This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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

Volume Author/Editor: Michael Grossman

Volume Publisher: UMI

Volume ISBN: 0-87014-248-8

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

Publication Date: 1972

Chapter Title: AN EMPIRICAL FORMULATION OF THE MODEL

Chapter Author: Michael Grossman

Chapter URL: http://www.nber.org/chapters/c3488

Chapter pages in book: (p. 39 - 54)

IV

AN EMPIRICAL FORMULATION OF THE MODEL

In the previous chapters, I have developed a framework that can be employed to predict the effects of certain variables on the demand for health and medical care. To test the implications of this framework, it is necessary to estimate demand curves for health and medical care and perhaps a gross investment production function as well. In the first section of this chapter, I explore the estimation of this set of equations in detail. In particular, I outline the structure and reduced form of the pure investment model. Although the formulation is oriented toward the investment framework, I offer two tests to distinguish the investment model from the consumption model. In the second section, I describe the measures of health used in the empirical analysis, discuss the data source from which these measures are obtained, and comment on the independent variables that enter the multiple regression estimates of the system.

The estimation of the investment model rather than the consumption model is stressed because the former model generates powerful predictions from simple analysis and innocuous assumptions. For example, if one uses the investment framework, then he does not have to know whether the production of health is relatively time-intensive to predict the effect of an increase in the wage rate on the demand for health. Again, he does not have to know whether shifts in education are commodity-neutral to assess the sign of the correlation between health and schooling. Moreover, the responsiveness of the quantity of health demanded to changes in its shadow price and the behavior of gross investment depend essentially on a single parameter—the elasticity of the MEC schedule. In the consumption model, on the other hand, three parameters are relevant—the own price elasticity of health, the elasticity of substitution between present and future health, and the wealth elasticity.

1. STRUCTURE AND REDUCED FORM

To derive estimating equations for the pure investment model, begin with the production function of healthy days utilized in Chapter II:

$$h_i = 365 - BH_i^{-C}. \tag{4-1}$$

Although the subscript i refers to age, it should be clear that H will vary across individuals as well as over the life cycle of a given individual. It has already been shown[1] that this production function generates the constant elasticity MEC schedule[2]

$$\ln \gamma_i = \ln BC - (C + 1) \ln H_i + \ln W_i - \ln \pi_i,$$

where $\varepsilon = 1/(1 + C)$. Solving the last equation for $\ln H_i$ and substituting $r - \tilde{\pi}_i + \delta_i$ for γ_i, one obtains the stock of health demand function

$$\ln H_i = B' + \varepsilon \ln W_i - \varepsilon \ln \pi_i - \varepsilon \ln (r - \tilde{\pi}_i + \delta_i), \tag{4-2}$$

where $B' = \ln BC/(1 + C)$. Suppose $\tilde{\pi}_i$ is positive and constant and the real-own rate of interest is equal to zero. Then equation (4-1) would reduce to

$$\ln H_i = B' + \varepsilon \ln W_i - \varepsilon \ln \pi_i - \varepsilon \ln \delta_i. \tag{4-3}$$

Although age and education do not appear explicitly on the right-hand side of (4-3), they are implicit in this equation because the rate of depreciation and the marginal cost of gross investment are not directly observable and expressions for them must be developed. It has been hypothesized that depreciation rates rise with age, at least after some stage in the life cycle, and vary among individuals of the same age as well. Let these factors be summarized by a depreciation rate equation of the form

$$\ln \delta_i = \ln \delta_0 + \tilde{\delta} i. \tag{4-4}$$

An equation for marginal cost can be developed from the household production function for gross investment. For analytical convenience, assume the production function is a member of the Cobb–Douglas class:

$$\ln I_i = r_H E + \alpha_1 \ln M_i + (1 - \alpha_1) \ln TH_i. \tag{4-5}$$

Here $\alpha_1 = 1 - K$ is the share of medical care in the total cost of gross investment or the elasticity of gross investment with respect to medical care. With this production function, the elasticity of substitution between medical services and own time equals one. Consequently, K is independent of the prices of these inputs.

[1] See Chapter II, footnote 11.

[2] Continuous time equilibrium conditions are utilized in this chapter. Hence, $\gamma_i = W_i G_i / \pi_i$, and the real-own rate of interest is $r - \tilde{\pi}_i$.

