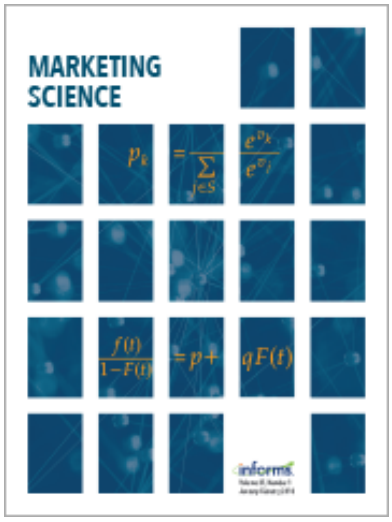

Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Addendum to "Cross-Validating Regression Models in Marketing Research" Vol. 12, No. 4, Fall 1993, pp. 415—427

Joel H. Steckel, Wilfried R. Vanhonacker,

To cite this article:
Joel H. Steckel, Wilfried R. Vanhonacker, (1996) Addendum to "Cross-Validating Regression Models in Marketing Research" Vol. 12, No. 4, Fall 1993, pp. 415—427. Marketing Science 15(1):109-109. https://doi.org/10.1287/mksc.15.1.109

Addendum to "Cross-validating Regression Models in Marketing Research"

Vol. 12, No. 4, Fall 1993, pp. 415–427[1]

Joel H. Steckel • Wilfried R. Vanhonacker
New York University
Hong Kong Institute of Technology

In our paper, "Cross-Validating Regression Models in Marketing Research," which appeared in the Vol. 12, No. 4, Fall 1993 issue of *Marketing Science*, pp. 415–427, we develop a formal statistical test for cross-validating regression models under the simple random splitting framework. In this framework, a data set is divided into two subsets, an estimation sample and a validation sample. In particular, Theorem 1 of this paper (p. 418) presents a statistic for testing whether the differences between the actual values of a dependent variable in the validation sample and the values predicted for the same validation sample by a regression model estimated on the estimation sample are "small enough" to believe that the proposed model is "correct." We then use this statistic to investigate what proportions of a data set should be optimally allocated to the estimation and validation samples.

As published, the statistic is

$$f = \frac{\dfrac{q}{N_2}}{\dfrac{s}{N_1 - k_1}},$$

where N_1 and N_2 are the sizes of the estimation and validation samples and k_1 is the number of independent variables in the regression model. The original publication neglected to present definitions for q and s. We present these now:

$$q = \mathbf{u}'[\mathbf{I} + \mathbf{X}_{21}(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{21}']^{-1}\mathbf{u}, \quad \text{and}$$

$$s = \mathbf{y}_1'\{\mathbf{I} - \mathbf{X}_{11}(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{21}'[\mathbf{X}_{21}(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{21}']^{-1} \times \mathbf{X}_{21}(\mathbf{X}_{11}'\mathbf{X}_{11})^{-1}\mathbf{X}_{11}\}\mathbf{y}_1.$$

The $\mathbf{X}$ terms define the values of the independent variables for the estimation ($\mathbf{X}_{11}$) and validation ($\mathbf{X}_{21}$) samples; $\mathbf{y}_1$ represents the dependent variable values for the estimation sample and $\mathbf{u}$ represents the prediction errors in the validation sample.

We apologize for any inconvenience and invite anyone to request the unabridged working paper for a more detailed development of the results than that which appears in the original publication.[1]

[1] We are grateful to Eltje Foelkens, Len Parsons, and Dick Wittink for inquiries that led to the discovery of our errors.

0732-2399/96/1501/0109$01.25