A statistic for allocating C_p to individual cases

by

Sanford Weisberg
University of Minnesota

Technical Report No. 296

August, 1977

ABSTRACT

Variable selection is primarially a global or aggregate statistical procedure since the techniques used depend on functions of the sufficient statistics, such as R^2, F and t tests, and the C_p statistic. In other parts of a regression analysis, it is common to examine an array of case statistics that have values for each of the n cases in a study. Case statistics include studentized residuals, fitted values, and various distance or influence measures.

In this paper, a subdivision of the C_p statistic into n components (one for each case) is developed, and the properties of this method are outlined. An example is given.

Keywords: Variable selection, linear regression, subset regression, C_p statistic, residual analysis.

Variable selection in multiple regression is fundamentally an aggregate procedure, since one bases the selection of subsets on aggregate or global statistics such as R^2, F or t tests, adjusted R^2, C_p statistics, or the like. All of these have the common thread that they depend on the observed data through sufficient statistics, and, in a sense, they model average behavior of the fit of a model to the data. In recent years, there has been increasing interest in the computation and use of "case statistics" --- statistics that have computed value for each case in a problem (following a suggestion by John Hartigan (1977), we use the word "case" rather than the ambiguous terms "observation" or "point", to correspond to the rows of a data matrix; the columns are called variables). The case statistics typically computed include fitted values, residuals, and other statistics, such as studentized residuals, various influence measures, including Cook's distance (Cook 1977a, 1977b), and the variances of the fitted values or Mahalanobis distances. The reason for examining case statistics is that global or aggregate behavior modeled by the aggregate statistics may not accurately portray the fit of a model in all regions of the observation space. In this paper, we develop a case statistic version of the C_p statistic, and show that this statistic can be used to help understand how the lack of fit measured by C_p is reflected in the individual cases in the study.

1. The C_p statistic.

We consider the problem of comparing a full rank linear model with n cases, given by

$$Y = X\beta + X_2\beta_2 + e \qquad (1)$$

where, in (1), X is n x p, X_2 is n x q and Cov(e) = $\sigma^2 I$, to a fixed subset model of the form

$$Y = X\beta + e'. \tag{2}$$

We assume that the goal of regression is the estimation of fitted values for the n cases. In general, (2) will provide biased estimates of fitted values but, as is well known (Hocking 1973), (2) may have smaller mean square error than (1) for estimating fitted values.

Suppose we let $V = X(X'X)^{-1}X'$, and define U by

$$U = (X\ X_2)\begin{pmatrix} X'X & X'X_2 \\ X_2'X & X_2'X_2 \end{pmatrix}\begin{pmatrix} X' \\ X_2' \end{pmatrix} . \tag{3}$$

The matrices U and V are fundamental in regression problems. Of particular interest are the diagonal entries of U and V, which we will denote by u_{ii} and v_{ii} respectively. For our purposes, it will be convenient to apply a linear transformation to X_2 so that, in the resulting parameterization, the variables not in the subset model are orthogonal to the variables in the subset model. To this end, define

$$Z = (I - V)X_2 , \tag{4}$$

so that Z is the projection of X_2 onto the orthogonal complement of X. The full model (1) can be rewritten as

$$Y = X\beta + Z\gamma + e , \tag{5}$$

where γ is an appropriately defined q x 1 parameter vector. All results in this paper will be derived in terms of (5) rather than (1), although, in practice, the transformation need not be computed.

Suppose that we let $W = Z(Z'Z)^{-1}Z'$. Then, one can easily show that $U = V + W$, or, for the purposes of computation, $W = U - V$, so that W need not be found explicitly. If we let w_{ii} be the diagonal elements of W, then we will have that $u_{ii} - v_{ii} = w_{ii} \geq 0$, since w_{ii} is a quadratic form.

When considering the subset model, let the subscript "p" refer to the use of the model (2), where p is the number of parameters. For example, RSS_p is the residual sum of squares for the subset model, $RSS_p = Y'(I-V)Y$. Fitted values for the subset model will be given by a lower case $\hat{y}_i$. For the full model (5), estimated quantities will have no subscripts, e.g., $\hat{\sigma}^2 = Y'(I - U)Y/(n-p-q)$. The fitted values for the full model will be denoted by capital letters, $\hat{Y}_i$.

The fitted value for the i-th case is given by $\hat{y}_i = x_i'\hat{\beta}$ for the subset model and by $\hat{Y}_i = x_i'\hat{\beta} + z_i'\hat{\gamma}$ for the full model, where x_i' and z_i' are the i-th rows of X and Z respectively. Because of the orthogonalization, $\hat{\beta}$ $(=(X'X)^{-1}X'Y)$ the least squares estimate of β, is the same for models (2) and (5); this is not true in general. Nevertheless, the estimates of the fitted values $\hat{Y}_i$ will be the same from model (1) or from model (5). The <u>expected bias</u> in the subset model (assuming that the full model is correct, with all relevant variables included in the proper scale) is given by $E(\hat{y}_i - \hat{Y}_i) = E(x_i'\hat{\beta} - x_i'\hat{\beta} - z_i'\hat{\gamma}) = -z_i'\gamma$. The variance of $\hat{y}_i$ is $Var(y_i) = \sigma^2 v_{ii}$. We now define the total expected squared standardized error, Γ_p, to be