Appendix D, Section 1, demonstrates that equations (4-3) and (4-4) generate the following reduced form demand curves for health and medical care:[3]

$$\ln H_i = (1 - K)\varepsilon \ln W_i - K\varepsilon \ln P + r_H \varepsilon E - \tilde{\delta}\varepsilon i - \varepsilon \ln \delta_0, \quad (4\text{-}6)$$

$$\ln M_i = [(1 - K)\varepsilon + K] \ln W_i - [(1 - K)\varepsilon + K] \ln P + r_E(\varepsilon - 1)E + \tilde{\delta}(1 - \varepsilon)i + (1 - \varepsilon) \ln \delta_0 + \ln (1 + \tilde{H}_i/\delta_i). \quad (4\text{-}7)$$

A demand curve for the time spent producing health could also be developed, but data pertaining to this input are, in general, not available. Equations (4-2) and (4-5) may be termed the basic structural relations of the investment model, while equations (4-6) and (4-7) are the ones that will be actually estimated to test the implications of the model. At this point, a number of comments concerning these latter two equations are in order.

If the absolute value of the percentage rate of net disinvestment were small relative to the rate of depreciation, the last term in (4-7) could be ignored.[4] Then (4-6) and (4-7) would express the two main endogenous variables in the system as functions of four variables that are treated as exogenous within the context of this model—the wage rate, the price of medical care, the stock of human capital, and age—and one variable that is unobserved, the rate of depreciation in the initial period. If P, the price of medical care, did not vary across the relevant units of observation, the estimating equations would become

$$\ln H_i = B_W \ln W + B_E E + B_i i + U_1 \quad (4\text{-}6')$$

$$\ln M_i = B_{WM} \ln W + B_{EM} E + B_{iM} i + U_2, \quad (4\text{-}7')$$

where $B_W = \varepsilon(1 - K)$, etc., $U_1 = -\varepsilon \ln \delta_0$, and $U_2 = (1 - \varepsilon) \ln \delta_0$. The investment model predicts $B_W > 0$, $B_E > 0$, $B_i < 0$, and $B_{WM} > 0$. In addition, if $\varepsilon < 1$, $B_{EM} < 0$, and $B_{iM} > 0$.

The variables U_1 and U_2 represent disturbance terms in the reduced form equations. These terms are present because depreciation rates vary among individuals of the same age, and such variations cannot be measured empirically. Provided $\ln \delta_0$ were not correlated with the independent variables in (4-6′) and (4-7′), U_1 and U_2 would not be

[3] Equations (4-5), (4-6), and (4-7) do not contain intercepts because all variables are expressed as deviations from their respective means.

[4] Chapter II indicated that if the stock of health falls over the life cycle, then the rate of depreciation must exceed the absolute value of the rate of net disinvestment. From equation (4-6), $\tilde{H}_i = -\tilde{\delta}\varepsilon < 0$. If $\tilde{H}_i$ is small relative to δ_i, $\tilde{H}_i/\delta_i$ approaches zero.

correlated with these variables. Therefore, the equations could be estimated by ordinary least squares.

The assumption that the real-own rate of interest equals zero can be justified along the following lines. A common empirical observation is that wage rates rise with age, at least during most stages of the life cycle. If W_i were growing at a constant rate $\tilde{W}$, then $\tilde{\pi}_i = K\tilde{W}$, all i. So the assumption implies $r = K\tilde{W}$. By eliminating the real rate of interest and postulating that $\tilde{H}_i$ is small relative to δ_i, $\ln H_i$ and $\ln M_i$ are made linear functions of age. If $r - \tilde{\pi}_i$ exceeded zero but $\tilde{H}_i/\delta_i$ were small, then

$$\tilde{H}_i = -\tilde{\delta} s_i \varepsilon$$
$$\tilde{M}_i = \tilde{\delta}(1 - s_i \varepsilon)$$
$$\tilde{H}_{ii} = \tilde{M}_{ii} = -\tilde{\delta}^2 s_i(1 - s_i)\varepsilon.$$

Since the curves relating $\ln H_i$ and $\ln M_i$ to age would be concave to the origin in this situation, the square of age might be included as an additional explanatory variable. This variable should have negative coefficients in the demand curves for health and medical care.

One could change the form of the gross investment production function without altering any of the parameters in (4-6) and (4-7) except the wage elasticity of medical care. For example, suppose medical care and own time were employed in fixed proportions. Then the elasticity of substitution between these two inputs would equal zero, and the wage elasticities of health and medical care would be equal.

