Hoerl A E & Kennard R W. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**:55-67, 1970.
[University of Delaware, Newark, and E.I. du Pont de Nemours & Co., Wilmington, DE]

In multiple regression, it is shown that least square parameter estimates can be unsatisfactory if the prediction vectors are not orthogonal. Proposed is a procedure based on adding small positive quantities to the diagonals of the normal equations to obtain estimates with smaller mean square error. [The *Science Citation Index*® (*SCI*®) and the *Social Sciences Citation Index*® (SSCI®) indicate that this paper has been cited in over 310 publications since 1970.]

Arthur E. Hoerl
Department of Mathematical Sciences
University of Delaware
Newark, DE 19711
and
Robert W. Kennard
Box 778
Route 1
Redwing Road
Groveland, FL 32736

July 7, 1982

"What we call ridge regression had its origins in the consulting work in statistics that we were doing in the latter half of the 1950s. At that time, we were both consultants in statistics at DuPont learning to apply statistics to a myriad of problems in manufacturing, research, and business. We would get together on Saturday mornings and hash over what we had been doing during the week. It would be great to report that we had profound discussions on the foundations of statistics, but such was not the case. Much of the time was spent trying to find ways to do regression computations economically and to come up with solutions that made engineering sense. We were charging $90/day for our time, but had to charge $450/hour for computer time on a Univac I that had 1,000 words of memory. With this machine, it took 75 processing minutes to invert a 40 × 40 matrix through a 4 × 4 partition of 10 × 10 submatrices, using magnetic tapes for temporary storage.

"In these discussions, we found that we had both encountered the same phenomenon, one that had caused some embarrassment with clients. We found that multiple linear regression coefficients computed using least squares didn't always make sense when put into the context of the process generating the data. The coefficients tended to be too large in absolute value, some would even have the wrong sign, and they could be unstable with very small changes in the data.

"Under the assumption that the computed regression coefficients were too large in absolute value, we set about to find ways to dampen them. One of us (Hoerl) had been working on the characterization of multidimensional quadratic response surfaces. The mathematics was the same for the residual sum of squares if the coefficients rather than the predictors were considered to be the variables. The other (Kennard) took the approach of bounding the coefficients using a weighting function for values of the coefficients; independent normals turned out to be tractable. Both methods came to the same end point, namely, adding a small positive quantity to the diagonals of the normal equations. This addition had the effect of dampening the magnitudes, but also allowed sign changes, and provided stability.

"This new approach proved quite useful in our practical problems. We then set about to give the method a solid theoretical basis and this is given in the publication.

"Since the method proposed attacked one of the sacred cows of linear regression—least squares—there was considerable resistance. However, the solid theoretical basis and the practical usefulness of the method gradually overcame most objections.

"The publication has been cited so often because it pointed out and gave reasons for difficulties in multiple linear regression, a data analysis method used in many fields, and demonstrated a simple method to overcome the difficulties. A survey of the present status of the theory and an annotated bibliography of important applications can be found in our 1981 publication."[1]

1. **Hoerl A E & Kennard R W.** Ridge regression—1980: advances, algorithms, and applications. *Amer. J. Math. Manage. Sci.* 1:5-83, 1981.