Gauging the robustness of regression estimates is especially important in small-sample analyses. Here, we examine recent developments in the detection and analysis of outliers and influential cases in multivariate studies. Specifically, we review five diagnostic procedures: partial regression plots, the "hat" matrix, studentized residuals, $DFITS_i$, and $DFBETAS_{ij}$. The main part of the article presents two empirical applications (drawn from recent cross-national studies) that show (a) how the diagnostic procedures can be incorporated into the research process, and (b) what we can learn from them. These applications serve to underscore the point that the diagnostics cannot be employed mechanically. Instead, once a case is diagnosed as influential, remedial action requires a firm substantive grounding. Although case deletion may be warranted in some circumstances, it is an extreme remedy of last resort that should not be routinely followed. The more fruitful approach is to ask why a given case is influential. As our applications indicate, the diagnostics can be helpful in isolating such problems as sample composition, specification error, and errors in measurement.

Regression Diagnostics

An Expository Treatment of Outliers and Influential Cases

KENNETH A. BOLLEN
University of North Carolina at Chapel Hill

ROBERT W. JACKMAN
Michigan State University

Regression analysis is a powerful tool in social research because it helps to identify and summarize relations between variables. The emphasis on generalization is critical: Among the many assumptions that statistical analysis involves is the idea that a minority of observations does not determine the obtained results.

AUTHORS' NOTE: *We would like to thank Brian Silver and the anonymous referees for their comments, and to acknowledge support from the National Science Foundation (SES-8112023). This work was done while Bollen was at Dartmouth College. Authors' names are listed alphabetically.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 13 No. 4, May 1985 510-542

We are justly skeptical of empirical results that are unduly sensitive to one case (or to a very small number of observations).

Whether or not regression estimates are sensitive to one or a few cases is in part a function of sample size. With very large samples, such as those typical of survey research, it is relatively unlikely that a few cases are responsible for the observed patterns. With smaller samples, however, the sensitivity of estimates to one or two cases can assume more significance. Even so, if the model is bivariate, it is possible to gain some perspective on the robustness of small-sample estimates by examining such items as the univariate distributions of the variables, bivariate scattergrams, and the residuals from the fitted model.

As the complexity of the model increases, simple diagnostic checks become more difficult. Most notably, when one is interested in a multivariate model, the examination of bivariate scatterplots is not always an optimal strategy. One alternative is to examine residuals, but even this strategy can be inadequate. For instance, one observation may have sufficiently extreme values on both the response variable and one or more of the regressors so that it has an overriding effect on the estimates, even though the residual for that observation is small.

The goal of this article is to provide an expository treatment of recent developments in the analysis of outliers and influential data points in the multiple regression framework. The term "influential data point" refers to an observation that either by itself or along with other observations, "has a demonstrably larger impact on the calculated values of various estimates (coefficients, standard errors, t-values, etc.) than is the case for most of the other observations" (Belsley et al., 1980: 11).

It is useful to distinguish influential cases from outliers. In a most general sense, outliers are observations that are distinct from most of the data points in a sample. This distinctiveness can assume a variety of forms. For example, an outlier may be a case with a residual that is large compared to the other residuals from a regression analysis. Alternatively, an outlier may be a case with an extremely high or low value on one variable. Yet these kinds of outliers are not necessarily influential cases. An outlier observation can only be called influential when its deletion from the analysis causes a pronounced change in one or more of the estimated parameters. Thus an influential case is a special kind of outlier.

It is also useful to remember that whether or not an observation is an outlier depends on the context of the analysis, that is, the sample of observations, the functional form of the model, the variables included in the model, and so on. An observation that is an outlier in one sample may no longer be an outlier after the sample has been supplemented with additional cases. Similarly, a change in functional form through the transformation of one or more variables may change an observation so that it is no longer an outlier.

Alternatively, a case may be an outlier because an important explanatory variable has been omitted from the model—indeed, in such a setting, the outlier may be the most important case in the analysis because it points to this form of specification error. And if the omitted variable is correctly identified and incorporated into the analysis, the former outlier case will no longer have the same status.

That the classification of a case as an outlier or not depends on the context of the model means that once a case is diagnosed as an outlier in a given analysis, a variety of possible remedial actions is available. And the choice among these remedies must be based on substantive knowledge of the model concerned. This is a point that our examples are intended to underscore.

We begin with a brief review of the diagnostics to be employed, which include graphic displays, measures based on residuals, and measures that highlight influential observations. In the main part of the article, these diagnostics are applied to two substantive examples from recently published contributions to the cross-national literature. As will become clear, we regard these procedures as a major tool in regression diagnostics that provide considerable information on the fit of a model to a given set of data, and therefore on the robustness of a given set of estimates. In this sense, they can be an important guide to understanding.

We also emphasize at the outset, however, that these are not procedures to be employed mechanically—they are an aid to, rather than a substitute for, careful statistical analysis. Indeed, our emphasis on the substantive examples is intended to highlight the variety of problems that the diagnostics can help identify. Without substantive knowledge of a given problem, the diagnostics take us nowhere. Auxiliary information is required before appropriate remedies can be evaluated.

DIAGNOSTICS

A variety of diagnostic procedures has recently been proposed, and we obviously have to be selective. Our choice of a set of diagnostics is governed by two main considerations. First, we want a graphic display alternative to the traditional bivariate scattergram. Second, we need measures that aid in general outlier detection, and we also need ways of finding out whether these outliers are influential cases. Procedures that meet these criteria are proposed in Belsley et al.'s (1980) work, which is the major treatment to date of regression diagnostics. Some or all of these procedures are available in standard software packages.[1] Given these considerations, we review five diagnostic procedures in this section: (1) partial regression plots, (2) the "hat" matrix, (3) studentized residuals, (4) $DFITS_i$, and (5) $DFBETAS_{ij}$.

PARTIAL REGRESSION PLOTS

In the case of one independent variable (X) and one dependent variable (Y), the bivariate scattergram of Y and X reveals a great deal. For instance, it can help identify curvilinearity, heteroskedasticity, or extreme observations (outliers). However, as Daniel and Wood (1980: 50-53) demonstrate, the bivariate scattergram is not ideal in the more common situation of multiple explanatory variables. Many potential problems that are detectable in the bivariate case are not visible with scattergrams when multivariate relations are estimated.

The partial regression plot is the multivariate analog of the bivariate scattergram that overcomes many of the scattergram's limitations. Like the usual scattergram, it is a plot of two variables. But unlike it, the variables plotted are residual variables. These plots may be simply illustrated using an equation with two explanatory variables. The equation for the population is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \qquad [1]$$

The Ordinary Least Squares (OLS) estimate of 1 for the sample is:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + e \qquad [2]$$

A partial plot for each explanatory variable (and the intercept) may be constructed. The partial plot for X_1 is formed, in part, from the residuals of the regression of Y on X_2. Call these residuals $Y._{[1]}$. The second set of residuals comes from the regression of X_1 on X_2. Call these residuals $X_1 \cdot_{[1]}$. The partial regression plot shows the relation of $Y._{[1]}$ to $X_1 \cdot_{[1]}$. In a similar manner, the partial regression plot for X_2 relates the residuals from the regression of Y on X_1 to the residuals of X_2 regressed on X_1.

These plots have several well-known properties. First, the regression of $Y._{[1]}$ on $X_1 \cdot_{[1]}$ leads to the same regression coefficient (i.e., b_1) as the regression coefficient for X_1 in the regression of Y on X_1 and X_2. Second, the residuals from the $Y._{[1]}$ and $X_1 \cdot_{[1]}$ regression are equal to e of equation 2. Third, the correlation between $Y._{[1]}$ and $X_1 \cdot_{[1]}$ is equal to the partial correlation of Y and X_1, controlling for X_2. Analogous results hold true for the partial regression plot of X_2. Among the most important characteristics of these plots is that they enhance an analyst's ability to see outliers not visible in the ordinary scattergram. Further, curvilinear relations and other possible violations of OLS assumptions are more apparent.

Partial regression plots are sometimes also labeled "partial plots," "partial-regression leverage plots," and "added variable plots."[2] Other discussions of these plots are in Belsley et al. (1980), Velleman and Welsch (1981), and Cook and Weisberg (1982). One of the attractive features of these plots from a practical point of view is that they are readily generated from most of the available regression computer packages; all that is required is the capacity to save and analyze OLS residuals.

