The Pakistan Development Review
28 : 4 Part II (Winter 1989) pp. 903–910

Latent Structure of Earnings Models†

TAYYEB SHABBIR*

1. INTRODUCTION

What determines individual earnings and its distribution in a population? The question is an important one both from an analytical as well as a public policy viewpoint. The possible answers to it have a strong bearing on issues of economic efficiency and social equity. Though social scientists have investigated the nature of income distribution and related matters for a long time, the fascination with the subject along with the list of unresolved questions has persisted.

Research regarding the determinants of earnings (or income) has enjoyed a long and venerable tradition in Economics. One such important influence has been the so-called 'human capital school'. The pioneering works of Schultz (1961); Becker (1964) and the follow up study by Mincer (1974) laid the foundations of the theory that investments in human capital such as schooling and on-the-job training enhance productivity which, in turn, leads to higher labour earnings. The basic idea of the human capital school can be expressed in terms of the following semi-log earnings function.

$$Ln\ Y = a + b\ S + c\ EXP + u$$

where $Ln\ Y$ is the natural logarithm of earnings or the wage rate, S = years of completed schooling and EXP = years of on-the-job experience.

In the above equation, the parameter b can be interpreted as the 'rate of return' to an additional year of schooling. A discussion of b's estimates is the major feature of this paper.

†Comments on this paper have not been received.

*The author is Research Economist at the Pakistan Institute of Development Economics, Islamabad.

Author's Note: This paper is a revised version of the one presented at the Meeting. The revision, however, is not expected to have affected the essential argument of the original paper which is available upon request. These papers have drawn on my recent research conducted as a Visiting Faculty Member of the Department of Economics, University of Pennsylvania, Philadelphia, U.S. This research has benefited greatly from comments by several people including Paul Taubman, Jere Behrman and Claudia Goldin. Of course, I alone bear responsibility for any shortcomings.

It is important to obtain unbiased estimates of *b*. If the above 'human capital specification' of an earnings function is the correct one, then regression estimates of *b* would be unbiased. However, in the last few years, the above specification of the earnings function has had to contend with the question of omitting certain important variables such as individual ability. Such exclusions will upwardly bias the regression estimates of the rate of return for schooling under the plausible conditions that the omitted variable is positively correlated with schooling and has a direct positive impact on earnings.

Particularly difficult problems arise if the omitted variable is an unmeasured or latent one. This paper focuses on a set of studies which have postulated such a latent variable to be familial i.e., shared by siblings in a family. If so, data on siblings can be used to obtain unbiased estimates of the rate of return to schooling. See Taubman (1977) for a good overview of the issues involved. Other relevant work includes [Behrman *et al.* (1980); Griliches (1979); Behrman and Wolfe (1984) and Shabbir (1987)].

The rest of this paper is organized as follows:

Section 2 describes the nature of ability as well as motivates its possible role in determining earnings.

Section 3 shows how the regression estimates of the rate of return to schooling may be biased if ability is not controlled for in an earnings equation. We presume ability to be a purely familial latent variable that affects both schooling and earnings.

Section 4 outlines a methodology to obtain unbiased estimates of the rate to return to schooling.

Section 5 discusses a sampling of the empirical estimates which are based on the above or closely related methodology to deal with the problem of relevant latent variables in earnings functions. Most of these estimates are for the U.S. or other developed countries. However, representative studies for the developing countries are noted as well.

Section 6 concludes this paper with some caveats and comments including the one about the implications of the above results for policies for economic development.

2. ABILITY, HUMAN CAPITAL INVESTMENTS AND EARNINGS

We interpret ability as everything that is shared by biological siblings that grow up together in the same family. Thus our measure is broader than the commonly understood notion of ability which is sometimes measured by IQ test scores. In fact, our notion of 'familial ability' may include such factors as innate ability, ambition, family background or an index of other environmental and/or genetic influences.

Ability may directly affect (presumably positively) earnings of an individual.

This sort of effect is not hard to motivate. Better environment or attitude (ambition ?) often would mean higher labour market earnings.

Besides direct effect, ability may also indirectly affects earnings by first affecting optimal schooling level of the individual. Let us elaborate this latter point by considering the following 'Ability-Augmented Human Capital Model'.

