Interpreting Semilogarithmic Regression Coefficients in Labor Research

ROBERT J. THORNTON and JON T. INNES*
Lehigh University, Bethlehem, PA 18015

Labor economists frequently misinterpret coefficients of variables in semilogarithmic regression equations. The proportional rates of change in the dependent variable that are implied by these coefficients are often erroneously assumed to be valid over arbitrarily large intervals. This note provides mathematical and empirical evidence on how serious the error can be. A simple formula is developed for making correct interpretations of semilog regression coefficients.

I. *Introduction*

Many economists commit an error when interpreting the coefficients of variables in semilogarithmic equations. The error seems to be exceptionally common among labor economists — perhaps because of their frequent use of Mincer-type earnings functions — but it is by no means restricted to them. While we would like to believe that most economists are aware of the error, the fact that it is so widely encountered in economics articles and textbooks and in the classroom leads us to believe otherwise.

We present here a brief mathematical demonstration to explain how the error arises and how serious it can be. In the process, we develop a simple formula that can be used to make correct interpretations of semilogarithmic regression coefficients. We close the note with several examples from the recent labor economics literature to point out just how widespread the confusion is.

II. *Semilog Regression Coefficients and Proportional Changes*

In semilogarithmic equations of the type

$$\ln Y = a + bX, \tag{1}$$

many economists unqualifyingly interpret b as the proportional change in Y (or $100b$ as the percentage change) resulting from a unit change in X. Although this interpretation may be approximately correct over a very small range, it is decidedly incorrect outside that small range.

*The authors thank David Greenaway for helpful comments.

We are not the first to point out at least one manifestation of this error, but the error is more fundamental than originally noted. In a 1980 article, Robert Halvorsen and Raymond Palmquist argued that the coefficients of *dummy* variables in semilog regression equations are frequently misinterpreted in this way and that as a result the estimated effects of those variables are incorrectly reported. Halvorsen and Palmquist, however, imply that such an interpretation is incorrect *only* for coefficients of dummy variables. As we emphasize here, this interpretation is incorrect for non-dummy variables as well. It is not the unique nature of a dummy variable that results in the error, but rather the fact that equating a logarithmic change with a proportional or percentage change is only approximately correct over a given range no matter what the nature of the independent variable.

To see this, consider again equation (1). The effect on $\ln Y$ of an *infinitesimal* change in X can be calculated by taking the differential of each side of the equation. This yields:

$$d\ln Y = b\,dX. \tag{2}$$

Because

$$d\ln Y = dY/Y, \tag{3}$$

bdX is thus seen to be equal to the proportional change in Y. But many economists interpreting a semilog regression also equate $b\Delta X$ — the product of the regression coefficient and a *non-infinitesimal* change in X — with the proportional change in Y. The true proportional change in Y (as it changes, say, from Y_o to Y_1) resulting from such a non-infinitesimal change in X, however, is equal to g, where

$$g = Y_1/Y_o - 1. \tag{4}$$

And it can be easily seen that $b\Delta X$ and g are not equal, because

$$b\Delta X = \ln Y_1 - \ln Y_o = \ln(Y_1/Y_o) = \ln(1 + g). \tag{5}$$

Therefore, to calculate the true proportional change in Y resulting from a non-infinitesimal change in X, one would have to calculate

$$g = \exp(b\Delta X) - 1. \tag{6}$$

For a unit change in X — the special case of a dummy variable — expression (6) becomes

$$g = \exp(b) - 1. \tag{7}$$

The gist of our argument is that the potential for erroneously equating regression coefficients with proportional changes exists in the case of such semilogarithmic regressions whether or not the independent variables are dichotomous. In each case the error takes the form of approximating a nonlinear relationship [exp $(b\Delta X) - 1$] with a linear relationship [$b\Delta X$]. As a result, the magnitude of the error is a function of both the value of the regression coefficient *and* the size of the X-interval over which the corresponding proportional change in the dependent variable is calculated.

Table 1 presents comparisons of actual vs. approximate proportional changes in the dependent variable associated with various values of $b\Delta X$. The table illustrates clearly that $b\Delta X$ is a close approximation to g only if $b\Delta X$ is also close to zero. The magnitude of the error can also be easily deduced from equation (6), from which it can be shown that

$$g = b\Delta X + \int_0^{b\Delta X} g d(b\Delta X). \tag{8}$$

In other words, the magnitude of the error is equal to the definite integral of g itself over the interval 0 to $b\Delta X$.

Table 1

Approximate vs. Actual Percentage Changes in the Dependent Variable Implied by Semilog Regression Coefficients

Approximate Percentage Change ($b\Delta X \times 100$)	Actual Percentage Change ($g \times 100$)
100	171.8
75	111.7
50	64.9
25	28.4
20	22.1
15	16.2
10	10.5
5	5.1
0	0.0
−5	−4.9
−10	−9.5
−15	−13.9
−20	−18.1
−25	−22.1
−50	−39.3
−75	−52.8
−100	−63.2

III. *How Widespread Is the Confusion?*

This commonly made error can be illustrated using just a few recent examples from the labor economics literature.[1] In his study of concentration and earnings, John Heywood (1986) regressed the natural logarithm of individuals' hourly wages on the four-firm concentration ratios for the industries in which the individuals are employed. In explaining the magnitude of the estimated regression coefficient (.0031), Heywood stated that a two standard deviation rise in the concentration ratio (a change of 36.6 percentage points) would result in an 11.35 percent increase in the average hourly wage. In fact, the correct interpretation implies an increase of over 12 percent.

