TABLE 3
The impact of point 3 on some regression results for the data in Stewart's Table 3[a]

Var.	ι_j		IMP_j		$\tilde{\beta}$		Var $(\tilde{\beta})$	
	F	R	F	R	F	R	F	R
X_1	.213	.181	.04	.07	2.19	1.89	.041	.129
X_2	.602	.470	.06	.26	1.15	.90	.004	.063
X_3	.104	.090	.05	.06	.76	.63	.029	.043
X_4	.171	.083	.04	.19	.49	.23	.003	.066
X_5	.001	.247	.06	.48	.02	10.02	1.065	95.988

[a] F denotes full data and R denotes reduced data (point 3 deleted). Values of ι_j and IMP_j are from Stewart's equations (5.1) and (5.2), respectively.

ACKNOWLEDGMENT

This work was partly supported by the United States Army Research Office through the Mathematical Sciences Institute of Cornell University.

ADDITIONAL REFERENCES

MASON, R. L. and GUNST, R. F. (1985). Outlier-induced collinearities. *Technometrics* **27** 401–407.

VELLEMAN, P. F. and VELLEMAN, A. Y. (1986). *The Data Desk Handbook.* Data Description, Ithaca, N. Y.

Rejoinder

G. W. Stewart

I would like to begin by thanking the commentators for giving my paper a fair and careful reading. Since the following remarks must necessarily focus on our differences, let me stress at the outset that I find much to agree with in their comments.

I am happy to acknowledge that Donald Marquardt knew of the connection between variance inflation factors and collinearity. My only quibble is that one must read a rather small section of his 1970 paper very carefully in order to see it. Marquardt never uses the word collinearity and only asserts that the variance inflation factors depend on the partial correlations, without explicitly stating the nature of the dependency. From his comment one can deduce that he takes a partial correlation near one as a synonym for collinearity and means for the reader to infer that the dependency is the same as the one he writes down for two variables. However, the passage can also be read as a vague afterthought, which is how I interpreted it on first reading.

On nomenclature, the difficulty with the term variance inflation factor is that it draws attention to one effect of near collinearity to the exclusion of other, equally important effects. It seems more natural to me to give a simple characterization of near collinearity and then show how it affects statistical procedures. Taking the square root of the variance inflation factors not only simplifies the formulas but stresses a useful connection with the condition number.

David Belsley's comments are practically a paper in themselves, and a complete response would amount to another. Here I will only make a few observations and trust the reader to sort out the issues.

Belsley would make a distinction between data and models, and in a sense I heartily agree. Numerical and statistical tricks are no substitute for a knowledge of the science underlying a problem. However, on close inspection his distinction appears elusive. Is a constant term model or data? How do we classify the design matrix for an unbalanced analysis of variance? Moreover, the term model has come to mean many things. Belsley's "rather exhaustive" survey evidently did not include Draper and Smith (1981, page 86) or Seber (1977, pages 42 and 43), who use the term model in much the same sense as I do. Attempting to preempt the word model is like trying to tell the tide where to come in.

I will save my comments on importance for the end of this rejoinder. Regarding centering, I will simply restate that centering is a change of variables, and the new ones are not equivalent to the old. There is nothing vague or "psychological" about this observation, and it is ironic that Belsley quotes at length from a passage that describes the psychological biases in the opposing view.

Belsley points out that the collinearity indices do not tell the dimension of the approximate null space and provide little help in selecting an independent subset: they can diagnose, but they cannot cure. Ronald Thisted makes the same point, and I wish I had done so too. The regularization (to borrow a term from the field of ill-posed problems) of nearly collinear regression problems is an open question. Many people have had a go at it, including myself along with my colleagues Gene Golub and Virginia Klema (1976). Unfortunately, our approach and all the others I have seen are sensitive to scaling, which for me renders them unsatisfactory. It may be a mistake to argue from lack of imagination, but I am coming to feel that the regularization problem can only be solved with additional information from the underlying problem—from the model in Belsley's sense of the word.

Ronald Thisted's observation on the condition number is a fine illustration of how scaling problems can insinuate themselves into even the most simple argument—here by way of the parenthetical statement "subject to the constraint, say, that $\| v \| = 1$." It should be observed that the quantity $\hat{\alpha} = v'\hat{\beta}$ is invariant under scaling (or equivalently under a change of units). For if the columns of X are scaled, then $\hat{\beta}$ inherits the scaling contravariantly, v inherits it covariantly, and everything cancels. However, the set $\{v: \| v \| = 1\}$ is not invariant under scaling; i.e., by scaling we inadvertently change our notion of the linear combinations that we consider important. It is considerations like this that have prompted me to search for scale invariant diagnostics, like $\mathrm{RE}_{\mathrm{bias}}$, $\mathrm{RE}_{\mathrm{lin}}$, and τ in (6.23). However, as I indicated at the end of the last paragraph, this approach can take us only so far.