$$\Gamma_p = \frac{\Sigma(z_i\gamma)^2}{\sigma^2} + \frac{\Sigma\ Var(\hat{y}_i)}{\sigma^2} = \frac{\Sigma(z_i\gamma)^2}{\sigma^2} + p \qquad (6)$$

since $\Sigma \text{ Var}(\hat{y}_i) = \sigma^2 \Sigma v_{ii} = \sigma^2 \text{ trace } (V) = \sigma^2 \text{ rank}(V) = \sigma^2 p$. This is a reasonable quantity to be interested in, as it measures the total expected error in estimating the fitted values. The C_p statistic is found by substituting estimates for $\Sigma(z_i'\gamma)^2$ and σ^2 into (6). Now, an obvious estimate of $z_i'\gamma$ is $z_i'\hat{\gamma} = \hat{Y}_i - \hat{y}_i$; but $\Sigma(z_i\hat{\gamma})^2 = \Sigma(\hat{\gamma}z_i z_i'\hat{\gamma}) = \hat{\gamma}'Z'Z\hat{\gamma} = Y'WY$, which is the sum of squares for regression on Z (in general, for Z after X) and

$$E(Y'WY) = q\sigma^2 + \Sigma(z_i\gamma)^2 \tag{7}$$

Letting $\hat{\sigma}^2 = Y'(I-U)Y/(n-p-q)$ estimate σ^2, and replacing $E(Y'WY)$ by $\Sigma(\hat{y}_i - \hat{y}_i)^2$ in (7), we substitute into (6) to get

$$C_p = \frac{\Sigma(\hat{y}_i - \hat{Y}_i)^2}{\hat{\sigma}^2} + p - q. \tag{8}$$

The usual form of C_p is found by noting that $\Sigma(\hat{y}_i - \hat{Y}_i)^2 = RSS_p - (n-p-q)\hat{\sigma}^2$ and, substituting for $\Sigma(\hat{y}_i - \hat{Y}_i)$ in (8),

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} + 2p - n. \tag{9}$$

The C_p statistic therefore estimates the total error (variance plus bias) in estimating the fitted values corresponding to the n cases. It bears a fundamental relationship to the F test (as noted by Spjotvoll (1977)) for the hypothesis that $\gamma = 0$ (in general, for $\beta_2 = 0$ given β), where the test is given by $F_p = \Sigma(\hat{y}_i - \hat{y}_i)^2/q\hat{\sigma}^2$. Under normality and the null hypothesis, F_p is distributed as central F, $F_p \sim F(q, n-p-q)$. Substituting into (8) one easily finds

$$C_p = p + q(F_p - 1) \tag{10}$$

From (10) we have that $F_p \leq 1$ if and only if $C_p \leq p$, $F_p \leq 2$ if and only if $C_p \leq p + q$ and, ignoring multiple test problems, the hypothesis H: $C_p = p$ can be tested with critical value given by (10) with F* substituted for F_p, with F* an appropriate percentage point of F(q, n-p-q).

The C_p statistic is widely used as a global measure of the adequacy of a subset model. For more on its use, see Gorman and Toman (1966), Daniel and Wood (1971), and Mallows (1973).

2. A Case version of C_p.

Now consider the i-th case. By analogy to the above development of C_p, $\mathrm{Var}(\hat{y}_i) = \sigma^2 v_{ii}$ and the expected bias in the subset model for the i-th case is given by $E(\hat{y}_i - \hat{Y}_i) = -z_i\gamma$. Define

$$\Gamma_{pi} = \frac{E(\hat{y}_i - \hat{Y}_i)^2}{\sigma^2} + \frac{\mathrm{Var}(\hat{y}_i)}{\sigma^2} = \frac{(z_i\gamma)^2}{\sigma^2} + v_{ii} \tag{11}$$

Now, $\hat{y}_i - \hat{Y}_i = z_i\hat{\gamma}$ is the observed bias, and

$$E(\hat{y}_i - \hat{Y}_i)^2 = \mathrm{Var}(z_i'\hat{\gamma}) + (z_i\gamma)^2 \tag{12}$$

$$= \sigma^2(w_{ii}) + (z_i\gamma)^2$$

$$= \sigma^2(u_{ii} - v_{ii}) + (z_i'\gamma)^2$$

Again replacing σ^2 by $\hat{\sigma}^2$ in (11) and $E(\hat{y}_i - \hat{Y}_i)^2$ by $(\hat{y}_i - \hat{Y}_i)^2$ in (12), we get

$$C_{pi} = \frac{(\hat{y}_i - \hat{Y}_i)^2}{\hat{\sigma}^2} + v_{ii} - (u_{ii} - v_{ii}). \tag{13}$$

Thus, C_{pi} measures the standardized error in estimating the i-th fitted value. Using (13) and (8), one can easily show that $\sum_{i=1}^{n} C_{pi} = C_p$, since $\Sigma v_{ii} = p$ and $\Sigma(u_{ii} - v_{ii}) = q$.

