Empirical Economics (2000) 25:15–34

EMPIRICAL ECONOMICS


# Schooling choices: Preferences, discount rates, and rates of return*

**Hessel Oosterbeek, Hans van Ophem**

University of Amsterdam, Department of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands (e-mail: hessel@fee.uva.nl; hvo@fee.uva.nl)

First version received: November 1997 / Final version received: February 1999

**Abstract.** This paper develops an empirical model to identify the structural parameters of schooling preferences and human capital production. Our model distinguishes between consumption and investment motives with regard to schooling. The results show that both motives matter. Preferences for schooling vary with social background and ability. Children from poorer social backgrounds and of lower ability have a lower preference for schooling. The discount rate that enters the net value of lifetime income varies with social background as well. The marginal rate of return to schooling decreases with ability and schooling. On average the marginal rate of return is 7.3 per cent, which can be contrasted with a 'Mincerian' rate of return equal to 4.8 per cent. This indicates that the usual OLS estimate underestimates the true rate of return.

**Key words:** Human capital, schooling preferences, wages, ability

**JEL classifications:** I20, J24

## 1. Introduction

In recent years a number of papers have appeared, providing new results on the effect of schooling on earnings. Following the lead from Angrist and Krueger (1991a), many of these papers have exploited data from what is typically presented as a natural experiment. It is widely recognized that OLS estimation of the conventional Mincerian earnings equation will give biased

* We gratefully acknowledge comments by Christian Dustmann, Joop Hartog, Arthur Schram, two anonymous referees, and participants at a seminar at the Tinbergen Institute, on an earlier draft of this paper.

estimates. Commonly cited reasons for this are: omitted variable bias, endogeneity bias and measurement error. Each of these problems lead to correlation between the observed amount of schooling and the error term of the earnings equation, thereby running the OLS assumption of uncorrelated error term and regressors invalid. One way to correct for this, is to use instrumental variable techniques. Suitable instruments are variables which are sufficiently correlated with the amount of schooling, and at the same time are uncorrelated with the error term of the earnings equation. This is where natural experiments enter the picture. The idea is that there is some sort of exogenous variation which affects the educational choices of some persons, while leaving others unaffected. Moreover, these variations are such that they will not have a direct influence on earnings. Some of the recent publications have presented some real ingenious instrumental variables.

Angrist and Krueger (1991a) use the quarter of birth times year and times state as the instruments, Angrist and Krueger (1991b) use as an instrument for schooling the year of birth times a lottery number. Card (1993) argues that proximity of a college in the county of residence is a good instrument, while Harmon and Walker (1995) use quarter of birth and several year dummies that relate to increases in the compulsory schooling age. Butcher and Case (1994) find that the sex composition of a woman's siblings affect her educational attainment (women with no sisters receive on average significantly more education), and use sibling sex composition as an instrument.

Other researchers rely on more conventional instruments such as the respondent's social background. Using Finnish data, Uusitalo (1999) instruments schooling by dummy variables for father's education and father's socioeconomic status. Blackburn and Neumark (1995) use as instruments, variables related to the respondent's sibling rank, parents' education, presence of newspapers and living with both parents at the age of 14.[1]

Almost all of these studies find returns to schooling which are at least as high as (and sometimes substantially higher than) the convential OLS estimate: OLS estimates are biased downwards. This conclusion is fortified by results from related studies which adopted other approaches to account for the correlation between schooling and the error term of the earnings equation. Angrist and Newey (1991) use panel data for that purpose, while (amongst others) Ashenfelter and Krueger (1994) use data of identical twins. Partially the downward bias of OLS estimates can be attributed to measurement error in the schooling variable; but many researchers also refer to Griliches' (1977) modelling of ability bias. Another explanation, offered by Card (1999); relates to heterogeneity of returns to schooling. Most of the instruments used in IV studies arguably affect only the schooling choices of individuals at the lower end of the schooling distribution who are likely to have above average returns to schooling.

Crucial issues when applying the instrumental variable approach are quality, validity and relevance of the instruments. The more recent papers in this line of research apply a battery of statistical tests addressing these issues. To examine whether the instruments are sufficiently correlated with the amount of schooling, Bound et al (1995) have proposed that researchers routinely report the partial $R$ square and the $F$ statistic of the identifying instru-

[1] An very insightful and up-to-date survey article on the effect of education on earnings is provided in Card (1999).

ments in the first-stage estimation. To inspect the validity of the instruments in terms of having no direct effect on earnings, an overidentification test is performed.[2] Finally, to test whether IV and OLS estimates are in fact different, a familiar Hausman *t*-test is performed.

