The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Cautionary Note about R^2

Tarald O. Kvålseth

Cautionary Note About R^2

TARALD O. KVÅLSETH*

The coefficient of determination (R^2) is perhaps the single most extensively used measure of goodness of fit for regression models. It is also widely misused. The primary source of the problem is that except for linear models with an intercept term, the several alternative R^2 statistics are not generally equivalent. This article discusses various considerations and potential pitfalls in using the R^2's. Specific points are exemplified by means of empirical data. A new resistant statistic is also introduced.

KEY WORDS: Coefficient of determination; Regression analysis; Resistant or robust methods.

INTRODUCTION

Whenever a data analyst is fitting models to some data, he or she is likely to use the coefficient of determination R^2 to assess the goodness of fit of the models. If the models are anything but linear with an intercept term, it is not unlikely that the analyst will use an inappropriate R^2 statistic and end up with possibly misleading results. Such mistakes may often appear to be attributable to careless use of regression program packages. The problem is not made any simpler by the fact that some regression packages compute R^2 inappropriately; I have encountered some of these and identify them in this article (e.g., Hewlett-Packard 1976; Texas Instruments 1977).

The underlying problem appears to be essentially twofold. First, the data analyst is faced with a possibly confusing variety of R^2 statistics. For the case of linear least squares regression models with an intercept term, this problem is not a serious one, since the majority of the R^2 statistics are equivalent. For other types of models, however, such as linear no-intercept models or nonlinear (in the parameters) models, the various R^2 statistics generally yield different values. Second, statisticians seem to be unaware of the extent of this problem, and most texts on regression analysis fail to address the problem at all. A few cautionary comments have been made in the statistical literature (e.g., Hahn 1973, 1977; Marquardt and Snee 1974; Montgomery and Peck 1982), but these appear to be confined to linear no-intercept models or their special case of so-called "mixture models" and cover only part of the problem.

It is the purpose of this article to (a) address the R^2 problem generally, (b) compare the various statistics for different types of models, (c) point out potential pitfalls and some of the more common mistakes, and (d) provide a recommendation for the most appropriate and generally applicable R^2 statistic. The various points raised are exemplified by means of empirical data. Finally, a new resistant statistic is introduced.

ALTERNATIVE STATISTICS

Consider that the dependent variable y is some function of the independent (predictor, exogenous) variables $x_1, x_2, \ldots, x_k$ with the set of parameters $\{\beta_j\}$ and a residual variable ε. Let $\hat{y}_i$ denote the fitted (calculated) value of y based on the set of estimated parameters $\{b_j\}$ and the values x_{ji} of the x_j (j = 1, 2, . . ., k) for $i = 1, 2, \ldots, n$. Furthermore, let $\bar{y}$ and $\bar{\hat{y}}$ denote the sample (arithmetic) mean of y_i and $\hat{y}_i$, respectively. Then the following expressions for R^2 appear throughout the literature:

$$R_1^2 = 1 - \sum(y - \hat{y})^2/\sum(y - \bar{y})^2 \tag{1}$$

$$R_2^2 = \sum(\hat{y} - \bar{y})^2/\sum(y - \bar{y})^2 \tag{2}$$

$$R_3^2 = \sum(\hat{y} - \bar{\hat{y}})^2/\sum(y - \bar{y})^2 \tag{3}$$

$$R_4^2 = 1 - \sum(e - \bar{e})^2/\sum(y - \bar{y})^2, \quad e = y - \hat{y}, \tag{4}$$

R_5^2 = squared multiple correlation coefficient between the regressand and the regressors (5)

R_6^2 = squared correlation coefficient between y and $\hat{y}$ (6)

$$R_7^2 = 1 - \sum(y - \hat{y})^2/\sum y^2 \tag{7}$$

$$R_8^2 = \sum\hat{y}^2/\sum y^2 \tag{8}$$

where all summations are from $i = 1$ to $i = n$ and the subscript i has been omitted to simplify the notation (as will be done throughout this article when there is no danger of ambiguity). The multiple correlation coefficient in (5) is based on the variables expressed in terms of the general linear regression model, which may involve certain transformations of some or all of the original variables $y, x_1, \ldots, x_k$. Thus R_5^2 is really a catchall term for the R_i^2 (i = 1, 2, 3, 4, 6) computed for the linearized model fitted by ordinary least squares regression, with the y and $\hat{y}$ replaced by the appropriately transformed variables y' and $\hat{y}'$ required for linearization. For $k = 1$, R_5 becomes the Pearson product–moment correlation coefficient (r) for the regressand and the regressor in the bivariate linear regression model. For the simple power model, for example, R_5 is the Pearson r for $y' = \log y$ and $x' = \log x$. The R_6 from (6) is the Pearson r for y and $\hat{y}$.

