A Guide to Using the Collinearity Diagnostics

David A. Belsley

December, 1989

Working Paper No. 190

A Guide to Using the Collinearity Diagnostics

David A. Belsley[1]

December 1, 1989

Abstract

The description of the collinearity diagnostics as presented in Belsley, Kuh, and Welsch's, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* is principally formal, leaving it to the user to implement the diagnostics and learn to digest and interpret the diagnostic results. This paper is designed to overcome this shortcoming by describing the different graphical displays that can be used to present the diagnostic information and, more importantly, by providing the detailed guidance needed to promote the beginning user into an experienced diagnostician and to aid those who wish to incorporate the collinearity diagnostics into a guided–computer environment.

Key Words

Regression Diagnostics, Ill–Conditioning, Condition Numbers, Condition Indexes, Guided–Computing, Variance–Decomposition Proportions

The collinearity diagnostics introduced by Belsley, Kuh, and Welsch in 1980 in *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (BKW) provide all users of linear regression with much useful information regarding the strength of a given set of data for estimating a specific linear model with ordinary least squares (OLS). The information can be used effectively to strengthen one's understanding of the reliability of a given regression model and its estimates for the

[1] Boston College and MIT. This paper was funded in part through a guest scholarship at the Rheinisch–Westfälisches Institut für Wirtschaftsforschung, Essen, West Germany, and in part through NSF Grant SES–8420614.

purposes of forecasting, research, or policy–making. Examples of its use in a wide variety of situations, both diagnostic and corrective, are to be found in various of the items listed in the bibliography. However, it appears that the "mass" of diagnostic material that results from a typical collinearity analysis can be overwhelming and indigestible to the new user, often inhibiting the further use and experimentation needed to gain the experience necessary to "read" the diagnostic evidence quickly and efficiently and to make effective use of the diagnostic procedure. Further, a lack of appropriate guidelines has made it difficult for those who have sought to incorporate these collinearity diagnostics into guided–computer environments. It is the purpose of this paper to supplement the presentation in BKW to correct for these shortcomings.

We begin in the first section with a very brief review of the diagnostic elements. It is, however, assumed that the reader is already familiar with Chapter 3 of BKW, so this review makes no great effort at motivation. In the second section a series of steps is described for carrying out the diagnostic procedure. This provides a useful revision of the corresponding elements of BKW. The final section, which is wholly new and is the essence of this paper, provides a set of hints for usage that detail how quickly to digest, interpret, and assess any given set of diagnostic results. This material is designed to promote the novice user to an experienced practitioner in short order.

1. A BRIEF REVIEW OF THE DIAGNOSTIC ELEMENTS

The collinearity diagnostics of BKW are based on two elements relevant to an $n \times p$ data matrix $\mathbf{X}$ used in a linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$: the scaled condition indexes and the variance–decomposition proportions. Both of these diagnostic elements are obtainable from the singular–value decomposition (SVD) of the matrix $\mathbf{X}$ as

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{T} \qquad (1)$$

where $U^TU = V^TV = I_p$ and D is diagonal with nonnegative diagonal elements $\mu_1, \ldots, \mu_p$, called the *singular values* of X.[2]

The condition indexes of the matrix X are simply

$$\eta_k \equiv \frac{\mu_{max}}{\mu_k} \qquad k = 1, \ldots, p, \tag{2}$$

where μ_{max} is the largest singular value. Naturally, $\eta_k \geq 1$, for all k, and it is shown in Chapter 3 of BKW that, for each large η_k, there is a near linear dependency among the columns of X, and the larger is η_k, the stronger is the corresponding near linear dependency.

The variance–decomposition proportions, which combine the information from U and D, arise from the fact that, using the SVD (1), the variance–covariance matrix of the least–squares estimator $b = (X^TX)^{-1}X^Ty$ may be written as

$$V(b) = \sigma^2(X^TX)^{-1} = \sigma^2VD^{-2}V^T, \tag{3}$$

where σ^2 is the variance of the disturbance term ε. Thus, the variance of the k–th regression coefficient, b_k, which is the k–th diagonal element of (3), is simply

$$\text{var}(b_k) = \sigma^2 \sum_j \frac{v_{kj}^2}{\mu_j^2}, \tag{4}$$

where the μ_j's are the singular values of X and $V \equiv (v_{ij})$.

Note that (4) decomposes $\text{var}(b_k)$ into a sum of components, each associated with one and only one of the p singular values μ_j of the n × p matrix X. Since these μ_j^2 appear in the denominator, other things being equal, those components associated with near dependencies — that is, with small μ_j — will be large relative to the other

[2] See, for example, Golub (1969), Golub and Reinsch (1970), Stewart (1973), and Becker et al. (1974).

components. This suggests, then, that an unusually high *proportion* of the variance of two or more coefficients concentrated in components associated with the same *small* singular value provides evidence that the variates (columns of X) corresponding to those coefficients are involved in the near dependency corresponding to that small singular value. The preceding says "two or more" because it is clear that it takes at least two variates to make a collinear relation.[3]

Define the k,j–th *variance–decomposition proportion* as the proportion of the variance of the k–th regression coefficient associated with the j–th component of its decomposition in (4). These proportions are readily calculated as follows:

First, let

$$\phi_{kj} \equiv \frac{v_{kj}^2}{\mu_j^2} \quad \text{and} \quad \phi_k \equiv \sum_{j=1}^{p} \phi_{kj} \qquad k = 1, \ldots, p. \tag{5}$$

Then, the variance–decomposition proportions are

$$\pi_{jk} \equiv \frac{\phi_{kj}}{\phi_k}, \qquad k, j, = 1, \ldots, p. \tag{6}$$

These variance–decomposition proportions are most easily digested when summarized in a Π matrix, that is, a table like that in Exhibit 1, although different forms will be discussed below. Here, each row corresponds to a given singular value, μ_j, or, equivalently, the associated condition index, $\eta_j \equiv \mu_{max}/\mu_j$. These rows can be ordered so that the condition indexes are in increasing (or decreasing) order. Naturally, the columns of π's should sum to one. Interest centers on patterns where two or more variates have large values associated with the same high condition index. We shall have a good deal more to say about how to view and interpret these tables as we proceed.

