The Interpretation of Dummy Variables in Semilogarithmic Equations

By ROBERT HALVORSEN AND RAYMOND PALMQUIST*

A number of recent articles on discrimination, education, and income misinterpret the regression of coefficients of some variables, and as a result report incorrect estimates of their effects. Examples appearing in this *Review* include articles by Eric Hanushek and John Quigley, David Hartman, Robert Lucas, James Smith and Finis Welch, Barry Chiswick, and Emily Hoffman. An analogous error appears in the literature on hedonic price indexes, including frequently cited articles by Zvi Griliches, Jack Triplett, and A. T. Court.

This common error involves the interpretation of the coefficients of dummy variables in semilogarithmic regression equations. The articles cited above assume that the coefficient of a dummy variable, multiplied by 100, is equal to the percentage effect of that variable on the variable being explained. However, it is easily shown that this interpretation, while correct for continuous variables, is not correct for dummy variables and can result in substantial errors in the reporting of results.

The general form of the equations estimated in the articles cited above is

$$(1) \qquad ln\, Y = a + \sum_i b_i X_i + \sum_j c_j D_j$$

where the X_i represent continuous variables and the D_j represent dummy variables. The coefficient of a continuous variable is

$$b_i = \frac{\partial ln\, Y}{\partial X_i} = \frac{1}{Y} \cdot \frac{\partial Y}{\partial X_i}$$

Thus the coefficient of a continuous variable, multiplied by 100, is equal to the percentage effect on Y of a small change in that variable.

Since a dummy variable enters the equation in dichotomous form, the derivative of the dependent variable with respect to the dummy variable does not exist. Instead, the coefficient of a dummy variable measures the discontinuous effect on Y of the presence of the factor represented by the dummy variable. The appropriate interpretation of the coefficient of a dummy variable can be shown directly by a transformation of equation (1).

For simplicity, it will be assumed that there is a single dummy variable. Equation (1) can then be written as

$$(2) \qquad Y = (1+g)^D \exp\left(a + \sum_i b_i X_i\right)$$

where g is the relative effect on Y of the presence of the factor represented by the dummy variable.[1] Thus the coefficient of the dummy variable in equation (1) is $c = ln(1+g)$. The relative effect on Y is $g = \exp(c) - 1$, and the percentage effect is equal to[2]

$$100 \cdot g = 100 \cdot \{\exp(c) - 1\}$$

The results reported in the studies cited above are based on the incorrect assumption that $c = g$. The relationship between c and g when the absolute value of g is less than one can be examined by expansion

*University of Washington. We are grateful to Taylor Dennen and Gerald Glandon for discussion of the topic of this note.

[1]Thus $g = (Y_1 - Y_0)/Y_0$ where Y_1 and Y_0 are the values of the dependent variable when the dummy variable is equal to one and zero, respectively.

[2]When time dummies are used to estimate hedonic price indexes, c is equal to the rate of change in price which, when continuously compounded, yields g, the relative change in price during the period. The value of the price index at the end of the period is equal to exp (c), not $1 + c$ as assumed in the hedonic studies cited above.

TABLE 1—RELATIONSHIP BETWEEN DUMMY VARIABLE COEFFICIENTS AND RELATIVE EFFECTS

Coefficient of the Dummy Variable (c)	Relative Effect (g)
1.50	3.48
1.25	2.49
1.00	1.72
0.75	1.12
0.50	0.65
0.25	0.28
0.00	0.00
−0.25	−0.22
−0.50	−0.39
−0.75	−0.53
−1.00	−0.63
−1.25	−0.71
−1.50	−0.78

of c,

$$c = ln(1+g) = g - \frac{1}{2}g^2 + \frac{1}{3}g^3 - \dots \quad |g| < 1$$

For small values of g, c is approximately equal to g. When g is positive, c is smaller than g, and when g is negative, c is algebraically smaller than g but larger in absolute value.

Table 1 shows the magnitude of the relative change g, implied by selected values for the coefficient of the dummy variable c. The errors involved in assuming that c is equal to g can be substantial for values of c within the range estimated in the cited studies. For example, Hanushek and Quigley (p. 74) report that a postgraduate degree increases the wages of a black worker by 64 percent, whereas the correct result implied by their regression is 90 percent.

REFERENCES

B. R. Chiswick, "Sons of Immigrants: Are They at an Earnings Disadvantage?," *Amer. Econ. Rev. Proc.*, Feb. 1977, *67*, 376–80.

A. T. Court, "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, New York 1939.

Z. Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in his *Price Indexes and Quality Change*, Cambridge 1971.

E. A. Hanushek and J. M. Quigley, "Implicit Investment Profiles and Intertemporal Adjustments of Relative Wages," *Amer. Econ. Rev.*, Mar. 1978, *68*, 67–79.

D. G. Hartman, "What Do Economics Majors Learn?," *Amer. Econ. Rev. Proc.*, May 1978, *68*, 17–22.

E. P. Hoffman, "Faculty Salaries: Is There Discrimination by Sex, Race, and Discipline? Additional Evidence," *Amer. Econ. Rev.*, Mar. 1976, *66*, 196–98.

R. E. B. Lucas, "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *Amer. Econ. Rev.*, Sept. 1977, *67*, 549–58.

J. P. Smith and F. R. Welch, "Black-White Male Wage Ratios: 1960–70," *Amer. Econ. Rev.*, June 1977, *67*, 323–38.

J. E. Triplett, "Automobiles and Hedonic Quality Measurement," *J. Polit. Econ.*, May/June 1969, *77*, 408–17.