Some sample references for these statistics are as follows: for R_1^2, Goldberger (1964, pp. 160, 166), Montgomery and Peck (1982, pp. 33–34, 146), and Theil (1971, pp. 175–178); for R_2^2, Draper and Smith (1981, pp. 33, 90–91) and Lewis-Beck (1980, pp. 20–25, 52–53); for R_3^2, Goldberger (1964, pp. 160, 166); for R_4^2, Theil (1971, pp. 164, 175–178); for R_5^2, Draper and Smith (1981, p. 90), Montgomery and Peck (1982, p. 146), and Theil (1971, p. 164) [these authors use R_5^2 for the linear model in (9) and not for linearized forms of nonlinear models]; for R_6^2, Draper and

*Tarald O. Kvålseth is Professor, Department of Mechanical Engineering, and Director of the Industrial Engineering Division, University of Minnesota, Minneapolis, MN 55455.

Smith (1981, p. 91); and for R_7^2 and R_8^2, Marquardt and Snee (1974), Montgomery and Peck (1982, pp. 38–43), and Theil (1971, pp. 163–192). The most frequently used forms of R^2 appear to be R_1^2, R_2^2, and R_5^2, but none of the R_i^2 seems to have been recommended for general usage. Which R^2 statistic to choose in any given situation may not be a trivial decision, and little guidance is offered to the data analyst by the statistical literature. The appropriate choice of R^2 must be based on a number of different considerations: (a) the types of models being formulated, (b) the model-fitting technique being used, (c) the purpose for which R^2 is used, and (d) the properties of R^2 considered to be desirable.

The choice of an R^2 statistic is relatively simple for the case in which (a) the model is a linear one of the general form

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \varepsilon \tag{9}$$

and (b) the ordinary least squares regression method is used for estimating the parameters β_j $(j = 0, 1, \ldots, k)$, which is the model-fitting method assumed throughout this article unless otherwise indicated. For this case, it can easily be shown that $R_1^2 = R_2^2 = \cdots = R_6^2$ (e.g., see Draper and Smith 1981, chaps. 1–2; Goldberger 1964, chap. 4). The common value of these six statistics, however, is generally different from the values of R_7^2 and R_8^2. The last two statistics have only been recommended for the case of a linear no-intercept model, that is, for $\beta_0 = 0$ in (9) (e.g., Marquardt and Snee 1974; Montgomery and Peck 1982, p. 38–43; Theil 1971, pp. 163–192). For this case $R_7^2 = R_8^2$, since $\Sigma(y - \hat{y})^2 = \Sigma y^2 - \Sigma \hat{y}^2$. The other R^2 statistics, however, are all generally different, as exemplified later in this article. Note that whereas $\bar{\hat{Y}} = \bar{y}$ and $\bar{e} = 0$ for the model in (9), these two equalities do not generally hold when β_0 is not used in the model. For the no-intercept case, and as will be illustrated empirically, R_2^2 and R_3^2 may exceed 1.

In addition, R_1^2 and R_4^2 may possibly be negative in a couple of situations to be discussed, that is, when a clearly inappropriate least squares regression model is used or when a model-fitting method such as robust regression is used for data containing outliers (influential observations). Both R_1^2 and R_4^2, however, will generally be nonnegative and less than or equal to 1. For the linear no-intercept case and for nonlinear models, it is recommended that R_4^2 be rejected, since a nonzero mean residual (error) $\bar{e}$ should be regarded as contributing to a reduction in the model fit rather than to an increase as implied by (4).

For nonlinear (in the parameters) models, arguments similar to those given for the no-intercept linear models apply. Thus the different R_i^2 statistics generally assume different values. The R_2^2 and R_3^2 may easily exceed 1, as will be exemplified for a simple power (geometric) model. Although R_1^2 for nonlinear models may conceivably be negative, as in the linear case, $0 \leq R_1^2 \leq 1$ in general. The R_5^2 measures the strength of the association between the regressand and the regressors when the nonlinear model is transformed into a linear regression model of the form in (9). Although the R_5^2 statistic provides an indirect indication of the fit of the original nonlinear model, however, it does not necessarily provide an accurate description of the strength or fit of the nonlinear model. Similarly, R_6^2 is based on a linear correlation measure, that is, correlation between the actual (observed) and the fitted (predicted) values of the original dependent variable (y). Such a correlation interpretation of R_6^2 would not seem to be particularly attractive or useful on intuitive grounds as a goodness-of-fit measure for nonlinear models. Furthermore, the use of R_6^2 for nonlinear models may produce potentially misleading results, since it is clearly possible for y and $\hat{y}$ to be highly correlated even if their corresponding values deviate substantially. Finally, although $0 \leq R_7^2 \leq 1$, R_8^2 may clearly be larger than 1 for nonlinear models, an example of which is given later.