[3] One can see from Belsley (1982), however, that a closely allied problem, short data, can affect a single variate.

Exhibit 1 Π matrix of variance–decomposition proportions

Condition	Proportions of			
Index	var(b_1)	var(b_2)	. . .	var(b_p)
η_1	π_{11}	π_{12}	. . .	π_{1p}
η_2	π_{21}	π_{22}	. . .	π_{2p}
.	.	.		.
.	.	.		.
.	.	.		.
η_p	π_{p1}	π_{p2}	. . .	π_{pp}

Care must be taken in forming the **X** matrix to which the collinearity diagnostics are applied. Data matrices that differ from one another only by the scale assigned their columns (matrices of the form **XB**, where **B** is a nonsingular diagonal matrix with positive diagonal elements) represent essentially equivalent information; it does not matter, for example, whether one specifies monetary data in dollars, cents, or billions of dollars. It is very clear from above, however, that such scale changes do affect the numerical properties of the data matrix and result in very different variance–decomposition proportions and condition indexes.[4] Without further adjustment, then, we have a situation in which near dependencies among structurally equivalent economic variates (differing only in their units of measurement) can result in greatly differing condition indexes. Clearly, the condition indexes can provide no stable information to the user of linear regression on the degree of collinearity among the **X** variates in such a case.

It is necessary, therefore, to standardize data matrices that correspond to equivalent model structures in a way that makes comparisons of condition indexes meaningful. A natural standardization is to scale each column to have equal length — column–equilibration. This scaling is natural because it transforms a data matrix **X** with mutually

[4] Such scale changes do not, however, affect the presence of exact linear dependencies among the columns of **X**, since, for any nonsingular matrix **B**, there exists a nonzero **c** such that $\mathbf{Xc} = 0$ if and only if $[\mathbf{XB}][\mathbf{B}^{-1}\mathbf{c}] \equiv \bar{\mathbf{X}}\bar{\mathbf{c}} = 0$, where $\bar{\mathbf{X}} = \mathbf{XB}$ and $\bar{\mathbf{c}} = \mathbf{B}^{-1}\mathbf{c}$. A more general discussion of the effects of column scaling is to be found in Belsley, Kuh, and Welsch (1980).

orthogonal columns, seemingly ideal data, into one whose condition indexes would all be unity, the smallest and most ideal possible condition indexes. Any other scaling would fail to reflect this desirable property. And, an important converse is true with column–equilibrated data; namely, when all condition indexes of a data matrix are equal to unity, the columns are mutually orthogonal. This is readily proved by noting that, if all condition indexes are equal to 1, then, in the singular–value decomposition of **X**, the **D** matrix must take the form $\mathbf{D} = \lambda\mathbf{I}$, for some λ. Hence, we have $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \lambda\mathbf{U}\mathbf{V}^T$, or $\mathbf{X}^T\mathbf{X} = \lambda^2\mathbf{V}\mathbf{U}^T\mathbf{U}\mathbf{V}^T = \lambda^2\mathbf{I}$, due to the orthogonality of **U** and **V**. This result is important, because it rules out the possibility that several high variance–decomposition proportions could be associated with a very low (near unit) condition index. Further, it is shown in Appendix 3B of BKW that column–equilibration results in condition numbers that are most meaningful as a collinearity diagnostic.

The exact length to which the columns are scaled is unimportant, just so long as they are equal, since the condition indexes are readily seen to be invariant to scale changes that affect all columns equally. But, as a matter of practice, we effect column–equilibration by scaling each column of **X** to have unit length. This scaling is similar to that used to transform the cross–products matrix $\mathbf{X}^T\mathbf{X}$ into a correlation matrix, which appears to cause some to confuse these two issues. Thus, it is worth re–emphasizing that (1) any equal length, not just unit, would suffice for our purposes here, and (2) whereas the columns have been scaled, they have not been centered to have zero means, as would be needed for correlations. Indeed, it is shown in Belsley (1984a and 1986) that centering, while useful for some purposes, is almost always inappropriate for analyzing collinearity, and indeed, if done, tends to produce misleading conditioning diagnostics.

By way of terminology, then, we continue to define the indexes $\eta_i(\mathbf{X})$ of a matrix **X** by (2), where the singular values are those of **X**. But we now introduce the *scaled condition indexes*, $\tilde{\eta}_i(\mathbf{X})$ of the matrix **X**, which are similarly defined, except that the singular values are those that result from applying the singular–value decomposition to $\mathbf{X}$ after its columns have first been scaled to have equal (unit) length.

