Send Orders for Reprints to reprints@benthamscience.ae

Open Access

On a Preferred Design in Regression

D. R. Jensen and D. E. Ramirez*

Department of Mathematics, University of Virginia, Charlottesville, VA 22903, USA

Abstract: Corrections are given to common misconceptions regarding the use of centered and uncentered regressors, and of meanings to be ascribed to $A-$, $D-$ and $E-$criteria in evaluating two such designs.

Keywords: Correction, centered and uncentered regressors, design preference.

PRINCIPAL FINDINGS

Maciag [1] recently considered the model $\{y_i = \alpha + \beta x_i + \varepsilon_i; 1 \le i \le n\}$ having regressors $\mathbf{x}' = [x_1, \ldots, x_n]$, together with "an alternative design based on the transformed explanatory variable $\mathbf{Tx}$, where $\mathbf{T} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$," also called *centered regressors*. Using moment matrices $\mathbf{M}(\mathbf{x})$ and $\mathbf{M}(\mathbf{Tx})$ with $ns_x^2 = \sum(x_i - \bar{x})^2$, their inverses, namely

$$\mathbf{M}^{-1}(\mathbf{x}) = \frac{1}{n^2 s_x^2}\begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix}, \quad \mathbf{M}^{-1}(\mathbf{Tx}) = \frac{1}{n^2 s_x^2}\begin{bmatrix} ns_x^2 & 0 \\ 0 & n \end{bmatrix}$$

are dispersion matrices for the respective OLS estimators, apart from a scalar σ^2. On considering linear functions $\Psi = c_1\alpha + c_2\beta$ and recalling that $Var(\hat{\Psi}) = \sigma^2 \mathbf{c}'\mathbf{M}^{-1}\mathbf{c}$ with $\mathbf{c}' = [c_1, c_2]$, Maciag concludes: "Since $ns_x^2 \le \Sigma x_i^2$ with the strict inequality unless $\bar{x} = 0$, the design induced by $\mathbf{Tx}$ is at least as good as one induced by $\mathbf{x}$, in the sense of minimal variance of the BLUE's of all parameters." A numerical example is given with $n = 5$, $\bar{x} = 1.96$, and matrices $\mathbf{M}^{-1}(\mathbf{x}) = \frac{1}{0.26}\begin{bmatrix} 19.26 & -9.80 \\ -9.80 & 5.00 \end{bmatrix}$ and $\mathbf{M}^{-1}(\mathbf{Tx}) = \frac{1}{0.260}\begin{bmatrix} 0.052 & 0 \\ 0 & 5.000 \end{bmatrix}$. The author concludes that "the variance of the BLUE's of α in the first and the second design are given by $74.0769\sigma^2$ and $0.2\sigma^2$, respectively."

These assessments are in error. As noted in Smith and Campbell [2; p.76]: "Because rewriting the model [in centered variables] does not affect any of the implicit estimates, it has no effect on the amount of information contained in the data." In fact, the errors here stem from a failure to recognize that the two models have different parameters. Rewriting $\{y_i = \alpha + \beta x_i + \varepsilon_i; 1 \le i \le n\}$ in centered regressors gives $\{y_i = \gamma + \beta(x_i - \bar{x}) + \varepsilon_i; 1 \le i \le n\}$, with $\gamma = (\alpha + \beta\bar{x})$. Indeed, Ψ now becomes $\Psi = d_1\gamma + d_2\beta$ in the transformed regressors, with $d_1 = c_1$ and $d_2 = c_2 - c_1\bar{x}$. Accordingly, the quadratic forms

$$\sigma^2\mathbf{c}'[\mathbf{M}^{-1}(\mathbf{x})]\mathbf{c} = \sigma^2(74.0769c_1^2 - 75.3846c_1c_2 + 19.2308c_2^2)$$

$$\sigma^2\mathbf{d}'[\mathbf{M}^{-1}(\mathbf{Tx})]\mathbf{d} = \sigma^2(0.2000d_1^2 + 19.2308d_2^2)$$

with $\mathbf{d}' = [d_1, d_2]$, are equivalent for $Var(\hat{\Psi})$ under the required constraints $\{d_1 = c_1, d_2 = (c_2 - 1.96c_1)\}$.

Specifically, the correct assessment is that $Var(\hat{\alpha}) = 74.0769\sigma^2$ and $Var(\hat{\gamma}) = 0.2\sigma^2$. This same error appears in Gunst [3], purporting to show that the uncentered *variance inflation factor* for $\hat{\alpha}$ is a genuine ratio of variances, namely, the price to be paid in variance for designing an experiment having $\bar{x} \ne 0$, in contrast to the alternative design having $\bar{x} = 0$.

The author [1] continues to assess A–, D–, and E–criteria in regard to $\mathbf{M}^{-1}$, in minimizing its trace, determinant, and largest eigenvalue, showing that D is identical and that A and E are smaller, thus preferable, for the centered design. In reference, Pukelsheim [4], as cited by the author, shows *via* the unimodular group that the D –criterion alone is invariant to reparametrization. In the numerical example, "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 $\mathbf{Tx}$ 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."

*Address correspondence to this author at the Department of Mathematics, University of Virginia, Charlottesville, VA 22903, USA.
Tel: 434-924-4934; Fax: 434-982-3084; E-mail: der@virginia.edu

Unfortunately, the comparative *A*– and *E*–criteria are meaningless here, as the *A*–criterion is the sum of variances in estimating different parameters in the two designs, thus not comparable. Moreover, spectral analysis shows the largest eigenvalue to be the variance in estimating the linear parametric function with coefficients as given by the corresponding eigenvector. But since these are functions of parameters differing between the two designs, their *E*–criteria are not comparable.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] K. Maciag, "An Unusual Preferences Among Regression Designs," *Stat. Probab. J.*, vol. 1, pp. 1–2, 2009.

[2] G. Smith and F. Campbell, "A Critique of Some Ridge Regression Methods," *J. Amer. Statist. Assoc.*, vol. 75, pp. 74–81, 1980.

[3] R. F. Gunst, "Comment: Toward a Balanced Assessment of Collinearity Diagnostics," *Amer. Statist.*, vol. 38, pp. 79–82, 1984.

[4] F. Pukelsheim, *Optimal Designs of Experiments*, New York: Wiley, 1993.

Received: February 13, 2014 Revised: February 28, 2014 Accepted: October 28, 2014