SOME COMMON MISTAKES

With so many potential candidates for the R^2 statistic whose values may differ substantially for no-intercept linear models and nonlinear models, it is perhaps no wonder that many data analysts make inappropriate choices and interpretations. When standard computer routines are used for model fitting, the user may have little or no choice of the R^2 statistic; but he or she still needs to exercise caution, since some of these computer programs have been found to use inappropriate R^2 expressions (such as the curve-fitting programs for the Hewlett-Packard and Texas Instruments programmable calculators discussed subsequently). Moreover, irrespective of the way in which R^2 is computed, it is an all too common practice to simply report the R^2 value without specifying which statistic has been used. Reported studies comparing the fits of alternative models in terms of R^2 values may clearly be suspect when failing to report the R^2 expression(s) being used.

One of the most frequent mistakes occurs when comparing the fits of a linear and a nonlinear model by using the same R^2 expression but different variables: the original variable y and the fitted $\hat{y}$ for the linear model and transformed variables for the nonlinear model. Thus for example, a power model or an exponential model may first be linearized by using logarithmic transformations and then fitted to empirical data by using ordinary least squares regression. The R^2 value is then often calculated using the data points ($\log y_i$, $\log \hat{y}_i$) and (a) interpreted as a measure of the goodness of fit of the nonlinear model and (b) compared with the fit of the linear model determined by the same R^2 expression but using the data points (y_i, $\hat{y}_i$). The R^2 value based on the transformed data points, however, provides a measure of fit for the linearized model and not for the nonlinear model. To make a sensible comparison between the fits of a linear and a nonlinear model to the same set of data, comparable data points (y_i, $\hat{y}_i$) and R^2 expressions have to be used; otherwise misleading results may be obtained.

Such R^2 computations based on incompatible data sets, when fitting linear and nonlinear models, are also incorporated in some of the more widely used programs for bivariate relationships marketed as standard regression programs by manufacturers of programmable calculators (Hewlett-Packard 1976; Texas Instruments 1977). The program for HP67/97 computes R^2 as the square of the Pearson correlation coefficient between the regressand and the regressor, that is, between y and x for the linear model, $\log y$ and log x for the power model, and log y and x for the exponential model. That is, R^2 is computed according to the R_5^2 statistic with $k = 1$ and irrespective of whether the original model is linear or nonlinear. However, such an R^2 value for a nonlinear model, which measures the goodness of fit of the linearized model and not the original nonlinear model, may not strictly be compared with the R^2 value for the linear model, since the two R^2 values are based on different data sets even though the same set of (x_i, y_i) values are used in both cases. A numerical example is presented later illustrating that when R^2 is computed in terms of R_5^2 for both a power model and a linear model, the power model will be selected as the best one although additional analysis of the residuals clearly favors the linear model.

Similarly, the TI58/59 program computes the unsquared R in the same inappropriate and potentially misleading way. The same argument applies to other computer program packages for regression analysis, such as SPSS (Statistical Package for the Social Sciences) and MULTREG (multiple regression package at the University of Minnesota), which routinely computes R_5^2 for both linear and nonlinear (intrinsically linear) models. Although these larger computer program packages are generally of sufficient flexibility to permit the user some choice of the R^2 statistic by using additional lines of code, no such flexibility is offered by the HP67/97 and TI58/59 programs; these require substantial program changes or hand calculations if the user is to calculate R^2 statistics other than R_5^2 for nonlinear models.

For no-intercept linear models, a couple of different R^2 mistakes appear frequently. First, such a model fit is sometimes reported in terms of the value of R_5^2, which is more appropriately a measure of fit for the intercept model rather than the no-intercept model. Second, an analyst may also occasionally (a) determine the fit of a no-intercept linear model by means of R_7^2 or R_8^2 and (b) compare such a fit with that of a linear intercept model using one of the other R^2 statistics, such as R_1^2. Any such comparison between the fits of alternative models, however, is only valid if comparable measures of fit are used; that is, the same R^2 statistic must be used for the models being compared.

Finally, the use of no-intercept models may also be subject to other types of pitfalls, so caution is necessary when using such models and interpreting the results (e.g., Hahn 1977; Montgomery and Peck 1982, pp. 38–43). Occasionally, an analyst may force such a model to be fitted to empirical data by relying purely on theoretical reasoning, whereas a careful analysis of the data may reveal that an intercept model is preferable. As a precautionary rule, no-intercept models should only be used when both theoretical justification and empirical data analysis indicate that they are indeed appropriate.

