Volume Title: The Effect of Education on Efficiency in Consumption

Volume Author/Editor: Robert T. Michael

Volume Publisher: NBER

Volume ISBN: 0-87014-242-9

Volume URL: http://www.nber.org/books/mich72-1

Publication Date: 1972

Chapter Title: The Engel Curve

Chapter Author: Robert T. Michael

Chapter URL: http://www.nber.org/chapters/c13549

Chapter pages in book: (p. 22 - 32)

3

The Engel Curve

THE PRECEDING ANALYSIS suggests that as a household's education level rises the composition of its consumption basket will shift in the same manner as it does when money income rises. To determine whether this predicted response is in fact observed, we must identify the nature of shifts in consumption that accompany a rising income and ascertain whether they also occur with a rising level of education, holding money income stationary. This is done by estimating income-expenditure curves (Engel curves) from cross-sectional data and observing the separate, partial effects of income and education on the expenditure patterns of households. The forms and variables used in the estimating equation are discussed in the following pages.

THE FORM OF THE FUNCTION

The Engel curve fitted to the cross-sectional data is of the general form

$$X_i = f_i(Y, E, F, A, R) \qquad (3.1)$$

where X_i is the household's expenditure on the market good i, Y is the measure of the household's income level, E is its level of education, and F, A, and R are family size, age, and geographical region, respectively. (The rationale for including each variable as well as a discussion of the specific variable used in each case are the subject of the latter part of this chapter.) By the usual multiple regression techniques estimates are obtained of the partial effects of these variables on the expenditure X_i. Of principal interest from the point of view of the model is the relationship between the income and education coefficients.

The relationship (3.1) between the expenditure on the market good and the level of income and several other variables is, as in all such studies, only a partial relationship. The actual level of the expenditures is influenced by factors not directly quantifiable by the economist. What is implied by the Engel curve analysis is not that income, education, and other specified factors fully determine the expenditure pattern, but, rather, that changes in the level of income, et cetera, are closely associated with changes in expenditures on goods. The other factors that determine the level of expenditures are presumed to be unrelated in any systematic way to the explanatory variables included here. Thus, systematic differences in expenditures between households are presumed to be related to these economic variables.

It should be noted that, while the analysis applies to the income elasticity of the commodity, η_i, the elasticity obtained from fitting (3.1) is η_{X_i}, the income elasticity of expenditure on the market good. If the market good x_i is used in the production of several commodities, it can be shown that, if each unit of x_i is used exclusively in the production of one Z, then

$$\eta_{X_i} = \sum_h k_h \eta_{Z_h}, \tag{3.2}$$

i.e., the income elasticity of the market good is equal to a weighted average of the income elasticities of the commodities which use that good, where the weights are the share of the good expended on each commodity.[1] If a unit of x_i is used jointly in the production of two or more commodities, relative prices of commodities are affected by the level of production of other commodities, and the distinction between environmental variables and direct inputs breaks down.[2] This issue will not be pursued here.

It should also be noted that the estimated income elasticities are gross elasticities. The effect of changes in the value of time accompanying

[1] Say x_i is used in the production of m commodities Z_h $(h = 1, 2, \ldots, m)$, so that x_{i1} is used in producing Z_1, x_{i2}, in producing Z_2, etc., and $\sum_h x_{ih} = x_i$, then, with a constant price of x_i,

$$\eta_{X_i} = \frac{dX_i}{dY}\frac{Y}{X_i} = \sum_h \frac{dX_{ih}}{dY}\frac{Y}{X_i} = \sum_h \eta_{Z_h}\left(\frac{dX_{ih}}{dZ_h}\frac{Z_h}{X_{ih}}\right)\frac{X_{ih}}{X_i},$$

but with linear homogeneous production functions and factors changing proportionately,

$$\left(\frac{dX_{ih}}{dZ_h}\frac{Z_h}{X_{ih}}\right) = 1$$

or

$$\eta_{X_i} = \sum_h \eta_{Z_h} k_h,$$

where k_h is the share of x_i used in the production of Z_h.

[2] For one application of this joint production problem, see Michael Grossman, "The Demand for Health: A Theoretical and Empirical Investigation," NBER, forthcoming.

changes in income will be reflected in these income elasticities. By the assumptions discussed in Chapter 2, however, the nonmarket education effect on real income is assumed to be a pure income effect, hence ideally the income elasticities would be estimated with the value of time held constant. An increase in the price of time raises the relative price of time-intensive commodities, and thereby induces substitution toward goods-intensive commodities as well as toward goods-intensive methods of producing commodities (see equation A. 20). The effect on the distribution of the observed income elasticities around their mean, and in one formulation on the relationship between the education elasticity and the observed income elasticity, depends upon the correlation between the pure income elasticity and the time-intensity of the commodities. As expenditure data become available which contain independent information on the value of time, the subsequent empirical implementation of the model will involve an important additional dimension.