That is, if $\mathbf{X} = [\mathbf{X}_1 \cdot \cdot \cdot \mathbf{X}_p]$, $s_i \equiv (\mathbf{X}_i{}^{T}\mathbf{X}_i)^{-1/2}$, and $\mathbf{S} \equiv \mathrm{diag}(s_1, \ldots, s_p)$, then

$$\tilde{\eta}_i(\mathbf{X}) \equiv \eta_i(\mathbf{XS}) \qquad i = 1, \ldots, p. \tag{7}$$

We can similarly refer to the scaled variance–decomposition proportions as those relevant to the scaled matrix $\mathbf{XS}$.

This terminology may at first appear confusing since a scaled condition index is not a condition index that has been scaled, but rather is a condition index of a matrix that has been scaled (column–equilibrated). However, this confusion is slight and does not outweigh the economy of terminology that otherwise results.

2. THE STEPS

Diagnosing any given data set for the presence and composition of near dependencies and assessing the potential harm that their presence may cause least–squares regression estimates is effected by a rather straightforward series of steps, the only problems of interpretation arising when there competing and dominating near dependencies. We begin here with a brief description of the steps to be followed and then provide various hints to make their enactment simpler and more transparent.

STEP 1. Determine $\mathbf{X}$.

STEP 2. Column–equilibrate $\mathbf{X}$.

STEP 3. Obtain scaled condition indexes and variance–decomposition proportions.

STEP 4. Determine number of near dependencies.

STEP 5. Determine variate involvement.

STEP 6. Determine auxiliary regressions.

STEP 7. Determine unaffected variates.

1. Determine **X**. Any **X** matrix can, in principle, be analyzed. However, if the data are being analyzed relative to estimating a linear regression model with least squares, it is assumed that the user has a specific parameterization β^* in mind and has transformed the raw data (if need be) to conform, so that the resulting **X** matrix is that relevant to the model in the form $\mathbf{y} = \mathbf{X}\beta^* + \varepsilon$. If an intercept is appropriate to the model, it should be made explicit, so that **X** has a column of ones. On a related matter, the data should not be centered. It is shown in Belsley (1984a or 1986) that, contrary to much popular opinion, centering the data does not get rid of "nonessential collinearity" and will rather generally mask the role of the constant in any underlying near dependencies and produce misleading diagnostic results. It is also shown that the diagnostic results are most meaningful when the **X** data are structurally interpretable. When this is the case, the corresponding model is said to be in *basic form.*

2. Column–Equilibrate **X.** Once a specific **X** matrix has been selected for analysis, it should be column–equilibrated; that is, each column should be scaled for equal Euclidean length. As noted above, unless this is done, the collinearity diagnostics can produce arbitrary results. The usual method is to scale each column $\mathbf{X}_i$ by its norm $\|\mathbf{X}_i\|$ so that the resulting $\mathbf{X}_i/\|\mathbf{X}_i\|$ has unit Euclidean length.

3. Obtain Scaled Condition Indexes and Variance–Decomposition Proportions. The best method for obtaining these fundamental pieces of diagnostic information is through the singular–value decomposition (1) of the column–scaled **X** matrix. Then,

a. the scaled condition indexes $\tilde{\eta}_k$ are the η_k as determined as in (2), and

b. the Π matrix of variance–decomposition proportions is determined as in (6) and Exhibit 1.

In the event that software for the singular–value decomposition is not available, the same formulas may instead be applied to the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and the square roots of the corresponding eigenvalues. For the reasons given in BKW, however, the use of the singular–value decomposition is to be preferred.

*4. **Determine the Number of Near Dependencies.*** Determine the number and relative strengths of the near dependencies by the scaled condition indexes exceeding some chosen threshold $\tilde{\eta}^*$, such as 30. The choice of this threshold is somewhat of an art form. It is not a classical significance level that must be chosen a priori, and is best chosen relativistically, depending on the pattern of scaled condition indexes that arises. More will be said on this shortly. The relative strengths of the scaled condition indexes are determined by their approximate position along the progression 1, 3, 10, 30, 100, 300, 1000, and so on.

*5. **Determine Variate Involvement.*** Rather generally, a variate is considered involved in, and its corresponding regression coefficient degraded by, at least one near dependency if the total proportion of its variance associated with the set of high scaled condition indexes exceeds some chosen threshold π^*, such as 0.5. Two cases are to be considered.

Case 1. Only one near dependency is present. Here, there is only one high scaled condition index and it is possible to determine variate involvement directly from the variance–decomposition proportions. A variate is considered involved (and its estimated coefficient degraded) if its variance–decomposition proportion associated with this single scaled condition index exceeds the threshold π^*.

Case 2. Coexisting or simultaneous near dependencies. Here there are several high scaled condition indexes. Variate involvement is now determined by aggregating the variance–decomposition proportions of each variate over the set of these several high condition indexes. Those variates whose *aggregate* proportions exceed the threshold π^* are involved in at least one of the near dependencies (and their corresponding estimated coefficients are degraded). If, in addition, the variance–decomposition proportion of a given variate exceeds π^* by itself in one of the several near dependencies, its involvement in that relation is typically indicated. This could be contradicted when there are competing near dependencies (several near dependencies with roughly equal condition indexes), so care should be taken here. Thus, when there are several near dependencies, the elements of the Π matrix can always be used to determine which variates are involved in at least one near dependency (and therefore which coefficients are degraded), but they cannot always be relied upon to determine which variates are involved in which specific near dependencies.