RECOMMENDED STATISTIC

To make a recommendation on the correct choice of R^2 statistic, it may be appropriate to impose some a priori conditions or requirements concerning the properties of a "good" statistic. These may be outlined as follows:

1. R^2 must possess utility as a measure of goodness of fit and have an intuitively reasonable interpretation.
2. R^2 ought to be independent of the units of measurement of the model variables; that is, R^2 ought to be dimensionless.
3. The potential range of values of R^2 should be well defined with endpoints corresponding to perfect fit and complete lack of fit, such as $0 \leq R^2 \leq 1$, where $R^2 = 1$ corresponds to perfect fit and $R^2 \geq 0$ for any reasonable model specification.
4. R^2 should be sufficiently general to be applicable (a) to any type of model, (b) whether the x_j are random or nonrandom (mathematical) variables, and (c) regardless of the statistical properties of the model variables (including residual ε).
5. R^2 should not be confined to any specific model-fitting technique; that is, R^2 should only reflect the goodness of fit of the model per se irrespective of the way in which the model has been derived.
6. R^2 should be such that its values for different models fitted to the same data set are directly comparable.
7. Relative values of R^2 ought to be generally compatible with those derived from other acceptable measures of fit (e.g., standard error of prediction and root mean squared residual).
8. Positive and negative residuals $(y_i - \hat{y}_i)$ should be weighted equally by R^2.

None of the R_i^2 $(i = 1, 2, \ldots, 8)$ statistics possesses all of these properties as indicated, although R_1^2 nearly does except for the lower bound requirement of property 3. Although R_1^2 is nonnegative for any linear least squares regression model with intercept, R_1^2 may clearly be negative when (a) gross model misspecification is used (e.g., when fitting a linear no-intercept model by means of least squares regression to data indicating that β_0 is substantially different from 0) or (b) a model-fitting technique other than linear least squares regression, such as a robust regression method, is used for grossly contaminated data [i.e., data with gross outlier(s)]. The first of these two cases is of no practical consequence, with $R_1^2 \leq 0$ simply indicating a complete lack of fit, whereas the second case may obviously present a problem, as will be discussed in a later section.

The intuitive derivation and interpretation of R_1^2 may be given as follows. If a prediction of the dependent variable y is to be made without using any knowledge about the independent variables x_j or about their relationship to y, the best prediction would be the sample mean $\bar{y}$ ["best" in the sense that $\Sigma_{i=1}^n (y_i - v)^2$ is minimized for $v = \bar{y}$], with the resulting deviation (error or variation) of $y_i - \bar{y}$ between the true y value, y_i, and the prediction. When such a prediction $\hat{y}_i$, however, is based on knowledge about the x_j and a model of the relationship between y and the x_j, the new deviation $y_i - \hat{y}_i$ is expected to be less than the initial deviation $y_i - \bar{y}$. For $i = 1, 2, \ldots, n$, measures of total deviation or variation giving equal emphasis to positive and negative individual deviations may be considered as $\Sigma_{i=1}^n (y_i - \bar{y})^2$ and $\Sigma_{i=1}^n (y_i - \hat{y}_i)^2$, where $\Sigma(y_i - \hat{y}_i)^2$ is the variation of y that is unexplained by the fitted model. Thus R_1^2 in (1) may be interpreted as the proportion of the total variation of y (about its mean $\bar{y}$) that is explained (accounted for) by the fitted model.

The statistic R_1 may also be used as a generalized correlation coefficient or index suitable for both linear and nonlinear models. This proposal would seem to be preferable to that made by some (e.g., Spiegel 1975) of using R_2 as a generalized (linear and nonlinear) correlation coefficient because of the favorable properties possessed by R_1^2 and hence by R_1 and the failure of R_2 to satisfy property 3 (i.e., R_2 may exceed unity for linear no-intercept models and nonlinear models, as exemplified in Table 1).

EXAMPLES

Bivariate Case

To exemplify the differences between the various R^2 statistics and some of the problems outlined earlier, three sets of data will be used. Consider first the case of one independent variable and the following two data sets: $(x, y) = (1, 15), (2, 37), (3, 52), (4, 59), (5, 83), (6, 92)$ and $(x, y') = (6, 3{,}882), (7, 1{,}266), (8, 733), (9, 450), (10, 410), (11, 305), (12, 185), (13, 112)$. The first data set is a fictitious one, whereas the second set is real data for total fatal accidents in the U.S. during 1978 for the age group 55–64, with $x_i = i + 5$ ($i = 1$ for the most frequent type of accident, $i = 2$ for the second most frequent accident type or category, etc.) and $y_i = y_i'/7{,}343$ is the frequency of the ith most frequent accident type (Kvålseth 1983).

Table 1 gives the values of the parameters and the various statistics for each of three different models—linear models with and without intercept and a power model—fitted by means of ordinary least squares regression and with the power model being linearized by taking logarithms. For the accident data, only a power model appeared to be a reasonable model specification. The last three statistics in Table 1 refer to the root mean squared residual or error (RMSE), mean absolute residual (MAE), and the mean squared residual (MSE) defined as

$$\text{RMSE} = \left[\sum (y - \hat{y})^2/n\right]^{1/2},$$
$$\text{MAE} = \sum |y - \hat{y}|/n,$$
$$\text{MSE} = \sum (y - \hat{y})^2/(n - p), \qquad (10)$$

where p denotes the number of model parameters.