An attractive feature of this new wave of "returns to schooling" studies is that in a short time span, a substantial number of provocative studies have emerged, all dealing with the same question. And there is no doubt that the question of the size of the returns to schooling is an important one. A possible disadvantage of this new wave of research is, however, that it has drawn attention away from another fundamental question related to schooling, namely "which factors determine individual schooling choices?". And since economics is sometimes even defined as the theory of choice, this issue should concern economists in the field of schooling to a large degree.

Economists have typically studied the determinants of schooling choices within the framework of structural models. Following the seminal paper by Willis and Rosen (1979), who estimate an endogenous switching regression model to analyze college choice, different studies have varied on this theme. Instead of treating education as a dichotomous choice variables as Willis and Rosen did, Kenny et al (1979) impose a tobit structure explaining the amount of post high school education. Hartog et al (1989) use a sequential logit framework thereby treating schooling choices as a sequence of dichotomous decisions. Studies in that line of research generate estimates of the effect of schooling on earnings and vice versa, of the effect of earnings prospects on schooling decisions.

It is our impression that the IV methodology to returns to schooling crowded out the estimation of such structural models. Structural models have been critized for the assumptions that have to be made with regard to the distribution of the error terms and exclusion restrictions that are required for identification. This paper's main focus are the determinants of schooling choices and we analyze these by developing a structural model. At the same time, however, we incorporate the lessons learned from the IV literature on returns to schooling. To be more precise, we will examine whether the variables we use to identify some of the causal effects in our structural model, pass the IV test battery.

This paper develops an empirical model to identify the structural parameters of schooling preferences and human capital production. Our model is a close empirical analogue of Becker's (1967) theoretical model. The schooling choice equation is derived from a Cobb-Douglas utility function that allows for investment as well as consumption motives with regard to schooling. This contrasts with most of the existing literature that operates on the assumption that individuals maximize lifetime income instead of utility. The human capital production function is quadratic in both schooling and ability and allows for interaction between schooling and ability. The results show that consumption motives with regard to schooling are important and depend on social background and ability. Children from poorer social backgrounds and

[2] The test statistic is calculated as $n$ times the $R^2$ from a regression of the IV residuals on the excluded instruments. This statistic follows a Chi-square distribution, where the number of degrees of freedom equals the number of instruments minus the number of instrumented variables (see Davidson and McKinnon 1993, p. 236).

of lower ability have a lower preference for schooling. The discount rate that enters the net value of lifetime income varies with social background as well, and ranges from 11.6 to 21.5 per cent. The human capital production function exhibits a decreasing rate of return to schooling for a given level of ability, whereas the profiles for different ability levels diverge if schooling increases.

As a result, the marginal rate of return to schooling decreases with ability and schooling. For persons with 6 years of schooling the rate of return is 9.0 per cent whereas for persons with more than 16 years of schooling it is only 3.3 per cent. On average the marginal rate of return is 7.3 per cent, which can be contrasted with a 'Mincerian' rate of return equal to 4.8 per cent. This indicates that the usual OLS estimate underestimates the true rate of return. This conclusion is in accordance with results from the recent IV papers.

The plan of this paper is as follows. Section 2 is devoted to the structural relations between schooling and earnings and to the estimation method that we apply in this paper. Section 3 introduces our dataset and discusses the choice of variables. Section 4 starts with presenting results we obtain when following the IV approach. This serves as a benchmark and also tests the validity of our choice of identifying variables. The remainder of section 4 presents and discusses the estimation results. Section 5 contains the conclusions.

## 2. Model and estimation method

The empirical model that we develop in this section is closely related to the theoretical structure that Becker (1967) presented in his famous Woytinsky lecture. Assume that individuals aim at maximizing the level of utility that they obtain from the net present value of lifetime earnings and from the mere consumption of schooling. As noted by Schultz (1963), the consumptive value of schooling relates both to present consumption (the joy of attending school) and to future consumption (for instance enjoying reading poems). We specify the utility function as a Cobb-Douglas function, i.e.

$$U(N, s) = \ln N + \alpha \ln s, \tag{1}$$

where $U$ is the utility index, $N$ is the net present value of lifetime earnings, $s$ the amount of schooling and $\alpha$ a measure of the weight in the utility function of $s$ relative to $N$. We may refer to $\alpha$ as the "taste for schooling" or the schooling preference parameter. Note that the formulation of the objective function is more general than the usual human capital formulation in which it is assumed that individuals only aim at the maximization of the net present value of lifetime earnings. That assumes that consumption motives play no role, i.e. $\alpha = 0$.[3] One of the main objectives of the present paper is to test whether this standard approach stands up against a statistical test.

Under some familiar simplifying assumptions (an infinite time horizon and direct schooling costs equal to zero[4]), the net present value can be expressed as

---

[3] Of course, a Cobb-Douglas utility function is arbitrary. We discuss our choice of utility function below.

