix} 19.26 & -9.8 \\ -9.8 & 5 \end{bmatrix}.$$

On the other hand, for the design based on the transformed explanatory variable

$$\mathbf{Tx} = (-0.16, -0.06, 0.04, 0.04, 0.14)',$$

we get

$$\mathbf{M}(\mathbf{Tx}) = \begin{bmatrix} 5 & 0 \\ 0 & 0.052 \end{bmatrix}, \quad \mathbf{M}^{-1}(\mathbf{Tx}) = \frac{1}{0.26}\begin{bmatrix} 0.052 & 0 \\ 0 & 5 \end{bmatrix}.$$

Hence the variance of the BLUE of α in the first and the second design are given by $74.0769\sigma^2$ and $0.2\sigma^2$, respectively. Besides, we see that the variance of the BLUE of β in both designs are the same and equal $19.2308\sigma^2$. Furthermore, the eigenvalues of $\mathbf{M}^{-1}(\mathbf{x})$ are 93.2665 and 0.04124, while the eigenvalues of $\mathbf{M}^{-1}(\mathbf{Tx})$ are 19.2308 and 0.2. Thus, the design based on **Tx** is better in the sense of minimal variance of the BLUE's of parameters and in the sense of *A-* and *E-criteria* of optimality.

ACKNOWLEDGEMENT

The author is grateful to Professor Czesław Stępniak for useful discussions and comments.

REFERENCES

[1] D. Hamilton, "Sometimes $R^2 > r_{yx_1}^2 + r_{yx_2}^2$", *Am. Stat.,* vol. 41, no. 2, pp. 129-132, May 1987.
[2] P. V. Bertrand, and R. L. Holder, "A quirk in multiple regression: the whole regression can be greater than the sum of its parts", *Statistician,* vol. 37, pp. 371-374, 1988.
[3] I. Currie, and A. Korabinski, "Some Comments on Bivariate Regression", *Statistician*, vol. 33, no. 3, pp. 283-293, Sep. 1984.
[4] R. J. Freud, "When is $R^2 > r_{yx_1}^2 + r_{yx_2}^2$? (Revisited)", *Am. Stat.,* vol. 42, no. 1, pp. 89-90, Feb. 1988.
[5] L. Friedman, and M. Wall, "Graphical Views of Suppression and Multicollinearity in Multiple Linear Regression", *Am. Stat.,* vol. 59, no. 2, pp. 127-136, May 2005.
[6] D. Hamilton (Reply to Freud and Mitra), *Am. Stat.,* vol. 42, no. 1, pp. 90-91, Feb. 1988.
[7] S. Mitra, "The Relationship Between the Multiple and Zero-Order Correlation Coefficients", *Am. Stat.,* vol. 42, no. 1, p. 89, Feb. 1988.
[8] R. B. Nelsen, "Correlation, Regression Lines and Moments of Inertia", *Am. Stat.,* vol. 52, no. 4, pp. 343-345, Nov. 1998.
[9] H. M. Schey, "The Relationship Between the Magnitudes of SSR(x_2) and SSR($x_2|x_1$): A Geometric Description", *Am. Stat.,* vol. 47, no. 1, pp. 26-30, Feb. 1993.
[10] G. Shieh, "The Inequality Between the Coefficient of Determination and the Sum of Squared Simple Correlation Coefficients, " *Am. Stat.,* vol. 55, no. 2, pp. 121-124, May 2001.
[11] A. Kagan, and L. A. Shepp, "A Sufficiency Paradox: An Insufficient Statistic Preserving the Fisher Information", *Am. Stat.,* vol. 59, no. 1, pp. 54-56, Feb. 2005.
[12] R. B. Bapat, *Linear algebra and linear models*, Delhi: Hindustan Book Agency, 1993, p. 38.
[13] E. P. Liski, N. K. Mandal, K. R. Shah, and B. K. Sinha, *Topics in Optimal Designs, Lecture Notes in Statistics*, 163, New York: Springer-Verlag, 2002, p. 23.
[14] F. Pukelsheim, *Optimal Designs of Experiments*, New York: Wiley, 1993, p. 114.
[15] A. Zaigraev, "Stochastic Design Criteria in Linear Models", *Aust. J. Stat.*, vol. 34, no. 2, pp. 211-223, 2005.

Received: October 5, 2008 Revised: October 23, 2008 Accepted: December 3, 2008