MULTICOLLINEARITY: A BAYESIAN INTERPRETATION

Edward E. Leamer *

I

THE problem of collinear data sets in the context of the univariate multiple regression model has generated a confusing array of papers, comments and footnotes. Perusal of this literature does not lead the reader to conclude that the problem has even been rigorously defined. The purpose of this paper is to offer several rigorous definitions and to suggest several substantive quantitative summaries of the degree of the problem.

Specifically we address our attention to the regression model $Y = X\beta + u$ where β is a k-dimensional parameter vector with elements $\beta_1, \beta_2, \dots \beta_k$, where Y and u are $T \times 1$ vectors and where X is a $T \times k$ matrix. Inferences are to be made about the vector β from observations of Y and X. When the columns of X are orthogonal, the design matrix $X'X$ is diagonal. Correlated columns of X imply a nondiagonal design matrix. The collinearity problem has to do with the differences in the inferences that may be drawn in these two situations.

The principal claim of this paper is that the most important aspects of the collinearity problem derive from the existence of undominated uncertain prior information which causes major problems in interpreting the data evidence. It is claimed here that if our a priori knowledge of parameter values were either completely certain or "completely uncertain" the aspects of the collinearity problem that most of us worry about would disappear.[1] As an empirical test of this proposition consider the situations when collinearity is identified as a culprit. Usually signs are wrong or point estimates are otherwise peculiar. Occasionally confidence intervals overlap unlikely regions of the parameter space. Yet to say these things is to say there exists undominated uncertain prior information.

Classical inference, with the possible exception of the pretesting literature, necessarily excludes undominated uncertain prior information. As a result most discussions of the collinearity problem miss a critical point. The textbook discussions including Theil (1971, p. 149), Malinvaud (1970, p. 218), and Goldberger (1964, p. 192), observe that when the design matrix $X'X$ becomes singular, the least squares estimator is non-unique and the sampling distribution has finite variance only for certain "estimable" functions. Thus extreme collinearity is implicitly defined as total lack of sample information about some parameters.

The case of less extreme collinearity is not dealt with so trivially since there is nothing in the least-squares theorems that is obviously dependent on the "near non-invertibility" of the design matrix. This fact has led Kmenta (1971, p. 391) to conclude "that a high degree of multicollinearity is simply a feature of the sample that contributes to the unreliability of the estimated coefficients, but has no relevance for the conclusions drawn as a result of this unreliability."

To put this another way, the problem of defining collinearity may be solved by identifying a distance function for measuring the closeness of the design matrix to some non-invertible matrix in which the collinearity problem is unambiguously extreme. Since the extreme case is associated with infinite marginal variances on the parameters, authors such as Theil (1971, p. 152), Malinvaud (1970, p. 218), and Goldberger (1964, p. 193) use a distance function informally related to the sampling variance of the coefficients. Collinearity is defined as large variances. The failure of this definition is that instead of defining a new problem, it identifies a new cause of an already well-understood problem — weak evidence. Although collinearity as a cause of the weak evidence problem can be distinguished from other causes such as small samples or large residual error variances, collinearity as

Received for publication November 7, 1972. Revision accepted for publication January 23, 1973.

* Research for this paper was supported by NSF Grant GS 31929. The author has benefited from conversations on the subject with Gary Chamberlain and Richard Kopcke. Both the discussion and the content have benefited significantly from a referee's comments. An earlier version was presented at the NBER-NSF Seminar in Bayesian Inference in Econometrics at the University of Minnesota, October, 1972.

[1] See Zellner (1971, chapter 2) for a discussion of the problem of defining complete uncertainty.

a problem is by this definition indistinguishable from the weak data problem in general. Thus Kmenta's conclusion that there is really nothing special about the collinearity problem is appropriate. Still a gnawing confusion remains. Goldberger (1964, p. 201) concludes this discussion with accurate ambiguity, ". . . when orthogonality is absent the concept of the contribution of an individual regressor remains inherently ambiguous."

The point of this paper is that there *is* a special problem caused by collinearity. This is the problem of *interpreting* multidimensional evidence. Briefly, collinear data provides relatively good information about linear combinations of coefficients. The interpretation problem is the problem of deciding how to allocate that information to individual coefficients. This will depend on prior information. A solution to the interpretation problem thus involves formalizing and utilizing effectively all prior information. The weak evidence problem will remain even when the interpretation problem is solved. The solution to the weak evidence problem is more and better data. Within the confines of the given data set there is nothing that can be done about weak evidence.

There are several threads in the literature which logically are part of the discussion that follows. Malinvaud (1970, p. 219) draws a two-dimensional confidence ellipse in the collinear case. His drawing suggests some trade-off in the sample evidence concerning the coefficients. This is quite clear in the textbook discussion of misspecification: dropping variables will drastically effect the estimates of other coefficients. The pretesting literature referenced in Raduchel (1971) and Leamer and Chamberlain (1972) explicitly exploits this fact. Constraints perhaps depending on preliminary tests may be used to improve the mean-square error of an estimator by reducing variance more than the square of the bias. But that literature usually concludes that a pretest estimator is desirable only when there is prior information that the constraint is relevant. These are exactly the two critical aspects of the problem as discussed below: (1) the existence of trade-offs in the sample evidence because of the peculiar shape of the likelihood contours, and (2) the existence of prior information that induces us to take advantage of those trade-offs.