As in most studies of Engel curves, the proper form of the equation is not suggested by our theory. Consequently, a number of forms were considered and no one function was unambiguously preferable to all others. The regressions were run in linear form, in double-log form for various combinations of the independent variables, and with certain cross-products or interaction effects. The form on which most emphasis was placed and for which most of the empirical results are reported below is one of the double-log forms, since this form (occasionally including some interaction effects) tended to have the highest explanatory power.[3] This conclusion of a generally superior

[3] It should be pointed out that this comparison is not strictly legitimate since in the linear case the residual is $u_2 = (X - \hat{X})$ and in the log case $u_1 = (\ln X - \widehat{\ln} X)$; thus, the variations are in different units and the $\bar{R}^2$'s are not comparable. An adjustment is possible (but was not made here) by computing $\widehat{\ln} X$, taking the antilog and correlating it with X. According to Prais and Houthakker, the adjustment "seems . . . to be of small effect for broad groups of commodities in which there are no low values" of expenditure (see S. J. Prais and H. S. Houthakker, *The Analysis of Family Budgets,* Cambridge, Cambridge University Press, 1955, p. 96).

A semilog form was not used since there appears to be little reason to presume that income elasticities fall as expenditures rise. The semilog form holds $dX/d\ln Y$ fixed, implying that $(dX_i/d\ln Y)/X_i = \eta_i$ falls as X_i rises. Now it can be argued that the elasticities may fall since the number of commodities rises as income increases. This would apply if detailed expenditure items were investigated, for higher income can lead to greater diversification in expenditures. But for the broad categories of goods studied here—clothing, travel, housing, et cetera—the increased diversification would take place within the expenditure class.

fit for a double-log form is consistent with the findings of Prais and Houthakker.[4]

Since we are interested in estimating an Engel curve for several market goods and it is clear that the separate equations in the system are not independent, we should, in principle, make use of the prior knowledge about the structure of the error terms in the system to obtain more efficient estimators. Zellner's method of estimating the whole system of equations simultaneously would seem appropriate. However, despite the fact that the disturbance terms in different equations may be correlated, this procedure collapses to a simple equation-by-equation estimation method whenever the same matrix of explanatory variables is used in each equation.[5]

A further complication arises from the fact that the model developed here suggests that the system of equations is restricted by a nonlinear constraint across the equations, involving two of the coefficients in each equation. That is, from equation (2.11)

$$\epsilon_{iE} = K(\eta_i - 1), \tag{3.3}$$

where ϵ_{iE} is the elasticity of expenditure on x_i with respect to education; η_i is the elasticity of expenditure on x_i with respect to income; and K is an unknown constant across all the x_i. The value of K is the estimated value of the elasticity of consumption income, $\epsilon_{Y_C E}$. Theil develops unbiased and efficient estimators for such a system when the constraint is a linear one, and in principle the nonlinear case would be analogous to it.[6] In practice, however, the procedures used in the present study were: (1) to estimate the system, equation by equation, without imposing the constraint and then to determine the average value of the coefficient K implied by the estimates of ϵ_{iE} and η_i; and (2) to impose the constraint with an assigned value for K and, by varying the value assigned, determine that K which minimized the overall weighted residual sum of squares.

[4] After considering a linear, inverse, semilog, log-inverse, and double-log function, they conclude (ibid., p. 103), "The double-logarithmic form gives a fairly satisfactory description of the curvature found in most commodities except for the difficulty of treating zero expenditures." This latter difficulty was encountered in the data used here and was circumvented by replacing the average expenditure of zero dollars a year by an average of one dollar or one cent, as indicated.

[5] Arnold Zellner, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, June 1962.

[6] H. Theil, *Economic Forecasts and Policy*, Amsterdam, North-Holland Press, 1961.

THE VARIABLES CONSIDERED

The dependent variable in the Engel curves is the expenditure on the market good. The principal reason for using expenditures rather than quantities purchased is the same as in most cross-sectional studies—data on quantities purchased are generally unavailable. Family budget studies, in particular, are concerned with expenditures on various items and the apportionment of family income. Much of the empirical investigation in the following chapters is based on two such studies.