According to property 3 requiring $0 \leq R^2 \leq 1$, it can be seen from Table 1 that R_2^2 and R_3^2 are inappropriate for both the no-intercept linear model and the power model; and that R_8^2 is inappropriate for the power model. If the quality of fit is measured in terms of R_5^2 or R_6^2 for all three models, as is sometimes done, the power model would be considered slightly superior to the other two. The statistics R_1^2, RMSE, MAE, and MSE, however, clearly favor the linear intercept model. If the choice between the intercept and no-intercept models is based on the two sets of statistics $\{R_1^2, \text{MSE}\}$ and $\{R_7^2\ (=R_8^2), \text{MSE}\}$, respectively, the no-intercept model would be selected; the fact that these criteria may sometimes lead to contradictory results is exemplified further in the subsequent analysis (Table 2). The relatively low MSE value for the no-intercept model is, of course, a consequence of the fact that this model has one less parameter than the other two models. Finally, it is observed from Table 1 that the various R_i^2 do not vary substantially in value for the first data set. For the power model fitted to the second data set, however, the R_i^2 values show considerable differences. The not infrequently but incorrectly used R_5^2 (based on logarithmic transformation) would clearly overestimate the fit of this model by as much as about 7% when compared with the more appropriate R_1^2.

Table 1. Parameter Estimates and Values of the Alternative Statistics for Different Models and Two Sets of Data

Parameters/Statistics	Data set 1: $\hat{y} = b_0 + b_1x$	Data set 1: $\hat{y} = b_1x$	Data set 1: $\hat{y} = b_0x^{b_1}$	Data set 2: $\hat{y} = b_0x^{b_1}$
b_0	3.3333	—	16.3757	594.6202
b_1	15.1429	15.9121	.9900	−4.0826
R_1^2	.9808	.9777	.9777	.9019
R_2^2	.9808	1.0836	1.0984	.5856
R_3^2	.9808	1.0830	1.0983	.5824
R_4^2	.9808	.9783	.9778	.9051
R_5^2	.9808	.9808	.9816	.9669
R_6^2	.9808	.9808	.9811	.9497
R_7^2	.9966	.9961	.9961	.9392
R_8^2	.9966	.9961	1.0232	.6879
RMSE	3.6166	3.9008	3.8982	.0500
MAE	3.5238	3.6520	3.6334	.0283
MSE	19.6196	18.2594	22.7939	.0033

NOTE: Alternative statistics are defined in the text. The data in set 1 are fictitious; in set 2, real.

Multivariate Case

As another example, consider the data for two independent variables given by Box et al. (1978, p. 462). The ordinary least squares parameter estimates and the values of the various statistics are given in Table 2 for each of three alternative models. All of the R_i^2 and the RMSE, MAE, and MSE tend to favor the linear model with intercept, with the power model being a close second. Perhaps the most striking observation from Table 2 is an apparent contradiction between some alternative measures of the goodness of fit for the linear intercept and no-intercept models. If as is

Table 2. Parameter Estimates and Values of the Alternative Statistics for Different Models Using Data by Box et al.

Parameters/Statistics	$\hat{y} = b_0 + b_1x_1 + b_2x_2$	$\hat{y} = b_1x_1 + b_2x_2$	$\hat{y} = b_0x_1^{b_1}x_2^{b_2}$
b_0	−1.7231	—	8.6089
b_1	1.4907	1.2073	.2178
b_2	8.9540	7.1230	.8513
R_1^2	.9657	.9247	.9653
R_2^2	.9657	.6170	.9640
R_3^2	.9657	.6153	.9640
R_4^2	.9657	.9263	.9653
R_5^2	.9657	.9657	.9500
R_6^2	.9657	.9656	.9653
R_7^2	.9977	.9950	.9977
R_8^2	.9977	.9950	.9950
RMSE	.3177	.4709	.3197
MAE	.2665	.3896	.2702
MSE	.2019	.3327	.2045

Source: Box et al. (1978, p. 462).

sometimes done, R_7^2 (or R_8^2) is considered for the no-intercept model and any one of the R_i^2 for $1 \leq i \leq 6$ is used for the intercept model, the no-intercept model would clearly be the preferred one, in spite of the additional fact that ⅔ of the individual residuals were found to have smaller absolute values for the intercept model.