There are several aspects of the collinearity problem that are explicitly not discussed in this paper. We assume that the model is completely specified and that there are no observation errors. We assume the existence of a computer with enough significant digits that any matrix of full rank can be inverted. We have before us the least squares estimate and variance-covariance matrix. Our only task is to interpret these sufficient statistics. We will not discuss in detail the weak data problem since it seems relatively straightforward.

Since we claim that the collinearity problem is associated with undominated uncertain prior information we will use an explicitly Bayesian approach. Bayesian analysis of the linear model by Raiffa and Schlaifer (1961) makes clear the fact that a proper prior distribution implies a proper posterior distribution. There could hardly be a problem of interpreting evidence in such a framework. Malinvaud (1970, pp. 246–249) points out that when the design matrix is non-invertible there is a linear subspace on which the prior and posterior coincide. This is the Bayesian analogue of the weak data problem which fails to distinguish a small sample from multicollinearity. Zellner (1971, pp. 75–81) provides a Bayesian analysis when the design matrix is singular and provides a Bayesian interpretation of a sampling theory approach using generalized inverses.

As is indicated by all of these authors, a Bayesian with a well-defined prior distribution can have no problem interpreting the sample evidence. A Bayesian with poorly defined priors or a wide readership may have extreme difficulties in reporting and interpreting evidence. This suggests the following definition:

Definition a: The *collinearity problem* will be said to affect a parameter $\hat{\beta}_i$ if the apparent sample evidence about β_i depends on uncertain prior information about other parameters. This will be made more precise below.

Since classical inference provides no assistance in using uncertain prior information, this definition does not apply directly to everyday "classical" inferences. The easiest thing to do is to neglect the off-diagonal terms of $(X'X)^{-1}$

and to proceed as if the sample evidence were generated by an orthogonal experiment. This may lead to significant misinterpretations of the data and suggests an alternative definition:

Definition 1: The *collinearity problem* is said to affect β_i if the sample evidence about β_i is distorted by an analysis which proceeds as if the data were orthogonal.

These two definitions are more fully explored in sections V and IV of this paper, respectively. Sections II and III provide the introductory groundwork. The reader is warned again that we are discussing the interpretation problem not the weak data problem, the latter being obvious.

II Conditional Bayesian Analysis of the Normal-Linear Model

Bayesian analysis of the linear model may be found in Raiffa and Schlaifer (1961) and Zellner (1971). Briefly, the assumptions are

(a) A conditional regression process

$$\begin{array}{ccccccc} (T\times 1) & & (T\times k) & (k\times 1) & & (T\times 1) \\ Y & = & X & \beta & + & u \end{array} \qquad (1)$$

with u distributed Normally with mean zero and variance $h^{-1}V$ with V assumed known and X independent of β and u, (Note: $h^{-1} = \sigma^2$)

$u \sim N(0, h^{-1}V)$.

(b) A conditional prior distribution for β

$\beta \sim N(b_1, h_1{}^{-1}N_1{}^{-1})$

with variance depending on an uncertain scale factor h_1.

(c) A joint distribution for h and h_1. This will be discussed in the next section.

Conditional on h and h_1 (and V) the posterior distribution is well known to be Normal with precision matrix and mean

$$H_2(h_1,h) = hN + h_1N_1 \qquad (2)$$

$$b_2(h_1,h) = (H_2(h_1,h))^{-1}\,(hNb + h_1n_1b_1) \qquad (3)$$

where

$N = X'V^{-1}X$

and b is any solution to

$Nb = X'V^{-1}Y$.

Note that the posterior mean is dependent only on the variance ratio $\rho = h_1/h$

$$b_2(\rho) = (N+\rho N_1)^{-1}(Nb+\rho N_1b_1), \qquad 0 \leqq \rho \leqq \infty. \qquad (4)$$

Let us now turn to the geometry of these equations. The likelihood contours are a family of concentric ellipsoids centered at the least squares estimates b

