Volume Title: The Demand for Health: A Theoretical and Empirical Investigation

Volume Author/Editor: Michael Grossman

Volume Publisher: NBER

Volume ISBN: 0-87014-248-8

Volume URL: http://www.nber.org/books/gros72-1

Publication Date: 1972

Chapter Title: Appendix D: Statistical Properties of the Model

Chapter Author: Michael Grossman

Chapter URL: http://www.nber.org/chapters/c3494

Chapter pages in book: (p. 97 - 101)

Appendix D

STATISTICAL PROPERTIES OF THE MODEL

1. STRUCTURE AND REDUCED FORM

Let the demand curve for the stock of health be given by

$$\ln H_i = \varepsilon \ln W_i - \varepsilon \ln \pi_i - \varepsilon \ln (r - \tilde{\pi}_i + \delta_i), \tag{D-1}$$

the depreciation rate function by

$$\ln \delta_i = \ln \delta_0 + \tilde{\delta} i, \tag{D-2}$$

and the gross investment production function by

$$\ln I_i = r_H E + \alpha_1 \ln M_i + (1 - \alpha_1) \ln TH_i, \tag{D-3}$$

where all variables are expressed as deviations from their respective means. Substituting (D-2) into (D-1) and assuming that $r - \tilde{\pi}_i = 0$, one gets

$$\ln H_i = \varepsilon \ln W_i - \varepsilon \ln \pi_i - \tilde{\delta} \varepsilon i - \varepsilon \ln \delta_0. \tag{D-4}$$

The total cost of gross investment can be written

$$C_i = \pi_i I_i = PM_i + W_i TH_i, \tag{D-5}$$

and in least-cost equilibrium

$$\frac{g - tg'}{g'} = \frac{P}{W_i} = \frac{\alpha_1}{1 - \alpha_1} \frac{TH_i}{M_i}. \tag{D-6}$$

Utilization of (D-3), (D-5), and (D-6) gives the marginal cost function

$$\ln \pi_i = K \ln W_i + (1 - K) \ln P - r_H E, \tag{D-7}$$

where $K = 1 - \alpha_1$ and $1 - K = \alpha_1$. Substitution of equation (D-7) into equation (D-4) generates the reduced form demand curve for the stock of health:

$$\ln H_i = (1 - K)\varepsilon \ln W_i - (1 - K)\varepsilon \ln P + r_H \varepsilon E - \tilde{\delta} \varepsilon i - \varepsilon \ln \delta_0. \tag{D-8}$$

Having obtained a stock demand curve, one can proceed to calculate a derived demand curve for medical care. If the production function equation is solved for $\ln M_i$, then

$$\ln M_i = \alpha_1^{-1} \ln I_i - (1 - \alpha_1)\alpha_1^{-1} \ln TH_i - \alpha_1^{-1} r_H E.$$

But $\ln I_i = \ln H_i + \ln(\tilde{H}_i + \delta_i)$. Hence,

$$\ln M_i = \alpha_1^{-1} \ln H_i + \alpha_1^{-1} \ln(\tilde{H}_i + \delta_i) - (1 - \alpha_1)\alpha_1^{-1} \ln TH_i - \alpha_1^{-1} r_H E.$$

Equations (D-6) and (D-8) can be employed to show that an alternative form of the last equation is

$$\ln M_i = [(1 - K)\varepsilon + K] \ln W_i - [(1 - K)\varepsilon + K] \ln P + r_E(\varepsilon - 1)E$$
$$-\tilde{\delta}\varepsilon i - \varepsilon \ln \delta_0 + \ln(\tilde{H} + \delta_i).$$

The last term on the right-hand side of this equation can be rewritten as $\ln \delta_i + \ln(1 + \tilde{H}_i/\delta_i)$. From (D-2),

$$\ln \delta_i + \ln(1 + \tilde{H}/\delta_i) = \ln \delta_0 + \tilde{\delta} i + \ln(1 + \tilde{H}_i/\delta_i).$$