*6. **Determine Auxiliary Regressions.*** Once the number of near dependencies has been determined in STEP 4 and some partial knowledge of variate involvement has been determined in STEP 5, auxiliary regressions among the indicated variates can be run to display the near dependencies in greater detail. As noted, these auxiliary regressions are not needed to determine either the number of near dependencies or the variates that are degraded by being in at least one of them, but they should be used if more detailed information is required to determine individual variate involvement among specific competing or dominating near dependencies. A simple procedure for forming these auxiliary regressions is described below.

*7. **Determine Unaffected Variates.*** A variate is considered uninvolved in any near dependency if the total proportion of its variance associated with the set of low scaled condition indexes exceeds the threshold π^*. The set of low scaled condition indexes is, of course, the set that does not exceed the threshold $\tilde{\eta}^*$ specified in STEP 4.

3. SOME HINTS ON USAGE

The mass of numbers presented by the Π matrix of variance–decomposition proportions in a collinearity diagnosis can be quite overwhelming until one learns what to look for — and what to ignore. This is particular true when p, the number of variates, gets large. Fortunately, this is an easy problem to correct, and we turn in this section to some hints for reading, interpreting and using the collinearity diagnostics. With a little instruction and practice, the salient features of the output of a collinearity diagnosis can be almost instantly digested, literally at a glance. We begin with a discussion of the possible formats that may be chosen for the output tableaux used to present the diagnostic information.

Possible Formats

The collinearity diagnostics are comprised of two basic blocks of information, the p scaled condition indexes and the p × p Π matrix of variance–decomposition proportions. There are two natural formats for displaying this information: the row–oriented format and column–oriented format. Each has its advantages and disadvantages.

The row–oriented format, which was chosen for the displays in BKW, is shown in Exhibit 2. In this format, the condition indexes and Π matrix are combined into a

Exhibit 2 Output tableau for the collinearity diagnostics, row–oriented format

Scaled	Proportions of			
Condition	X_1	X_2		X_p
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	. . .	var(b_p)
$\tilde{\eta}_1$	π_{11}	π_{12}	. . .	π_{1p}
$\tilde{\eta}_2$	π_{21}	π_{22}	. . .	π_{2p}
.	.	.		.
.	.	.		.
.	.	.		.
$\tilde{\eta}_p$	π_{p1}	π_{p2}	. . .	π_{pp}

$p \times (p+1)$ matrix having the condition indexes displayed as the first column so that they act essentially as row headings.[5] It is clearly advantageous that the condition indexes be ordered, but it makes little difference whether the order is ascending or descending. The former has been chosen in this monograph. Each row in this format, then, corresponds to a near dependency, and, since it is the existence and composition of these near dependencies that is the main focus of a collinearity diagnosis, it is the structure and patterns of these rows that are the main focus of the analysis — hence the term "row–oriented" format.

Each column in this format can be associated equally well with a variate or a variance. Viewed strictly as a collinearity diagnostic, the πs are indirect evidence of the relevance of the variates to the various near dependencies. Within this interpretation, each column of the Π matrix is associated with a column of $\mathbf{X}$, or a variate. Viewed, however, as a regression diagnostic — where we are concerned also with the strength of the given data for estimating a regression model by least squares — each column can be associated with the variance of the estimated coefficient of the corresponding variate. The column headings in a tableau like Exhibit 2 then, can either be a variance name, or the name of the corresponding variate, or both.

The column–oriented format transposes the preceding to give a tableau like Exhibit 3. Here, the scaled condition numbers become the column heads so that each column corresponds to a near dependency. Each row, which must sum to one, now corresponds to a variate and/or a variance.

There are several advantages to this format. First, when p is large, since there can be more rows on a page than columns, the columns associated with each near dependency are more likely to stay together on a page, rather than being broken across several pages. Visually, this makes assessment of the near dependencies easier. Second,

[5] In Belsley, Kuh, and Welsch (1980), a column of singular values is also included in the display tableaux. The singular values, however, are quite unnecessary, and their presence just creates unneeded confusion.

Exhibit 3 Output tableau for the collinearity diagnostics, column–oriented format

Variate or variance	Scaled Condition Index			
	$\tilde{\eta}_1$	$\tilde{\eta}_2$		$\tilde{\eta}_p$
X_1	π_{11}	π_{21}	. . .	π_{p1}
X_2	π_{12}	π_{22}	. . .	π_{p2}
.	.	.		.
.	.	.		.
.	.	.		.
X_p	π_{1p}	π_{2p}	. . .	π_{pp}

if the scaled condition numbers are placed in descending order, only the first several columns, those corresponding to the largest condition indexes, may need to be printed. Third, this format makes it somewhat easier to produce clean computer output on line printers that cannot print sideways on the page. In more flexible computer environments, however, reasonable programming can make either the row– or column–oriented formats equally effective. The main disadvantage of this format is that its presentation is less appealing to the intuition.

Keep the Numbers Simple

Regardless of the format, the numbers that are printed should be kept simple. It is, for example, completely unnecessary to print any fractional part of the scaled condition indexes; only the integer part is ever needed. Recall that these magnitudes tend to group along the 1, 3, 10, 30, 100, 300 progression. Thus, scaled condition indexes of 28 and 35 are essentially the same, as are ones of 109 and 92. It is clear that, under these circumstances, the value of seeing a condition index of 34.859 rather than 35 is nil. Correlatively, when viewing scaled condition indexes, it is unnecessary to take the trouble mentally to digest the full number; merely an appreciation for the relative order of magnitude is adequate.