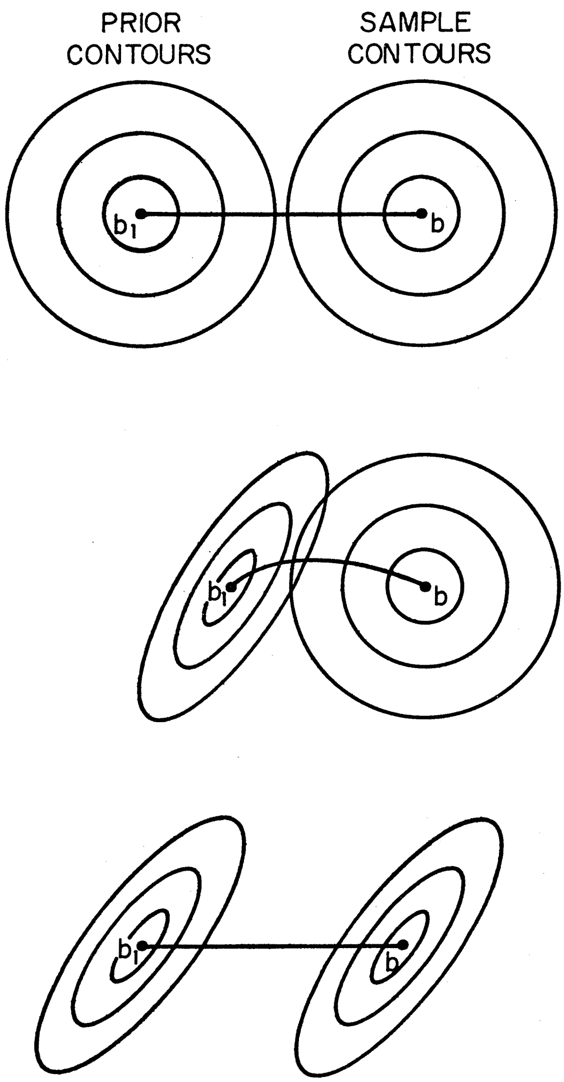
$Q(\beta,b,N) = (\beta-b)'N(\beta-b) = k$.

The conditional prior contours are also a family of concentric ellipsoids governed by the equation

$Q(\beta,b_1\,N_1) = k^*$.

Examples are graphed in figure 1.

FIGURE 1. — CURVES DÉCOLLETAGE



The points of tangencies between these ellipsoids may be found mathematically by maximizing one of the quadratic forms subject to the other being a constant. This is a

Lagrangean problem requiring the derivatives of

$$Q(\beta,b,N) - \lambda[Q(\beta,b_1,N_1) - k]$$

to be set to zero. That is,

$$2N(\beta-b) - 2\lambda N_1(\beta-b_1) = 0$$

which solves to

$$\begin{aligned}\beta &= (N+\lambda N_1)^{-1}(Nb+\lambda N_1 b_1)\\ &= (N+\rho N_1)(Nb+\rho N_1 b_1) \qquad (5)\end{aligned}$$

with suitable choice of ρ depending on λ. Note that equation (5) is identical to the posterior mean equation (4).

A definition and several observations may now be made: *Definition*: The posterior mean $b_2(\rho)$ as a function of the variance ratio ρ sweeps out a curve in k-space as ρ varies from zero to ∞. This curve is affectionately called the *curve décolletage* by Dickey (1971).

Observations: The *curve décolletage*

(a) is the locus of tangencies between the prior elliptical contours and the likelihood elliptical contours.

(b) may include *any* point in the space excluding points on the line through b and b_1, but including the line segment $[b,b_1]$.[2]

(c) is constrained to the hypersphere with diameter $[b,b_1]$ when N and N_1 commute, $NN_1 = N_1N$.[3]

(d) is a straight line when N is proportional to N_1.

(e) includes all modes of the conditional posterior distribution.

(f) lies in the rectangular solid with corners b and b_1 if N and N_1 are diagonal.

III Marginal Analysis

The previous section dealt with inference conditioned on the sample scale parameter h and the prior scale parameter h_1. The marginal posterior for β requires that we integrate out these parameters

$$f(\beta|Y) \propto \int_h \int_{h_1} f(\beta|Y,h,h_1)\, f(h,h_1|Y)\, dh_1 dh.$$

Independent of any assumptions about the prior, the marginal posterior mean is a weighted average of points on the curve décolletage

$$E(\beta|Y) = \int_h \int_{h_1} E(\beta|h,h_1,Y)\, f(h,h_1|Y)\, dh_1 dh.$$

In the absence of meaningful, generally-acceptable priors for h and h_1, we may prefer to analyze and to report the curve décolletage in its entirety. In the next section, we shall define multicollinearity in terms of the shape of the curve décolletage.

For reference, the textbook assumptions about priors for h and h_1 are

(1) Known process variance, conjugate prior; h and h_1 known. (See Raiffa and Schlaifer (1961).)

(2) Unknown process variance, conjugate prior; $h = h_1$, h has a gamma distribution. (See Raiffa and Schlaifer (1961).)

(3) Unknown process variance, marginal Student prior for β independent of h; h and h_1 have independent gamma distributions. (See Dickey (1971).)

[2] See Leamer (1972) or Chamberlain and Leamer (1973).

[3] See Pratt (1971).