Assume schooling to be the only kind of human capital. Then a person maximizes V, the present discounted value of the lifetime earnings (as given at "birth") i.e.

$$Max\ V(S, A) = \int_S^N Y(S, A)\, e^{-it}\, dt = Y(S, A)\ (\tfrac{1}{i})\ (e^{-is} - e^{-iN}) \quad \ldots \quad (1)$$

where S is years of schooling, A is ability, i is the fixed discount rate, N is the fixed retirement date and $Y(S, A)$ is the constant level of income for a given contribution of schooling and ability.

Let us assume that foregone earnings are the only cost of schooling, there are no non-pecuniary returns to schooling, hours worked are exogenous, there is perfect competition and perfect information (or risk neutrality).

Then, in order to maximize $V(S, A)$, set $\partial V/\partial S = 0$ which gives us the following Equation:

$$[1 - e^{\{-i(N-S)\}}]\ (Y'/Y) = i \qquad \ldots \qquad \ldots \qquad \ldots \qquad (2)$$

where $[1 - e^{\{-i(N-S)\}}]$ is the finite life correction; Y'/Y or $r(S, A)$, the marginal internal rate of return with $r(S, A) \geqslant 0$, $\partial r(S, A)/\partial S < 0$ and i is the discount rate (or the market interest rate).

If we assume that $N \to \infty$ then (2), in fact, is the stopping rule which determines the optimal amount of schooling level for the kth individual, $\hat{S}_k\ (A_k)$ i.e. $r(\hat{S}_k, A_k) = i$.

Using the above framework,[1] one can now study the effects of interpersonal

[1]Incidentally, using the above model, one can quickly show how to derive the familiar semi-log earnings function mentioned in the text. In terms of our notations, we can interpret the Becker-Mincer treatment as implicitly assuming that everyone has the same A (ability) and faces the same i as well. Then, the relative supplies of labour to alternative occupations must be infinitely elastic, depending only on the financial opportunity costs of entry associated with schooling. In other words, relative wages in each occupation adjust so as to equalize the associated earnings streams everywhere i.e.

$$V(S) = Y(S)\ e^{-is}/i = V_o, \text{ for every } S$$

Letting $Y_o = iV_o$, the above implies $Y(S) = Y_o\, e^{(is)}$ or taking natural logs, $Ln\ Y(S) = ln\ Y + iS$. Since in equilibrium $i = r$, one can also express the final results as $ln\ Y(S) = ln\ Y + rS$.

differences in ability on the optimal human capital investments (and thus indirectly on earnings). It will be useful to recast our analysis of the optimal investments in the following manner which is suggested by the Woytinsky Lecture framework of Becker (1967).

Let I = dollar investment in schooling. A given individual, then, may be characterized by a set of Schooling Demand and Supply Functions given respectively as $DD = f(r, A)$ and $SS = g(i)$ where r = marginal rate of return, i = marginal interest rate or cost of financing and A = Ability.

In Figure 1, DD_1 and SS_1 pertain to a given individual (i.e. given ability). The demand curve indicates diminishing returns to investment (I). However, the cost of borrowing is constant with respect to I.[2] The intersection of DD_1 and SS_1 determines optimal level $\hat{I}_1$ where $r(\hat{I}_1, A_1) = i$. The DD curve shifts upwards and to the right for individuals with higher ability. Thus for individual 2 in Figure 1, the relevant curves are DD_2 and SS_2 ($= SS_1$, by assumption) with $\hat{I}_2$, given by $r(\hat{I}_2, A_2) = i$. Note that $\hat{I}_2 > \hat{I}_1$. Thus, the higher earnings of the more able individual reflect not only greater investment in human capital but also greater ability. Thus, if one follows the traditional Becker-Mincer type of earnings function specification which excludes ability, then, some of the returns to schooling are really "returns" to ability thus leading to an upward bias in the regression estimate of the schooling's rate of return.[3]

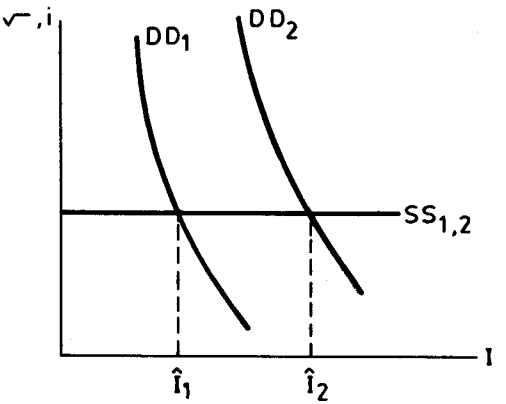

Fig. 1.