There are two empirical procedures for assessing whether the investment model gives a more adequate representation of people's behavior than the consumption model. In the first place, the wage rate would have a positive effect on the demand for health in the investment model as long as K were less than one. On the other hand, it would have a positive effect in the consumption model only if health were relatively goods-intensive ($K < \bar{K}$), a somewhat more restrictive requirement. So if the computed wage elasticity turns out to be positive, then the larger its value the more likely it is that the investment model is preferable to the consumption model. Of course, provided the production of health were relatively time-intensive, the wage elasticity would be negative in the consumption model. In this case, a positive and statistically significant estimate of B_W would lead to a rejection of the consumption model.

In the second place, suppose the rate of interest does not depend on wealth. Then health would have a zero wealth elasticity in the investment model. It would, however, have a positive wealth elasticity in the consumption model if it is a superior commodity. This suggests that $\ln R$, the logarithm of full wealth, should be added to the set of independent variables in the demand curves for health and medical care so that these equations would become

$$\ln H_i = B_R \ln R + B_W \ln W + B_E E + B_i i + U_1 \qquad (4\text{-}6'')$$

$$\ln M_i = B_{RM} \ln R + B_{WM} \ln W + B_{EM} E + B_{iM} i + U_2. \qquad (4\text{-}7'')$$

Computed wealth elasticities of H and M that do not differ significantly from zero would tend to support the investment model. Although the investment framework could rationalize positive wealth elasticities in terms of a negative correlation between R and r,[5] this correlation is not likely to be very large. Regardless of the size of the correlation between R and r, the wealth effect would be small if the share of depreciation in the cost of health capital were relatively large.[6]

In addition to fitting equations (4-6″) and (4-7″) to the data, the gross investment function given by equation (4-5) might also be estimated. By estimating the production function, the hypothesis that the more educated are more efficient producers of health could be tested directly. Note that the production function contains two variables for which no data exist—gross investment and the own time input. But since $\ln I_i = \ln H_i + \ln (\tilde{H}_i + \delta_i)$ and since $\tilde{H}_i/\delta_i$ has been assumed to be small, one could fit[7]

$$\ln H_i = \alpha \ln M_i + r_H E - \tilde{\delta}_i i - \ln \delta_0. \qquad (4\text{-}8)$$

The trouble with the above procedure is that it requires a good estimate of the gross investment production function. Unfortunately, equation (4-8) cannot be fitted by ordinary least squares because $\ln M_i$ and $\ln \delta_0$, the disturbance term, are bound to be correlated. It is clear from the demand curve for medical care that

$$\text{Cov}(\ln M_i, \ln \delta_0) = (1 - \varepsilon)\sigma^2 \ln \delta_0,$$

[5] In addition, the real rate of interest would have to be positive.

[6] Health would also have a positive wealth elasticity in the investment model for people who are not in the labor force. For such individuals, an increase in wealth would raise the ratio of market goods to consumption time, the marginal productivity of consumption time, and its shadow price. Hence, the monetary rate of return on an investment in health would increase. Since the empirical work in the next section is limited to members of the labor force, a pure increase in wealth would not change the shadow price of their time.

[7] If factor prices do not vary as more and more health is produced, a 1 percent increase in medical care would be accompanied by an equal percentage increase in own time. Therefore, the regression coefficient α in equation (4-7) would reflect the elasticity of gross investment or health capital with respect to both inputs. Given constant returns to scale, the true value of α should be unity.

where Cov means covariance and $\sigma^2 \ln \delta_0$ is the variance of $\ln \delta_0$. So, given $\varepsilon < 1, \ln M_i$ and $\ln \delta_0$ would be positively correlated. Since an increase in the rate of depreciation decreases the quantity of health capital demanded, the coefficient of medical care would be biased *downward*. If wealth were included in the set of exogenous variables, the production function would be "overidentified" and could be fitted by two-stage least squares.[8] Using this technique, one would first estimate the demand curve for medical care. He would then compute the predicted values of $\ln M$, which by definition are not correlated with $\ln \delta_0$. Finally, he would use these predicted values to estimate the production function.