Now consider a test of the hypothesis of no bias for the i-th case, H: $z_i'\gamma = 0$. Under normality, the likelihood ratio statistic for this hypothesis is given by $t_{pi}^2 = (z_i'\hat{\gamma})^2/\hat{\sigma}^2(u_{ii} - v_{ii})$, where t_{pi}^2 is distributed as F(1, n-p-q) (e.g. t_{pi} is a t-statistic). Substituting t_{pi}^2 into (13), we can write (for $u_{ii} \neq v_{ii}$)

$$C_{pi} = v_{ii} + (u_{ii} - v_{ii})\ (t_{pi}^2 - 1) \tag{14}$$

as a striking parallel to (10) (if $u_{ii} = v_{ii}$, then $C_{pi} = v_{ii}$). As with the results following (10), we have $C_{pi} \leq v_{ii}$ if and only if $t_{pi}^2 \leq 1$; $C_{pi} \leq u_{ii}$ if and only if $t_{pi}^2 \leq 2$. Also, the critical value for a test that $C_{pi} = v_{ii}$ is given by

$$C_{pi}^* = v_{ii} + (u_{ii} - v_{ii})(t_p^{*2} - 1) \tag{15}$$

where t_p^{*2} is the appropriate percentage point of F(1,n-p-q), $C_{pi} \geq C_{pi}^*$ indicating bias.

Properties of C_{pi}. The statistic C_{pi} depends on $y_1, \ldots, y_n$ only through the sufficient statistics Y'Y, Y'1 X'Y and Z'Y. Thus, C_{pi} is influenced by outliers in Y only through the influence of the outlier on the estimates $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\sigma}^2$, and essentially measures the error at the point (x_i', z_i').

From (14) it is easy to find the mean and variance of C_{pi}, since under normality, t_{pi}^2 is, in general, distributed as a non-central F. We find

$$E(C_{pi}) = v_{ii} + (u_{ii} - v_{ii})\left[\frac{2}{n-p-q-2} + \frac{n-p-q}{n-p-q-2}\ (z_i'\gamma)^2\right], \tag{16}$$

or, ignoring terms of $O(n^{-1})$,

$$E(C_{pi}) = v_{ii} + (u_{ii} - v_{ii})(z_i'\gamma)^2 \tag{17}$$

Hence, the excess of C_{pi} over v_{ii} measures bias, but if $(u_{ii} - v_{ii})$ is small, even large bias may have little effect on C_{pi}. If C_{pi} is large, $(u_{ii} - v_{ii})$ must also be examined. The variance of C_{pi} is given, for $n-p-q > 4$, by

$$\text{Var}(C_{pi}) = 2(u_{ii} - v_{ii})^2 \left(\frac{n-p-q}{n-p-q-2}\right)^2 \left(\frac{1}{n-p-q-4}\right)\left(\frac{(1+(z_i\gamma)^2)^2}{n-p-q-2} + 1 + 2(z_i'\gamma)^2\right). \tag{18}$$

If we ignore terms of $O(n^{-2})$,

$$(n-p-q)\ \text{Var}(C_{pi}) = 2(u_{ii} - v_{ii})^2(1 + 2(z_i'\gamma)^2). \tag{19}$$

For a fixed value of C_p, we can find minimum and maximum possible values for C_{pi}. The minimum will occur when $t_{pi}^2 = (\hat{y}_i - \hat{Y}_i)^2 = 0$, which will happen (using model (5)) only if $z_i = 0$ or if $\hat{\gamma} = 0$. The case of $z_i = 0$ implies $u_{ii} = v_{ii}$ and $C_{pi} = v_{ii}$. If $\hat{\gamma} = 0$, then C_{pi} is bounded below by $v_{ii} - (u_{ii} - v_{ii})$, which may be negative, $v_{ii} - (u_{ii} - v_{ii}) \geq -(n-2)/n$, if the constant is in the subset model. The maximum value of C_{pi} is found by equating t_{pi}^2 to its maximum value of $(C_p - p + q)/(u_{ii} - v_{ii})$. Combining results, we find

$$v_{ii} - (u_{ii} - v_{ii}) \leq C_{pi} \leq C_p - (p-q) + (v_{ii} - (u_{ii} - v_{ii})). \tag{20}$$

In particular, C_{pi} may be greater than C_p if $q \geq p$. Thus, in some problems most or all of the lack of fit measured by C_p may be due to just a few of the cases.

Consider, for example, the artificial data given in Table 1, $n = 10$, $p = 2$, $q = 1$. Easy computations show that, using model (5) $\hat{\gamma} = (y_{10} - y_1)/2$. The values of C_{pi} for all cases are given in the table, and $C_p = \Sigma C_{pi} = 1 + (y_{10} - y_1)^2/2\hat{\sigma}^2$. All of the increase of C_p over 1 can be attributed to the difference $(y_{10} - y_1)$ and this increase will be reflected only in the C_{pi} for the first and tenth cases.