Ali Hadi and Paul Velleman maintain that the distinction between $\mathrm{RE}_{\mathrm{bias}}$ and $\mathrm{RE}_{\mathrm{lin}}$ is artificial, and looking over my paper I see that I did not provide adequate motivation. Briefly, the argument goes as follows. The coefficient $\tilde{\beta}_p$ in (6.5) depends quadratically on the elements of h_p, but linearly on γ_p. Consequently, as ε_p becomes small, γ_p must ultimately have the greater effect. Now there is reason for believing that if h_p can be ignored and $\mu = 0$, then the effects of γ_p will not be particularly harmful (see the subsection "The diagnostics in terms of τ_j" in Section 6 and also David and Stewart (1986)). It is therefore important that the analyst have one diagnostic to say when h_p is negligible and another to assess the effect of γ_p.

Two technical points. The reason for the difference in the denominators of the definitions of $\mathrm{RE}_{\mathrm{bias}}$ and $\mathrm{RE}_{\mathrm{lin}}$ is that for the range of values we are interested in it does not make much difference, and we may as well choose the denominator that yields the simplest formula. The anomaly reported at the end of their first section is due to replacing γ_p by its mean instead of by ε_p.

The remainder of Hadi and Velleman's comment is a nice supplement to my paper. I agree that the detection of hidden collinearities is an important problem and am glad that they have undertaken to provide procedures for resolving it.

The late Jim Wilkinson, a leading authority on numerical linear algebra, used to say, "This is what I do in practice. But if you want to argue about it, I would rather be on your side." I feel much that way about the section on significance and importance, which I included with some reluctance. Thisted seems to have the best treatment, and therefore I will deal with his points.

The thesis of Section 5 is that assessing the effects of collinearity on significance testing requires more information that is contained in X and y. The notion of importance is one way of allowing the analyst to provide such information. Thisted's objection amounts to saying that this is no substitute for thinking hard about the problem, since the importance of a variable is not as easy to interpret as its simple definition might suggest.

In the first place, collinearity can create importances greater than one. A numerical analyst, who sees cancellation daily, is not fazed by a part being greater than the whole; but this will be disconcerting to many others. In any event, I think Thisted will agree that models with importances greater than one should be viewed with suspicion.

Thisted argues that the variables in his model (3.1) are enormously important in only a very narrow sense. To establish his point he transforms the second variable, leaving the first unaltered, and observes that the importance of the first decreases. However, Thisted greases the skids by writing β_1 in both his (3.1) and (3.2), when in fact these are different numbers. One could even argue that x_1 is a different variable in the two models, since the coefficients are different. But many people would not accept such an argument, and to that extent Thisted has a valid point.

If, however, one can interpret importances for the model at hand in such a way as to make reasonable selections of the tolerances λ_j, then the rest is unexceptionable algebra. Note that the procedure does not require one to estimate importances from imprecise data; instead one calculates the diagnostics IMP_j and compares them with the λ_j arrived at from other considerations. Since as a practical matter one would select λ_j's considerably less than one, say 0.1, and since collinearity inflates the IMP_j, there does not seem to be much danger of collinearity causing us to accept a poor model. For example, in Thisted's model (3.1) we have $\mathrm{IMP}_1 = \mathrm{IMP}_2 = 0.59$, which is quite large. For (3.2) the values are 0.34, which suggests that transforming the model has not improved things by much.

To summarize, where the notion of importance can be interpreted, it leads to a useful measure of the effects of collinearity on significance testing. However, it is only a beginning, and time may produce less refractory diagnostics.

Again I would like to thank my commentators for their trouble. I have found the processes of responding both educational and pleasurable.

ADDITIONAL REFERENCES

DAVID, N. and STEWART, G. W. (1986). Hypothesis testing with errors in the variables. Technical Report TR-1735, Dept. Computer Science, Univ. Maryland.

DRAPER, N. R. and SMITH, H. (1981). *Applied Regression Analysis*, 2nd ed. Wiley, New York.

GOLUB, G. H., KLEMA, V. and STEWART, G. W. (1976). Rank degeneracy and least squares problems. Technical Report TR-751, Dept. Computer Science, Univ. Maryland.