"HAT" MATRIX

Consider the general linear regression model:

$$Y = XB + \epsilon \qquad [3]$$

where Y is an $n \times 1$ vector of values for the dependent variable, X is an $n \times p$ matrix of the explanatory variables (and intercept), B is a $p \times 1$ vector of regression coefficients, and ϵ is an $n \times 1$ vector of disturbances. The number of observations is n, and the number of columns in X is p, and X is assumed to be of full column rank. The usual assumptions of OLS regression are made, namely, that ϵ is distributed independently of X, $E(\epsilon) = 0$, and $E(\epsilon\epsilon') = \sigma^2 I$.

The OLS estimator of B is

$$b = (X'X)^{-1}X'Y. \qquad [4]$$

The predicted value of the dependent variable is formed as

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y. \qquad [5]$$

By defining a new matrix H, as $X(X'X)^{-1}X'$, equation 5 may be written as

$$\hat{Y} = HY. \qquad [6]$$

The $\hat{Y}$ (Y-hat) vector equals the observed dependent variable, Y, premultiplied by the $n \times n$ matrix H. In other words, Y is transformed into $\hat{Y}$ with the H matrix. Because of this, H is referred to as the "hat matrix" (Hoaglin and Welsch, 1978: 17).

The diagonal elements of H, usually called h_i (or h_{ii}), are important to regression diagnostics. They possess a number of useful properties. Each h_i gives the "leverage" exerted on $\hat{Y}_i$ by Y_i. The h_i value is bounded by 0 and 1 ($0 \leq h_i \leq 1$).[3] The closer h_i is to 0, the less Y_i's leverage on $\hat{Y}_i$. On the other hand, $\hat{Y}_i$ is completely determined by Y_i when h_i is 1. In general, h_i will fall somewhere between these extremes.

Another property of these diagonal elements is:

$$\sum_{i=1}^{n} h_i = p. \qquad [7]$$

That is, the sum of all n h_i elements is equal to p, the number of variables in X. From 7, we can see that the mean (average) value of

h_i is p/n. Hoaglin and Welsch (1978: 18) suggest that a reasonable rule of thumb for large h_i is $h_i > 2p/n$.[4] Observations associated with h_i values greater than 2p/n have high leverage and therefore require further investigation.

It is instructive to consider the simple regression case, $Y_i = b_0 + b_1X_i + e$, to gain further insight into the meaning of the h_i elements. The leverage value for the ith observation is (Hoaglin and Welsch, 1978: 18):

$$h_i = 1/n + (X_i - \bar{X})^2 / \sum_{k=1}^{n} (X_k - \bar{X})^2 \qquad [8]$$

Equation [8] shows that the further X_i is from the mean ($\bar{X}$), the larger is h_i. The closer X_i is to $\bar{X}$, the less is the leverage of the data point. This illustrates the "distance" interpretation of h_i. The same distance interpretation holds for more than a single explanatory variable. The greater the h_i value, the further a case is from the mean of all the explanatory variables.

STUDENTIZED RESIDUALS

In OLS regression, the disturbance of the population, ϵ, is assumed to have the same variance for each observation ($E(\epsilon\epsilon') = \sigma^2 I$). However, the sample residuals, e, are generally not homoskedastic. Rather, their variance is a function of σ^2 and h_i:

$$\mathrm{var}(e_i) = \sigma^2 (1 - h_i). \qquad [9]$$

In equation [9], we see that those observations with the greatest leverage (the largest h_i's) have errors, e_i, with the smallest variance. Observations with the least leverage have the largest variance.

Because of these differences in the variances of the sample residuals, it is useful to standardize e_i. This helps to determine whether the residual is large relative to its variance. Belsley et al. (1980: 20) recommend use of studentized residuals (RSTUDENT):

$$e_i^* = e_i / \left(\sqrt{s^2(i)(1 - h_i)}\right) \qquad [10]$$

In equation 10, $s^2(i)$ is the sample estimate of the disturbance variance when the ith case is removed.

The studentized residuals are useful for several reasons. In many practical situations, Belsley et al. (1980) suggest that e^*_i is distributed closely to a t-distribution with n-p-l degrees of freedom (although like e_i's, the e^*_i's are not independent). Also, unlike the untransformed residuals (e_i), the studentized residuals e^*_i have equal variances. Finally, if a dummy variable is coded one for the ith observation and zero elsewhere, and a t-statistic is computed for the regression coefficient of this dummy variable, the resulting t-statistic is equal to e^*_i.

Despite these useful properties, caution must be observed in using studentized residuals. Unfortunately, the term "studentized residuals" is sometimes applied to transformations of e_i other than that described in 10. For instance, Weisberg (1980: 105) forms "studentized residuals" by using the sample-estimated disturbance variance (s^2) including the ith case, rather than s^2 (i) as in 10. In this alternative form, the numerator (e_i) and the denominator ($\sqrt{s^2[1 - h_i]}$) are not independent and are not distributed as a t-distribution.[5] Other residuals referred to as "standardized residuals" may be computed in alternative ways, such as e_i/s (Velleman and Welsch, 1981: 238). Thus the researcher using statistical packages should be sure to check the formula used to compute the "studentized" or "standardized" residuals.

More important than these issues is that reliance solely on residuals for detecting influential points may be misleading. An observation with a small residual may still be quite influential in the sense that its removal may seriously affect regression estimates. For example, this would occur in the bivariate case where the bulk of the observations exhibit no association, but a single point distant from the rest of the data pulls the regression line through it. While the residual associated with the observation is small, removal of the observation leads to a radical shift in the regression estimates. For this reason, we cannot rely solely on residuals.

$DFITS_i$[6]

In the two previous sections, we reviewed methods designed to detect observations with high leverage (i.e., large h_i values) and large studentized residuals (i.e., e^*_i). However, these two indicators of problem cases need not overlap. We may find observations with large h_i values but small e^*_i's, or vice versa (Hoaglin and Welsch, 1978: 20). Ideally, we should employ a measure that is affected by both deviant residuals and extreme leverage points. DFITS_i is one such measure:

$$\text{DFITS}_i = (\sqrt{(h_i/[1 - h_i])})\,(e_i/\sqrt{(s^2[i][1 - h_i])}) \qquad [11]$$

The $\sqrt{(h_i/[1 - h_i])}$ term in equation 11 is greatest for points with the greatest leverage. The remaining expression on the right-hand side of 11 is the formula for the studentized residual discussed in the previous section. Thus DFITS_i may be affected by large leverage points or by large studentized residuals. As a rough cutoff point, Belsley et al. (1980: 28) suggest that DFITS_i greater than $2\sqrt{(p/n)}$ require more investigation.

An alternative but equivalent representation of DFITS_i is:

$$\text{DFITS}_i = (\hat{Y}_i - X_i b_{(i)})/\sqrt{(s^2(i)h_i)} \qquad [12]$$

The numerator of equation 12 is the predicted Y_i value for the full sample minus the Y_i value predicted from the regression coefficients obtained with the ith observation omitted. The denominator of 12 is the standard deviation of the fit $\hat{Y}_i = X_i b$, with the disturbance variance (σ^2) estimated by s^2 (i). Combining the numerator and denominator, DFITS_i may be interpreted as a scaled measure of the change in fitted Y_i values. A change in any of the regression coefficients for the X variables may affect DFITS_i, because the fitted Y values depend on all the regression coefficients.[7]

$DFBETAS_{ij}$

DFITS_i takes into account changes in all the regression coefficients that result when a single observation is removed. It seems natural to have a measure of how individual coefficients change when a case is omitted. DFBETAS_{ij} is one such measure:

$$\text{DFBETAS}_{ij} = (b_j - b_j(i))/\sqrt{(s^2\,(i)(X'X)_{jj}^{-1})} \qquad [13]$$

The i subscript refers, as before, to the measure obtained when the ith observation is removed. The second subscript, j, is needed to index the jth component in the b vector. For instance, the regression coefficient associated with the third column of X would be indexed with j equal to 3.

The numerator of 13 is the difference in the regression coefficients for the jth variable estimated for the full sample and for the sample removing the ith observation. The denominator is the estimated standard error of the jth regression coefficient, with the disturbance variance (σ^2) estimated by s^2 (i). The double j subscript indicates the jth diagonal element of $(X'X)^{-1}$.

Large positive or negative values of $DFBETAS_{ij}$ indicate observations that lead to large changes in the jth regression coefficient. Belsley et al. (1980: 28) suggest a size-adjusted cutoff for $DFBETAS_{ij}$ as $2/\sqrt{n}$.