DISCUSSION

Model-Fitting Technique

It has been assumed throughout this article that linear least squares regression is being used. Thus the discussion of nonlinear models is strictly confined to those that are so-called "intrinsically linear"; that is, they can be converted, by suitable transformations of the variables, into the standard linear model form of (9). For the power models fitted to the three data sets used earlier, this assumption requires that the error (residual) term is reasonably considered to be multiplicative. Of course, although this article is only concerned with sample measures of model fit and not with statistical inferences about population parameters, any such inferences based on parametric (vs. nonparametric or distribution-free) methods would necessarily require additional assumptions about the error term (zero mean, constant variance, independence, and normality for the transformed additive error).

Although R_1^2 is the recommended R^2 statistic for linear models with or without intercepts, for nonlinear models that are intrinsically linear, and when linear least squares regression is used, there does not appear to be any particular reason why R_1^2 should not also be the preferable R^2 statistic for models that are intrinsically nonlinear and fitted by nonlinear methods (e.g., see Draper and Smith 1981 for such nonlinear estimation). The same recommendation in favor of R_1^2 is made for linear and linearized models fitted by methods other than ordinary regression, although the potential limitation of lack of resistance may pose a problem for R_1^2, as will be discussed subsequently.

No-Intercept Models

It has been proposed (e.g., Hahn 1977; Marquardt and Snee 1974; Montgomery and Peck 1982) that R_7^2 (or R_8^2) should be used for linear no-intercept models and R_1^2 (or $R_2^2, \ldots, R_6^2$) for linear models with intercepts. Two such R^2 statistics, however, are not generally comparable and may, if the analyst fails to realize this, produce misleading results, as illustrated by two of the preceding examples (see especially Table 2). One obvious exception is the very special case in which the sample means of y and the x_j ($j = 1, 2, \ldots, k$) are all 0, so the two models' parameter estimates and R_i^2 values coincide exactly.

This last fact may also be used for fitting a linear no-intercept model to any set of n data points $\{(x_{1i}, x_{2i}, \ldots, x_{ki}, y_i); i = 1, 2, \ldots, n\}$ by means of some regression program package without a no-intercept option. Thus (as also suggested by Hawkins 1980) by entering the original data points and then the same points a second time but with opposite signs, the ordinary least squares estimates $b_0 = 0$ and b_j ($j = 1, 2, \ldots, k$) are the correct ones for the no-intercept model. The computed R^2 value based on these $2n$ data points, whichever R_i^2 ($i = 1, 2, \ldots, 8$) is used by the program package, is the same as the value of R_7^2 (or R_8^2) for the no-intercept model fitted to the n original data points, which would seem to be another reason for rejecting the R_7^2 and R_8^2 statistics. Furthermore, the two statistics RMSE and MAE computed for the intercept model based on the $2n$ data points are the same as those for the no-intercept model based on the n original data points. A correction factor is needed, however, for the MSE statistic. It is easily seen that the correct MSE for the no-intercept model is obtained from the MSE for the intercept model and the $2n$ data points by multiplying the latter MSE by the correction factor $c = (2n - k - 1)/2(n - k)$. Similarly, for testing the hypothesis H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ for the no-intercept model, the F_0 statistic computed for the intercept model using the $2n$ data points must be multiplied by the factor $1/2c$.

Resistance

All of the R_i^2 ($i = 1, 2, \ldots, 8$) are clearly sensitive to extreme values (outliers), since they contain terms involving squared residuals or squared values of the dependent variable or both. That is, these R_i^2 statistics have a relatively low degree of resistance, as do RMSE, MAE, and MSE. Similarly, the least squares regression method itself lacks resistance. Thus the fitted regression model and the various goodness-of-fit statistics discussed so far may all be unduly influenced by one or more outliers in the data set. Resistant or robust model-fitting techniques (e.g., Emerson and Hoaglin 1983; Huber 1981; Montgomery and Peck 1982, p. 364–383) avoid this difficulty with respect to the fitted model itself but obviously not relative to the R_i^2. Although R_1^2 is generally still considered to be the preferable one of the eight R_i^2 statistics when resistant model-fitting techniques are being used, there may clearly be cases when R_1^2, as well as the R_i^2 ($i = 2, \ldots, 8$), will lead to potentially misleading results because of lack of resistance. Thus for example, a resistant technique may produce a perfectly reasonable model with an apparently good fit to all data points except for a single gross outlier, but R_1^2 may be negative because of this single data point.

When using a resistant model-fitting method, it may thus be desirable to use an R^2 statistic that also possesses the resistance property. Such a statistic may, for example, be derived by simply replacing the preceding arithmetic means by sample medians. Specifically, by considering the median of the absolute deviations of the y_i from the model predictions $\hat{y}_i$ and from the naive prediction $\bar{y}$, a resistant R^2 statistic may be defined as

$$R_9^2 = 1 - (M\{|y_i - \hat{y}_i|\}/M\{|y_i - \bar{y}|\})^2 \qquad (11)$$

and derived from R_1^2 as follows. The numerator and denominator of the sums of squares ratio of R_1^2 may both be divided by n, and the resulting arithmetic means of the two sets of squared deviations may then be replaced by the medians $M\{(y_i - \hat{y}_i)^2\}$ and $M\{(y_i - \bar{y})^2\}$. But these two medians equal $(M\{|y_i - \hat{y}_i|\})^2$ and $(M\{|y_i - \bar{y}|\})^2$ exactly when n is odd and approximately when n is even (exactly when the median is based on two tied absolute values), so (11) results. Based on the new variability measures in (11), R_9^2 may still be given the same interpretation as that of R_1^2. That is, R_9^2 is a measure of the proportion of the total variability of y that is explained (accounted for) by the fitted model.

