HUMAN CAPITAL AND THE ENGEL CURVE FOR HEALTH INSURANCE*

Douglas A. Chiriboga
Research Associate

Cleve E. Willis
Associate Professor
Department of Food and Resource Economics
University of Massachusetts, Amherst

Insurance is generally found whenever uncertainty presents itself in economic activity. Adaptation to the existence of uncertainty of treatment and incidence of illness had by 1970 produced 77 percent and 57 percent of the individuals in the United States purchasing hospital and doctor visit insurance, respectively. Corresponding figures for the rural farm population were approximately 10 percent lower in each category [5, p. 53]. The recognition that a substantial proportion of individuals have failed to protect themselves voluntarily from the financial consequences of illness has lead researchers to develop models of decision-making which explain individual reluctance to protect from adverse consequences in the event of illness.

The work by Michael [6] suggests that as the educational level of a household rises, the composition of its consumption bundle shifts in the same direction as when money income rises. Without repeating his rather lengthy development of this postulate, we simply note that the result hinges on the model suggestion that if education (assumed Hicks-neutral) increases the productivity of household production functions, with fixed money incomes, real incomes rise. He then cites two reasons for expecting the effect of education on efficiency to be positive. First, since education tends to raise productivity in the labor market, it is not unreasonable to hypothesize that it would do likewise in non-market production. Second, he argues that since technology represents the implementation of new knowledge and the level of technology influences productivity, then likewise, if education means possession of more knowledge and cultivation of a positive disposition in using new information, then education may be expected to increase household productivity.

The present purpose is to test whether this hypothesis holds in the purchase of private health insurance for households headed by individuals

*Massachusetts Experiment Station Paper No. 2114. The paper also emanates from the work of Regional Project NE-77.

under sixty-five in rural areas, based upon cross-sectional data gathered as a part of Regional Project NE-77 for the Northeast. The following section develops the empirical model, the subsequent part sets out some results, and a final section draws some conclusions.

Empirical Model

To test whether the proposition, that "as the educational level of a household rises, the composition of it's consumption bundle shifts in the same direction as when income rises", holds in the purchase of private health insurance, the shift in consumption resulting from rising income must be estimated as must the effect of rising educational levels holding income, demographic, and other environmental variables constant. This will be attempted through estimation of an Engel curve (a function which relates equilibrium quantity purchased to the level of money income) to determine the partial effects of education and income on consumption patterns (purchase of private health insurance).

The Engel function fitted is of the form:

$$(1)\quad Y = Y(X_2, X_3, X_4, X_5, X_6, X_7^*),$$

where:

Y denotes the dependent variable indicating whether the household purchased some form of private health insurance (Y = 1 if yes, 0 otherwise),

X_2 denotes the educational attainment of the household head,

X_3 represents income class for the household,

X_4 is size of the household,

X_5 is age of the household head,

X_6 is sex of the household head (1 if male, 0 otherwise), and

X_7^* is a group of several environmental variables other than education.

In our application, X_7^* contains two regional measures of health status—percentage deaths in resident population of the county to which the household belongs and infant mortality as a percent of live births in the county. The set X_7^* also contains a measure of the hygienic conditions of each household as indicated by the household's water and sewage source (municipal or private), availability of piped water and toilet and shower facilities. This last set of variables is assumed, at least in part, to characterize the health status of the household.

Ideally, the proper income measure to use in the Engel curve would be the household's permanent income; what was used, however, was income class with one denoting less than \$3,000 and nine denoting the top class of over \$25,000. Education is represented by seven different levels each containing successively more years of formal education by the household head. Ideally one would have accounted for quality of education and education of other household members. The relationship observed may be weaker than the "true relation" for these reasons. Family size was included to remove the effect it has on income—i.e., for a given level of income as family size rises, real income falls and consumption tends to shift from luxuries toward necessities. The age structure of the household should be used to account for the depreciation of education and health and additions to human capital through experience (again the head of household's age was used). The net effect of age on efficiency is not unambiguous <u>a priori</u>. Households with different educational levels may respond differently to a given change in income and conversely. Likewise, households of different sizes may respond differently to a given change in income, and given the age structure of a family, it may respond differently to a change in educational level. These three reciprocal effects are formulated by introducing an "interaction term" (a multiple of the product of two explanatory variables) for each. A dummy variable representing the sex of the household head is included to remove the effect it may have on income and education.