Similarly, the variance–decomposition proportions need not be carried out to numerous digits. The three digits to the right of the decimal shown for the π_{ij} in the displays later in this paper are wholly adequate. Leading zeros should also be suppressed; they tend only to clutter the display with unnecessary ink and to distract the eye from seeing quickly the one instance in which a digit will meaningfully be to the left of the decimal, namely in the interesting and unusual case where the variance–decomposition proportion is 1.000. Furthermore, the displays are most easily read when each π_{ij} has the same number of digits to the right of the decimal. Thus, trailing zeros are to be kept. And finally, FORTRAN–like exponential formats should be avoided at all costs. This formatting technique, so useful in some other circumstances, renders the tableaux of the collinearity–diagnostics almost unreadable.

Look at the High Scaled Condition Indexes First

Given, then, the diagnostic output, the first thing to do is mentally to blot from vision everything but the scaled condition indexes, and indeed to focus first only on their high end. The largest condition index, which is also the scaled condition number of the data matrix, tells you immediately the worst with which you must contend. It defines for you relativistically what is "large" in the context of the given data. If "large" is itself absolutely small, for example, 5 or 10, then, although further analysis may be of interest, collinearity is not really a major problem besetting these data. If "large" is moderate, say 30 to 100, then there are collinearity problems and further analysis is definitely of interest. If, however, "large" is immense, say, 1000 or 3000, then condition indexes that are several orders of magnitude smaller, even ones like 30 that might, by themselves, be of concern, are small by comparison and are not necessarily of major concern.

The next thing to look for is the possible presence of a gap in the progression of the condition indexes. This gap may occur because of a separation between a scaled condition index which is absolutely small from one that is large (a gap of the first kind), or it may occur between large condition indexes that are separated by several orders of magnitude along the usual progression (a gap of the second kind). Such a gap provides a natural starting place to determine the number of near dependencies.

In a progression such as 1, 3, 5, 30, the gap is of the first kind, between 30, which is high, and 5, which is absolutely low; one near dependency is indicated here. This situation is even easier to analyze in a case like 1, 3, 5, 100, where the gap is across several orders of magnitude. Likewise, in 1, 3, 5, 30, 100, there is little reason not to interpret this again as a gap between 5 and 30, indicating two near dependencies. Seemingly trickier is a progression like 1, 3, 5, 24, 32, 100. However, since 24 and 32 are really of the same order of magnitude, this is really quite similar to the preceding case except that three near dependencies are called for.

The real problems occur in highly unbalanced sequences and in very smoothly graded ones. In the case 1, 3, 5, 24, 32, 107, 1427, 3456, for example, a gap of the first kind exists between 5 and 24, indicating five near dependencies, while a gap of the second kind exists between 107 and 1427, indicating two. Clearly, this latter interpretation is taking liberties, for a scaled condition index of size 107 is indeed quite large. However, it is often advantageous to begin the analysis as simply as possible, examining only the worst cases. It is true that there may be other near dependencies that are thereby ignored, but their effects are relatively less, even if not absolutely without importance. Thus, one might begin here by assuming two near dependencies, only making refinements, if need be, at a later stage. There may, for example, be little reason to go further if it can be shown that all the variates of interest, or the relations of interest, are included in the two strongest near dependencies. On the other hand, if a variate of interest is shown not to be involved in the strongest near dependencies, its role in the next strongest set can then be investigated.

The most frustrating progressions are those that give no hint of a natural break. Consider, for example, the sequence, 1, 3, 5, 10, 22, 29, 39, 45, 72, 95, 129, 245, 373, 498. These nasty situations are more likely to arise when p, the number of variates, is large. My sympathies are often limited when confronted with cases like this with p = 25 or, in some cases, 50, for, by and large, it has been my experience that regression equations with more than 10 variates tend to be misspecified, resulting from ad hoc specifications that should properly be modelled as part of a simultaneous system of smaller equations. But not always, and that quip does not really answer the question of how to interpret the above sequence of scaled condition indexes, which could arise legitimately. My inclination is to begin the analysis by picking the top three to five, not ignoring anything astronomically large, but trying to keep the number of near dependencies below some reasonable proportion of p, say 25–40%. In this case, I would probably first pick the top three.

Thus, as noted above, picking the number of near dependencies is occasionally an art form. Sometimes it can be done quite mechanically, but often some degree of judgement is required. It helps somewhat to realize that the user is not required here to state a priori what the cutoff will be, as he would be, for example, in properly setting the test size for a classical test of hypothesis. The cutoff value is best determined relativistically according to the needs of the analysis at hand and can be changed as the analysis proceeds. The only absolute is that very small values for the scaled condition index, values of 5 or 10 and less, will rarely be of interest.

Train the Eyes to See Only the Important Variance-Decomposition Proportions

Having determined the number of near dependencies, one next looks at the variance–decomposition proportions that correspond to them, beginning with the strongest. In the row–oriented displays like those below, this means beginning with the bottom row. In the column–oriented format of Exhibit 3, this would be the first column. The eye is easily trained to ignore the elements in the other rows, and indeed to search first only for the large numbers in the given row, values like .8 and .9. There will almost always be some values like this in the row corresponding to the largest scaled condition index.[6] These, mentally, should stand out like they were in bold face. The columns these numbers are in indicate those variates that are definitely involved in the strongest near dependency, bearing in mind that there may be others as well, which are also being determined simultaneously in other near dependencies.

Now let the eye pick up the variance–decomposition proportions associated with the next largest scaled condition index (the penultimate row in the displays given here), and examine them for large values that indicate obvious variate involvement in this near dependency. Evidence should now also be sought as to simultaneous involvement of variates between this and strongest near dependency. This would obviously take the form of variance–decomposition proportions distributed across the two, so that their sum is large even if no single part is. Rough sums are all that is needed here. Continuing in this way, one can build a story of the collinear structure.