The new R_9^2 statistic possesses the same types of properties as those of R_1^2 given previously and has the additional resistance property. The R_9^2 also has an upper bound of unity and may possibly assume negative values when model specification is clearly inappropriate. It may also be noted that for symmetric distributions and large n, the unsquared variability ratio in (11) will be approximately equal to a ratio of the corresponding interquartile ranges. Finally, it may be observed that if the mean of the y_i in R_9^2 was replaced by their median, the unsquared denominator of R_9^2 would become the more commonly used median absolute deviation measure.

As an example of the resistant nature of R_9^2 and the lack of this property by R_1^2, consider the artificial data illustrated in Figure 1, where the resistant line has been fitted by means of Tukey's method (McNeil 1971). The R_9^2 is computed to be .98 and is not unduly influenced by the single outlier, which, however, has a dramatic effect on R_1^2, with $R_1^2 = .75$. For the least squares regression line in Figure 1, it is found that $R_1^2 = .78$ and $R_9^2 = .95$; with the outlier excluded, $R_1^2 = R_9^2 = .99$, indicating that R_9^2 is more resistant than R_1^2 even for least squares regression. In general, however, when the nonresistant method of least squares regression is used, R_9^2 is not expected to be substantially and consistently more resistant than R_1^2 (Dzubay 1984).

Degrees of Freeedom and Model Specification

None of the statistics defined so far, except for MSE, incorporate explicitly the degrees of freedom of the relevant quantities. It may sometimes be preferable to adjust R^2 for the appropriate degrees of freedom. Care must be exercised,

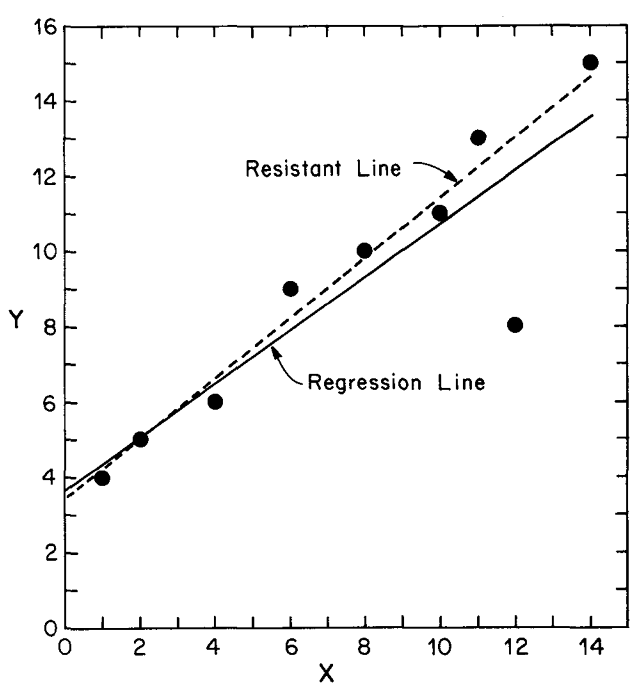

Figure 1. Example of Model Fitting by Tukey's Resistant Method and the Ordinary Least Squares Regression Method When One Outlier Is Present. Respective fitted lines: $\hat{y} = 3.40 + .80x$ and $\hat{y} = 3.66 + .71x$.

however, when using such adjusted R^2 to avoid the same types of problems as discussed earlier for the unadjusted R^2, especially when the fits of alternative types of models are being compared. The most common adjustment for linear intercept models is to divide $\Sigma(y_i = \hat{y}_i)^2$ and $\Sigma(y_i = \bar{y})^2$ of R_1^2 by their respective degrees of freedom so that the adjusted R_1^2 statistic may be defined as

$$\begin{aligned} R_{1a}^2 &= 1 - a\sum(y_i - \hat{y}_i)^2 \Big/ \sum(y_i - \bar{y})^2 \\ &= 1 - a(1 - R_1^2), \end{aligned} \tag{12}$$

with the adjustment factor $a = (n - 1)/(n - k - 1)$ for the linear intercept model.