Aside from the practical consideration of availability, expenditures have two other useful properties. First, they enable us to aggregate the different goods into whatever composite seems appropriate. Second, to the extent that variations in price reflect varying quality of goods, the use of expenditures permits aggregation over various qualities and expresses the purchase in terms of some standard unit. Particularly since we are viewing the market goods as inputs in the production of commodities, and since changes in quality reflect changes in the number of some standard units of the input, these variations in quality should not be disregarded in our estimate of the commodity's income elasticity.

Using expenditures also involves some disadvantages, particularly if there are price differences that do not reflect quality variations. If such price variations are purely random they do not affect the consistency of the parameters estimated. On the other hand, if, say, prices are systematically lower in one geographical region than in another, then the inclusion of some variable that catches these price variations can increase the explanatory power of the equation and remove the effect from other independent variables. (As discussed later, there is evidence that prices are systematically lower in the Southern states than in the North and this is one rationale for the South–non-South region dummy which has been included.)

A second type of price variation can result from price dispersion in the market place. Under certain conditions these price differences may be correlated with the household's income and lead to biases in the estimate of the income effect. Mincer has shown that a search model suggests "lower prices are paid by the rich for 'luxury' goods, and by the poor for necessities." But the difficulty introduced by this relationship is not limited to the use of expenditures as the dependent variable. On the contrary, the biases discussed by Mincer exist when the dependent variable is the quantity purchased, while the bias in *expenditure* elasticities is greater the further the price elasticity is from unity.[7]

Turning to the independent variables, equally interesting and difficult problems are raised in considering the proper income measure to use in the Engel curve. Conceptually, the proper variable would be the household's long-run level of income. But one's permanent level of income, or standard of living, is related in a complex way to one's past income history, current income, and expectations about future income, so measured current income may not be the best available indicator of the household's long-run income position. Furthermore, it is well known that, if measured disposable income is used in an estimating equation as a proxy for permanent income, the income coefficient obtained is biased toward zero, and the bias is greater the larger the variance of transitory income relative to the variance of the permanent component. Also, an upward bias results in the coefficient of any other independent variable if that variable is positively correlated with permanent income.[8]

An alternative proxy for permanent income is the household's total consumption expenditure. This is commonly used since it is argued that the transitory component in expenditure is smaller than the transitory part of measured disposable income (as households attempt to smooth out their consumption expenditure by allowing savings to absorb much of the temporary fluctuation in income). Nevertheless, there are at least two difficulties in using this variable. First, under a certain specification of the model discussed by Liviatan, the use of total consumption expenditure as a proxy for permanent income involves a bias in the estimate of the true coefficient.[9] To circumvent this problem Liviatan suggests the use of a particular instrumental variable that can be shown to give a consistent estimate. Alternatively, he shows that a consistent estimate is obtained by grouping the data by measured income and using the average total consumption expenditure of each group as the independent variable. This is the method, used

[7] This model is presented in Jacob Mincer's "Market Prices, Opportunity Costs, and Income Effects," *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Palo Alto, Stanford University Press, 1963. The relevance of this search model in interpreting the estimated coefficients is discussed on p. 42.

[8] See E. Malinvaud, *Statistical Methods of Econometrics*, Chicago, Rand McNally and Company, 1966, p. 132; and Nissan Liviatan, "Errors in Variables and Engel Curve Analysis," *Econometrica*, July 1961, p. 359.

[9] N. Liviatan, "Errors in Variables and Engel Curve Analysis," p. 338.

by Liviatan himself and also suggested by Friedman, which has been followed in this study.[10]

A second difficulty with using total consumption as the independent variable involves expenditures for major durable goods. The household that purchases a durable good during the survey period—an automobile, a home, a major appliance—may be expected to exhibit a higher level of total consumption and a higher level of expenditure on that durable than it would otherwise; thus its total consumption could overstate its permanent income level and lead to an upward bias in the estimate of the income elasticity of durables. Since the estimates of the income elasticities are not independent, this could also lead to a downward biased estimate of some other items. The tendency to purchase durables with consumer credit would alleviate this problem somewhat. Also, Prais and Houthakker suggest that it "is to some extent reduced by the tendency for households to offset such expenditures by lower expenditures elsewhere."[11] Furthermore, it seems clear that the use of averages from grouped data should further reduce this problem since individual idiosyncrasies and exceptional purchases will be averaged out or possibly offset. In view of these considerations, total consumption expenditure of households (grouped by measured disposable income and other variables) was chosen as the income measure in most of the empirical work presented below. Some tests were made using measured disposable income and the Liviatan instrumental variable, but the results were not considered of sufficient interest to merit inclusion in this study. The consumption variable was, for one test, purged of durable expenditures and thus represented a more nearly pure current consumption expenditure or proxy for the total service flow from all purchased market goods; these results are given in Appendix C.