While the two-stage least squares technique is employed in the next chapter, there are a number of difficulties with it and major reliance is placed on the calculations of the reduced form. These difficulties are discussed when the production function estimates are presented. A production function taken by itself tells nothing about producer or consumer behavior, although it does have implications for behavior, which operate on the demand curves for health and medical care. Thus, they serve to rationalize the forces at work in the reduced form and give the variables that enter the equations economic significance. Because the reduced form parameters can be used to explain consumer choices and because they can be obtained by conventional statistical techniques, their interpretation should be pushed as far as possible. Only then should one resort to a direct estimate of the production function.

2. MEASUREMENT OF HEALTH AND VARIABLES CONSIDERED

The equations formulated in Section 1 have been fitted to data contained in the 1963 health interview survey conducted by the National Opinion Research Center and the Center for Health Administration Studies of the University of Chicago. The NORC sample is an area probability sample of the civilian noninstitutionalized population in which each family had the same probability of inclusion. Data were obtained from 2,367 families containing 7,803 persons.[9]

[8] The production function is overidentified because the number of variables excluded from it (R and W) exceeds the number of endogenous variables in the system (H and M) less one by a factor of one. If wealth were not an endogenous variable, the number of excluded variables would equal the number of endogenous variables less one. In this situation, the production function would be "exactly identified" and could still be estimated by two-stage least squares.

[9] For a complete description of the sample, see Ronald Andersen and Odin W. Anderson, *A Decade of Health Services: Social Survey Trends in Use and Expenditure*, Chicago, 1967.

The stock of health, like the stock of knowledge, is a theoretical concept, one that is difficult to quantify empirically. On the other hand, the healthy time output produced by health capital can be measured in a straightforward fashion. If TL_i is time (in days) lost from market and nonmarket activities due to illness and injury, then $h_i = 365 - TL_i$. Therefore, consider the results that would be obtained if healthy time or its complement served as the dependent variable in the demand curve. The production function of healthy days given by equation (4-1) implies

$$-\ln TL_i = -\ln B + C \ln H_i. \tag{4-9}$$

Substitution of equation (4-6) for $\ln H_i$ yields[10]

$$-\ln TL_i = CB_R \ln R + C(1 - K)\varepsilon \ln W_i + Cr_H \varepsilon E - C\tilde{\delta}\varepsilon i - C\varepsilon \ln \delta_0. \tag{4-10}$$

While equation (4-6) gives a demand curve for the stock of health, equation (4-10) gives a demand curve for the flow of services yielded by health capital. The flow coefficients can be estimated by regressing the negative of the natural logarithm of sick time on the relevant set of exogenous variables. The coefficients obtained would exceed, equal, or fall short of the corresponding coefficients in the stock demand curve as C exceeds, equals, or falls short of 1. The formulation of the flow demand curve suggests that $-\ln TL_i$, and not $\ln h_i$, should be the dependent variable. Consequently, the increase in healthy time for a one unit increase in education *falls* as education increases.[11] Thus, although education has an increasing marginal product in the gross investment production function, the model still implies diminishing returns to this variable in terms of its impact on healthy time.

Two variants of sick time are available in the NORC sample. These are the number of restricted-activity days reported by persons in 1963 and the number of work-loss days. A restricted-activity day (*RAD*) is a day on which a person is kept away from his usual activities because of

[10] The derivation of (4-10) assumes that wealth is one of the independent variables in (4-6). In addition, all variables are expressed as deviations from their respective means.

[11] Let $\hat{B} = Cr_H\varepsilon$. Then from equation (4-10),

$$\frac{\partial h}{\partial E} = \hat{B} \exp(-\hat{B}E),$$

and

$$\frac{\partial^2 h}{\partial E^2} = -\hat{B}^2 \exp(-\hat{B}E) < 0.$$

illness or injury. A work-loss day (*WLD*) is a day on which a person would have gone to work but instead lost the entire day because of illness or injury. Work-loss days are, of course, relevant only for members of the labor force. For such individuals, every *WLD* is an *RAD*, but the converse is not true. This follows because an *RAD* might occur on a Saturday, Sunday, holiday, or vacation day or because an individual might go to work even though he does not feel well but might cut down on his nonmarket activities.

Although *RAD* is a more encompassing measure of sick time than *WLD*, both variables have been employed because it seems likely that the latter is a more objective concept than the former. In other words, respondents can probably recall and identify *WLD* with more precision than they can recall and identify *RAD*. To compare the results obtained with the two measures, only members of the labor force are included in the regressions computed with the NORC sample. The labor force consists of all people who reported their current status on the date the sample was taken (early 1964) as working full time, working part time, or unemployed. People who fell into one of these three categories but who failed to report the number of weeks they worked in 1963 or who said they worked no weeks in that year were excluded from the regressions.