Empirical Results

As with the hypothesis under consideration, analysis of surveys of household samples frequently requires estimating the relationship of a dependent variable to a set of independent variables. The dependent variable may be a measure reflecting some household behavior or decision; the independent variables typically represent characteristics over which the household has less control in the short run. When the dependent variable takes on a large number of possible values, multiple regression is an appropriate statistical model. Oftentimes, what is of interest are applications of models developed to explain variation in discrete random variables, in our case variations of a dichotomous variable—the household has purchased some form of private health insurance or it has not. A rather standard approach has been to assign the dependent variable a value of one if the answer is yes and zero if no and use ordinary least squares (OLS) to estimate the unknown parameters. In most cases, OLS is not appropriate for these dichotomous situations.1/ As developed in [7], Probit analysis is an attractive estimating technique in some of these cases.

1/ For detail on the shortcomings of OLS for dichotomous dependent variables and on Probit analysis, refer to [7] and the references cited (continued)

Tables 1 and 2 report the results of OLS and Probit estimation of the marginal effects of the various independent variables on the "probability"2/ of purchase of private health insurance in rural areas in the Northeast. The first column of each table records the estimated effect ($\hat{\beta}$) for each explanatory variable and the second column depicts the ratio of each of these effects to its associated standard error estimate ($S_{\hat{\beta}}$). This term is distributed Students-t for the general linear model with continuous dependent variable and the usual assumptions regarding the stochastic disturbance. For the discrete case, this is not true—however, the ratio can serve, at least in a relative sense, to indicate significance and precision of estimation.

The proposition that as the educational level of the household rises, the composition of its consumption bundle (for the case of health insurance purchase) shifts in the same direction as when income rises appears to be supported by the empirical evidence. For both the OLS and the Probit estimations the estimated effects of income and education were both positive, and rather precisely measured (the estimated effects were four or five times as large as their standard errors). As expected, the estimated sign of the family size parameter was positive, though not precisely measured. Furthermore, those households headed by a male, all other characteristics equal, appear to have a higher probability of purchasing health insurance. This is partially explained by the fact that full-time workers are far more likely to have some private health insurance coverage than part-time workers [4, p. 37]. The remaining variables which were to give an indication of health status through regional differences in mortality and household hygienic conditions are subject to the same sorts of interpretations.

The estimated values of β in Tables 1 and 2 can also be used to "predict". That is, given values of the explanatory variables, say X', one can estimate the conditional probability of that sort of household purchasing insurance (alternatively, the proportion of households expected to purchase health insurance given those characteristics described by X').

1/ (continued)
there. For present, we merely note the following conclusions. Since the dependent variable is dichotomous, the disturbance takes one of only two possible values. If the expected value of the disturbance is zero, the disturbance is heteroscedastic and hence least squares estimators are not efficient. If the expected value of the disturbance is not zero, the least squares estimators are biased. Finally, as will be shown below, OLS can provide nonsensical estimates of "conditional probabilities" of the event to be explained which are negative or greater than one.

2/ The conditional expectation of Y given the values of the explanatory variables (X_i) and their marginal effects (β_i) can be regarded as "probabilities" since Y ranges from zero to one. For more detail, refer to [7].

Table 1

OLS Estimates

Exogenous Variables		$\hat{\beta}$	$\frac{\hat{\beta}}{S_{\hat{\beta}}}$	X'	$\hat{\beta}_i$ X'	B"	$\hat{\beta}_i$ X"
Education	(X_2)	.1278	4.8746	1	.1278	7	.8946
Income	(X_3)	.0716	5.3532	1	.0716	5	.3580
Age	(X_4)	.0054	2.9591	20	.1080	60	.3240
Family Size	(X_5)	.0111	0.9630	1	.0111	6	.0666
(Ed) (Income)	$(X_2 X_3)$	-.0116	5.4747	1	-.0116	35	-.4060
(Inc) (Family Size)	$(X_3 X_5)$	-.0029	1.4831	1	-.0029	30	-.0870
(Ed) (Age)	$(X_2 X_4)$	-.0009	1.6932	20	-.0180	420	-.3780
Sex	(X_6)	.1077	3.3565	0	0	1	.1077
% Deaths	(X_7)	-.0561	0.8737	1.25	-.0701	.7	-.0392
Infant Mortality	(X_8)	.0120	0.7968	1.20	.0144	1.5	.0180
Water	(X_9)	-.0060	0.2500	0	0	1	-.0060
Sewage	(X_{10})	.0271	1.1078	0	0	1	.0271
Piped Water	(X_{11})	.0669	0.9311	0	0	1	.0669
Toilet	(X_{12})	.2674	2.4810	0	0	1	.2674
Shower-Tub	(X_{13})	-.1139	1.1658	1	-.1139	1	-.1139
Constant	(X_{14})	-.0880	0.6192	1	-.0880	1	-.0880