The second explanatory variable is education. Ideally, this variable would be a vector of the formal educational attainment of each family

[10] See N. Liviatan, *Consumption Patterns in Israel,* Jerusalem, Falk Project for Economic Research in Israel, 1964; and Milton Friedman, *A Theory of the Consumption Function,* Princeton, Princeton University Press for NBER, 1957, p. 207.

[11] They argue that given its level of total income, the household determines its expenditure on current consumption, and having made this decision proceeds to distribute this amount among the several desired goods, so that "the distribution of expenditures among the various commodities depends only on the level of total expenditure." See Prais and Houthakker, *The Analysis of Family Budgets,* p. 81. See also S. J. Prais, "A Comment," *Econometrica,* January 1959.

member, adjusted for quality and similar factors. The particular measure used here was the level of formal education of the head of the household. Presumably, there is a high correlation between the education of the head and that of other family members. However, the correlation between educational attainment of family members is not perfect and for this reason the observed relationship may be weaker than the theoretical analysis implies.

The model suggests that in order to understand household expenditure patterns one should include both permanent money income and education (or some other proxy for nonmarket productivity). But these two explanatory variables—educational attainment and, say, total consumption expenditures—are positively correlated both statistically (in one body of data, the simple correlation between $\ln C$ and $\ln E$ is $+0.59$) and in our a priori notions of what they represent, market and nonmarket productivity. Including both of these variables is intended to separate the effect of money income from the effect of nonmarket productivity.

But one might ask, "What is the intuitive sense of holding income constant and raising the education level as the multiple regression technique is intended to do, and what biases if any are introduced by this procedure?" The latter question is discussed in Appendix B. The former might be rephrased as, "Why, if two family heads have the same amount of schooling, might one household have considerably higher permanent money income?" There are a variety of possible explanations: different amounts of property income, different relative degrees of labor shortage or abundance in different occupations, different degrees of monopsony power or of union strength, different innate ability, different qualities of schooling, different amounts of on-the-job training, health or other forms of human capital, luck, and so forth. Ceteris paribus, an increase in education raises one's permanent income or one's wealth, but looking across households, that "ceteris paribus" does not hold. Permanent money income and education, though undoubtedly positively correlated, might not be highly correlated as a result of rational choices, native endowments, market conditions, or chance. The separate effects of permanent money income and nonmarket efficiency may be identified if some of these differences are operative. Appendix B explores the implications of attributing a low correlation to other differences, such as differences in ability.

The rationale for including the remaining explanatory variables—family size, age, and region—is twofold. In the first place, these variables clearly influence the household's expenditures (outlays for household durables, education, children's clothing, and sporting goods are obvious examples). So, in order to remove these influences from the income and education coefficients and to improve the explanatory power of the estimating equation, these variables deserve consideration. A more important reason for our interest in these three variables is that each may be interpreted as an additional efficiency parameter. Although the theoretical chapters of this paper were developed in terms of human capital, or education, the same basic analysis could be made for any other environmental variable which affects the productivity of the production functions.

One possible interpretation of the family-size variable is along these lines: For a group of households with the same money income, education, and age, families with more children may have less knowledge about or be less proficient in using birth control information. This may reflect a general inefficiency in acquiring and using many forms of information. By this argument, the family-size variable can be considered a proxy for the ability or inclination to obtain and make use of information, and so is negatively correlated with efficiency. The predicted direction of the effect of family size on expenditures would therefore be opposite from the predictions for education—holding money income and education fixed, an increase in family size should shift expenditures toward necessities if in fact the increase in family size reflects a decrease in efficiency.

This is not the only possible interpretation of the family-size effect in the Engel curve. One alternative is to argue that family size is an endogenous variable determined by choice, and its inclusion in the Engel curve should be given roughly the same interpretation as one would give the coefficient for, say, automobiles or any other durable good in Engel curves. That is, relative prices held fixed, one would expect positive coefficients for complements to children (or automobiles) and negative coefficients for substitutes and insignificant coefficients for all other items. Since the efficiency argument implies that luxuries will have negative family-size coefficients and necessities will have positive ones, and since this latter argument implies no correlation between η_i and the family-size coefficient, ϵ_{iF} (as long as there is no correlation between income elasticities and complementarity with children), the empirical results should distinguish between these two interpretations.