SUMMARY

By way of summary to this point, Table 1 lists the five procedures that we have discussed along with the approximate cutoff criteria for identifying "unusual" cases suggested by Belsley et al. (1980). We emphasize that these are approximate criteria that in some instances (most notably, when p and n are small) may be a little severe. For example, the suggested cutoff for h_i of 2p/n is useful with moderate sample sizes ($p > 10$ and $n - p > 50$), but "for small p, 2p/n tends to call a few too many points to our attention" (Belsley et al., 1980: 17). Thus Velleman and Welsch (1981: 234-235) suggest that for smaller p and n, 3p/n is more appropriate. Similarly, Velleman and Welsch (1981: 236-237) suggest a more conservative cutoff point for $DFITS_i$ of $\sqrt{p}$ (as opposed to the $2\sqrt{(p/n)}$ in Table 1).

Now that we have reviewed these five regression diagnostics, it is important to examine their behavior in an empirical setting to gain a practical sense of what we can learn from them. Both of our empirical applications are drawn from cross-national comparative research, and each illustrates a different use for the diagnostics. The

TABLE 1
Approximate Cutoff Criteria for Identifying "Unusual" Cases

Diagnostic	Cutoff*
Partial Regression Plot	Based on visual inspection of plot
Diagonal of Hat Matrix (h_i)	2p/n
Studentized Residuals (e^*_i) (RSTUDENT)	t-distribution** (df = n-p-1)
$DFITS_i$	$2\sqrt{(p/n)}$
$DFBETAS_{ij}$	$2/\sqrt{n}$

*Cutoff points are based on recommendations of Belsley et al. (1980).
**Unless suspected outliers are identified in advance and then tested, the usual critical t-values are not appropriate. See Weisberg (1980: 116-117) for discussion.

first cross-national example centers on the relation between voting turnout and income inequality in industrial societies, whereas the second is based on an analysis of economic dependency and political democracy.

Typically, published research reports the end-product of a long research process rather than showing each step of the analysis that led to the final results. This, of course, presents a distorted picture of how empirical research is actually done. And because we want to show how these diagnostics can be incorporated into the data analysis process, we start with an initial model and proceed step by step toward a more refined end product.[8]

CASE 1: VOTING TURNOUT AND INCOME INEQUALITY

The effect of democratic political systems and voter participation on the distribution of a society's income has long been of interest to social scientists. Hewitt (1977) and Stack (1979) provide two recent empirical analyses of this relationship. Hewitt selected a sample of industrial democratic societies with varying years of "democratic experience," and found that democratic experience had no effect on income inequality. Stack criticized Hewitt and suggested that voter turnout in national elections (as a percentage of the adult population) would be a superior measure of democracy. Stack then provided evidence that when turnout statistics are used, they have a significant negative effect on inequality.

Stack's (1979) analysis is useful from an expository point of view for several reasons. First, it employs a simple multivariate analysis confined to just two explanatory variables. Second, it is a small-sample analysis based on only eighteen countries. Third, examination of at least the simple bivariate scatterplots suggests that Stack's analysis suffers from a severe outlier problem (Jackman, 1980); whether such a conclusion is warranted in light of the more complete diagnostic procedures reviewed above remains an open question.

Following Stack, when inequality is regressed on voting turnout and energy consumption per capita, the estimates (with their t-ratios in parentheses) are as follows (Jackman, 1980: Table 1):

$$\underset{}{\mathrm{INEQ}} = \underset{(8.9)}{10.31} - \underset{(5.8)}{.081}\mathrm{TURNOUT} - \underset{(2.1)}{.0003}\mathrm{ENCAP} \qquad [14]$$

The overall equation has an $\bar{R}^2$ of .666 and an F-ratio of 18.0.[9]

Figures 1a and 1b display the partial plots for these estimates. The first figure contains the plot for turnout, and arrays the residuals obtained from regressing INEQ on ENCAP (vertical axis) against those from the regression of TURNOUT on ENCAP (horizontal axis). The second figure shows the corresponding partial plot for energy consumption. Inspection of these graphs suggest that there are two apparently influential observations in the analysis. Specifically, South Africa appears as an extreme outlier in the northwest corner of Figure 1a. Indeed, it is evident from this figure that South Africa could well be responsible for the negative estimated coefficient for TURNOUT reported above. In a similar but less extreme vein, the United States appears as an influential outlier in the southeast corner of Figure 1b that could be largely responsible for the negative estimated coefficient reported above for ENCAP.[10]

To pursue these issues further, Table 2 displays the remaining diagnostics by country. Overall, the entries in this table reinforce the

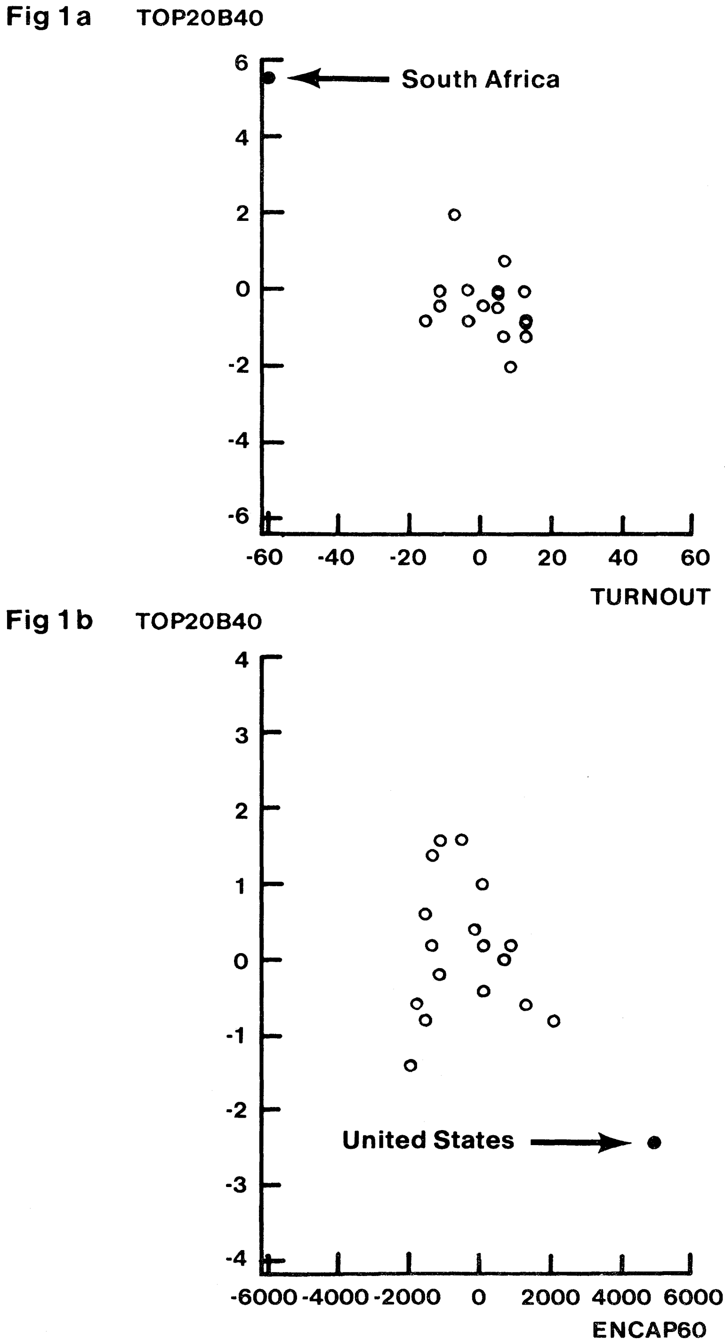

Figure 1: Partial Regression Plots for (a) TURNOUT and (b) ENCAP (N = 18)

TABLE 2
Regression Diagnostics for Estimates in Equation 14, n = 18

Country	Diagonal of Hat Matrix h_i	RSTUDENT e^*_i	$DFITS_i$	$DFBETAS_{ij}$ TURNOUT	$DFBETAS_{ij}$ ENCAP
Argentina	.15	-2.55*	-1.06*	.52*	.71*
Australia	.12	-.25	-.09	-.05	-.05
Denmark	.09	.35	.11	.07	.01
Finland	.09	1.43	.45	.16	-.21
France	.07	1.64	.44	-.17	-.10
West Germany	.08	.57	.17	.05	.08
Israel	.12	-1.25	-.46	-.17	.26
Italy	.15	.21	.09	.05	-.05
Japan	.11	-1.16	-.40	.05	.28
Netherlands	.10	1.20	.40	.26	.03
Norway	.07	-.39	-.11	-.05	-.01
Puerto Rico	.09	-.22	-.07	.00	.04
South Africa	.75*	2.22*	3.89*	-3.73*	-.74*
Sweden	.07	.25	.07	.02	.03
Trinidad & Tobago	.07	-.59	-.18	.08	.08
United Kingdom	.15	-.06	-.02	.00	-.02
United States	.64*	-1.70	-2.27*	.36	-2.07*
Venezuela	.06	.41	.11	.03	.00
Cutoff points (n = 18, p = 3)	.33	2.1	.82	.47	.47

*Exceeds cutoff point in absolute value.

conclusions drawn earlier. Looking first at the diagonal elements of the hat matrix (i.e., the h_i in the first column of the table), most of the values are small. However, the h_i for South Africa of .75 is large, while the h_i for the United States (.64) is only slightly smaller. Both figures easily exceed the 2p/n cutoff of .33 and, indeed, both are substantially larger than even the more conservative cutoff of 3p/n (= .50). Remember that h_i is bounded by 0 and 1, and a value of 1 means that the fitted value of Y_i is wholly determined by Y_i. The two-starred h_i values in Table 2 are clearly much closer to 1 than they are to 0.