At this point, it will be useful to discuss four methodological issues that arise when the complement of sick days, and especially when the complement of *WLD*, is used as a measure of the services yielded by the stock of health. First, part of the variation in *WLD* might simply reflect variations in weeks worked among individuals. To cite a rather extreme example, someone who worked only one week in 1963 might have reported many fewer work-loss days than someone who worked fifty weeks. If weeks worked were correlated with some of the independent variables in (4-10), the estimates of the parameters of this equation would be biased. To take account of this problem, *WLD* is adjusted for variations in weeks worked (*WW*).[12] Adjusted *WLD* is given by[13]

$$WLD1 = (52/WW)(WLD).$$

Second, sick time should not be confused with the time input in the gross investment function. In Chapter II, I pointed out that in the absence

[12] Weeks worked are defined as weeks actually worked plus paid vacation time plus work-loss weeks.

[13] The correlation coefficient between unadjusted and adjusted work-loss is .97. Regression coefficients obtained with unadjusted work-loss as the dependent variable (not shown) are very similar to the coefficients presented in the next chapter.

of variations in depreciation rates, these two variables would be *negatively* correlated. I also demonstrated that if depreciation rates did vary, the correlation might well be *positive*. But provided δ_0 were independent of the exogenous variables in (4-10), the coefficients of this equation could be interpreted in the manner that has been suggested.[14]

A third methodological issue arises if consumers face a probability distribution of depreciation rates in each period of their lives. As stated in Chapter II, individuals who insure against this uncertainty in part by acquiring market insurance would have smaller stocks of health and more sick days than those who rely solely on self-insurance. The latter group protect themselves against potential losses by holding excess stocks in relatively desirable states of the world—excess in the sense that the marginal efficiency of health capital might be extremely small and even zero in some cases. To standardize for the effects of uncertainty, a variable that indicates the presence or absence of disability insurance (insurance that finances earnings lost due to illness or injury) is included in some of the regressions run. This insurance variable might be related to observed sick time not only because of its effect on potential losses but also because of its effect on the probability that a given loss will occur. The theory of "moral hazard" suggests that if the insurance premium paid by an individual is fixed, then he may have an incentive to increase his probability of loss.[15]

Finally, some investigators have argued that the number of work-loss days reported by a person is determined almost entirely by the presence or absence of disability insurance and informal sick leave arrangements. These investigators claim that to a large extent, measured work-loss is simply one component of "leisure time." Hence, this variable is an unreliable indicator of the health status of individuals.[16]

[14] For the contrary view, see Morris Silver, "An Economic Analysis of Variations in Medical Expenses and Work-Loss Rates," in Herbert E. Klarman (ed.), *Empirical Studies in Health Economics*, Baltimore, 1970, and reprinted as Chapter 6 in Victor R. Fuchs (ed.), *Essays in the Economics of Health and Medical Care*, New York, NBER, 1972. Part of Silver's analysis is based on the assumption that sick time and the time input are identical. Suppose δ_0 were correlated with some of the independent variables that Silver and I use in our regressions. Then his view that the two types of time are equivalent would be partially correct, my view that they are different would also be partially correct, and the truth would lie somewhere between these two extremes. The empirical results in the next chapter seem to indicate, however, that my assumption concerning sick time is very plausible.

[15] See Isaac Ehrlich and Gary S. Becker, "Market Insurance, Self-Insurance and Self-Protection," *Journal of Political Economy*, 80, No. 4 (July/August 1972).

[16] For a summary of this argument, see Philip E. Enterline, "Sick Absences in Certain Western Countries," *Industrial Medicine and Surgery*, 33, No. 10 (October 1964).

The NORC data tends to contradict this view since the observed correlation between medical care and sick time is positive and very significant. The correlation between M and $WLD1$ is .356 and the correlation between M and RAD is .409. These correlations reflect the positive relationship between medical care and the depreciation rate. Apparently, this relationship is so strong that it swamps the positive effect of an increase in medical care on health (or the negative effect on sick time) that would be observed if the depreciation rate were held constant. These correlations substantiate the common-sense point of view that illness and utilization of medical services are positively associated. They also indicate that the number of work-loss days reported by the members of the sample measures their illness levels rather than a fraction of their leisure time.