To avoid any problems and confusion on the part of the data analyst, it seems a safe recommendation to use R_{1a}^2 consistently for any type of model with the appropriate a value, rather than adjusting any of the other R_i^2 ($i = 2, \ldots, 8$). For a no-intercept model, for example, the proper adjustment factor for (12) is clearly $a = n/(n - k)$. Suggestions such as using (12) for linear intercept models and, for example, $(n/(n - k))\Sigma(y_i - \hat{y}_i)^2/\Sigma y_i^2$ for no-intercept models (Marquardt and Snee 1974) lead to the problem that the two adjusted R^2's are not comparable.

The adjusted R_9^2 statistic may be defined as

$$\begin{aligned} R_{9a}^2 &= 1 - aM^2\{|y_i - \hat{y}_i|\}/M^2\{|y_i - \bar{y}|\} \\ &= 1 - a(1 - R_9^2), \end{aligned} \tag{13}$$

where a is the same adjustment factor as that of R_{1a}^2. For model specification analysis based on the use of some resistant model-fitting technique, and although R_9^2 may be an appropriate statistic for selecting the "best" type of model for a given set of independent variables, the resistant R_{9a}^2 may possibly serve as the preferable statistic for selecting the "best" subset of independent variables, since R_{9a}^2 takes into account the number of independent variables or model parameters. Accordingly, the best variable selection is the one for which R_{9a}^2 is maximized. Thus although the R_{9a}^2 criterion has the advantage of being resistant, since it is based on the use of resistant medians, the established model specification criteria of R_1^2, R_{1a}^2, MSE, and C_p (e.g., Montgomery and Peck 1982, pp. 249–255) are all based on the sum of squares of the residuals and consequently lack resistance against extreme residual values.

CONCLUSION

The various considerations, potential pitfalls, and variations in numerical values of the alternative R^2 statistics lead to the conclusion that R_1^2 or the adjusted R_{1a}^2 ought to be used consistently for any type of model- and curve (surface)-fitting technique. There does not appear to be any compelling reason for preferring any of the other R_i^2 ($i = 2, 3, \ldots, 8$) statistics. The data analyst is advised against using different R^2 statistics for different types of models, as this may lead to incompatible and potentially misleading results. In addition, care must be exercised when using computer program packages for model fitting.

A possible exception to the recommendation of using R_1^2 consistently is the case in which a resistant or robust model-fitting technique is used for highly contaminated data. In this case, the new resistant R_9^2 statistic or the degree-of-freedom adjusted R_{9a}^2 may be used as a possible alternative or in addition to R_1^2 or R_{1a}^2.

Finally, it ought to be emphasized that although R_1^2 and R_9^2 may serve as useful summary statistics for measuring model adequacy, uncritical use of and sole reliance on such statistics may fail to reveal important data characteristics and model inadequacies. Additional detailed analyses of the residuals is strongly recommended as an investment of time and effort with potentially significant return to the data analyst.

[*Received July 1983. Revised May 1984.*]

REFERENCES

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, New York: John Wiley.

Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis* (2nd ed.), New York: John Wiley.

Dzubay, A. E. (1984), *A Possible Alternative to the Established Model Specification Criteria*, unpublished M.S. project paper, University of Minnesota, Dept. of Mechanical Engineering.

Emerson, J. D., and Hoaglin, D. C. (1983), "Resistant Lines for y Versus x," in *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: John Wiley.

Goldberger, A. S. (1964), *Econometric Theory*, New York: John Wiley.

Hahn, G. J. (1973), "The Coefficient of Determination Exposed," *Chemtech*, 3, 609–612.

——— (1977), "Fitting Regression Models With No Intercept Term," *Journal of Quality Technology*, 9, 56–61.

Hawkins, D. M. (1980), "A Note on Fitting a Regression Without an Intercept Term," *The American Statistician*, 34, 233.

Hewlett-Packard (1976), *HP-67 Standard Pack*, Corvallis, OR: Author.

Huber, P. J. (1981), *Robust Statistics*, New York: John Wiley.

Kvålseth, T. O. (1983), "Rank–Frequency Distribution of Accident Statistics," *Journal of Safety Research*, 14, 173–181.

Lewis-Beck, M. S. (1980), *Applied Regression: An Introduction*, Sage University Paper series on Quantitative Applications in the Social Sciences, No. 22, Beverly Hills, CA: Sage.

Marquardt, D. W., and Snee, R. D. (1974), "Test Statistics for Mixture Models," *Technometrics*, 16, 533–537.

McNeil, D. R. (1971), *Interactive Data Analysis: A Practical Primer*, New York: John Wiley.

Montgomery, D. C., and Peck, E. A. (1982), *Introduction to Linear Regression Analysis*, New York: John Wiley.

Spiegel, M. R. (1975), *Probability and Statistics*, New York: McGraw-Hill.

Texas Instruments (1977), *TI Programmable 58/59 Applied Statistics*, Dallas, TX: Author.

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