Therefore, the reduced form demand curve for medical care is

$$\ln M_i = [(1 - K)\varepsilon + K] \ln W_i - [(1 - K)\varepsilon + K] \ln P + r_E(\varepsilon - 1)E$$
$$+ \tilde{\delta}(1 - \varepsilon)i + (1 - \varepsilon) \ln \delta_0 + \ln(1 + \tilde{H}/\delta_i). \tag{D-9}$$

2. THE EFFECTS OF ERRORS OF MEASUREMENT

Suppose P does not vary across the relevant units of observation, $\tilde{H}_i/\delta_i$ is small, and wealth is excluded from the set of exogenous variables.[1] Then the reduced form would be

$$\ln H_i = B_1 + B_W \ln \overline{W} + B_E E' + B_i i + u_1 \tag{D-10}$$

$$\ln M_i = B_2 + B_{WM} \ln \overline{W} + B_{EM} E' + B_i i + u_2, \tag{D-11}$$

where $\overline{W}$ is the true wage at age i and E' is a measure of efficiency in nonmarket production. The investment model predicts $B_W > 0$, $B_{E'} > 0$, $B_i < 0$, and $B_{WM} > 0$. In addition, if $\varepsilon < 1$, $B_{E'M} < 0$ and $B_{iM} > 0$. The problem of errors of measurement arises in estimating these equations for two reasons. First, at any given age, a person's observed wage may contain a transitory component due to factors such as unexpected unemployment

[1] The analysis of errors of measurement would be considerably more complicated if wealth or family income were one of the independent variables. None of the basic conclusions reached in this section, however, would be substantially altered.

and response errors. Second, the variable E' in (D-10) and (D-11) is not education per se. Instead, it is a more general index of efficiency in nonmarket production. This index depends on education, E, but also depends on a vector of other variables, u_3, that reflects variations in nonmarket ability across individuals with the same amount of formal schooling.

Let the relation between the true wage and the observed (measured) wage be

$$\ln \overline{W} = \ln W - \ln W^*, \tag{D-12}$$

where $\ln W^*$ is the transitory component. In addition, let the equation for nonmarket efficiency be

$$E' = a_1 + a_2 E + a_3 u_3, \tag{D-13}$$

where $a_2 > 0$ and $a_3 > 0$. Substitution of (D-12) and (D-13) into (D-10), yields

$$H = B_1 + B_E a_1 + B_W W + B_E a_2 E + B_i i + u_1 - B_W W^* + B_E a_3 u_3$$

or

$$H = B_1^* + B_W W + B_E E + B_i i + u_4, \tag{D-14}$$

where $u_4 = u_1 - B_W W^* + B_E a_3 u_3$.[2] If (D-14) is fitted by ordinary least squares, the estimated equation would be

$$H = b_1 + b_W W + b_E E + b_i i + e,$$

or in matrix notation

$$H = Xb + e.$$

To determine whether the regression coefficients in the vector b are unbiased estimates of the true population parameters, write

$$b = (X'X)^{-1}X'H.$$

The expected value of this vector is

$$E(b) = E[(X'X)^{-1}X'(XB + u_4)]$$

$$E(b) = B + (X'X)^{-1}E(X'u_4).$$

[2] The abbreviation for natural logarithm, ln, is omitted before H, W, $\overline{W}$, and W^* from now on.

Note that $(X'X)^{-1} = \sigma^2 b/\sigma^2 u_4$. Note also that

$$(X'X)^{-1} = \begin{bmatrix} K_W & -r_{WE}K_{WE} & -r_{Wi}K_{Wi} \\ -r_{EW}K_{EW} & +K_E & -r_{Ei}K_{Ei} \\ -r_{iW}K_{iW} & -r_{iE}K_{iE} & +K_i \end{bmatrix} \tag{D-16}$$

where, for example, $K_W = \sigma^2 b_W/\sigma^2 u_4 > 0$, $K_{WE} = (\sigma b_W \sigma b_E/\sigma^2 u_4) > 0$, and r_{WE} is the partial correlation coefficient between W and E, with i heid constant.[3] Thus, the sign of a typical off-diagonal element in $(X'X)^{-1}$ depends solely on the partial correlation between the two relevant independent variables.