The dependent variable in the demand curve for medical care is personal medical outlays. Medical expenditures include outlays on doctors, dentists, hospital care, prescribed and nonprescribed drugs, nonmedical practitioners, and other medical care—chiefly appliances like eyeglasses. Expenditures exclude health insurance premiums but contain benefits paid for by insurance.

Dollar outlays are a more desirable measure of medical care than quantity indexes for two reasons. First, the former allows one to combine the various components of medical care into an aggregate index of the utilization of this care in a simple way. Second, part of the variation in price across individuals reflects variations in the quality of services purchased instead of true differences in the price of standard units of service. If these variations in quality were ignored, the true quantity of medical services would not be accurately measured. Of course, the price of standard units of service would not be constant if, as has frequently been alleged, doctors either discriminate in price according to wealth or derive psychic benefits from treating the poor. Given this type of variation, outlays would be positively correlated with price and hence would overstate the quantity of services purchased provided the elasticity of the MEC schedule were less than unity. But note that there is a factor at work that tends to counteract the effect of price discrimination. Since the average federal income tax rate rises with income, the value of tax deductions allowed for medical expenditures is greater for wealthier individuals.

The price of medical care might also vary across consumers because of the existence of health insurance. This would be the case if an individual's premium depended not on the size of his potential loss (medical outlays) in unfavorable states of the world and on his probability of incurring the loss but on the expected outlays and expected probability of a large group of persons. In this situation, the relevant price would be cP, where c is the fraction of medical expenditures not financed by insurance. If c were correlated with the independent variables in the demand curves for health and medical care, their coefficients would be biased. These biases would be mitigated to the extent that c does not vary greatly or to the extent that premiums are not entirely fixed.

Although the stock of health is difficult to define and measure empirically, a proxy for it is available in the NORC sample. Persons in the sample were asked whether their health status was, in general, poor, fair, good, or excellent. Their response to this question is utilized as an index of the amount of health capital they possess. This measure of H suffers from the defect that it depends on an individual's *subjective* evaluation of the state of his health: what one person considers to be excellent health may be viewed as good or only fair health by another. Moreover, it is not obvious how to quantify the four possible responses. That is, one must determine exactly how much more health capital a person in, say, excellent health has compared to someone in poor health.

Nothing can be done about the subjective nature of the health status variable, but it is possible to construct a particular scaling scheme. The procedure employed is based on two propositions. First, since the units of health capital are unknown, it seems reasonable to view the four quantities assigned to health status as measures of H in index number form. Thus, if $H = 1$ for people whose health status is poor, the three other quantities would express a person's stock relative to the stock of those in poor health. Second, the observed gross correlation between medical outlays and sick time is positive. It has already been indicated that this correlation reflects the positive relationship between medical care and the depreciation rate.

Using the second proposition, assume that the gross relation between the stock of health and medical expenditures is

$$H = aM^{-b},\ a \text{ and } b > 0 \tag{4-11}$$

and let M_P, M_F, M_G, and M_E be mean outlays by people in poor, fair, good, and excellent health. Then to express the stock of health in index number form with $H_P = 1$, write $H_F/H_P = (M_P/M_F)^b$, etc. In the NORC sample, $M_P/M_F = 1.7$, $M_P/M_G = 2.3$, $M_P/M_E = 5.0$.[17] Thus, the health capital series is 1, 1.7, 2.3, and 5.0. It should be clear that a multiplicative

[17] These ratios pertain to whites in the labor force who reported positive sick time. See the end of this section and the beginning of Chapter V, Section 1, for the reasons why this group was used.

relation between H and M was selected to free the computed health series of units.[18]

Since the numerical magnitudes of health capital depend on medical expenditures, one might think at first that the coefficients of the stock demand curve would be related to those of the medical care demand curve. For example, if B_{XM} were the coefficient of variable X in the medical care demand curve, then equation (4-11) would seem to imply that the regression coefficient of $\ln H$ on X would be $-bB_{XM}$. This analysis, however, is not correct because equation (4-11) is not applied to individual observations. Instead, it is used to construct four quantities of H from data grouped by health status. Hence, the only purpose of this equation is to derive a health series in which increases in the stock of health reflect improvements in health status.[19] To emphasize this point and to show the effect of selecting different values for health status, two other sets of scales are employed in some of the regressions run in the next chapter These are 1 = poor, 2 = fair, 3 = good, 4 = excellent; and 0 = poor, 206 = fair, 290 = good, 411 = excellent. The first is an arbitrary set that has no relation to medical expenditures. The second is based on the *differences* between M_P and the outlays of those in the other three groups instead of on the ratios.