Table 1 about here

3. The v_{ii} and the u_{ii}

By examination of (14), we see that C_{pi} depends on two fixed quantities, v_{ii} and $(u_{ii} - v_{ii})$ and one random component, t^2_{pi}. The fixed quantities arise in many contexts in case statistics, for example, Cook (1977b) derives a distance measure to model the influence of the i-th case on the estimation of β_2 in (1) that depends only on q, v_{ii}, $u_{ii} - v_{ii}$ and the i-th studentized residual. The v_{ii} also have interest in their own right (see, for example, Bhenken and Draper (1972)).

To understand the C_{pi}, one should carefully examine the v_{ii} and $u_{ii} - v_{ii}$. Consider first v_{ii}. Suppose that the mean is included in the subset model and let S be the (p-1) x (p-1) cross product matrix for the p-1 remaining variables in X. Let $\bar{x}$ be the (p-1) x 1 vector of means of the (p-1) x's and redefine x_i to the (p-1) x 1 vector of x's for the i-th case (deleting the constant). Then, we can rewrite

$$v_{ii} = \frac{1}{n} + (x_i - \bar{x})' S^{-1} (x_i - \bar{x}) \qquad (22)$$

Now, consider a spectral decomposition of S: Let $S = P' \Lambda P$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda p-1)$, such that $\lambda_1 \geq \lambda_L \geq \ldots \geq \lambda_{p-1}$, and $P'P = I$. Then the λ_j's are the eigenvalues, and the columns of P, say P_j, are the eigenvectors of S. Since $S^{-1} = P' \Lambda^{-1} P = \Sigma P_j P_j'/\lambda_j$, (22) can be rewritten as

$$v_{ii} = \frac{1}{n} + (x_i - \bar{x})' \left(\sum_j \frac{P_j P_j'}{\lambda_j} \right) (x_i - \bar{x}) \qquad (23)$$

$$= \frac{1}{n} + \sum_j \left(\frac{P_j'(x_i - \bar{x})}{\sqrt{\lambda_j}} \right)^2$$

Let Θ_{ji} be the angle between the j-th eigenvector P_j and x_i. Then (23) becomes

$$v_{ii} = \frac{1}{n} + (x_i-\bar{x})'(x_i-\bar{x}) \sum_j \frac{1}{\lambda_j} \text{Cos}^2(\Theta_{ji}) \tag{24}$$

Thus, v_{ii} can be large if <u>either</u> $(x_i-\bar{x})'(x_i-\bar{x})$ is large - that is, the i-th case is far removed from the center of the rest of the data, or if $(x_i-\bar{x})$ is nearly parallel to an eigenvector corresponding to a small eigenvalue ($\text{Cos}^2\theta_{ji} \tilde{=} 1$ for j near p-1). Thus, cases in an "unusual" direction will have large values of v_{ii}. If we consider only the (p-1) dimensional space spanned by the p-1 variables in X, then v_{ii} is related to the Mahalanolis distance, in this space, MD_i, by the simple equation $MD_i = (n-1)(v_{ii} - \frac{1}{n})$.

Now consider the u_{ii}. By the orthogonalization (1.4) we can find matrices Q and Δ, with $\Delta = \text{diag}(\delta_1, \delta_2, \ldots \delta_q)$ and $Q'Q = I$ such that $Z'Z = Q'\Delta Q$ and the u_{ii} can be written

$$u_{ii} = (x_i-\bar{x})'(x_i-\bar{x}) \sum_{j=1}^{p-1} \frac{1}{\lambda_j} \text{Cos}^2(\Theta_{ji}) + (z_i'z_i) \sum_{j=1}^{q} \frac{1}{\delta_j} \text{Cos}^2(\eta_{ji}) \tag{25}$$

where η_{ji} is the angle between Q_j and z_i. Thus the difference $u_{ii} - v_{ii}$ may be large or small, relative to u_{ii}, depending on whether or not the small eigenvalues (if any) are in the (p-1) "in" variables or in the q "out" variables, and depending on $z_i'z_i$.

Further insight can be gained by considering a bivariate regression model $y_i = \beta_o+\beta_1 x_{ii} + \beta_2 x_{2i} + e_i$, i = 1,...n. Using model (1), contours of constant u_{ii} are given by ellipsoids in the (x_1, x_2) space as shown in Figure 1. u_{ii} is essentially the squared distance from (x_{1i}, x_{2i}) to $(\bar{x}_{1+}, \bar{x}_{2+})$ Now suppose we consider the subset model obtained by deleting x_2 from the model.

The v_{ii} corresponding to a given (x_{1i},x_{2i}) pair is found by projecting (x_{ii},x_{ii}) onto the x_1 axis: contours of constant v_{ii} now consist of two points equidistant from $\bar{x}_{1+}$. We can distinguish those points with fixed u_{ii} that will have relatively large $u_{ii}-v_{ii}$ as shown in Figure 1. It is interesting to note that the points with large $u_{ii} - v_{ii}$ for one subset model may be small for another subset model, as can be seen by projecting in Figure 1 on the x_2 axis.