In their analysis of wildcat strike incidence, Byrne and King (1986) interpreted the coefficient of their National Labor Relations Board (NLRB) variable (−.167) as indicating that an increase of one complaint per thousand union workers filed with the NLRB reduces wildcat strikes by 16.7 percent (p. 396). The true effect, though, is −15.4 percent, about 8 percent lower than Byrne and King claimed.

In their article explaining the determinants of Political Action Committee (PAC) funds, Wilhite and Theilmann (1986) interpreted the coefficient (.174) of a representative's AFL rating in the following way: "If a representative's rating increases by a point, his or her [PAC] funds increase by 17 percent"(p. 182). The actual increase inplied by the Wilhite-Theilmann coefficient, however, is just over 19 percent.

In analyzing the determinants of the number of applicants for federal job openings, Alan Krueger (1988, p. 573) stated that a coefficient of 6.49 meant that "a one percentage point increase in the unemployment rate . . . is associated with a 6.49 percent increase in the application rate. . . ." Krueger has, in fact, underestimated slightly the true association between the two variables.

In their study of union-nonunion wage differentials, Schulenburger, McLean, and Rasch (1982, p. 253) interpreted a coefficient of −.501 in this way: "Taken literally, it implies that the union-nonunion wage ratio was 50.1 percent smaller when 100 percent of the workforce was organized." As can be seen from Table 1, the true ratio was only about 39 percent smaller.

Still other examples of this unqualified equating of semilogarithmic regression coefficients with proportional changes in the dependent variable can be found in basic econometrics textbooks. Gujarati (1978), for example, asserted that the regression coefficient in an equation similar to (1) "measures the constant relative or proportional change in Y for a given absolute change in X" (p. 54). Likewise, Studenmund and Cassidy (1987) invited misinterpretation when

[1] In selecting these articles to illustrate the interpretation error, we by no means intend to impugn their overall quality. Quite to the contrary, these works can be considered to be examples of otherwise solid empirical research.

they wrote: "If X_1 changes by one unit, Y changes by β_1 (times 100) percent . . ." (p. 149).

Finally, it is worth noting that, despite Halvorsen and Palmquist's 1980 warning, errors in the interpretation of *dummy* variable coefficients also continue to occur with disturbing frequency in labor research. For instance, John Richards and Alan Carruth (1986) use a dummy variable to explain the effect of a coal strike on the number of workers on short-time. They interpreted the coefficient of 1.18 as implying that the effect is "to boost the figures by 118% . . ." (p. 54). In fact, the true effect of the strike is to raise the figures by 225 percent — nearly twice the magnitude claimed by Richards and Carruth.[2]

As common as these kinds of errors are, we believe that their incidence would be even greater were it not for the widespread practice in economics of simply *reporting* regression coefficients and allowing the readers to interpret the results for themselves. In any case, we hope that the reminder expressed here and the contents of Table 1 will serve as an effective antidote.

REFERENCES

Byrne, D. and R. King. "Wildcat Strikes in U.S. Manufacturing, 1960-77." *Journal of Labor Research* 7 (1986): 387-401.

Gujarati, D. *Basic Econometrics*. New York: McGraw Hill Book Company, 1978.

Halvorsen, R. and R. Palmquist. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70 (1980): 474-75.

Heywood, J. "Labor Quality and the Concentration-Earnings Hypothesis." *Review of Economics and Statistics* 68 (1986): 342-46.

Krueger, A. "The Determinants of Queues for Federal Jobs." *Industrial and Labor Relations Review* 41 (1988): 567-81.

Moore, W. and J. Raisian. "The Level and Growth of Union/Nonunion Relative Wage Effects, 1967-77." *Journal of Labor Research* 4 (1983): 65-79.

Richards, J. and A. Carruth. "Short-Time Working and the Unemployment Benefit System in Great Britain." *Oxford Bulletin of Economics and Statistics* 48 (1986): 41-59.

Schulenburger, D., R. McLean and S. Rasch. "Union-Nonunion Wage Differentials: A Replication and Extension." *Industrial and Labor Relations Review* 21 (1982): 248-55.

Studenmund, A. and H. Cassidy. *Using Econometrics*. Boston: Little, Brown and Company, 1987.

Wilhite, A. and J. Theilmann. "Unions, Corporations, and Political Campaign Contributions: The 1982 House Elections." *Journal of Labor Research* 7 (1986): 175-86.

[2]Ironically, some researchers have been careful to interpret correctly their dummy-variable coefficients only to then misinterpret the coefficients of their non-dummy variables. For example, see the study of union/nonunion relative wage effects by William Moore and John Raisian (1983, p. 73). It should be noted that although Moore and Raisian misinterpret the proportional changes in wages implied by their trend coefficients, the error involved is trifling.