The remainder of this section discusses the independent variables that enter the demand curves for health and medical care. Age is simply given by the age of the individual. Education is the number of years of formal schooling completed by the head of the household. That is, if two or more members of a family are in the labor force, E is the same for each one. Since the head of the household is in most cases the husband, E is more accurately measured for males. Therefore, at one point in the empirical analysis, separate regressions are run for males and females. Moreover, to test the hypothesis that females might be more efficient producers of health than males (or vice versa), a sex dummy (1 = female) is included in all regressions computed with persons of both sexes.

[18] If a linear relationship were used, then $H = a - bM$, and $H_F - H_P = b(M_P - M_F)$, etc. This series would not be free of units. Note also that the dependent variable in the stock demand curve should be $\ln H_j/H_P$ ($j = F, G, E$). Therefore, the use of $\ln M_j/M_P$ as the dependent variable would generate coefficients that would exceed, equal, or fall short of the true coefficients as b exceeds, equals, or falls short of 1. But because b is a constant, the t ratios associated with these coefficients would be unaffected.

[19] For some empirical evidence that the estimated regression coefficients in the demand curve for health capital are *not* linear transformations of the regression coefficients in the demand curve for medical care, see Chapter V, Tables 1 and 3.

A difficulty arises when education is used to measure human capital. The stock of human capital possessed by a person who has completed his formal education is not constant over the remainder of his life cycle. Instead, it tends to increase at first due to on-the-job training and then decrease due to depreciation. To the extent that age is correlated with human capital, its regression coefficient might reflect forces other than the depreciation rate. But since this correlation is positive during the early stages of the life cycle and negative during the later stages, its net effect on the age coefficient is not clear.

The wage rate variable is defined as follows. Let EA be actual earnings reported by a person and N be net earnings lost due to work-loss (gross earning lost minus disability insurance payments). Then $W = (52/WW)(EA + N)$. Put differently, the relevant wage rate variable is the full-time annual equivalent of the weekly wage rate adjusted for variations in net earnings lost per work-loss week.[20]

The weekly wage can be written as

$$\frac{W}{52} = \frac{EA}{WW^*}\frac{WW^*}{WW} + \frac{N}{WLW}\frac{WLW}{WW},$$

where WW^* is the number of weeks actually worked, WLW is the number of work-loss weeks, and $WW = WW^* + WLW$. Note that if the weekly wage were not adjusted for variations in net earnings lost per work-loss week, a spurious negative relation would be created between the wage and sick time with the causality running from sick time to the wage instead of the other way around. This would occur if the wage were measured by EA/WW. In theory, the most desirable wage variable is the hourly wage rate, but unfortunately, hours of work per week were not available. To the extent that people with higher stocks of health work more hours per week than those with lower stocks, the causal relationship that goes from health to the wage has not been entirely eliminated. Note also that with all other variables held constant, a reduction in net earnings lost per work-loss week, N/WLW, would reduce $W/52$. Hence, the actual wage rate used in the regression analysis takes some account of the effect of informal sick leave arrangements and disability insurance on the value of the marginal product of health capital.[21]

[20] Since ln W is the dependent variable in all regressions, its coefficient would be the same whether or not the weekly wage rate were multiplied by 52.

[21] A more complete discussion of this point appears in the last subsection of Chapter V, Section 1. There I indicate why it would be inappropriate to use N/WLW as an index of the benefits from reducing sick time.

As in most cross-sectional studies, the NORC sample contains data on income but not wealth. Hence, the former is used as a proxy variable for the latter. Regardless of whether wealth or income measures a person's command over real resources, since a family pools its resources, family wealth or family income (Y) is the appropriate variable to enter in the demand functions.[22] This follows even though the dependent variables—H, $WLD1$, RAD, and M—pertain to a particular individual in a given family.