The main technique, then, for absorbing the information of the variance–decomposition proportions is to train the eye first to see only the most obvious few values: the large πs in the rows corresponding to the large $\tilde{\eta}$s. Once this information is digested and preliminarily interpreted, then allow the focus of the eye to broaden to encompass the more refined structure that occurs due to dominance and competition.

An Illustration

Thus, consider the sequence of tableaux given in Exhibits 4a through 4c, where actual bold–face type simulates the eye's discriminatory powers. Looking first at only the first column of Exhibit 4a, the following story begins to unfold: "There are three

[6] An exception could occur when there is strong competition and/or simultaneity among the "strongest" near dependencies. Here the variance–decomposition proportions for the variates belonging to these near dependencies could be spread across them, producing values no higher than .4 or .5 in each cell of the Π matrix. The eye also is readily trained to pick up these unmistakable patterns.

Exhibit 4a Interpretative illustration: the high condition indexes and the strongest near dependency

Scaled Condition	Proportions	of						
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	var(b_3)	var(b_4)	var(b_5)	var(b_6)	var(b_7)	var(b_8)
1	.000	.000	.001	.000	.003	.000	.000	.000
2	.000	.000	.005	.000	.070	.000	.000	.000
5	.002	.000	.000	.003	.027	.001	.000	.000
7	.000	.000	.013	.000	.044	.001	.000	.000
11	.000	.000	.657	.001	.020	.001	.002	.000
35	.000	.000	.076	.835	.007	.852	.000	.000
153	.970	.001	.071	.083	.000	.073	.973	.000
455	.028	**.999**	.177	.078	**.829**	.072	.025	**1.000**

near dependencies, one very strong, one strong, and one moderately strong." Now, picking up the large elements in the last row, "The strongest near dependency involves variates 2, 5 and 8." And, allowing our eyes to pick up subsequent lines as in Exhibit 4b, the story continues, "The next strongest dependency involves variates 1 and 7, but could well involve 2, 5, or 8, as well, due to dominance. The third strongest near dependency could contain any of the previously involved variates, but certainly involves variates 4 and 6."

At this point it is usually interesting to stop and put real names and ideas on the variates to see what all this means. For example, variate 2 might be GNP while variate 8 is lagged consumption and variate 5 is lagged investment. Then we would know that

Exhibit 4b Interpretative illustration: all three near dependencies

Scaled Condition	Proportions	of						
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	var(b_3)	var(b_4)	var(b_5)	var(b_6)	var(b_7)	var(b_8)
1	.000	.000	.001	.000	.003	.000	.000	.000
2	.000	.000	.005	.000	.070	.000	.000	.000
5	.002	.000	.000	.003	.027	.001	.000	.000
7	.000	.000	.013	.000	.044	.001	.000	.000
11	.000	.000	.657	.001	.020	.001	.002	.000
35	.000	.000	.076	**.835**	.007	**.852**	.000	.000
153	**.970**	.001	.071	.083	.000	.073	**.973**	.000
455	.028	**.999**	.177	.078	**.829**	.072	.025	**1.000**

autocorrelation typical of economic time series, along with an identity in the unlagged values, is causing troubles. Such considerations can often stimulate useful ideas regarding the process that generated the given data set, ideas that can possibly give aid in carrying out appropriate corrective action. This latter is illustrated in Belsley (1984b).

Next, the eye can examine what variates might be unaffected by the collinear relations. Exhibit 4c highlights the information that indicates that variate 3 seems to be associated mainly with the smaller scaled condition indexes. Its weak involvement in the stronger near dependencies is not altogether to be ignored, but these data seem well–conditioned relative to variate 3.

Exhibit 4c Interpretative illustration: noninvolvement

Scaled Condition	Proportions of							
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	var(b_3)	var(b_4)	var(b_5)	var(b_6)	var(b_7)	var(b_8)
1	.000	.000	.001	.000	.003	.000	.000	.000
2	.000	.000	.005	.000	.070	.000	.000	.000
5	.002	.000	.000	.003	.027	.001	.000	.000
7	.000	.000	.013	.000	.044	.001	.000	.000
11	.000	.000	**.657**	.001	.020	.001	.002	.000
35	.000	.000	.076	.835	.007	.852	.000	.000
153	.970	.001	.071	.083	.000	.073	.973	.000
455	.028	.999	.177	.078	.829	.072	.025	1.000

An Alternative Transformation for the Scaled Condition Indexes

It is found in BKW that the scaled condition indexes tend to progress along the sequence 1, 3, 10, 30, 100, 300, 1000, 3000, . . . , as the the corresponding near dependencies become tighter and tighter. Both the nonlinearity of this progression and the large values that occur in its more removed terms can be eliminated by using a logarithmic transformation. Thus, consider defining the set of condition indexes

$$\zeta_k \equiv \log_{10}(\eta_k) \qquad k = 1, \ldots, p. \qquad (8)$$

The progression 0.0, 0.5, 1.0, 1.5, 2.0, 2.5, . . . , in the ζs corresponds to a progression of 1, 3, 10, 32, 100, 316, 1000, 3162, . . . , in the ηs, which is, of course, essentially the same as the familiar progression given above. Thus, condition indexes transformed as (8) are an excellent alternative means for presenting the scaled condition indexes, particularly when a linear scale is desirable, as would be the case, for example, for a graphic display of the condition indexes. This has been used to advantage in the system designed by Oldford and Peters (1984) for providing guided use of these collinearity diagnostics. The graphics that result are particularly effective in helping to determine the "gaps" in the progression described above.[7] In this scale, values of (transformed) scaled condition indexes $\tilde{\zeta}$ below 1.0 are of little concern, while those above 1.5 are typically worthy of notice. It is unnecessary to present these values with more than one place beyond the decimal point.