Figure 1 here

Example. We illustrate the use of C_{pi} on a complex data set given by Narula and Wellington (1977). The data, reprinted here as Table 2, relates selling price of n = 28 houses to 11 potential predictor variables, with the eventual goal of developing a prediction equation based on a subset. We will not give a complete analysis here, but merely indicate possible uses of C_{pi}.

We begin by computing the regression for the full model, and then examining several useful case statistics (the computations for the full model given in Narula and Wellington are incorrect). Included in Table 3 are the studentized residual and its t-distribution transform, Cook's distance D_i, and the diagonal elements of $U = X(X'X)^{-1}X'$ for each of the 28 cases. While none of the cases appear to be outliers (in the sense of large t-values), the 28th case has $D_{28} = 1.06$, indicating that this case is relatively influential in estimating parameters. D_{28} is large because of the unusual value of X_1 (= taxes) for this case. A prudent approach at this point, therefore, is to do variable selection twice, once including the 28-th case, and once excluding it. To this end, the 10 subsets with the lowest values of C_p for the 28 case data set and for the 27 case data set are given in Table 4. Computations were done using the Furnival and Wilson (1974) algorithm. It is interesting to note that the best model for the 27 case data is not among the best 10 for the 28 case data, and the best two variable model for the 28 case data, (X_1, X_4), is not among the best models for the 27 case data. In the remainder of this discussion, we will look only at the models (X_1, X_2), (X_1, X_4) and (X_1, X_2, X_4) for the 27 and 28 case data sets.

Tables 2, 3 and 4 about here

The C_{pi} and values of C^*_{pi} from (15) with t^{*2}_p equal to the .05 point of the appropriate F distribution are given in Table 5 for the three models under consideration and both the 27 and 28 case data sets. Consider first the model (X_1, X_2). The relatively large value of C_p for this model in the 28 case data is largely due to cases 12, 19 and 28, all of which have values of C_{pi} exceeding C^*_{pi}. Case 28 alone, with $C_{p,28} = 4.40$ appears to dominate the lack of fit. The fit of this model to the 27 case data is somewhat better although still not ideal: the C_{pi} for case 21 is actually greater than the overall C_p, so that the fit of the model is not uniform throughout the observation space even with case 28 excluded.

Table 5 about here

The model (X_1, X_4) is more satisfactory overall than is the model considered above, as none of the C_{pi} in either the 27 or 28 case data sets are greater than the corresponding C^*_{pi} although the values for cases 11 and 21 are somewhat large. The same qualitative judgements hold for the model (X_1, X_2, X_4), where case 21 remains troublesome. At this point, the last two models should be examined in greater detail, with careful attention given to case 21. Tentatively, one would wish to leave case 28 in the data set to obtain estimates, since the C_{pi} appear to be more uniform for the two models under consideration in the 28 case data.

I am indebted to Christopher Bingham and to R. Dennis Cook for many illuminating discussions on case statistics. The data used in the example was suggested by Kinley Larntz.

REFERENCES

Behnken, D.W. and Draper, N.R. (1972). Residuals and their variance patterns, Technometrics 14, 102-11

Cook, R.D. (1977a). Detection of Influential Observations in Linear Regression, Technometrics 19, 15-19

Cook, R. D. (1977b). Isolating the effects of influential observations, Univ. of Minnesota, School of Statistics Technical Report #283

Daniel, C. and Wood F. (1971). Fitting Equations to Data, Wiley, New York

Furnival, G.M., and R.W. Wilson (1974). Regression by leaps and bounds. Technometrics 16, 499-512

Gorman, J.W. and Toman, R.J. (1969). Selection of variables for fitting equations to data, Technometrics 8, 27-51.

Hartigan, John (1977). Unpublished talk at the Spring 1977 IMS Meeting, Dallas, Texas.

Hocking, R.R. (1974). Misspecification in regression, American Statistician 28, 39-40.

Mallows, C.I., (1973). Some Comments on C_p, Technometrics 15, 661-75.

Narula, S.C. and J.F. Wellington (1977). Prediction, linear regression and the minimum sum of relative errors, Technometrics 19, 185-190.

Spjotvoll, E. (1977). Alternatives to plotting C_p in multiple regression. Biometrika 64, 1-8.

Figure 1. Contour of constant u_{ii} for a bivariate regression problem. The points on the contour within the cross hatched area will have relatively small values of v_{ii} for the subset model with x_2 deleted, giving large values of $u_{ii} - v_{ii}$.