To free family income of transitory components associated with variations in weeks worked and net earnings lost, it must be adjusted in the manner that is employed to adjust earnings. Certain difficulties arise here because the number of wage earners is not the same in every family. Since labor force participation of married women is inversely related to husband's income, families with one wage earner might have more full income (defined as full potential family earnings plus family property income) but less reported income than families with two wage earners. To deal with this problem, four concepts of family income are utilized: $Y1$, $Y2$, $Y3$, and $Y4$. The first two income variables are partially adjusted for variations in weeks worked, the third is fully adjusted, and the fourth is unadjusted. These variables are defined as follows. Let V be family property income, n be the number of members in a given family who are in the labor force, and the subscript 1 identify the member of the family who worked the most number of weeks in 1963. Then

$$Y1 = V + (52/WW_1)(EA_1 + N_1)$$

$$Y2 = Y1 + \sum_{j=2}^{n} (EA_j + N_j)$$

$$Y3 = V + \sum_{j=1}^{n} (52/WW_j)(EA_j + N_j)$$

$$Y4 = V + \sum_{j=1}^{n} (EA_j + N_j).$$

Obviously, $Y1 = Y2 = Y3$ for families with only one wage earner. In addition, note that $Y4$ does take account of net earnings lost by all family members.

The variables just described place an upper and lower limit on potential family income. The upper limit is given by $Y3$, the lower limit

[22] As I point out later, family income per capita might be used. In any case, one would not want to simply enter the income reported by a given family member.

by $Y1$ or $Y4$, and a value that lies between these two extremes by $Y2$.[23] Hopefully, the income elasticities computed with these four variables will bracket the true parameter.[24]

The regression equations fitted to the NORC data contain family income, an individual's weekly wage rate, and his level of education as independent variables. Since Y, W, and E are all positively correlated, one might be puzzled at first by the interpretation to be given to the procedure of parceling out the separate effects of each via the multiple regression technique. Variations in property income could explain why two people with different full-time earnings have the same income. But how could two persons with the same amount of education have different wage rates when it is by this time well-accepted that an increase in education raises market productivity? Robert Michael has considered this question and has concluded that there are a variety of possible answers: "different relative degrees of labor shortage or abundance in different occupations, different degrees of monopoly power or of union strength, different innate ability, . . . different amounts of on-the-job training . . . or other forms of human capital, [and] luck."[25] Thus, it is possible at a conceptual level to raise W, Y, or E with the other two fixed. Of course, if these variables were subject to errors of measurement, their coefficients would be biased; the sources and directions of these biases are discussed in Appendix D, Section 2.

The last explanatory variable in the demand functions is family size (FS). This variable is included in the regressions for two reasons. First, the number of children in a family and the health stocks of its adult members might be complements. This is a plausible hypothesis because the lower the amount of sick time the more time there is available for childrearing activities. Second, since the dependent variables pertain to individuals and not families, per capita income might be a more appropriate measure of command over real resources than family income. This is easily accomplished if family income and family size enter the regressions.[26]

[23] Although $Y1$ exceeds $Y4$ for one wage earner families, the mean of $Y4$ exceeds that of $Y1$ in the NORC sample.

[24] Since income is used as an empirical proxy for wealth, the term income elasticity is substituted for wealth elasticity from now on.

[25] *The Effect of Education on Efficiency in Consumption*, New York, NBER, Occasional Paper 116, 1972, p. 29.

[26] If $\ln H = B_{Y/FS} \ln Y/FS + B_{FS} \ln FS$, then $\ln H = B_{Y/FS} \ln Y + (B_{FS} - B_{Y/FS}) \ln FS$. Hence, the use of family income instead of per capita income alters the coefficient of family size alone.

One final comment on the regression analysis is in order. Whites in the NORC sample reported more sick days than nonwhites, which contradicts data in the U.S. National Health Survey.[27] Since there were few nonwhites in the sample, it was felt that the data for them might be unreliable. Consequently, it was decided to restrict the analysis to whites in the labor force. The sample size of this group is 1,770.

3. GLOSSARY

α_1	Share of medical care in total cost of gross investment or elasticity of gross investment with respect to medical care
RAD	Restricted-activity days
WLD	Work-loss days
WW	Weeks worked
*WLD*1	Work-loss days adjusted for variations in weeks worked
EA	Earnings
N	Net earnings lost due to work-loss
*Y*1, *Y*2, *Y*3, *Y*4	Various measures of family income
FS	Family size

[27] See Geraldine A. Gleeson and Elijah L. White, "Disability and Medical Care Among Whites and Nonwhites in the United States," *Health, Education, and Welfare Indicators* (October 1965). The U.S. National Health Survey is a continuing probability sample that contains approximately 40,000 households. It was not used in this study because data from it are available only at a fairly aggregate level.