Table 2

Probit Estimates

Exogenous Variables		$\hat{\beta}$	$\frac{\hat{\beta}}{S_{\hat{\beta}}}$	X'	$\hat{\beta}_i X'$	X''	$\hat{\beta}_i X''$
Education	(X_2)	.5785	3.8747	1	.5785	7	4.0495
Income	(X_3)	.3346	4.6408	1	.3346	5	1.6730
Age	(X_4)	.0235	2.4229	20	.4700	60	1.4100
Family Size	(X_5)	.0481	0.8698	1	.0481	6	.2886
(Ed) (Income)	$(X_2\ X_3)$	-.0535	4.4215	1	-.0535	35	-1.8725
(Inc) (Family Size)	$(X_3\ X_5)$	-.0155	1.5196	1	-.0155	30	- .4650
(Ed) (Age)	$(X_2\ X_4)$	-.0035	1.1667	20	-.0700	420	-1.4700
Sex	(X_6)	.4245	2.8509	0	0	1	.4245
% Deaths	(X_7)	-.4095	1.1597	1.25	-.5119	.7	- .2867
Infant Mortality	(X_8)	.0702	0.8798	1.20	.0842	1.5	.1053
Water	(X_9)	-.0548	0.4391	0	0	1	- .0548
Sewage	(X_{10})	.1840	1.4100	0	0	1	.1840
Piped Water	(X_{11})	.2456	0.7789	0	0	1	.2456
Toilet	(X_{12})	.7355	1.5429	0	0	1	.7355
Shower-Tub	(X_{13})	-.4596	0.9829	1	-.4596	1	- .4596
Constant	(X_{14})	-2.4541	3.4366	1	-2.4541	1	-2.4541

For example, X' (column 3) describes a twenty year old single female from the lowest education and income classes. The inner product of the $\hat{\beta}$ (transposed) vector and the X' vector produces an estimated "probability" of purchasing health insurance for that individual of 0.028. Following procedures summarized in [7], the Probit counterpart for X' is a very similar 0.020.

However, if one looks at an opposite extreme (X'')—a family of six with a sixty year old male head of the household with high income and educational status—OLS produces a conditional "probability" in excess of one (1.012) and a Probit estimate of 0.979. Although the two estimates differ by small amounts, the OLS estimate for X'' gives a nonsense probability larger than one, a mishap not produced by the Probit model. Finally, this is an extreme projection—for this set of data and model OLS does fairly well. This is not always the case however.

Conclusions

Traditionally, education has been thought of as an important policy variable for assuring equality of opportunity. As an outgrowth of the Human Capital theory adapted by [6] for analyzing the effect of education on efficiency of consumption, the hypothesis that "education and money income shifts the composition of a consumption bundle in the same direction" was advanced for the purchase of private health insurance. To test whether this predicted response is observed, an Engel curve was estimated from cross-sectional data. The data for the dependent variable was dichotomous and as such, estimating the parameters of the function by OLS involves a misspecification of the statistical model. Therefore, an alternate model for estimating dichotomous dependent variable functions (Probit) was also used in estimating the Engel curve. The empirical results support the hypothesis.

As a word of caution, however, the results may be somewhat overstated when consideration is given to money market imperfections, the nature of the skills associated with schooling, and genetic-family background endowments of individuals. The Human Capital theory suggests that individuals exhibiting their subjective rate of time preference, when confronting job possibilities requiring certain skills, choose the form and quantity of investment in personal development. Money market imperfections introduce substantial friction into what is presented as a straightforward choice situation. The work by [1, 2] on the nature of the skills associated with education, indicates that the income effect of schooling is reduced only slightly when cognitive skills are held constant, suggesting that the association between schooling and income is largely unrelated to differences in cognitive skills. The work by [3] implies that genetics and common environment determine nearly four-fifths of explained variance in years of education. It is with these points in mind that the efficiency effect of education on consumption in the Human

Capital tradition suggested in the work of Michael and the empirics presented in this paper must be evaluated.

References

1. Bowles, S. and H. Gintis, Capitalism and Education in U.S., 1974 draft, University of Massachusetts, published by Harper and Row, 1975.

2. Bowles, S. and H. Gintis, "The Problem with Human Capital Theory", American Economic Review, Volume LXV, No. 2, May 1975.

3. Behrman, J. and P. Taubman, "Intergenerational Transmission of Income and Wealth", American Economic Review, Volume 66, No. 2, May 1976.

4. Davis, Karen, National Health Insurance, The Brookings Institute, 1975.

5. Health U.S. 1975, DHEW Publication No. (HRA) 76-1232, National Center for Health Statistics.

6. Michael, Robert, The Effect of Education on Efficiency in Consumption, NBER Occasional Paper 116, New York, 1972.

7. Willis, C. and C. Crawford, "Methodological Issues in Analyses of Use and Adequacy of Community Services in Rural Areas: With an Application to Legal Services in the Northeast", JNAEC, 5:2:29-39, October 1976.