The second column in Table 2 displays the studentized residuals (the e^*_i). With $n = 18$, $p = 3$, the appropriate cutoff here ($p = .05$) is 2.15.[11] Again, the value for South Africa of 2.27 is large and exceeds this cutoff. The e^*_i for the United States is somewhat smaller than the cutoff point, but it remains the third largest e^*_i in the table. Note, too, that Argentina has the largest e^*_i, but for reasons discussed earlier, we are reluctant to identify Argentina as an influential case solely on the basis of studentized (or any other) residuals.

The $DFITS_i$ are displayed in the third column of Table 2. It is clear that those for South Africa and the United States are extreme: Both exceed the $2\sqrt{(p/n)}$ criterion and even the less stringent $\sqrt{p}$ criterion. Indeed, the value of 3.89 for South Africa is more than twice the size of even the latter criterion. Thus the $DFITS_i$ figures reinforce the impression from the other diagnostics that these two cases are highly influential. Note also that the $DFITS_i$ for Argentina is larger than the $2\sqrt{(p/n)}$ criterion, although it remains smaller than the $\sqrt{p}$ criterion; this case again seems somewhat influential, but its influence is much smaller than the other two observations.

Finally, the $DFBETAS_{ij}$ are reported in the fourth and fifth columns of the table for TURNOUT and ENCON, respectively. These figures show the influence of individual cases on each of the two coefficients (rather than the overall fitted values), and reinforce our earlier reading of the partial plots. The TURNOUT DFBETA for South Africa is -3.73, almost eight times the size of the $2/\sqrt{n}$ cutoff value of .47, and indicates that South Africa has a major effect on the TURNOUT coefficient. The ENCAP DFBETA for the United States is somewhat smaller, but remains over four times the size of the cutoff value, which suggests that this case substantially influences the coefficient for ENCAP. Note also that both DFBETAS for Argentina are larger than the cutoff, but much smaller than the two others just discussed.

Taken as a whole, these diagnostics clearly identify South Africa and the United States as problem cases that are substantially influencing the estimates. They also suggest that Argentina behaves somewhat unusually. We can further gauge the impact of these outliers by examining the metric regression coefficients and the $\bar{R}^2$s gen-

TABLE 3
Regression Estimates Removing, in Turn, South Africa, the United States, and Argentina [coefficients (t-ratios)]

Country removed	Intercept	TURNOUT	ENCAP	$\bar{R}^2$	F-ratio	N
South Africa	6.457 (3.2)	-.035 (1.4)	-.0002 (1.6)	.083	1.7	17
United States	10.048 (9.1)	-.086 (6.4)	-.00003 (0.1)	.712	20.7	17
Argentina	11.130 (10.7)	-.088 (7.2)	-.0004 (3.1)	.769	27.6	17

erated when each case is dropped. As is evident in Table 3, there are several differences between the resulting sets of estimates and those for the full sample in equation 14 above. First, South Africa has a substantial influence on the estimates. Removing just this one case drastically reduces both the $\bar{R}^2$ (from .666 to .083) and the coefficient for TURNOUT (from −.081 to −.035). In fact, with South Africa excluded, the coefficients for both TURNOUT and ENCAP have t-ratios considerably less than 2.0.

The United States seems to be the next most influential case. When it is removed, a drastic drop toward zero occurs for the ENCAP coefficient, which has a t-ratio of 0.1 (compared to 2.1 in the full sample). The absolute values of the TURNOUT coefficient and the $\bar{R}^2$ marginally increase. The last influential case is Argentina. Removing this case increases the $\bar{R}^2$ from .666 to .769, and leads to a slight increase in the absolute value of the coefficients for TURNOUT and ENCAP.

These regression estimates reinforce the diagnostics reported in Table 2. South Africa and the United States are influential cases, with Argentina having lesser effects. Now that these cases have been highlighted, we need to ask, why are they so influential? Here we must draw on our substantive and empirical knowledge.

First, consider South Africa and Argentina. If we recall that the sample of countries is supposed to consist of industrial democracies, it is rather odd to find South Africa and Argentina included in this group. Although they may meet the criteria for industrialization set forth by Hewitt (1977: 450), they did not have democratic political systems in the early 1960s.[12] Examining the other countries in the sample, the same argument can be made for Venezuela, which, in the period to which the analysis refers, was not particularly democratic.

These arguments raise questions about the appropriateness of including South Africa, Argentina, and Venezuela in a sample of industrial democracies and, indeed, one could make a strong case for excluding these three countries from the analysis. However, for the purpose of further illustrating these regression diagnostics, we do not do this. Instead, we retain Argentina and Venezuela in the sample because of their relatively minor influence on the estimates, but we do provisionally remove South Africa, given its extreme outlier position. We will return to South Africa's influence on the regression estimates shortly.

What other corrective action remains, given that the United States remains an influential data point? Note that the estimates under consideration are based on raw ENCAP figures, that these raw scores are badly skewed, that the ENCAP value for the United States of 8047 is an extreme value (even within this subset of industrial countries) that is over 60% higher than the next value of 4907 for the United Kingdom, and that it is the ENCAP coefficient that is most affected by the inclusion of the United States. This suggests that instead of excluding the United States from the analysis, we might profitably consider a simple transformation of ENCAP to reduce the skewness of its distribution: A natural logarithmic transformation would appear appropriate.

The next step in the analysis is to reestimate the model with South Africa excluded and ENCAP transformed, and then to repeat the diagnostic checks on the revised estimates. With such a procedure, we can withhold judgment on the status of Argentina and Venezuela in the analysis until we examine the revised diagnostics.

The revised estimates are:

$$\underset{}{\text{INEQ}} = \underset{(2.0)}{7.97} - \underset{(1.1)}{.026}\text{TURNOUT} - \underset{(0.9)}{.359}\,ln\text{ENCAP} \qquad [15]$$

With n = 17 and t-ratios in parentheses, this overall equation has an $\overline{R}^2$ of .000 and an F-ratio of only 0.8, which is statistically in-significant at any meaningful test level. It is evident that these estimates are similar to the Table 3 estimates removing South Africa, but radically different from the full-sample estimates in equation 14. Where the equation 14 figures indicated that both TURNOUT and ENCAP accounted for about two-thirds of the variance in INEQ, the revised estimates in equation 15 imply that neither of these two explanatory variables influences INEQ, at least among the relatively wealthy countries under consideration here. Note, too, that the size of the primary coefficient of interest (that for TURNOUT) drops to a third of its original value (from –.0813 to –.0259).

Figures 2a and 2b display the partial plots for these new estimates: Figure 2a shows the revised partial plot for TURNOUT; Figure 2b has the corresponding plot for *ln*ENCAP. Comparing these partial plots with the originals indicates that one case (France) stands a little apart from the other points in Figure 2a and, to a lesser extent, in Figure 2b. However, compared to the outlier status of South Africa in Figure 1a, and even of the United States in Figure 1b, it is evident that France is best seen as a minor outlier in both Figures 2a and 2b.[13]

Of more significance is the pattern followed by the residuals of Figure 2b. Specifically, these indicate that equation 15 misspecifies the relation between ENCAP and INEQ, and suggest instead that an inverted U-shaped curve may be more appropriate, even for this relatively wealthy subset of countries. Such a curve was originally proposed by Kuznets (1955), and has been discussed by many others since (e.g., Ahluwalia, 1976). However, Kuznets was concerned with countries at all levels of economic development, and we did not initially anticipate such a pattern for the wealthier subset of countries under consideration here.

