COMMENTARIES

Commentaries are informative essays dealing with viewpoints of statistical practice, statistical education, and other topics considered to be of general interest to the broad readership of *The American Statistician*. Commentaries are similar in spirit to Letters to the Editor, but they involve longer discussions of background, issues, and perspectives. All commentaries will be refereed for their merit and compatibility with these criteria.

Another Cautionary Note About R^2: Its Use in Weighted Least-Squares Regression Analysis

JOHN B. WILLETT and JUDITH D. SINGER*

A recent article in this journal presented a variety of expressions for the coefficient of determination (R^2) and demonstrated that these expressions were generally not equivalent. The article discussed potential pitfalls in interpreting the R^2 statistic in ordinary least-squares regression analysis. The current article extends this discussion to the case in which regression models are fit by weighted least squares and points out an additional pitfall that awaits the unwary data analyst. We show that unthinking reliance on the R^2 statistic can lead to an overly optimistic interpretation of the proportion of variance accounted for in the regression. We propose a modification of the estimator and demonstrate its utility by example.

KEY WORDS: Coefficient of determination; Multiple R^2.

1. INTRODUCTION

Several recent papers have commented on the utility of the coefficient of determination, R^2, as a measure of goodness of fit in ordinary least-squares (OLS) regression. Draper (1984) suggested that R^2 was misleading in data sets in which there were replicate data points and suggested that the Box–Wetz (1973) criterion was more appropriate, although this view was modified later (Draper 1985). Healy (1984) commented that R^2 was an unsatisfactory measure of an OLS regression relationship, and that "an absolute rather than a relative measure is to be preferred" (p. 608).

Kvålseth (1985) discussed the use of R^2 as a measure of goodness of fit in OLS regression. He noted that the several alternative definitions of R^2 that abound in the statistical and data-analytic literature are not, in general, equivalent. He suggested one definition of R^2 as being superior and recommended its use consistently in data analysis:

$$R^2_{\text{OLS}} = 1 - \left[\frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol\beta})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol\beta})}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}\right], \qquad (1)$$

where $\mathbf{Y}$ is an $(n \times 1)$ vector of observations with $\bar{Y}$ as their mean, $\mathbf{X}$ is an $(n \times k)$ matrix of measurements on k predictors (including an intercept), $\hat{\boldsymbol\beta}$ is the $(k \times 1)$ vector of OLS parameter estimates, and $(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol\beta})$ is the $(n \times 1)$ vector of residuals from the OLS fit. Kvålseth (1985) argued that, whereas other estimators are flawed under particular combinations of no intercept and nonlinearity, this particular estimator successfully continues to offer, across a wide variety of contexts, an intuitive interpretation in terms of the "proportion of the total variation of Y (about its mean $\bar{Y}$) that is explained (accounted for) by the fitted model" (p. 281). A robust alternative to the estimator in (1) was presented and its behavior examined empirically.

We extend the earlier discussion to situations in which regression models are fitted by a weighted least-squares (WLS) strategy, rather than an OLS strategy. We show that unthinking reliance on an R^2 statistic, even the superior estimator in (1), can lead the unwary data analyst to an overly optimistic interpretation of the proportion of variance accounted for in the regression. We propose a modification of the estimator in (1) and demonstrate its utility by example.

2. WEIGHTED LEAST-SQUARES REGRESSION ANALYSIS

As Mosteller and Tukey (1977, p. 346) suggested, the action of assigning "different weights to different observations, either for objective reasons or as a matter of judgement" in order to recognize "some observations as 'better' or 'stronger' than others" has an extensive history. Whether the investigator wishes to downplay the importance of data points that are intrinsically more variable at specific levels of the predictor variables, or simply to decrease the effect on the fit of remote data points, the strategy is the same. The model under consideration is usually of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol\beta + \boldsymbol\epsilon, \qquad (2)$$

where $\boldsymbol\beta$ is a $(k \times 1)$ vector of unknown parameters, $\boldsymbol\epsilon$ is an $(n \times 1)$ vector of unobserved random errors with $\boldsymbol\epsilon \sim (\mathbf{O}, \sigma^2\mathbf{W})$, and $\mathbf{W}$ is a known $(n \times n)$ diagonal matrix with $w_{ii} > 0$. Usually, in an empirical analysis, $\mathbf{W}$ is not known and has to be generated by a "combination of prior knowl-

*John B. Willett and Judith D. Singer are Assistant Professors of Quantitative Methods, Graduate School of Education, Harvard University, Cambridge, Massachusetts 02138. The order of the authors' names was determined by randomization. The authors thank two anonymous referees for their helpful comments on earlier drafts of the article.

edge, intuition, and evidence" (Chatterjee and Price 1977, p. 101). Often, the required evidence is derived from inspection of residuals obtained from an initial unweighted (OLS) regression analysis.