Forming the Auxiliary Regressions

When there are several near dependencies, we have seen that it is not always possible to determine from the variance–decomposition proportions alone exactly which variates are involved in which near dependencies. Of course, it is always possible to determine which variates are involved in at least one near dependency, but the presence of dominating and competing dependencies can obscure individual variate involvement. When this happens, a simple procedure can be used to form a set of auxiliary regressions that typically will display structure of the near dependencies in greater detail. The basic idea is to use the variance–decomposition proportions to identify one variate known to be in each near dependency, and then to regress each variate in this set on the remainder. This is exemplified in Exhibit 5, a variance–decomposition $\mathbf{\Pi}$ matrix for a seven–column data matrix $\mathbf{X}$.

[7] Indeed, cluster analysis could be used to mechanize this procedure.

Exhibit 5 Forming the auxiliary regressions

Scaled							
Condition				Proportions of			
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	var(b_3)	var(b_4)	var(b_5)	var(b_6)	var(b_7)
1	.000	.002	.000	.000	.000	.000	.000
2	.000	.000	.001	.000	.000	.120	.000
6	.000	.024	.010	.000	.002	.002	.000
8	.000	.000	.051	.005	.000	.616	.008
32	.000	.103	**.602**	.000	.097	.056	.088
78	.438	**.856**	.312	.092	.105	.141	.002
118	.562	.015	.024	.903	.796	.065	**.902**

Here we see that there are three near dependencies among the seven variates, associated with the scaled condition indexes 32, 78, and 118. Hence, in a sort of reduced form, we can express three of the seven variates in terms of the remaining four; that is, we can select three variates to regress on the remaining four to produce three descriptive auxiliary regressions displaying the three near dependencies. The t–statistics that accompany these regressions can then be used descriptively (not inferentially, because we are making no claim that these relations in fact are relevant to the process that actually generates the **X** data) to signal variate involvement.

To choose the three variates to act as the "dependent" variates for these auxiliary regressions, we need a process that will guarantee that we pick for each near dependency a variate known to be involved in it. Beginning, then, with the strongest near dependency ($\tilde{\eta}$ = 118), look along its row to find large variance–decomposition proportions to signal variates known to be in this near dependency. In this case, both C4 (π = .903) and C7 (π = .902) appear initially to be equally good candidates. C7 is chosen, however, because it has the remainder of its variance determined in more removed near dependencies, thereby minimizing the possibility that the values for the variance–decomposition proportions are distorted through competing near dependencies. The π value for C7 in this last row is made bold to indicate that C7 has been picked to be the "dependent" variate in the auxiliary regression corresponding to $\tilde{\eta}$ = 118. For the next strongest near dependency ($\tilde{\eta}$ = 78), there is a clear winner in C2 (π = .856). And C3 (π = .602) looks good for the weakest near dependency ($\tilde{\eta}$ = 32). Hence, we pick C7, C2, and C3 as the "pivots" to regress separately on the remaining variates C1, C4, C5, and C6.

The procedure just described works extremely well in a wide variety of cases but often has some troubles when there are many near dependencies or when there are several dominating near dependencies that cover up the involvement of all the variates in the weaker near dependencies. Consider, for example, the case in Exhibit 6.

Here again there are three near dependencies ($\tilde{\eta}$s of 1035, 367, and 32), and there is no question of C7's involvement in the strongest of the three. But there is little information to be learned about which variates are unmistakably involved in the weaker two near dependencies. One might make a good guess about C1 in the second strongest near dependency, but what about the third? C2 and C6 show themselves to be effectively uninvolved in any of the near dependencies, and the strengths of the top two obscure all information relative to the third.

Exhibit 6 Forming the auxiliary regressions: problems with dominant dependencies

Scaled Condition				Proportions of			
Index, $\tilde{\eta}$	var(b_1)	var(b_2)	var(b_3)	var(b_4)	var(b_5)	var(b_6)	var(b_7)
1	.000	.002	.000	.000	.000	.000	.000
2	.000	.019	.000	.000	.000	.120	.000
6	.000	.853	.000	.000	.002	.002	.000
8	.000	.000	.001	.005	.000	.616	.008
32	.000	.103	.000	.000	.097	.056	.088
367	.347	.008	.175	.092	.105	.141	.002
1035	.653	.015	.824	.903	.796	.065	.902

An extension of the preceding procedure that works well in a case like this is the following: First, make a good guess, based on what information is available, to pick a set of variates equal in number to the number of near dependencies, r. Here, for example, we might pick the three variates ℂ7, ℂ1, and ℂ5. This latter variate is picked because there is a strong possibility that ℂ5 is involved in the third strongest near dependency but it is being masked by its involvement in the two stronger ones. This might also be true of ℂ4, but ℂ4's π of .000 in this third strongest near dependency is not encouraging. Second, find the condition number of the matrix composed of the remaining p–r variates — in this case, ℂ2, ℂ3, ℂ4, and ℂ6. If the condition number of this matrix is of the same order of magnitude as the next smallest condition index (the p–rth — whose value is 8 in this example), then one has found a subset of the variates that possesses no near dependences. This subset is therefore associated only with the "background" conditioning of the original data matrix and can be used as the set of regressands for the auxiliary regressions. If, on the other hand, the condition number of this matrix is of a larger order of magnitude, then try a different subset of p–r variates until a matrix with a condition number with the appropriate order of magnitude is found. The variates comprising this matrix can be used as the auxiliary regressands, and the complementary r variates become the auxiliary regressors. It has not yet formally been proved that a submatrix can always be found in this way whose condition number is of the same order of magnitude as the next smallest condition index, but this is an excellent conjecture; I have never found a counterexample.