Accordingly, we estimate a revised model that includes a quadratic term for *ln*ENCAP, again excluding South Africa:

$$\begin{aligned}\text{INEQ} = &-\underset{(2.9)}{100.19} - \underset{(2.7)}{.060}\text{TURNOUT} + \underset{(3.1)}{28.094}ln\text{ENCAP} \\ &- \underset{(3.1)}{1.818}(ln\text{ENCAP})^2\end{aligned} \qquad [16]$$

With n = 17, and t-ratios in parentheses, this overall equation has an $\bar{R}^2$ of .371 and an F-ratio of 4.1, which is statistically sig-

Fig 2a

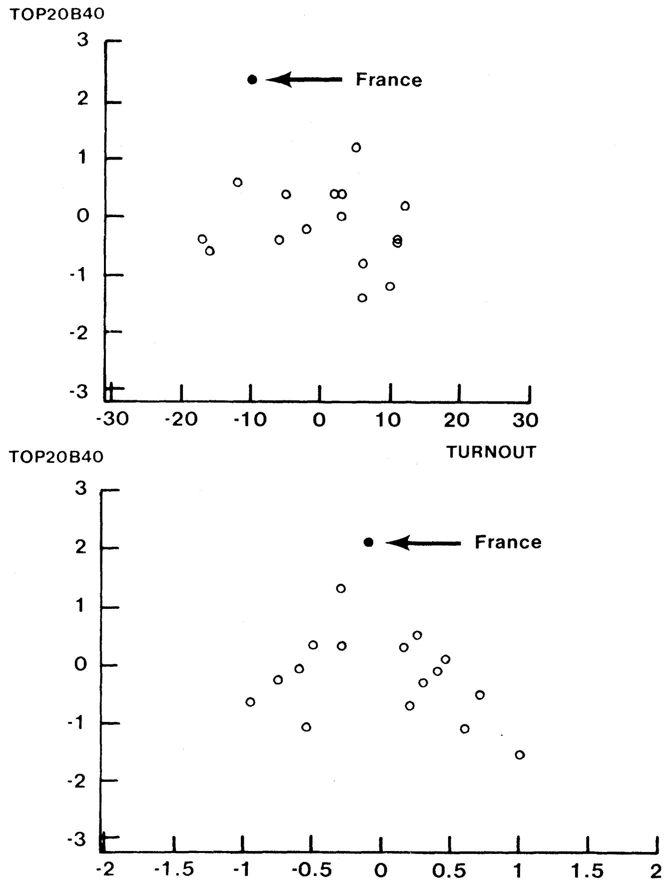

Figure 2: Revised Partial Regression Plots for (a) TURNOUT and (b) lnENCAP (N = 17)

nificant at the .03 level. Comparing these figures with those in equation 15, which are estimated for the same cases, we see that the misspecification of ENCAP suppressed the effects of both TURNOUT and ENCAP, even with South Africa excluded.

The partial plots associated with equation 16 are displayed in Figure 3: Figure 3a has the plot for TURNOUT; Figures 3b and 3c show the plots for *ln*ENCAP and (*ln*ENCAP)^2, respectively. Com-

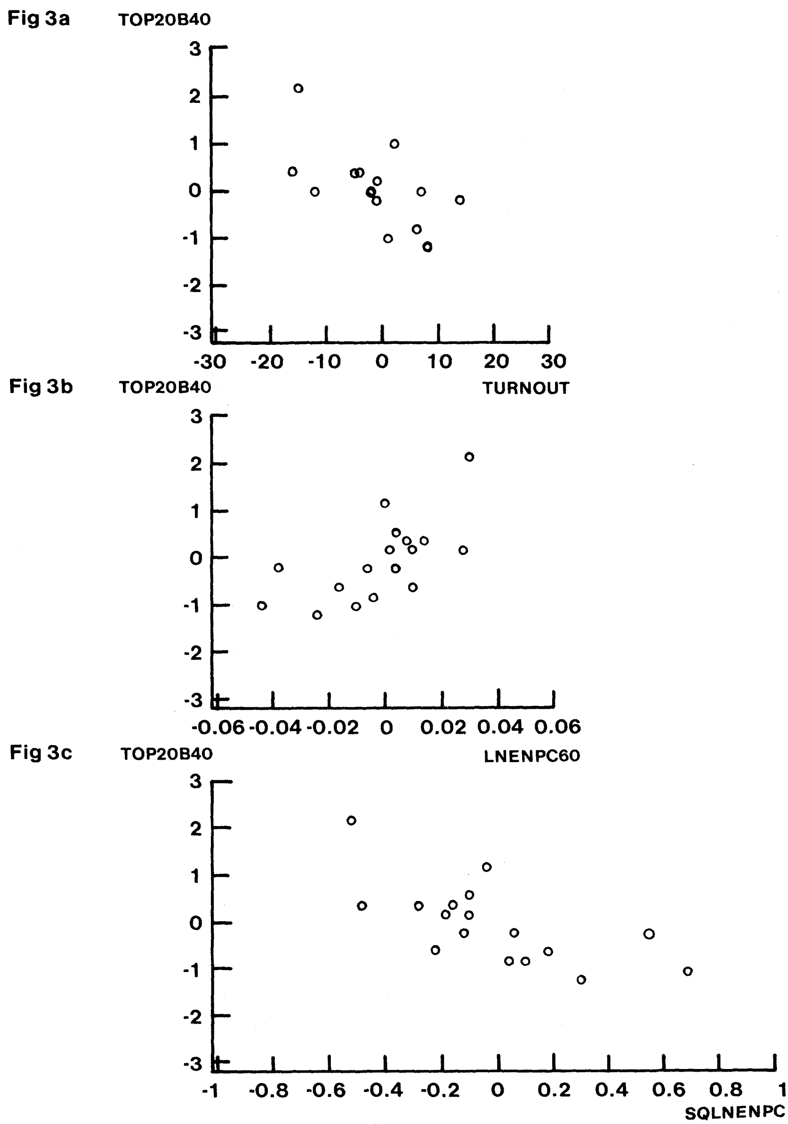

Figure 3: Revised Partial Regression Plots for (a) TURNOUT, (b) lnENCAP, and (c) $(\ln ENCAP)^2$ (N = 17)

pared with the earlier partial plots, those in Figure 3 are relatively well-behaved. First, there are no extreme outliers, as there were in Figure 1, although one case does stand out a little in each of the panels of the figure. The case involved is, again, France. Second, there is no evidence of nonlinearity, which contrasts markedly with the pattern evident in Figure 2b. In sum, the plots in Figure 3 do not identify any major problems.

The remaining diagnostics are displayed by country in Table 4, with the diagonal elements of the hat matrix (the h_i) in the first column. All except one are smaller than the 2p/n criterion of .47, but the h_i for the United States is larger than even the less severe cutoff criterion of 3p/n. Evidently, the United States remains a high leverage observation. The second column of Table 4 contains the studentized residuals. All cases but one are smaller than 2.2, but the e^*_i for France is 2.51. Thus France is flagged by both the e^*_i and the partial plots.

It is interesting that both the United States and France have comparatively large values in the remaining columns of the table. The $DFITS_i$ for both countries are larger than the $2\sqrt{(p/n)}$ criterion of .97, although only the United States has a value that exceeds the less stringent $\sqrt{p}$ criterion of 2.0. Similarly, both countries have large $DFBETAS_{ij}$ for *ln*ENCAP and its square, and France also has a large $DFBETA_{ij}$ for TURNOUT. In addition, the diagnostics in the last four columns for Italy are larger than the cutoff points (but smaller than those for France and the United States). Note, too, that the first two diagnostics do not suggest that Italy is an excessively influential observation.