Although WLS estimates of $\boldsymbol{\beta}$, var($\hat{\boldsymbol{\beta}}$), and σ^2 are usually computed directly, it is informative to transform (2) in order to create a model that can be fitted by OLS. Multiplying (2) throughout by $\mathbf{W}^{-1/2}$ gives

$$\mathbf{W}^{-1/2}\mathbf{Y} = \mathbf{W}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{-1/2}\boldsymbol{\epsilon}$$

or

$$\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\epsilon}_*, \quad (3)$$

where $\mathbf{Y}_* = \mathbf{W}^{-1/2}\mathbf{Y}$, $\mathbf{X}_* = \mathbf{W}^{-1/2}\mathbf{X}$, and $\boldsymbol{\epsilon}_* = \mathbf{W}^{-1/2}\boldsymbol{\epsilon}$. In the transformed model, var($\boldsymbol{\epsilon}_*$) = $\sigma^2\mathbf{I}$; therefore, the assumptions of OLS regression are met. Then, providing that ($\mathbf{X}'\mathbf{X}$) is nonsingular, estimates for $\boldsymbol{\beta}$, var($\hat{\boldsymbol{\beta}}$), σ^2, and the regression analysis-of-variance table can be obtained directly by replacing $\mathbf{Y}$ and $\mathbf{X}$ by $\mathbf{Y}_*$ and $\mathbf{X}_*$, respectively, in standard regression computations (Draper and Smith 1981, pp. 85–96).

Thus, in the weighted analysis, Kvålseth's coefficient of determination becomes

$$R^2_{\text{WLS}} = 1 - \left[\frac{(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_*)'(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_*)}{\mathbf{Y}'_*\mathbf{Y}_* - n\bar{Y}^2_*}\right], \quad (4)$$

where $\hat{\boldsymbol{\beta}}_*$ is the WLS estimate of $\boldsymbol{\beta}$. The denominator of the second term in (4) is the sum of squares of the weighted Y values about their mean, and the numerator is the sum of squares of the weighted residuals

$$\hat{\boldsymbol{\epsilon}}_* = \mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}}_*. \quad (5)$$

Therefore, R^2_{WLS} in (4) is the coefficient of determination in the *transformed* data set. It is a measure of the proportion of the variation in *weighted Y* that can be accounted for by *weighted X*, and it is the quantity that is output as "R^2" by the major statistical computer packages when a WLS regression is performed.

Weighted least-squares regression analysis minimizes the sum of squared residuals (and therefore maximizes the coefficient of determination) with respect to the *transformed* variables, whereas OLS regression analysis minimizes the sum of squared residuals (and maximizes the coefficient of determination) with respect to the *original* variables. Providing that the weighting scheme has been chosen appropriately to counteract the heteroscedastic nature of the random errors, a better fit will be achieved by WLS in the transformed world. Thus the coefficient of determination obtained unthinkingly from a statistical computer package under WLS regression is frequently much larger than the value obtained under the corresponding OLS fit. To the naive consumer of computer output, this apparent increment to the coefficient of determination can represent a considerable improvement in fit and is displayed prominently in any account of the analysis, whereas closer inspection reveals that the increment reflects, in part, the success of the weighting in solving the problem of heteroscedasticity.

From an applied perspective, however, it is more appropriate and less misleading to continue to report the "proportion of variance explained" in the *original* metric, and not in the transformed world. Attention should focus not on R^2_{WLS} but on a modification of (1) that incorporates the correct residuals for the model in (2),

$$\hat{\boldsymbol{\epsilon}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*, \quad (6)$$

and not the *weighted* residuals in (5). Therefore, an appropriate modification of Kvålseth's coefficient of determination becomes

$$\text{pseudo}\, R^2_{\text{WLS}} = 1 - \left[\frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*)'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_*)}{\mathbf{Y}'\mathbf{Y} - n\bar{Y}^2}\right], \quad (7)$$

a statistic that will necessarily be of smaller magnitude than the coefficient of determination obtained in the initial unweighted analysis. But since the OLS and WLS estimates of β remain unbiased even when var($\boldsymbol{\epsilon}$) $\neq \sigma^2\mathbf{I}$, it is likely that, except in particularly idiosyncratic data sets, the WLS estimate of β will not be much different from the OLS estimate and, therefore, the difference between R^2_{OLS} and pseudo R^2_{WLS} will not be great.

A simple formula for computing the pseudo R^2_{WLS} statistic can be obtained by summing and squaring the corrected residuals in Equation (6), and by computing the sample variance of the dependent variable, s^2_Y:

$$\text{pseudo}\, R^2_{\text{WLS}} = 1 - \left[\frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{(n-1)s^2_Y}\right]. \quad (8)$$