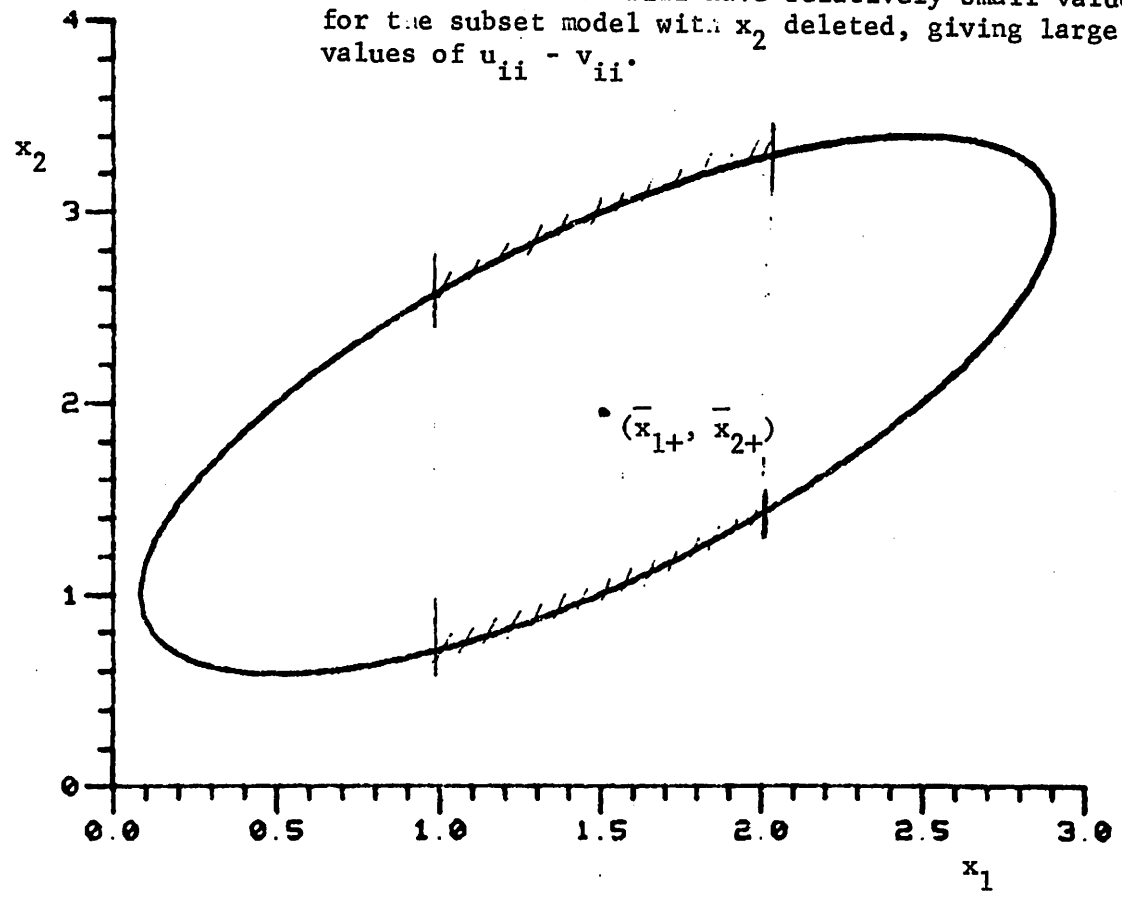

Table 1. Artificial Data

full model: Subset Model			Y	v_{ii}	w_{ii}	C_{pi}
1	-4	-1	y_1	$\frac{22}{60}$	1/2	$-\frac{8}{60} + (y_{10} - y_1)^2/4\hat{\sigma}^2$
1	-3	0	y_2	$\frac{15}{60}$	0	$\frac{15}{60}$
1	-2	0	y_3	$\frac{10}{60}$	0	$\frac{10}{60}$
1	-1	0	y_4	$\frac{7}{60}$	0	$\frac{7}{60}$
1	0	0	y_5	$\frac{6}{60}$	0	$\frac{6}{60}$
1	0	0	y_6	$\frac{6}{60}$	0	$\frac{6}{60}$
1	1	0	y_7	$\frac{7}{60}$	0	$\frac{7}{60}$
1	2	0	y_8	$\frac{10}{60}$	0	$\frac{10}{60}$
1	3	0	y_9	$\frac{15}{60}$	0	$\frac{15}{60}$
1	4	1	y_{10}	$\frac{22}{60}$	1/2	$-\frac{8}{60} + (y_{10} - y_1)/4\hat{\sigma}^2$