Taken as a group, then, the diagnostic procedures in Figure 3 and Table 4 consistently isolate France and the United States as potentially influential cases. What implications does this have for the estimates in equation 16? Our own additional analyses indicate that the TURNOUT coefficient is especially sensitive to the inclusion of France. When this case and South Africa are excluded (n = 16), the TURNOUT coefficient drops to –.037 (t-ratio: 1.8), which is a drop of about 40% below the corresponding figure in equation 16. At the same time, the $\overline{R}^2$ drops by almost one-half to .205, with an F-ratio of 2.3. (The additional exclusion of the United States does not influence the TURNOUT coefficient.) In contrast, the exclusion of France and the United States produces *ln*ENCAP coeffi-

TABLE 4
Regression Diagnostics for Estimates in Equation 16, N = 17

COUNTRY	DIAGONAL OF HAT MATRIX h_i	RSTUDENT e^*_i	$DFITS_i$	$DFBETAS_{ij}$ TURNOUT	lnENCAP	$lnENCAP^2$
Argentina	.38	-1.23	-.96	.54*	.16	-.14
Australia	.18	-1.13	-.54	.33	.05	-.06
Denmark	.15	- .45	-.19	-.10	-.04	.04
Finland	.10	1.92	.63	.19	.04	-.05
France	.28	2.51*	1.57*	-1.24*	1.16*	-1.16*
West Germany	.11	.24	.09	.00	.03	-.03
Israel	.23	1.08	-.59	-.32	.33	-.32
Italy	.43	1.52	1.33*	.93*	-.91*	.89*
Japan	.21	.29	-.15	.02	.05	-.05
Netherlands	.16	.65	.29	.17	.05	-.05
Norway	.11	-1.39	-.49	-.09	-.23	.23
Puerto Rico	.12	.16	.06	-.02	.01	-.01
Sweden	.11	-.24	-.09	.00	-.04	.03
Trinidad & Tobago	.31	-.85	-.57	.47	-.37	.38
United Kingdom	.17	-.05	-.02	.00	.00	.00
United States	.83*	.94	2.06*	-.22	-1.20*	1.24*
Venezuela	.11	-.08	-.03	.00	-.02	.02
Cutoff points (N = 17, p = 4)	.47	2.2	.97	.49	.49	.49

*Exceeds cutoff point in absolute value.

cients of similar magnitude to those in equation 16. This seems to suggest that the estimates for *ln*ENCAP in that equation are relatively robust, but that the estimate for TURNOUT is somewhat less reliable.

There are two other points about the estimates for the quadratic model worth noting. First, the parallel estimates with South Africa included (n = 18) are:

$$\begin{aligned} \text{INEQ} = {} & -119.49 - .083\,\text{TURNOUT} + 33.556\,ln\text{ENCAP} \\ & \quad (3.9) \qquad (8.0) \qquad\qquad\qquad (4.2) \\ & - 2.170(ln\text{ENCAP})^2 \\ & \quad (4.3) \end{aligned} \tag{17}$$

This equation has an $\bar{R}^2$ of .814 and a highly-significant F-ratio of 25.8. Thus, even with the revised specification, South Africa remains a highly influential observation in the sense that excluding it reduces the $\bar{R}^2$ by over one-half. However, its impact on the TURNOUT coefficient is less extreme than it was in the original specification, involving a reduction of about 30% (from .083 to .060), and the TURNOUT DFBETA for South Africa is −2.24, which, though large, is smaller than the figure of −3.73 reported in Table 2. Although we have continued to exclude South Africa given both its influence on the estimates and its distinctive political arrangements, it is clear that its extreme influence on the estimates of equation 14 is in part due to the misspecification of ENCAP in that equation.

Second, we should note that inspection of the diagnostics obtained with both South Africa and France excluded indicates that other cases were becoming mildly influential. For example, in addition to the United States, the DFBETAS_{ij} for Italy and Israel exceeded the cutoff points of Table 1, although they were of opposite sign. Assuming for the moment that removal of cases is the only remedy for influential cases and that the cutoff points are to be applied mechanically, one might be tempted to repeat the process iteratively, calculating the diagnostics, removing apparently influential cases, recalculating the diagnostics, and so on. We hope that these diagnostics are not misused in this way. For one thing, in a purely practical sense, such a strategy has to generate an increasing number of "influential" observations.

Beyond this, as we have emphasized throughout, these diagnostics are not procedures to be employed mechanically—indeed, with a small n (as in the example here), the cutoff points summarized in Table 1 are likely to be too severe and therefore to flag too many observations. Additionally, the idea that case removal is the only remedy for apparently influential cases is quite misleading. In fact, case removal is the most severe remedy—sometimes, as in the present example, less severe action such as a change in functional form is optimal, while at other times, no "remedy" at all is called for.

There are at least three general lessons to be learned from our analysis. First, the diagnostics helped identify problems of *sample composition,* by identifying South Africa and Argentina as influential cases. Neither these two countries nor Venezuela belongs in a sample of industrial democracies for the period under investigation, and this is true even though Venezuela did not appear as an outlier. If we take the arguments of Hewitt and Stack seriously, then all three countries should be excluded from the analysis because none was especially democratic during the period under investigation. Alternatively, if we are concerned with a more diverse sample of countries, these cases may in fact be seen as providing useful information rather than as outliers. If this view is correct, then our analyses would suggest a need to obtain data on additional nondemocratic countries to broaden the nature of the sample by ensuring that a reasonable portion of the sample does represent such cases. An important implication is that the status of an observation depends in good part on the substantive definition of the sample (and population) being studied; a case may be an outlier in one setting but not in another.

Second, the diagnostics helped identify a *misspecification of functional form.* This possibility was indicated by the outlier status of the United States in the earlier part of our analyses. The partial plots, combined with previous research, indicated that the inverted U-shaped relation between level of economic development and inequality suggested in other research held even within the context of an "industrialized" group of countries. Correcting the misspecification increased the fit of the model considerably and, of more interest, indicated that the misspecification itself was responsible for suppressing some of the effects of voter turnout on inequality. This indicates that the diagnostics we have discussed are useful tools in the analysis of specification error.

Third, we hope it is evident from the foregoing that a *synergy* of statistical diagnostics and substantive knowledge is required. Specifically, the diagnostics did not by themselves somehow magically "show" the problems of sample composition or of misspecification in functional form. Rather, they helped to highlight abnormalities in the data that demanded explanation. In the next section, we will illustrate a different application.

CASE 2: WORLD SYSTEM POSITION AND POLITICAL DEMOCRACY

World system theory (Wallerstein, 1974; Chirot, 1977) has had a major impact on the analysis of international development. Basic to this view is the division of the world system into three sets of countries: the core, the semiperiphery, and the periphery. The core consists of the industrialized powerful nations, the periphery contains the poor and relatively weak ones, and the semiperiphery is made up of countries inbetween that are striving for core status. The problem that our next empirical example addresses is whether periphery or semiperiphery status has a negative impact on political democracy. The details of the theoretical arguments suggesting such a relationship are detailed in Bollen (1983). Here we are concerned with the narrower question of whether regression diagnostics help us to analyze the relationship.

Following Bollen (1983: Table 1), when political democracy (POLDEM) is regressed on dummy variables indicating semiperipheral (SEMPER) and peripheral (PER) positions, and economic development (*ln*ENCAP), the following results are obtained (t-ratios in parentheses):

$$\underset{}{\text{POLDEM}} = \underset{(0.5)}{7.29} + \underset{(5.7)}{10.05}\,ln\text{ENCAP} - \underset{(0.4)}{2.71}\text{SEMPER} - \underset{(0.8)}{6.76}\text{PER} \qquad [18]$$

With an n of 100, this equation has an $\bar{R}^2$ of .453 and an F-ratio of 28.3.[14]

Given these regression results, it is tempting to conclude that *ln*ENCAP has a positive and significant effect on political democracy, while SEMPER and PER have no meaningful effects at all. But, as we shall see, such a judgment is premature.

Figure 4 displays the partial plot for SEMPER, and reveals a minimum of three, and perhaps as many as six, outliers. Specifically, Spain, Portugal, and South Africa (represented with black dots in the lower left of Figure 4) have large negative values for both axes. These countries would tend to pull the line fitted to these points toward zero. Three less obvious outliers (see the lower right portion of Figure 4) are Taiwan, Iraq, and Saudi Arabia. The position of these cases would tend to magnify the negative slope of the line fitted to

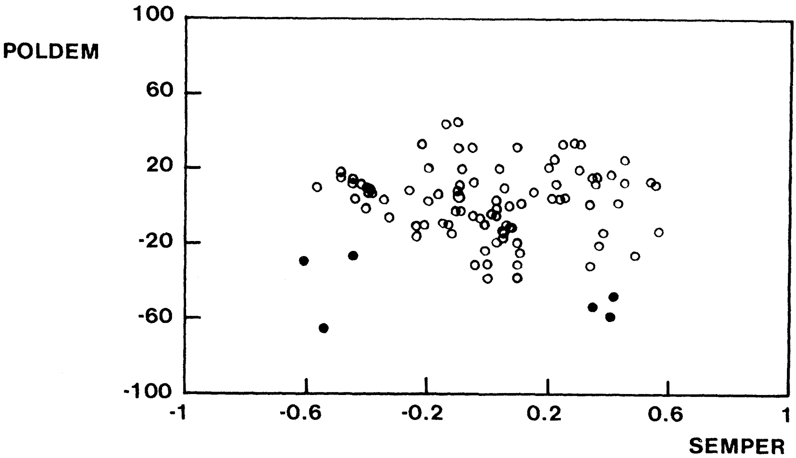

Figure 4: Partial Regression Plot for SEMPER (N = 100)

the data. A similar pattern occurs in the partial plot for PER (not shown here).