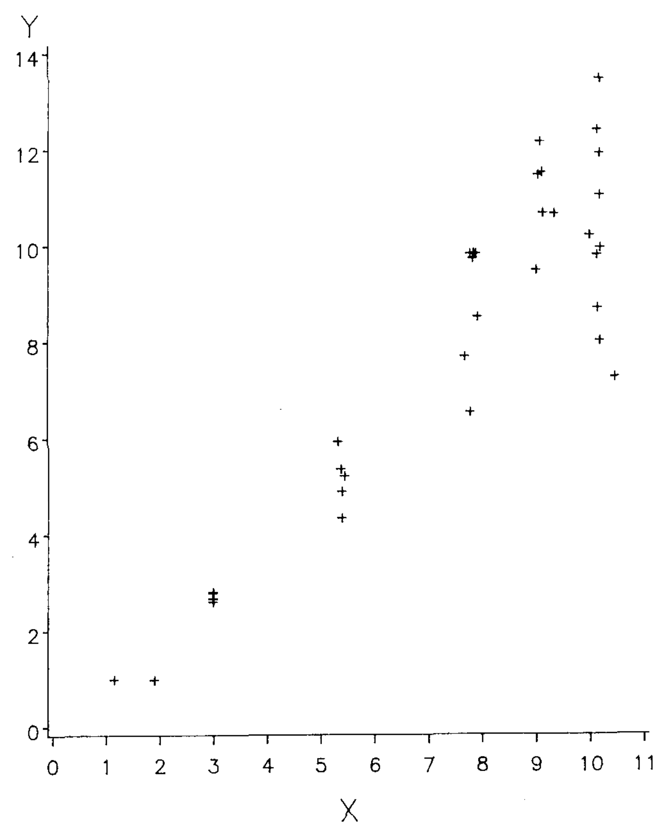

Figure 1. Scatterplot of the Example Data Set: Dependent Variable Y Versus Independent Variable X for a Sample of 35 Cases. The plot was created from data presented by Draper and Smith (1981, table 2.1).

Table 1. Ordinary Least-Squares (OLS) and Weighted Least-Squares (WLS) Parameter Estimates, Standard Errors, and Coefficient of Determination for the Example Data Set

Type of fit	Estimated intercept (standard error)	Estimated slope (standard error)	Coefficient of determination (percent)
OLS	−.579 (.679)	1.135 (.086)	84.0
WLS	−.889 (.300)	1.165 (.059)	92.1

NOTE: The estimated weight for Case 10 was misprinted as 6.70574 in Draper and Smith (1981, table 2.1). Our analysis made use of the correct value of 6.78574, which was recomputed according to directions given by Draper and Smith (1981, p. 115).

3. AN EXAMPLE

In this section we present a reanalysis of a small data set provided by Draper and Smith (1981, table 2.1). This data set contains information for 35 cases on a single outcome variable, Y, and a single predictor, X. The data are plotted in Figure 1. Inspection of the figure reveals a point cloud in the familiar "right-opening megaphone" shape typical of heteroscedastic random errors. Draper and Smith (1981, pp. 112–115) described in considerable detail the estimation of the elements of the **W** matrix. By inverting the diagonal elements of this matrix, we can obtain the estimated weights needed to fit a linear model to these data by WLS regression analysis.

We have used the Draper and Smith data to fit the linear model in Equation (2) (with a single predictor) using OLS regression analysis and WLS regression analysis, respectively. The results of these analyses are presented in Table 1, with the coefficients of determination estimated by Equations (1) and (4), respectively. Notice that, as anticipated, there has been an increase in the estimated value of the coefficient of determination from 84.0% to 92.1%. That this apparent increase is easy to misinterpret is illustrated by estimating the pseudo R^2_{WLS} statistic in Equation (7). This leads to an estimate of .839 for pseudo R^2_{WLS}, a value that is only marginally lower than the magnitude of the coefficient of determination in the original OLS regression.

4. CONCLUSION

The small difference between R^2_{OLS} and pseudo R^2_{WLS} raises the following question: Why estimate pseudo R^2_{WLS} at all? The answer is simply that such a computation emphasizes, for the naive user, that the goodness of fit (as documented by the coefficient of determination) has not been dramatically improved by the WLS regression but, in fact, has deteriorated slightly. This serves to refocus attention on other aspects of the analysis, particularly the increased precision of the estimates of β. Therefore, our results reinforce the view that statistics other than the coefficient of determination are of primary interest in both OLS and WLS regressions, and they support Kvålseth's contention that "sole reliance on [the coefficient of determination] may fail to reveal important data characteristics and model inadequacies" (1985, pp. 282–284).

[Received October 1986. Revised February 1988.]

REFERENCES

Box, G. E. P., and Wetz, J. (1973), "Criteria for Judging Adequacy of Estimation by an Approximating Response Function," Technical Report 9, University of Wisconsin, Dept. of Statistics.

Chatterjee, S., and Price, B. (1977), *Regression Analysis by Example*, New York: John Wiley.

Draper, N. R. (1984), "The Box–Wetz Criterion Versus R^2," *Journal of the Royal Statistical Society*, Ser. A, 147, 100–103.

——— (1985), "Corrections: The Box–Wetz Criterion Versus R^2," *Journal of the Royal Statistical Society*, Ser. A, 148, 357.

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

Healy, M. J. R. (1984), "The Use of R^2 as a Measure of Goodness of Fit," *Journal of the Royal Statistical Society*, Ser. A, 147, 608–609.

Kvålseth, T. O. (1985), "Cautionary Note About R^2," *The American Statistician*, 39, 279–285.

Mosteller, F., and Tukey, J. W. (1977), *Data Analysis and Regression*, Reading, MA: Addison-Wesley.