[2]In a more general case, SS could be upward sloping as well. In addition, we could introduce an SS related shift factor such as family background. Then, however, we will need to consider the case where both SS and DD may vary across individuals. However, that general analysis is still consistent with the point we are making in this paper particularly so if we assume, as e.g., Chiswick (1974) does, that there is a positive correlation between ability and family based financial cost advantages.

[3]In addition, the endogeniety of schooling issue may lead to simultaneity bias problem. For an elaboration of this issue see Rosen (1976).

3. BIAS DUE TO THE OMISSION OF ABILITY

The above discussion raises the prospect of there being 'left out' unobserved variable that may be a relevant determinant of the earnings of an individual. Consider the following earnings function:

$$Ln\ Y = a + b\,S + d\,A + u \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (3)$$

where A represents our measure of 'ability' – a latent variable which is familial. In light of the above discussion, Ability (A) can have direct effect on $ln\ Y$ (i.e. $d > 0$) as well as corr $(S, A) > 0$.

If A is not controlled for in (3), it is the standard left out variable case where the bias in $\hat{b}$, the regression estimate of b, is given as follows:

$$E(\hat{b}) - b = Cov\,(S, (dA + u))/var\,(S) \quad \ldots \quad \ldots \quad \ldots \quad (4)$$

The above implies that $\hat{b}$ will be biased as long as the $Cov\ (S, A) \neq 0$ and $d \neq 0$. (We assume S and u to be uncorrelated).

4. METHODOLOGY TO CONTROL FOR THE OMITTED LATENT VARIABLE

One possible strategy to deal with the problem of a biased regression estimate of b is to re-specify (4) in terms of the deviations for each of the appropriate variables from the corresponding family means. Then, instead of estimating the 'levels' equation $Y = a + b\,S + u$, we will be estimating the following one:

$$\Delta Y = \overline{a} + b\,\Delta S + d\,\Delta A + \Delta u \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

where $\Delta Y = Ln\ Y_{if} - Ln\ Y_{f}$, $\Delta S = S_{if} - S_{f}$, $\Delta A = 0$ and $\Delta u = u_{if} - u_{f}$

Note that the subscript '*if*' refers to the ith sibling in the fth family and the subscript f used by itself refers to the family mean for that variable. Let $\tilde{b}$ denote the OLS estimate from regressing the deviation-form Equation (5).

Then, the bias will be given by

$$E(\tilde{b}) - b = Cov\,(\Delta S, (d\Delta A + \Delta u))\ /\ Var\ \Delta S \quad \ldots \quad \ldots \quad (6)$$

The bias in $\tilde{b}$ will be zero since $\Delta A = 0$. (A is purely familial and thus shared identically by all the siblings.) We are also assuming that the usual OLS assumptions hold (in particular, no measurement error for the variables and ΔS and Δu are uncorrelated).

Thus, $\tilde{b}$ would give us the unbiased estimate of the true rate of return which can then be compared with the (upwardly) biased $\hat{b}$ that comes out of the level estimates.

5. OVERVIEW OF THE EMPIRICAL RESULTS

Griliches (1979) has a detailed comparison of schooling coefficient estimates with and without controls for the latent 'ability' of the type we represented with variable A. Additional studies have been reviewed in Behrman *et al.* (1980). These studies are mostly for the U. S. Here we just want to summarize the main point of some of the representative studies.