In view of both the partial regression plots and the other regression diagnostics (see Bollen, 1983), these six countries appear to be outliers that may influence the regression results.[15] In fact, when these six cases are removed, the regression estimates are:

$$\underset{(1.9)}{\text{POLDEM} = 23.48} + \underset{(6.1)}{8.97 ln\text{ENCAP}} - \underset{(0.6)}{3.90\text{SEMPER}} - \underset{(2.5)}{17.84\text{PER}} \qquad [19]$$

Equation 19 has an $\bar{R}^2$ of .610 and an F-ratio of 49.4. Note that with the 6 outliers removed, the coefficient for PER has nearly tripled in size over that reported in equation 18—from –6.76 to –17.84—indicating a substantial effect of peripheral position compared to core position. The coefficient for SEMPER is larger than in 18, but still not significant, while the $\bar{R}^2$ for equation 19 is considerably larger than that for 18: .610 versus .453. Thus the estimates in 19 reinforce the impression from the partial plots that these 6 countries are influencing our assessment of the effects of world system position.

Knowing that our estimates are affected by outliers raises the next question: What should we do with these observations? If we simply leave them in the sample and do nothing, we are left with a distorted picture of the relationship that characterizes the bulk of the countries. If, on the other hand, we remove them from our sample, we are ignoring the fact that half a dozen of our cases are not fitted by the model.

However, this "drop or keep" strategy does not exhaust the potential remedies and, indeed, oversimplifies the alternatives. It is much more instructive to ask why these countries appear as deviant cases. In the present context, consideration of a number of possible explanations of the outlier behavior suggested classification error as the most probable source of the problem. Specifically, Spain, Portugal, and South Africa were classified by Snyder and Kick (1979) as core countries, but semiperipheral status seems far more reasonable for the period of these data (circa 1965).[16] Similarly, although Taiwan, Iraq, and Saudi Arabia were placed in the semiperipheral category by Snyder and Kick, peripheral status seems more appropriate.

Details on the arguments to support this reclassification of these six countries are reported in Bollen (1983), and we do not repeat them here. But we do want to show the results obtained when the same model is estimated with these six countries reclassified on the world system measure. If the reclassification arguments are correct, then these six countries should be less likely to appear as outliers. Equation 20 reports the regression estimates after reclassification:

$$\underset{}{\text{POLDEM}} = \underset{(2.6)}{37.69} + \underset{(4.4)}{7.23}\,ln\text{ENCAP} - \underset{(1.8)}{13.35}\text{SEMPER} - \underset{(3.3)}{25.98}\text{PER} \quad [20]$$

The R^2 is .507 and the F-ratio is 34.9.

Examining the new partial regression plots associated with equation 20, Spain is the only country of the six that still exhibits a distinctive pattern. However, the removal of Spain does not substantially change the estimates in equation 20. Thus the evidence is consistent with the idea that the six countries were outliers in the original data because of measurement error. Once this error is corrected, negative relations between PER and POLDEM, and to a lesser extent, SEMPER and POLDEM are found, leading to quite different conclusions than would be reached from the "naive" original regression estimates of equation 18.

Leaving to one side the substantive implications of this example, Case 2 illustrates two additional features of the regression diagnostics. First and most important, it indicates that the diagnostics can be a useful tool in the analysis of measurement error, because such error may make observations apparently influential. Second, it is evident from this example that a very small minority of observations can be influential even when sample size increases. That one or two cases can be influential with a basic n of 18 may be little cause for surprise, but our second example shows that similar problems can occur with an n of 100.

CONCLUSIONS

We have focused on a number of diagnostic procedures that provide important information on the robustness of regression estimates, especially those generated from small samples. One could, of course, ignore these procedures and employ a naive form of analysis that involves estimation with no checks for robustness. However, as our examples make clear, such a mode of analysis can lead us to false acceptance or rejection of hypotheses. Consider the relation between economic development and inequality in Case 1 or that between world system position and political democracy in Case 2.

Although the analysis of bivariate scatterplots and regression residuals offers some improvement over the naive approach, such analyses have major shortcomings. For example, the bivariate scatterplot between semiperiphery status (a binary variable) and political democracy does not identify the problem suggested by the partial plot in Figure 4. Similarly, residuals by themselves do not necessarily help identify influential cases, as is evident, say, from the status of South Africa in Figure 1a (an outlier despite its relatively small residual).

Compared to these simple procedures, the diagnostics we have discussed provide much more systematic procedures for evaluating the robustness of regression estimates. The question that remains is this: What corrective action should be taken when one or more cases have been identified as influential?

Under certain circumstances, we may be justified in dropping the case(s) involved. However, this involves radical surgery, and can be recommended as a last resort only when there are good substantive reasons for it.[17] Instead of routinely following such a procedure, it is more fruitful to determine why the observation is an outlier. As our examples indicate, there are many different ways that this outcome can occur. From Case 1 we saw that the diagnostics can be helpful in identifying problems of sample composition and of misspecification of functional form. Case 2 illustrated a way in which the diagnostics helped detect measurement error problems. And there are other problems, such as heteroskedasticity, that these diagnostics can help identify. Remember, too, that apparent outliers may be the most important cases of some analyses in the sense that they help identify omitted variables and other forms of specification error. Once detected, corrective action can be taken for problems like these, and such action will typically fall far short of the extreme procedure of case removal.

Other diagnostic procedures in addition to those we have discussed have been proposed (e.g., Cook and Weisberg, 1982).[18] To date, no consensus has emerged on the single best way to detect outliers and influential cases. Although we have found the five techniques discussed here very helpful in our own research, we encourage others not to rule out alternative procedures. We need far more experience with these detection procedures before we can decide which are the most useful.

Finally, as we have stressed throughout, these procedures are an important aid to careful empirical analysis, not a substitute for it. Like any other techniques, these can be misapplied to produce a degenerate form of number-crunching. Without some substantive knowledge of the problem at hand, any statistical analysis becomes vacuous, concealing more than it reveals.

APPENDIX A
Raw Data for the Inequality Example[a]

Country	INEQ.	TURNOUT	ENCAP
Argentina	2.960	61.8	1088
Australia	1.940	85.3	3918
Denmark	2.734	86.8	2829
Finland	4.441	82.1	1650
France	5.653	66.5	2419
West Germany	3.435	77.6	3673
Israel	1.950	84.1	1243
Italy	2.916	89.2	1135
Japan	3.007	72.3	1166
Netherlands	3.457	87.9	2691
Norway	2.440	81.9	2740
Puerto Rico	3.693	73.3	1453
South Africa	9.410	14.3	2338
Sweden	3.143	78.1	3491
Trinidad & Tobago	3.888	64.7	1935
United Kingdom	2.876	72.4	4907
United States	2.296	56.8	8047
Venezuela	3.515	78.8	2623

a. These data and their sources are described in Jackman (1980: 344-345).

NOTES

1. Descriptions of software to compute various diagnostics can be found in Velleman and Welsch (1981: 239-241) and Cook and Weisberg (1982: 355-356). The most recent versions of SAS, SPSS, and MINITAB have procedures to provide all or some of the regression diagnostics described in this article. Our figures were computed using SAS.

2. "Partial residual plots" or "component-plus-residual plots" are an alternative discussed by Larsen and McCleary (1971) and Daniel and Wood (1980).

3. The proof that $0 \leq h_{ii} \leq 1$ follows from H being an idempotent matrix (i.e., $[X(X'X)^{-1}X'][X(X'X)^{-1}X'] = [X(X'X)^{-1}X']$. Because of this,

$$\begin{aligned} h_{ii} &= \Sigma\, h_{ij}^2 \\ &= h_{ii}^2 + \sum_{i \neq j} h_{ij}^2 \end{aligned}$$

This may be rewritten as

$$h_{ii}^2 - h_{ii} + \sum_{i \neq j} h_{ij}^2 = 0$$

The summed term in this expression must be a nonnegative number because all of its components are squared. Thus $h_{ii}^2 - h_{ii}$ must be ≤ 0. For this latter condition to hold, h_{ii} must be less than or equal to one, and greater than or equal to zero (see Hoaglin and Welsch, 1978: 18).

