

Farrar D E & Glauber R R. Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Statist.* 49:92-107, 1967.
[Sloan School of Management, Massachusetts Institute of Technology, and Harvard Business School, Cambridge, MA]

This paper reviews the nature of multicollinearity among independent or predictor variables in regression analysis. It develops certain statistical measures which may be used to determine the presence of multicollinearity in a set of independent variables and, if present, the variables most seriously affected and the pattern of interdependence among them. [The *Science Citation Index*® (SCI®) and the *Social Sciences Citation Index*® (SSCI®) indicate that this paper has been cited over 215 times since 1967.]

Donald E. Farrar
Benham Management Corporation
755 Page Mill Road
Palo Alto, CA 94304

December 3, 1981

"This paper was the first scholarly project I undertook on completing my doctoral dissertation. It grew out of my exposure to regression analysis and to standard multiple regression computer programs as a young economist in the early-1960s, together with my interest at the time in factor analysis as a technique for analyzing sets of highly interdependent variables in other empirical studies.

"Certain statistics which I did not understand, such as the determinant of the correlation matrix of independent variables and diagonal elements of the inverse correlation matrix, were produced routinely by standard multiple regression computer programs at the time. I was told they helped to diagnose multicollinearity among independent variables, but did not understand how or why. In an effort to understand this witchcraft, Robert Glauber and I immersed ourselves in the literature on distributional properties of closely related multivariate statistics, until finally we were able to develop transformations of the variables of interest (determinant and diagonal elements of inverse correlation matrix) which we

could interpret and whose distributional properties were known. We then added additional statistics through which one could identify patterns of interdependence among multicollinear variables.

"The paper has become something of a classic over the years for a number of reasons. First, multicollinearity is an important and ubiquitous problem faced by any economist in applying the most fundamental of econometric tools, multiple regression analysis. Our paper was the first which attempted to help practitioners understand and deal with the problem. Second, the statistics we proposed for diagnostic purposes were easily available by-products of standard calculations performed routinely by computers during the course of regression computations. Hence, they were easily and economically obtained and were incorporated in a number of standard computational routines. And third, the paper was reasonably well written. It is capable of being understood by most of the persons who would have an interest in the subject.

"I am not sure which of the above reasons is most important. I believe that being first in an important field helps. I'd also guess that most persons read the paper after encountering the statistics it suggests in reams of computer output rather than the other way around. And, perhaps, the exposition also helped.

"As a matter of passing interest, I recall that the paper was turned down by at least one journal and was accepted only grudgingly by the *Review of Economics and Statistics*, where it was the last rather than the lead article in the issue which carried it. I also recall that a senior colleague of my co-author did not like the paper, as it was not sufficiently Bayesian to satisfy his tastes. As a result, Glauber gave serious consideration at the last minute to withdrawing his name from the paper. Publication was not particularly important at the time for junior faculty at the Harvard Business School, but disapproval of one's work by a senior professor could be.

"I always liked the paper, however, and am pleased that it has left its mark in the literature and in the computing centers where work in our profession is done. A more recent review in this field has been written by Belsley, Kuh, and Welsch."¹

1. Belsley D A, Kuh E & Welsch R E. *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley, 1980. 292 p.