IV Multicollinearity: A Bayesian Analysis of the Classical Problem

Although it is possible to make enlightened use of prior information through interpretative searches, perhaps as suggested by the pretesting literature, we shall assume that a researcher has before him only the sufficient statistics and no computer, as would be the case of a reader of a technical report. Off-diagonal terms of N^{-1} may not be reported and, even if they are, classical inference provides no very clear way of interpreting them. Instead, many of us in this situation would proceed as if N^{-1} were diagonal. Furthermore, when prior information on the coefficients is not dominated, it may be entered into the analysis informally on a coefficient-by-coefficient basis; that is, N_1^{-1} is also treated as though it were diagonal. This informal Bayesian analysis implies a posterior mean that lies on the *diagonalized curve décolletage*

$$b^*_2(\rho) = (D + \rho D_1)^{-1}(Db + \rho D_1 b_1) \qquad (6)$$

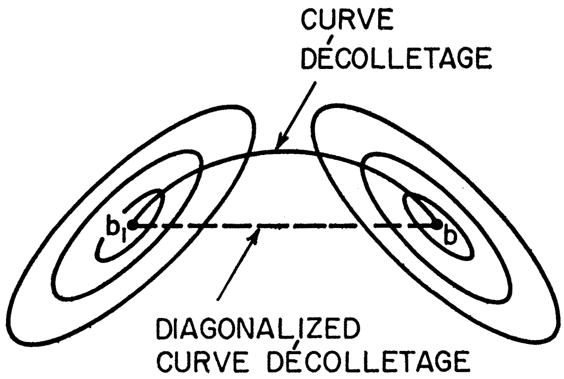
where D^{-1} and D_1^{-1} are diagonal matrices formed by setting the off-diagonal elements of N^{-1} and N_1^{-1} to zero.

If N and N_1 are diagonal, Bayes rule may be applied coefficient-by-coefficient and the resultant conditional posterior mean is given by (6). For N or N_1 nondiagonal, the true curve décolletage

$$b_2(\rho) = (N + \rho N_1)^{-1}(Nb + \rho N_1 b_1) \qquad (7)$$

may deviate substantially from $b^*_2(\rho)$ and our careless but easy use of prior information may cause major data misinterpretations and ultimately unnecessary expected losses (see figure 2). Collinearity thus provides an incentive to use prior information more carefully.

FIGURE 2. — CURVE DÉCOLLETAGE AND DIAGONALIZED VERSION



This incentive, that is the difference between $b_2(\rho)$ and $b^*_2(\rho)$ may be assessed in several ways:

(1) *Rectangle Test.* One aspect of a univariate problem is that the posterior mean is necessarily between the prior mean and the sample mean. Thus, the location change induced by the sample evidence has an unambiguous sign and is limited in distance by the sample point. The diagonalized curve décolletage also has this property, coefficient-by-coefficient; that is, it lies in the rectangular solid with diagonal $[b_1,b]$. The true curve décolletage need not have this property and the sign of the elements of $b_2(\rho) - b_1$ and $b_2(\rho) - b$ may be ambiguous. Thus, it may not be possible to say whether the data suggest positive or negative revisions to opinions about some coefficient or to limit the distance of the revision until a full prior distribution is specified. This suggests the following definition.

Definition 1′. The collinearity problem in the *sign sense* with respect to a prior with mean b_1 and variance $h_1{}^{-1}N_1{}^{-1}$ is said to affect a coefficient β_j if, for some value of ρ, $(b_j-b_{1j})(b_{2j}(\rho) - b_{1j}) < 0$, that is, if the location change is just the opposite in sign as that suggested by the location of the sample point. Symmetrically, the collinearity problem in the *distance sense* is said to affect a coefficient β_j, if for some value of ρ, $(b_j-b_{1j})(b_j-b_{2j}(\rho)) < 0$, that is if the location change satisfies the sign test but the distance of change exceeds that suggested by the location of the sample point. The collinearity problem in the *rectangle sense* is said to affect a coefficient β_j if the coefficient suffers from collinearity in either the sign or distance sense, that is if the curve décolletage travels outside the slab $b_j \leqq b_{2j}(\rho) \leqq b_{1j}$. The collinearity problem in the rectangle sense affects no parameters if the curve décolletage lies everywhere in the rectangular solid with diagonal $[b,b_1]$.

The curve décolletage may in general lie anywhere. Even with orthogonal data, that is with N diagonal, it need not be restricted to the appropriate rectangular solid. *Thus orthogonal data is not sufficient to prevent the collinearity problem. We must also restrict the class of priors.* If *both*, the sample information *and* the prior are orthogonal, N and N_1 are both diagonal and from (*f*) in the previous section there is no collinearity problem. If N is proportional to N_1, the curve décolletage is a straight line and there is no collinearity problem in the rectangle sense.

Since the case of orthogonal data is ordinarily thought to be free of the collinearity problem I would like to construct a "realistic" example that will illustrate the discussion above. Suppose that we have the following distributed lag process

$$Y_t = \beta_1 X_t + \beta_2 X_{t-1} + u_t$$

with fortuitously orthogonal data and with a least squares estimate (1,1). We will restrict our class of priors to those with mean (2,1). Is there a collinearity problem? Since the identity matrix commutes with any square matrix, the curve décolletage is restricted to the circle with diameter [(1,1),(2,1)]. Thus $E(\beta_1|Y)$ necessarily satisfies $1 \leqq (E(\beta_1|Y) \leqq 2$ and β_1 is *not* affected by collinearity in the rectangle sense. Since both the prior location and the sample location for β_2 are equal to one, we must have $E(\beta_2|Y) = 1$ for β_2 to be free of collinearity. This need not be the case. Figure 3 illustrates situations in which the sample evidence about β_2 leads us to increase our "estimate" (3a) or to decrease our "estimate" (3b). Case (a) applies when we know relatively much about the sum of the coefficients, or the long-run impact; Case (b) applies when we know relatively much about the shape of lag distribution. In Case (a) the sample is "used to estimate" the shape of the lag (both coefficients equal) and β_2 is increased to make it more like β_1. In Case (b), the sample is "used to estimate" the long run response $(\beta_1+\beta_2=2)$ and β_1 is reduced so that the sum is less.