4. Belsley et al. (1980) show that when the explanatory variables follow an independent multinormal distribution, then $(n - p)[h_i - (1/n)]/(1 - h_i)(p - 1)$ is distributed as F, with $p - 1$ and $n - p$ degrees of freedom. For $p > 10$ and $n - p > 50$, the 95% value for F is less than 2. Thus twice the average h_i value ($2p/n$) is a useful cutoff. As Belsley et al. (1980: 17) state, multinormality and independence are often not found in practice, so this rule must be regarded as an approximation.

5. Weisberg (1980: 106) suggests that as a first approximation, these transformed residuals may be treated as if they are distributed as standard normal variables.

6. Belsley et al. (1980) refer to this measure as DFFITS. We drop the extra F as suggested by Velleman and Welsch (1981: 236).

7. A popular alternative to DFITS_i is Cook's D (Cook, 1977). Although these measures are very similar, Cook's D uses s^2 as an estimate of the disturbance variance whereas DFITS_i uses $s^2(i)$. For further comparisons, see Velleman and Welsch (1981: 236-237) and Cook and Weisberg (1982).

8. Of course, this strategy means that significance levels are not fully accurate, which makes replication even more important.

9. Inequality is defined as the ratio of the income received by the wealthiest population quintile to that received by the poorest two quintiles. The raw data, which are for around 1960, are listed by country in the Appendix.

10. Note that the extreme scores for these two cases could have been detected from simple histograms for ENCAP and TURNOUT. However, such histograms would not have helped identify the two cases as potentially influential. In addition, observations can be influential without having extreme values on one or more variables.

11. Remember that this is an approximate cutoff. Because we did not specify in advance the cases to be discussed as possible outliers, the probability level should be somewhat larger than .05, and we have, if anything, identified too many residuals as outliers (on this point, see Weisberg, 1980: 116-117). With only two cases exceeding our cutoff, we are not concerned about examining too many cases using this criterion.

12. South Africa, in particular, is unique in the way it formally restricts the political participation of an overwhelming majority of its adult population.

13. In this connection, care is needed in the visual comparison of plots such as these. Most obviously, the plotting routines contained in statistical computing packages generally adjust the range of both vertical and horizontal axes to reflect the observed range of the variables being plotted. In our case, the exclusion of South Africa drastically influences the variance of both INEQ and TURNOUT (the standard deviations of both variables are almost halved by the exclusion of this one case). As a result, when comparing Figures 1a and 2a, it is important to note the changes in both axes: Where the vertical axis in Figure 1a ranges from -6 to $+6$, that in Figure 2a ranges from -3 to $+3$; the range of the horizontal axis is reduced by the same factor when we go from Figure 1a to Figure 2a. The exclusion of South Africa coupled with the transformation of ENCAP produces parallel differences between Figures 1b and 2b. These differences need to be borne in mind because it is important that one's visual interpretation and comparison of plots not be influenced unduly by simple changes in the metrics of either or both axes.

14. For this equation, the core is the omitted category of world-system dummy variables so that the coefficients for SEMPER and PER should be interpreted as deviations from the core countries. The N of 100 includes all noncommunist countries with valid data on all variables. To conserve space, we do not display the data here, but they are readily available from the following sources: Bollen (1980) for POLDEM; Taylor and Hudson (1972) for ENCAP; and Snyder and Kick (1979) for SEMPER and PER. The logarithmic transformation of ENCAP (*ln*ENCAP) is used following Jackman (1973).

15. At this point, a comprehensive examination of the DFITS_i, RSTUDENTS, and the other diagnostics for each of the 100 countries could be displayed. While these other diagnostics inform the following analysis, we do not display them by country to save space. Instead, we concentrate on the six cases that are found to be influential. On a more general level, note that as the n increases, so does the number of combinations of potentially influential observations. This again underscores the point that the regression diagnostics cannot be applied mechanically to a set of estimates. Indeed, such routine applications can readily produce numbers that are difficult, if not impossible, to interpret. Instead, the diagnostics can only be used in conjunction with substantive information about the problem at hand.

16. On this point, compare the Snyder and Kick (1979) classification with that in Wallerstein (1976).

17. It is unfortunate that other treatments of regression diagnostics often fail to consider any remedy other than case deletion. See, for example, two articles that appeared after the present article was completed (Chatterjee and Wiseman, 1983; Stevens, 1984).

18. As an alternative to these diagnostics, one can use robust regression procedures that are less sensitive to outliers and influential cases (see Weisberg, 1980: 237-238). However, as we have already indicated, outliers and influential cases can provide useful information about model specification problems, so that it is not always optimal to downplay them.

REFERENCES

AHLUWALIA, M. S. (1976) "Inequality, poverty and development." J. of Development Economics 3: 307-412.

BELSLEY, D. A., E. KUH, and R. E. WELSCH (1980) Regression Diagnostics: Identifying Influential Data and Sources of Colinearity. New York: John Wiley.

BOLLEN, K. A. (1983) "World system position, dependency, and democracy." Amer. Soc. Rev. 48 (August): 468-479.

———(1980) "Issues in the comparative measurement of political democracy." Amer. Soc. Rev. 45 (June): 370-390.

CHATTERJEE, S. and F. WISEMAN (1983) "Use of regression diagnostics in political science research." Amer. J. of Pol. Sci. 27 (August): 601-613.

CHIROT, D. (1977) Social Change in the Twentieth Century. New York: Harcourt Brace Jovanovich.

COOK, R. D. (1977) "Detection of influential observations in linear regression." Technometrics, 19: 15-18.

———and S. WEISBERG (1982) "Criticism and influence analysis in regression," pp. 313-361 in S. Leinhardt (ed.) Sociological Methodology 1982. San Francisco: Jossey-Bass.

DANIEL, C. and F. S. WOOD (1980) Fitting Equations to Data: Computer Analysis of Multifactor Data. New York: John Wiley.

HEWITT, C. (1977) "The effect of political democracy and social democracy on equality in industrial societies: a cross-national comparison." Amer. Soc. Rev. 42 (June): 450-464.

HOAGLIN, D.C. and R.E. WELSCH (1978) "The hat matrix in regression and ANOVA." Amer. Statistician 32 (February): 17-22.

JACKMAN, R. W. (1980) "The impact of outliers on income inequality." Amer. Soc. Rev. 45 (April): 344-347.

———(1973) "On the relation of economic development to democratic performance." Amer. J. of Pol. Sci. 17 (August): 611-621.

KUZNETS, S. (1955) "Economic growth and income inequality." Amer. Econ. Rev. 45 (March): 1-28.

LARSEN, W. A. and S. J. McCLEARY (1971) "The use of partial residual plots in regression analysis." Technometrics 14: 781-790.

LENSKI, G. (1966) Power and Privilege. New York: McGraw-Hill.

SNYDER, D. and E. L. KICK (1979) "Structural position in the world system and economic growth, 1955-1970." Amer. J. of Sociology 84 (March): 1096-1126.

STACK, S. (1979) "The effects of political participation and socialist party strength on the degree of income inequality." Amer. Soc. Rev. 44 (February): 168-171.

STEVENS, J. P. (1984) "Outliers and influential data points in regression analysis." Psych. Bull. 95 (March): 334-344.

TAYLOR, C. L. and M. C. HUDSON (1972) World Handbook of Political and Social Indicators. New Haven, CT: Yale Univ. Press.

VELLEMAN, P. F. and R. E. WELSCH (1981) "Efficient computing of regression diagnostics." Amer. Statistician 35 (November): 234-242.

WALLERSTEIN, I. (1976) "Semi-peripheral countries and the contemporary world crisis." Theory and Society 3: 461-484.

———(1974) The Modern World System: Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century. New York: Academic.

WEISBERG, S. (1980) Applied Linear Regression. New York: John Wiley.

Kenneth A. Bollen is an Associate Professor of Sociology at University of North Carolina at Chapel Hill. His major research interests are in international development and statistical methodology.

Robert W. Jackman is a Professor of Political Science at Michigan State University. His research interests are in political sociology and comparative politics, and he is author of Politics and Social Equality *(1975) and coauthor of* Class Awareness in the United States *(1983).*