There are, of course, many other ways in which a set of auxiliary regressions could be chosen. Since these auxiliary regressions are in no way intended to specify and estimate a model describing the way the given **X** data were actually generated, any reasonable choice can serve the purpose of descriptively illuminating the nature of the linear near dependencies that just happen to have occurred in the given data matrix. Indeed, the specific context of any given data set may suggest a natural set of pivots, and the "automatic outcome" of the above procedure may not be appropriate. Interest may center, for example, in the possibly simultaneous role a specific variate plays in several of the near dependencies, and so it would be important to choose that variate as a regressand entering in all the auxiliary regressions and not as a regressor appearing in only one. This occurs in the study given in Belsley (1984b).

In the absence of any other considerations, however, the procedure just described for constructing auxiliary regressions has the advantages that (1) it is simple to employ, (2) it picks as a "dependent" variate for each auxiliary regression one that is known to be strongly involved in the underlying near dependency, and (3) the set of regressands (the right–hand–side variates) will necessarily be relatively well conditioned, so the auxiliary regressions should not themselves be subject to the problems of ill–conditioning.

4. CONCLUSION

Applying the collinearity diagnostics of Belsley, Kuh, and Welsch is a straightforward procedure, easily learned and mastered. With the techniques described here, the eye is trained readily to focus only on the most important information presented in a diagnostic tableau and quickly to digest its meaning. Examples of the use of these techniques with real data in actual studies, including those using the diagnostics for harmful collinearity and short–data, are to be found in numerous of the author's works cited in the bibliography.

BIBLIOGRAPHY

Becker, R., N. Kaden, and V. Klema (1974), "The Singular Value Analysis in Matrix Computation," Working paper 46, Computer Research Center, National Bureau of Economic Research, Cambridge, MA.

Belsley, D. A. (1976), "Multicollinearity: Diagnosing Its Presence and Assessing the Potential Damage It Causes Least–Squares Estimation," *Working Paper # 154,* Computer Research Center, National Bureau of Economic Research, Cambridge, MA.

Belsley, D. A. (1982), "Assessing the Presence of Harmful Collinearity and Other Forms of Weak Data through a Test for Signal–to–Noise," *Journal of Econometrics* 20 211–253.

Belsley, D. A. (1984a), "Demeaning Conditioning Diagnostics through Centering," with accompanying comments and author's reply,*The American Statistician* 38 73–93.

Belsley, D. A. (1984b), "Collinearity and Forecasting," *Journal of Forecasting* 3 183–196.

Belsley, D. A. (1986), "Centering, the Constant, First–Differencing, and Assessing Conditioning," in E. Kuh and D. A. Belsley (eds.), *Model Reliability,* MIT Press: Cambridge, MA.

Belsley, D. A. (1987), "Comment: Well–Conditioned Collinearity Indices?" *Statistical Science,* 2 86–91.

Belsley, D. A. (1988), "Conditioning in Models with Logs," *Journal of Econometrics* 38 127–143.

Belsley, D. A. and W. R. Oldford (1986), "The General Problem of Ill Conditioning and Its Role in Statistical Analysis," *Computational Statistics & Data Analysis* 4 103–120.

Belsley, D. A. and R. E. Welsch (1988), "Comment: Modelling Energy Consumption: Using and Abusing Regression Diagnostics," *Journal of Business and Economic Statistics* 6 442–447.

Belsley, D. A., E. Kuh, and R. E. Welsch (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity,* John Wiley and Sons: New York.

Businger, P. and G. H. Golub (1965), "Linear Least Squares Solutions by Householder Transformations," *Numerische Mathematik* 7 269–276.

Golub, G. H. (1969), "Matrix Decompositions and Statistical Calculations," *Statistical Computation,* R. C. Milton and J. A. Nelder, Eds., Academic Press: New York, 365–397.

Golub, G. H. and C. Reinsch (1970), "Singular Value Decomposition and Least–Squares Solutions," *Numerische Mathematik* 14 403–420.

Golub, G. H. and C. F. Van Loan (1983), *Matrix Computations,* Johns Hopkins University Press: Baltimore.

Johnston, J. (1984), *Econometric Methods,* 3rd Edition, McGraw–Hill: New York.

Oldford, R. W. and S. Peters (1984), "Building a Statistical Knowledge Based System with Mini–Mycin", *Technical Report No. 42,* Center for Computational Research in Economics and Management Science, MIT.

Oldford, R. W. and S. Peters (1985), "DINDE: Towards More Statistically Sophisticated Software," *Technical Report No. 55,* Center for Computational Research in Economics and Management Science, MIT.

Silvey, S. D. (1969), "Multicollinearity and Imprecise Estimation," *Journal of the Royal Statistical Society, Series B* 31 539–552.

Stewart, G. W. (1973), *Introduction to Matrix Computations,* Academic Press: New York.

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