FIGURE 3. — INTERPRETATION OF DISTRIBUTED LAG EVIDENCE

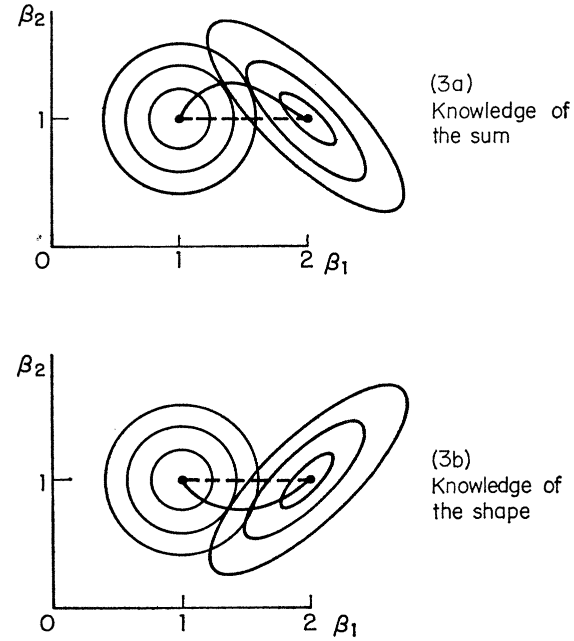
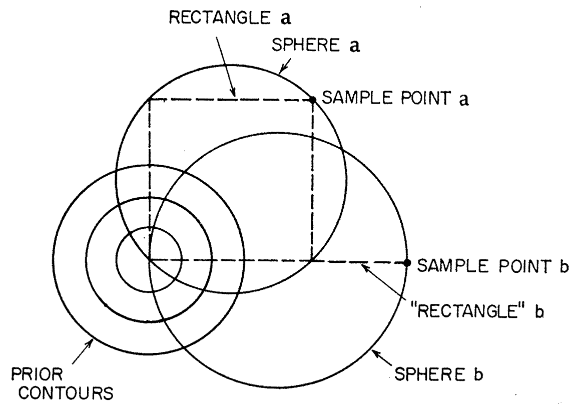


FIGURE 4. — FEASIBLE REGIONS



Another illustration is given in figure 4. Relative to a prior with fixed location and diagonal variance matrix, the likelihood of a collinearity problem depends on the sample location. The curve décolletage is constrained to the circle with radius $[b,b_1]$. This sphere is almost fully contained in the rectangle with diagonal $[b,b_1]$ for sample point a, but almost completely outside for sample point b. Thus the least amount of non-orthogonality of the evidence located at b will induce the collinearity problem in the rectangle sense, whereas very substantial non-orthogonality at point a is required to induce the collinearity problem.

(2) *Angle Test.* Collinearity may be absent in the rectangle sense but the true and the diagonalized curves décolletage may be quite different. One item of interest is the angle between the directional derivatives of b^*_2 and b_2 at b. These derivatives are

$$\left.\frac{\partial b^*_2(\rho)}{\partial \rho}\right|_{\rho=0} = D^{-1}D_1(b_1-b) = d^*$$

$$\left.\frac{\partial b_2(\rho)}{\partial \rho}\right|_{\rho=0} = N^{-1}N_1(b_1-b) = d$$

and the suggested collinearity measure is the cosine of the angle between d and d^*

$$c_1(N,N_1,b_1-b) = (d^{*\prime}d)/\sqrt{d'd}\,\sqrt{d^{*\prime}d^*}. \qquad (8)$$

This summary c_1 takes on the value one when $d = d^*$, that is when $b_2(\rho)$ and $b^*_2(\rho)$ coincide locally around b. In words, points of peculiar interest around b are clearly indicated by the diagonalized curve décolletage. When c_1 is 0, d and d^* are orthogonal. In words, points of interest around b suggested by $b^*_2(\rho)$ form a right angle with the points we should be interested in. Collinearity is clearly a problem. When c_1 is negative our approximate analysis takes us in a very undesirable direction and collinearity is extreme.

To re-emphasize a point of this paper: the quantitative measure of collinearity c_1 depends on prior information and the problem is personal. An impersonal bound may be found by minimizing c_1 over all priors.