A final interpretation of the family-size variable involves the question of economies or diseconomies of scale within the household.[12]

[12] The notion of scale effects involves either shifts in the production functions of the commodities as the scale of output changes or changes in the factor prices

If increasing income and family size by the same percentage has no effect on per capita expenditures, there are no externalities of scale, and per capita expenditure and per capita income are the relevant variables for the Engel curve; if the two explanatory variables are included separately their elasticities should sum to unity in this case. This interpretation also has certain specific implications regarding the relationship of the two elasticities.[13] Whichever interpretation of the empirical findings one chooses, one may say that if expenditures shift toward necessities as family size rises, households are behaving as if their real incomes were falling; the proper explanation for this decline in real income remains an open issue.

Another efficiency parameter that has been included is the age of the head of the household. Ideally, the age of each family member should be incorporated into some age index, but in the absence of more

as the level of purchases changes. While it is likely, perhaps, that the scale effects differ for different commodities, the discussion here assumes the same scale effect for all commodities. This assumes away any relative price effects that would result from unequal scale effects.

Prais and Houthakker discuss these unequal "specific" scale effects in a somewhat different context. By estimating the overall economies of scale from a quality-income relationship, they infer from the separate scale parameters for each item the specific scale effects. See their *Analysis of Family Budgets,* Chapter 10.

[13] Define δ_i as the sum of η_i and ϵ_{iF}. Then if

$$\delta_i \gtreqless 1 \quad \text{as} \quad \eta_i \gtreqless 1,$$

per capita expenditures shift toward luxuries as incomes and family size rise proportionately, and households behave as if their real incomes rose; in this case we might say there is evidence of "economies of scale." If

$$\delta_i \lesseqgtr 1 \quad \text{as} \quad \eta_i \gtreqless 1,$$

per capita expenditures shift toward necessities, or analogously there is evidence of "diseconomies of scale." Finally, if

$$\delta_i = 1 \quad \text{as} \quad \eta_i \gtreqless 1,$$

there is no evidence of effects of scale—per capita expenditures are unaffected. These three conditions can be summarized as

$$\gamma_i = \left(\frac{-\epsilon_{iF}}{\eta_i - 1}\right) \lesseqgtr 1 \quad \text{implies} \quad \begin{array}{l}\text{economies of scale}\\ \text{no effects of scale}\\ \text{diseconomies of scale.}\end{array}$$

If the latter condition does not hold with the same inequality for all commodities, we might conciude that there is no consistent evidence of any scale effects.

specific information regarding the age structure of the household the age of the head has been used. The effect of age on productivity is not unambiguous. This may be seen by considering the effect of age on one's total stock of human capital. If we believe that knowledge is acquired by experience, age may contribute to human capital through experience—a form of on-the-job training in consumption, so to speak. Experience clearly contributes information about markets, prices, and so forth and could be considered an investment in "search" to the extent the information or experience acquired at one period is relevant at some future time.

On the other hand, casual empiricism suggests that after some point one's health deteriorates with age (we observe that age-specific death rates rise and that older persons tend to have more serious and more frequent illnesses and longer periods of convalescence). Similarly, it is often suggested that the investment in human capital in the form of schooling is subject to depreciation with age.[14] Thus, while age is positively related to one form of human capital through experience, after some point it is negatively related to the stock of human capital through the depreciation rates on that stock. The average net effect of age on efficiency (or on the stock of human capital) is not clear, a priori.[15]

Finally, the region variable (a South–non-South dummy) could reflect changes in efficiency resulting from systematic differences in climate or in the quality of educational training. Here, again, the region with the higher efficiency would be expected to exhibit a systematic shift in expenditures toward luxuries. An important alternative interpretation of the region effect exists since there appear to be systematic price differences between regions, and a shift toward luxuries could be expected in the region having the relatively lower cost of living in terms of market prices.[16]

[14] See, for example, Yoram Ben-Porath, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, August 1967.

[15] Since one's wealth is the present value of an income stream over a lifetime, changes in efficiency with age do not affect one's permanent income, only the relative prices of commodities between age intervals. Thus, the predictions of the model with respect to education's effect vis-à-vis the income effects are not strictly analogous to the predicted effects of age on expenditure patterns. Ceteris paribus, if age increases efficiency, thus reducing the prices of commodities over one's lifetime, we would expect increased consumption of *commodities* with advancing age and the effect on the expenditure for goods would depend upon the substitution elasticities between periods.

[16] See Chapter 4, footnote 5.