In terms of the magnitude of the relative upward bias in the schooling coefficient when 'ability' is not controlled for, we basically have two groups. In general, both these groups agree as to the presence of such a bias but they differ in terms of its magnitude. Thus on the one hand, studies like Taubman (1977) and Behrman *et al.* (1980) belong to the group that reports upwards bias in the 30–60 percent range. On the other hand, studies such as Chamberlain and Griliches (1977) do not find as large a bias. Their estimates range from 10–15 percent. My own estimates Shabbir (1987), obtained using sibling data for the U.S., fall in the former group's range. The above divergence in the estimates may be due to sample specific characteristics in addition to there being different model specifications across studies. Efforts to resolve the above differences have had only mixed results. In any event, while there may not be a consensus as to degree of such bias, in general, one can contend that not controlling for the omitted latent variable would bias the regression estimate of the schooling coefficient upwards.

Incidentally, as mentioned earlier, most of these sibling studies have been done only for the developed countries. However, Behrman and Wolfe (1984) is one of the few sibling studies for a developing country (991 15–45 years old Nicaraguan women and their adult sisters). Basically, their results are in line with the Behrman *et al.* (1980) ones for the U.S. (Upward bias of 33 percent in the Household Income if deviation form is used). In general, in fact, the mean family background effect on Household income is even higher than in the U.S.

6. CAVEATS/IMPLICATIONS FOR DEVELOPMENT POLICY

1. One important caveat is concerning the assumption that A is purely familial. In fact, there may be individual specific components which would require more complicated estimation techniques than those that have been suggested in this paper. Some of these techniques have been discussed in Griliches (1979) or Behrman *et al.* (1980). However, it is my contention that such relatively more complicated models that are able to ask finer questions often can do so only after

making correspondingly more heuristic assumptions.

2. As mentioned earlier, many of the sibling studies have been done only for the developed countries, in fact, mostly for the U.S. Again, till recently such studies were based only on brothers. For an example of two studies using data on sisters as well as on brothers [see Bound *et al.* (1986) and Shabbir (1987)].

3. Implication for Economic Development Policy.

To the extent that there is an upward bias in the schooling coefficient, this would imply a reduced potential for education subsidies to influence individual earnings. Increasing education opportunity has typically been thought of as an important part of any development strategy. This may still be valid. However, if the upward bias is truly in the 30–60 percent range, then the relative efficacy of increasing schooling as a policy variable to affect earnings would have to be re-evaluated. Perhaps improvements in the family environment are relatively more beneficial specially when such improvements occur early enough in an individual's life cycle.

REFERENCES

Becker, Gary S. (1964) *Human Capital.* New York: National Bureau of Economic Research (NBER).

Becker, Gary S. (1967) Human Capital and the Personal Distribution of Income. *Woytinsky Lecture No. 1.* Ann Arbor: University of Michigan.

Behrman, Jere R., Z. Hrubec, Paul Taubman and T. Wales (1980) *Socioeconomic Success.* New York: North Holland.

Behrman, Jere R., and Barbara L. Wolfe (1984) The Socioeconomic Impact of Schooling in a Developing Country. *Review of Economics and Statistics* 65 : 2.

Bound, J., Z. Griliches and B. H. Hall (1986) Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ? *International Economic Review* 27 : 1.

Chamberlain, Gary, and Zvi Griliches (1977) More on Brothers. In Paul Taubman (ed) *Kinometrics.* Amsterdam: North Holland Publishing Company.

Chiswick, Barry (1974) Income Inequality. National Bureau of Economic Research. Columbia University Press.

Griliches, Zvi (1979) Sibling Models and Data in Economics: Beginning of a Survey. *Journal of Political Economy* 87 : 5.

Mincer, Jacob (1974) Schooling, Experience and Earnings. New York: National Bureau of Economic Research.

Rosen, Sherwin (1976) Human Capital: A Survey of Empirical Research. Paper presented at the Third World Congress of the Econometric Society, Toronto.

Shabbir, Tayyeb (1987) Across and Intrahousehold Effects in a Model of Earnings

and Schooling with Controls for Latent Factors. Unpublished Ph.D. Dissertation. Philadelphia: Department of Economics, University of Pennsylvania.

Schultz, T. (1961) Investment in Human Capital. *American Economic Review* 51 : 1.

Taubman, Paul (ed) (1977) *Kinometrics: Determinants of Socioeconomic Success within and between Families.* New York: North Holland Publishing Company.