In order to make this measure invariant to scale transformations let us standardize the variables so that N^{-1} has ones down the diagonal and $D = I$. The collinearity bound becomes

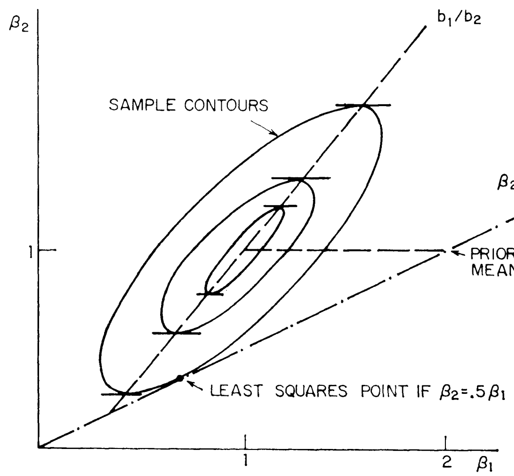
$$c_2 = \min_{b_1,N_1} c_1(D^{-1/2}ND^{-1/2},N_1,b_1-b).$$

Perverse negative values for c_1 can occur with highly correlated priors. For example, figure 5 illustrates

$c_1 = - |\rho|/1 + \rho^2$

where ρ is the correlation coefficient between the pair of coefficients. This occurs when the prior and the sample point have a component in common, and thus the diagonalized curve décolletage is a straight line. An appropriate analysis would take us in a direction more than ninety degrees from this line. To further explain why this can happen suppose in figure 5 that we "almost" know that $\beta_2 = .5\beta_1$ with a prior mean of $(\beta_1 = 2, \beta_2 = 1)$. If the sample point (1,1) is corrected for prior information on a coefficient-by-coefficient basis the resultant point is of the form $(w + (1-w)2,1)$ whereas the constrained least squares point and the rest of the relevant curve décolletage lie in the "opposite" direction.

FIGURE 5. — AN EXTREME COLLINEARITY PROBLEM



Note by the way that orthogonal data implies a value of zero for c_1, not the collinearity-free value of one. We might prefer to call this a prior collinearity problem and restrict data collinearity to cases when the prior is orthogonal. Then we would get the positive measure

$c_2 = 2\sqrt{\lambda_1\lambda_m}/(\lambda_1 + \lambda_m)$

where λ_1 and λ_m are the minimum and maximum eigenvalues of $D^{1/2}N^{-1}D^{1/2}$. This is proved in an appendix.

Loss Test

A measure such as c_1 or c_2 has one clear disadvantage. Although it tells us the angular error in the approximate curve décolletage, it does not indicate anything about the distance over which we make that error. A large angular error over a small distance may be less important than a small angular error over a large distance. Any attempt to include this distance notion in the collinearity measure requires a distance function, and that seems to mean a loss function.

Suppose, for example, that we wish to estimate β with quadratic loss

$L(\beta,\hat{\beta}) = (\beta - \hat{\beta})\, Q(\beta - \hat{\beta})$

where Q is symmetric positive definite. The posterior expected loss is

$$\begin{aligned} EL(\beta,\hat{\beta}) &= E(\beta-\bar{\beta})\, Q(\beta-\bar{\beta}) \\ &\quad + (\hat{\beta}-\bar{\beta})\, Q(\hat{\beta}-\bar{\beta}) \\ &= \mathrm{tr}QV(\beta) + (\hat{\beta}-\bar{\beta})\, Q(\hat{\beta}-\bar{\beta}) \end{aligned}$$

where $\bar{\beta}$ is the posterior expected mean of β. This is minimal for $\hat{\beta} = \bar{\beta}$.

Conditional on the variance ratio ρ the posterior mean is a point on the curve décolletage $\beta^*(\rho)$. The percentage increase in the mean squared error loss due to the use of the diagonalized curve is

$c_3 = (\beta(\rho) - \beta^*(\rho))'Q(\beta(\rho) - \beta^*(\rho))/\mathrm{tr}QV(\beta)$

where

$$\begin{aligned} \beta(\rho) - \beta^*(\rho) &= \rho[(N+\rho N_1)^{-1}N_1 \\ &\quad - (D+\rho D_1)^{-1}D_1](b_1 - b) \end{aligned}$$

and

$V(\beta) = (hN + hN_1)^{-1}$.

This measure of collinearity is a number between zero and infinity. Values above one suggest severe collinearity since estimation losses are at least double what they need be. The absence of a collinearity problem occurs at c_3 equal to zero. This can happen when

(a) N and N_1 are diagonal (orthogonal data and prior).

(b) ρ or N_1 are zero (prior information is dominated by the sample).

(c) $b = b_1$ (the least squares estimate and the prior location are in agreement; thus there is no need to select the posterior location different from b).

(d) $N = N_1$ (prior information has the same structure as sample information).

This third measure seems to capture the principal practical aspects of the collinearity problem as we informally know them. It furthermore highlights the importance of both the prior distribution and the loss function.

V Multicollinearity: The Bayesian Problems

Given a prior distribution, the posterior distribution is fully defined and there is no ambiguity about measures of location and thus no collinearity problem in the sense of the previous section. A Bayesian therefore apparently has no special difficulty working with collinear data. This, however, ignores difficulties in selecting an acceptable prior distribution.

When collinearity is present, the posterior distribution may be highly sensitive to changes in the prior and apparently innocuous differences in the prior may be amplified into significant differences in the posterior distribution. Thus the collinearity problem is transformed from a problem of characterizing and interpreting a multi-dimensional likelihood function into a problem of characterizing and interpreting a multidimensional prior distribution.