Equation (D-16) allows one to compare the expected value of any regression coefficient with its true value. For instance, the term in the matrix $E(b)$ that corresponds to the regression coefficient of the wage rate is

$$E(b_W) = B_W + K_W E[\mathrm{Cov}\,(Wu_4)] - r_{WE}K_{WE}E[\mathrm{Cov}\,(Eu_4)] \\ -r_{Wi}K_{Wi}E[\mathrm{Cov}\,(iu_4)].$$

To determine the biases introduced by errors of measurement, the correlation between u_4 and each of the independent variables must be evaluated. It is reasonable to suppose that age and education are independent of u_4. It is also reasonable to assume that the wage rate is positively correlated with u_3 because *market* and *nonmarket* ability are positively correlated. In other words, an increase in u_3, with education held constant, reflects an increase in nonmarket efficiency. On the other hand, an increase in the wage, with age and education fixed, may be viewed as an increase in market efficiency. Then if W^* were independent of $\overline{W}$,

$$E[\mathrm{Cov}\,(Eu_4)] = E[\mathrm{Cov}\,(iu_4)] = 0$$
$$E[\mathrm{Cov}\,(Wu_4)] = -B_W\sigma^2 W^* + B_E a_3 \,\mathrm{Cov}\,(Wu_3),$$

and

$$\begin{aligned} E(b_W) &= B_W - K_E B_W \sigma^2 W^* + K_W B_E a_3 \,\mathrm{Cov}\,(Wu_3) \\ E(b_E) &= B_E + r_{EW}K_{EW}B_W\sigma^2 W^* - r_{EW}K_{EW}B_E a_3 \,\mathrm{Cov}\,(Wu_3) \\ E(b_i) &= B_i + r_{iW}K_{iW}B_W\sigma^2 W^* - r_{iW}K_{iW}B_E a_3 \,\mathrm{Cov}\,(Wu_3). \end{aligned} \tag{D-17}$$

[3] For a derivation of equation (D-16), see Yoel Haitovsky, "Simplified Formulae for Covariance B," New York, NBER, mimeographed, 1968. See also Robert T. Michael, *The Effect of Education on Efficiency in Consumption*, New York, NBER, Occasional Paper 116, 1972, Appendix B. Much of my discussion of the effects of errors of measurement is based on Michael's formulation of the problem.

Equation (D-17) suggests that two types of biases operate on the estimated coefficients. The first is due to measurement error in the wage variable and is represented by the variance of the transitory wage, $\sigma^2 W^*$. The second is due to the positive correlation between market and nonmarket ability and is represented by the term Cov (Wu_3). Since $B_W > 0$, the presence of random errors of observation biases the wage coefficient downward. Since $B_E a_3 > 0$ and Cov $(Wu_3) > 0$, the ability effect biases this coefficient upward. So the net impact of these two biases on the wage coefficient is not certain.

In the NORC sample, $r_{EW} = .418$ and $r_{iW} = .135$.[4] Since both these partial correlation coefficients are positive, measurement error biases the age and education coefficients upward, while ability biases them downward. Consequently, the important conclusion is reached that none of the coefficients in the demand curve for health is biased in an *obvious* direction.

The situation is very different in the demand curve for medical care. In that demand curve, the expected value of the regression coefficient of education—or of nonmarket ability in general—would be negative if the elasticity of the MEC schedule were less than unity. Given this condition, both the measurement error and the ability effects would bias the wage coefficient *downward*. Moreover, both these effects would bias the education and age coefficients *upward*. The important point to stress is that the two sources of bias operate in *opposite* directions on any given coefficient in the demand curve for health but operate in the *same* direction on the corresponding coefficient in the demand curve for medical care.

[4] These partial correlations pertain to whites with positive sick time. They hold constant sex and family size as well as age or education.