$$C_p = 1 + (y_{10} - y_1)^2/2\hat{\sigma}^2$$

Table 2. Data from Narula and Wellington

Case	Variable Number											
	1	2	3	4	5	6	7	8	9	10	11	Y
1	4.9176	1.0	3.4720	0.9980	1.0	7	4	42	3	1	0	25.9
2	5.0208	1.0	3.5310	1.5000	2.0	7	4	62	1	1	0	29.5
3	4.5429	1.0	2.2750	1.1750	1.0	6	3	40	2	1	0	27.9
4	4.5573	1.0	4.0500	1.2320	1.0	6	3	54	4	1	0	25.9
5	5.0597	1.0	4.4550	1.1210	1.0	6	3	42	3	1	0	29.9
6	3.8910	1.0	4.4550	0.9880	1.0	6	3	56	2	1	0	29.9
7	5.8980	1.0	5.8500	1.2400	1.0	7	3	51	2	1	1	30.9
8	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	1	1	0	28.9
9	15.4202	2.5	9.8000	3.4200	2.0	10	5	42	2	1	1	84.9
10	14.4598	2.5	12.800	3.0000	2.0	9	5	14	4	1	1	82.9
11	5.8282	1.0	6.4350	1.2250	2.0	6	3	32	1	1	0	35.9
12	5.3003	1.0	4.9883	1.5520	1.0	6	3	30	1	2	0	31.5
13	6.2712	1.0	5.5200	0.9750	1.0	5	2	30	1	2	0	31.0
14	5.9592	1.0	6.6660	1.1210	2.0	6	3	32	2	1	0	30.9
15	5.0500	1.0	5.0000	1.0200	0.0	5	2	46	4	1	1	30.0
16	5.6039	1.0	9.5200	1.5010	0.0	6	3	32	1	1	0	28.9
17	8.2464	1.5	5.1500	1.6640	2.0	8	4	50	4	1	0	36.9
18	6.6969	1.5	6.9020	1.4880	1.5	7	3	22	1	1	1	41.9
19	7.7841	1.5	7.1020	1.3760	1.0	6	3	17	2	1	0	40.5
20	9.0384	1.0	7.8000	1.5000	1.5	7	3	23	3	3	0	43.9
21	5.9894	1.0	5.5200	1.2560	2.0	6	3	40	4	1	1	37.5
22	7.5422	1.5	4.0000	1.6900	1.0	6	3	22	1	1	0	37.9
23	8.7951	1.5	9.8900	1.8200	2.0	8	4	50	1	1	1	44.5
24	6.0931	1.5	6.7265	1.6520	1.0	6	3	44	4	1	0	37.9
25	8.3607	1.5	9.1500	1.7770	2.0	8	4	48	1	1	1	38.9
26	8.1400	1.0	8.0000	1.5040	2.0	7	3	3	1	3	0	36.9
27	9.1416	1.5	7.3262	1.8310	1.5	8	4	31	4	1	0	45.8
28	12.0000	1.5	5.0000	1.2000	2.0	6	3	30	3	1	1	41.0

X_1 = Taxes (100's of dollars)
X_2 = No. of baths
X_3 = Lot size/1000 ft^2
X_4 = Living space/1000 ft^2
X_5 = No. of garages
X_6 = No. of rooms
X_7 = No. of bedrooms
X_8 = Age of house
X_9 = Construction type (coded 1,2,3,4)
X_{10} = Style (coded 1,2,3)
X_{11} = No. of fireplaces
Y = Sale price (1000's of dollars)

Table 3. Case statistics for the full model

CASE	Y	RESIDUAL	STUD. RES	u	DISTANCE	T
1	25.90	1.685	.6976	.6503	.0754	.69
2	29.50	-.1648	-.0555	.4720	.0002	-.05
3	27.90	1.170	.3249	.2222	.0025	.32
4	25.90	-2.947	-.8143	.2145	.0151	-.81
5	29.90	2.296	.5904	.0931	.0030	.58
6	29.90	6.998	1.9910	.2593	.1157	2.22
7	30.90	.6188	.1918	.3759	.0018	.19
8	28.90	-1.812	-.5828	.4206	.0205	-.57
9	84.90	4.002	1.8457	.7181	.7233	2.01
10	82.90	2.895	1.2593	.6832	.2849	1.28
11	35.90	5.326	1.5937	.3302	.1044	1.68
12	31.50	-2.532	-.7797	.3676	.0295	-.77
13	31.00	2.620	.8216	.3901	.0360	.81
14	30.90	.5920	.1744	.3090	.0011	.17
15	30.00	.6312	.2187	.5005	.0040	.21
16	28.90	-1.812	-.5828	.4206	.0205	-.57
17	36.90	-5.386	-1.6595	.3684	.1339	-1.77
18	41.90	.5884	.2251	.5902	.0061	.22
19	40.50	1.390	.4054	.2955	.0057	.39
20	43.90	3.390	1.1999	.5215	.1308	1.22
21	37.50	1.167	.4359	.5701	.0210	.42
22	37.90	-3.435	-1.0855	.3994	.0653	-1.09
23	44.50	-.9019	-.2736	.3486	.0033	-.27
24	37.90	-3.791	-1.2766	.4712	.1210	-1.30
25	38.90	-5.621	-1.6577	.3107	.1032	-1.76
26	36.90	-3.434	-1.2091	.5164	.1301	-1.23
27	45.80	-.1528	-.0478	.3876	.0001	-.05
28	41.00	-3.380	-1.8199	.7932	1.0587	-1.98

RESIDUAL SS = 266.8499856
PRESS = 1147.330450

Table 4. 10 Best regressions for n = 28 and n = 27

n = 28 cases

BEST 10 REGRESSION WITH Y DEPENDENT, USING CP

P		C(P)	R2(ADJ)	R**2	RSS	VARIABLES
3	*	2.620	.9180	.9241	410.6	1 4
4	*	1.940	.9239	.9324	365.9	1 2 4
4	*	3.013	.9202	.9291	383.8	1 4 9
4	*	3.202	.9195	.9285	387.0	1 4 11
5	*	3.026	.9239	.9352	350.7	1 2 4 10
5	*	3.134	.9235	.9349	352.5	1 2 4 7
5	*	3.147	.9235	.9348	352.7	1 2 4 9
5	*	3.169	.9234	.9348	353.1	1 2 4 8
5	*	3.187	.9233	.9347	353.4	1 2 4 11
6	*	3.285	.9271	.9406	321.6	1 4 8 9 11

CP TIME USED IS .531 SECONDS.

n = 27 cases (case #28 deleted.)