In the pure orthogonal case with N and N_1 diagonal, the posterior distribution of any coefficient is conditionally independent of the prior distributions of the other coefficients. This suggests the following slightly ambiguous definition of the collinearity problem: *Definition a*. The collinearity problem is said to affect parameter β_1 if the interpretation of the sample evidence about β_1 depends meaningfully on uncertain prior information about the other parameters $\beta_2, \beta_3, \ldots, \beta_k$. The *interpretation of the sample evidence* about β_1 is a mapping of marginal prior distributions for β_1 into marginal posterior distributions. The phrase *depends meaningfully* can be interpreted in terms of both the location change and the scale change induced by the sample evidence. As in the previous section, we will restrict ourselves to the location change.

In general, the interpretation of the sample evidence about one coefficient will be sensitive to the prior about others due to prior correlations. In order to make sense out of this definition we shall thus have to define meaningful classes of priors within which to perform the sensitivity analysis.

The sensitivity of the posterior mean to variation's in the prior mean holding other things constant is indicated by the matrix of derivatives

$$\partial b_2(\rho)/\partial b_1 = \rho(N+\rho N_1)^{-1}N_1 \tag{9}$$

The off-diagonal elements of this matrix indicate the extent to which the conditional posterior mean of one coefficient depends on the prior mean of the others. These will be zero for N and N_1 diagonal, for N proportional to N_1, and also for N_1 dominated by N.

We may also be interested in the sensitivity of the posterior mean to changes in the prior variances. Let us write

$$N_1^{-1} = DRD$$

where D is a diagonal matrix with $\sqrt{V}(\beta_i)$ on the diagonal and R is the matrix of correlation coefficients. A change in the prior variances induces a change in the prior precision matrix according to the formula

$$dN_1 = -(N_1D^{-1}dD + D^{-1}dDN_1),$$

and a change in N_1 induces a change in b_2 according to the formula

$$db_2(\rho) = (N+\rho N_1)^{-1}(\rho dN_1)(N+\rho N_1)^{-1} N(b_1-b).$$

In the diagonal case, the i^{th} element of $db_2(\rho)$ will depend only on the i^{th} differential $\{dD_i\}$. Otherwise changes in the prior variance of one coefficient will induce changes in the posterior means of other coefficients.

A somewhat expanded version of definition a would include sensitivity to prior information about the variance ratio ρ:

Definition a′: The collinearity problem is said to affect parameter β_1 if the sample evidence depends meaningfully on prior information about the variance ratio ρ.

In the orthogonal case, the posterior mean (the curve décolletage) lies everywhere in the rectangular solid with diagonal $[b,b_1]$. In that case the sign and maximum distance of the mapping from prior to posterior mean are unambiguous and we could say that the sample evidence does not depend meaningfully on prior information about β. When the curve décolletage travels outside this region, the sample evidence does become ambiguous and collinearity is the culprit. Note by the way that this is mathematically the same as the collinearity problem in the rectangle sense that was discussed in the previous section, although the interpretations are quite different. In that section the curve décolletage was assumed to lie in the relevant rectangular solid and when it did not, a major data misinterpretation occurred. Here, we know where the curve décolletage lies but we are uncertain whether particular points on the curve outside the rectangular solid are relevant since our prior information about the variance ratio is ambiguous.

Global sensitivity analysis of this form is discussed in more detail in Chamberlain and Leamer (1973). One interesting result is that if only the prior mean is specified the curve décolletage is constrained to lie in an ellipsoid passing through b and b_1 with center at $(b + b_1)/2$ and with axes the same as the sample ellipsoid. The projection of this bound onto the i^{th} axis is an interval centered at $(b_{1i} + b_i)/2$ with length $|b_{1i} - b_i|\sqrt{\chi^2}/Z_i$ where χ^2 is the chi-square statistic for testing the restriction $\beta = b_1$ and Z_i is the Normal statistic for testing $\beta_i = b_{1i}$. This compares with the rectangular restriction centered at $(b_{1i} + b_i)/2$ with length $|b_{1i} - b_i|$. The statistic

$$c_4 = \sqrt{\chi^2}/Z_i$$

thus indicates the extent to which collinearity amplifies the difficulties of interpreting evidence about β_i. (Note except for a degrees of freedom adjustment this is just $\sqrt{F}/t$.)

VI Summary

The most discussed aspect of collinearity is the weak data problem associated with large standard errors of estimated coefficients and, in a Bayesian analysis, the coincidence of prior and posterior distributions on certain subspaces. As Kmenta suggests there is nothing special about this problem in the collinearity context. What is special in the collinearity context is the major problem of interpreting the evidence. We have attempted to analyze these problems.

When prior information is fully specified and unique, both personally and publicly, the posterior mean and hence the interpretation of the evidence are unambiguous. The diagonalizations of the data and the prior that some of us implicitly perform may, however, lead to very poor approximations to the posterior mean. Qualitative and quantitative summaries of the error of approximation provide one way of assessing the collinearity problem.