BEST 10 REGRESSION WITH Y DEPENDENT, USING CP

P		C(P)	R2(ADJ)	R**2	RSS	VARIABLES
3	*	1.535	.9363	.9412	317.9	1 2
4	*	2.293	.9371	.9444	300.4	1 2 11
4	*	2.394	.9369	.9441	301.8	1 2 4
4	*	2.917	.9353	.9428	309.2	1 2 6
4	*	3.201	.9345	.9420	313.2	1 2 9
4	*	3.371	.9340	.9416	315.6	1 2 7
5	*	2.978	.9383	.9478	281.8	1 2 4 11
5	*	3.029	.9382	.9477	282.6	1 2 4 6
5	*	3.379	.9371	.9468	287.5	1 2 6 11
6	*	3.029	.9417	.9529	254.3	1 2 4 6 11

CP TIME USED IS .479 SECONDS.

Table 5. C_{pi} and C^*_{pi} for 3 models

	(X_1, X_2)				(X_1, X_4)				(X_1, X_2, X_4)			
	n = 28		n = 27		n = 28		n = 27		n = 28		n = 27	
Case	C_{pi}	C^*_{pi}	C_{pi}	C^*_{pi}	C_{pi}	C^*_{pi}	C_{pi}	C^*_{pi}	C_{pi}	C^*_{pi}	C_{pi}	C^*_{pi}
1	.45	2.12	-.12	2.24	-.37	2.10	-.46	2.20	-.18	2.07	-.27	2.17
2	-.28	1.50	-.23	1.52	.30	1.40	-.02	1.27	.03	1.34	-.09	1.19
3	-.05	.59	-.07	.59	-.07	.60	-.06	.59	-.03	.59	-.06	.58
4	.05	.57	.01	.64	-.03	.56	-.02	.60	-.04	.56	-.02	.60
5	.08	.18	.03	.19	.02	.18	.02	.19	.03	.18	.03	.19
6	.51	.64	.27	.63	-.08	.69	-.08	.71	.13	.60	.21	.63
7	-.27	1.18	-.22	1.19	-.21	1.20	-.14	1.21	-.25	1.18	-.23	1.19
8	-.27	1.34	-.31	1.43	.44	1.29	.55	1.32	.07	1.20	.06	1.20
9	.32	1.58	.09	1.55	.72	1.30	.42	1.34	.55	1.28	.32	1.32
10	.76	1.50	.56	1.52	1.06	1.59	1.32	1.62	.57	1.49	.63	1.52
11	-.22	1.02	-.22	1.06	-.23	1.04	-.24	1.08	-.23	1.02	-.22	1.06
12	1.18	1.15	.35	1.37	-.17	1.04	-.12	1.11	-.00	.94	.14	.99
13	.25	1.21	-.23	1.66	-.18	1.11	-.15	1.37	-.18	1.11	-.15	1.37
14	-.20	.95	-.22	1.02	-.19	.94	-.16	.97	-.17	.93	-.15	.96
15	-.35	1.61	.09	1.81	.16	1.58	.76	1.77	-.01	1.57	.41	1.75
16	-.27	1.34	-.31	1.43	.44	1.29	.55	1.32	.07	1.20	.06	1.20
17	.07	1.16	.42	1.18	-.27	1.18	-.09	1.18	-.15	1.14	.23	1.13
18	-.36	1.77	-.39	2.08	.79	1.97	-.18	2.30	.06	1.69	-.37	2.02
19	1.05	.88	.19	1.14	-.07	.89	.03	1.00	.04	.65	.13	.73
20	.57	1.10	.25	.93	-.11	1.60	.71	1.40	.17	1.05	.24	.91
21	1.45	1.86	2.21	1.99	.73	1.88	1.60	2.02	1.17	1.85	2.24	1.99
22	-.08	1.22	-.30	1.37	-.31	1.29	-.20	1.44	-.17	1.22	-.29	1.37
23	-.24	1.10	-.19	1.13	-.21	1.10	-.22	1.13	-.21	1.10	-.21	1.12
24	.08	1.22	-.08	1.62	.80	1.45	.33	1.80	.18	1.22	-.13	1.60
25	-.20	.96	-.16	.98	-.22	.97	-.21	.98	-.20	.96	-.17	.97
26	1.22	1.35	-.08	1.35	-.42	1.68	.19	1.72	.32	1.28	-.08	1.35
27	-.25	1.22	.21	1.22	-.19	1.22	.17	1.21	-.22	1.22	.13	1.21
28	4.40	2.07			.50	1.28			.58	1.27		
:												
C_p =	9.41		1.54		2.62		4.29		1.94		2.39	