Thus, the principal implication of collinearity is that data evidence cannot be interpreted in a parameter-by-parameter fashion. The informal use of non-data based prior information by practicing classical statisticians almost necessarily implies a parameter-by-parameter analysis, and consequently the data misinterpretation described above. The great benefit of a Bayesian approach is that it provides instruction on how to deal with prior information in a multi-parameter problem. For example, the posterior mean is under suitable assumptions, a *matrix* weighted average of the prior mean and the sample estimate, not a simple average.

Although the Bayesian approach appropriately spotlights the fundamental source of the collinearity problem — personal prior information — it necessarily leaves to the individual resolution of the problem through the construction of a personal prior distribution. Difficulties in constructing a personal prior and/or variation in opinions among intended readers may cause major difficulties in analyzing and reporting collinear evidence. Thus the problem of collinearity from a Bayesian viewpoint concerns the sensitivity of the posterior distribution to changes in the prior distribution, and quantitative measures of that sensitivity may be used to summarize the nature of the problem.

The principal claim of this paper is that the collinearity problem concerns the way in which sample evidence fits together with prior information. If prior information dominated sample evidence in all directions there would be no collinearity problem. When there is a collinearity problem classical inference which excludes undominated uncertain prior information fails as a method of interpreting evidence. Peculiarities in the likelihood surface make the BLUE (least squares) estimate almost irrelevant. A fuller exploration of the likelihood contours informally directed by prior information is difficult and rarely convincing, especially when the number of dimensions of prior information is more than one. Although a Bayesian approach cannot provide a complete cure, it does indicate the source of the disease and offers reasonable suggestions for improvement.

APPENDIX

Theorem: If $d^* = N_1(b_1 - b)$ and $d = N^{-1}N_1(b_1 - b)$, with N_1 a diagonal matrix and N symmetric positive definite then

$$c_2 = \min_{b_1,N_1} d^{*\prime}d/\sqrt{d^{*\prime}d^{*}d^{\prime}d}$$
$$= 2\sqrt{\lambda_1\lambda_m}/\lambda_1 + \lambda_m$$

where λ_1 and λ_m are the minimum and maximum eigenvalues of N^{-1}.

Proof: Consider the case when $N_1 = I$ and let

$$N^{-1} = \sum_i \lambda_i^2 P_i P'_i$$

where P_i and λ_i are eigenvectors and eigenvalues of N^{-1}. Then

$$N^{-1}N^{-1} = \sum_i \lambda_i^2 P_i P'_i$$
$$(b_1 - b) = \Sigma \propto_i P_i$$

and

$$c_1 = \Sigma\lambda_i \propto_i^2/\sqrt{\Sigma\propto_i^2\Sigma\propto_i^2\lambda_i^2}$$
$$= \Sigma\lambda_i w_i/\sqrt{\Sigma\lambda_i^2 w_i}$$

for

$$w_i = \propto_i^2/\Sigma\propto_i^2.$$

The derivative of this expression with respect to w_j may be set to zero which yields (with $w_1 = 1 - \Sigma w_j$)

$$2\Sigma\lambda_i^2 w_i - (\lambda_j + \lambda_1)\Sigma\lambda_i w_i = 0$$

which can be satisfied in two dimensions only. This leads to solutions on the edges of the simplex

$$w_1 = \lambda_j/(\lambda_1 + \lambda_j)$$
$$w_j = \lambda_1/(\lambda_1 + \lambda_j)$$
$$w_{j'} = 0; \quad j' \neq j$$

which implies

$$c_1 = 2\sqrt{\lambda_1\lambda_j}/(\lambda_1 + \lambda_j).$$

This is positive but minimal for the minimum and maximum eigenvalues.

For $N_1 \neq I$ but diagonal we have

$$d^* = N_1(b_1 - b) = z$$
$$d = N^{-1}N_1(b_1 - b) = N^{-1}z$$

for arbitrary $z = D^{-1}D_1(b_1 - b)$. This is the same problem as above.

REFERENCES

Chamberlain, G., and E. Leamer, "Matrix Weighted Averages and Double Student Densities," 1973 (mimeo).

Dickey, J., "The Bayesian Alternatives to the F-Test," SUNY at Buffalo Research Report no. 50, 1971.

Goldberger, A., *Econometric Theory* (New York: John Wiley and Sons, Inc., 1964).

Kmenta, J., *Elements of Econometrics* (New York: The Macmillan Company, 1971).

Leamer, E., "Another Note on Posterior Means," 1971 (mimeo).

Leamer E., and G. Chamberlain, "A Bayesian Theorem on Pretesting," 1972 (mimeo).

Malinvaud, E., *Statistical Methods of Econometrics* (Amsterdam: North-Holland Publishing Company, 1970).

Pratt, J., "A Note on Posterior Means," 1970 (mimeo).

Raduchel, W. J., "Multicollinearity Once Again," Harvard Institute of Economic Research Paper no. 205, 1971.

Raiffa, H., and R. Schlaifer, *Applied Statistical Decision Theory* (Boston: Graduate School of Business Administration, Harvard University, 1961).

Theil, H., *Principles of Econometrics* (New York: John Wiley and Sons, Inc., 1971).

Zellner, A., *An Introduction to Bayesian Inference in Econometrics* (New York: John Wiley and Sons, Inc., 1971).