[4] Notice that this does not imply that schooling is completely free as the larger component of total schooling costs (forgone earnings) are included in the model.

a function of the amount of schooling (cf. Willis 1986, p. 532):

$$N(s) = w(s)e^{-rs} \int_0^{\infty} g(t)e^{-rt}\, dt, \qquad (2)$$

with $w(s)$ the initial wage rate as a function of schooling, $r$ the discount rate and $g$ the wage growth rate as a function of the amount of work experience $t$. The initial wage rate $w$ is a function of an individual's amount of human capital. The amount of human capital comes from two sources. One source is innate ability; the other is schooling. Combining innate ability and schooling produces a person's stock of human capital. The function that describes this process is the human capital production function or structural earnings equation: $\ln w = f(s, a)$, where $a$ represents innate ability. For empirical purposes, a specific functional form for this equation has to specified. A first natural assumption is that this function is concave with respect to schooling (*cf.* Willis 1986, p. 553). This assumption is based on elementary notions about the technology for the production of human capital. An additional unit of schooling adds to a person's stock of human capital (and therefore increases the log wage rate), but the more schooling the person has already received in the past, the smaller is the relative gain from an extra unit of schooling. A second natural assumption is applicable to the relation between human capital and innate ability; greater innate ability raises the log wage rate, but at a decreasing rate. Thirdly, the functional form should be able to capture the possibility of a non-zero cross-effect between schooling and innate ability. A positive cross-effect implies that the pay-off on education increases with ability of the student. Such a positive cross-effect is a necessary condition for abler students to opt for longer schooling. Theoretically, abler students may opt for shorter schooling if the higher pay-off does not compensate for higher opportunity costs of schooling borne by more able individuals (e.g. Hartog 1994, p. 10/1). Consequently, whether the cross-effect is positive or negative is an important empirical question.

An expression that captures these notions parsimoniously is a quadratic log wage-schooling-ability relation:[5]

$$\ln w = \beta_0 + \beta_1 s + \beta_2 s^2 + \beta_3 a + \beta_4 a^2 + \beta_5 sa. \qquad (3)$$

This relation – the human capital production function – describes the technology. As such this function represents the constraint which an individual with a given level of innate ability faces.[6] Note that as a wage equation this function does not deal with ability bias by simply adding a linear ability term to it. Instead, ability is added in a quadratic form, and more important, the product of schooling and ability allows for interaction effects.[7]

[5] This form can also be regarded as a second order Taylor approximation to a more complex functional form of $f(s, a)$.

[6] In his pioneering book, Mincer (1974) derives the wage equation from as an optimality condition. As Willis (1986) notes, however, this derivation either represents a human capital production function or is simply a tautology which follows from the definition of present value (p. 550/1).

[7] In our dataset we do not observe initial wages. Instead wages are observed from a cohort of individuals who are all about 43 years of age. The stocks of human capital of these persons have also been affected by work experience. We capture the wage growth due to experience by adding the usual linear and quadratic experience terms to the wage equation that we estimate.

Maximization of equation (1) with respect to schooling subject to equations (2) and (3) yields the optimal amount of schooling:[8]

$$s^* = \frac{r - \beta_1 - \beta_5 a - \sqrt{(\beta_1 + \beta_5 a - r)^2 - 8\beta_2 \alpha}}{4\beta_2}. \tag{4}$$

An individual's optimization problem can be expressed in a diagram as in Figure 1 (which is similar to the figure employed by Kodde and Ritzen 1984). The line NN pictures the net value of lifetime earnings as a function of schooling; assuming that $\beta_2 < 0$. The maximum of NN is reached at $s' = (r - \beta_1 - \beta_5 a)/2\beta_2$; which is the optimal amount of schooling if the individual aims at earnings maximization ($\alpha = 0$). If $\alpha > 0$, the indifference curves have a negative slope, and the optimum solution ($s^*$) is somewhere to the right of $s'$.

It is easy to see that if $\beta_2 < 0$, the optimum amount of schooling in the pure investment model ($\alpha = 0$), decreases with the value of the discount rate, while it increases with ability if and only if $\beta_5 > 0$. Hence, a necessary and sufficient condition for more able persons to attain higher levels of schooling than less able persons is that the human capital production functions (equation 3) for persons with different ability levels diverge.

Equation (4) is derived using a Cobb-Douglas utility function $U = \ln N + \alpha \ln s$, which can be rewritten as $U = \ln w - rs + \alpha \ln s$ (where we deleted a constant term). An alternative for the Cobb-Douglas form is the quasi-linear utility function $U = \ln N + \mu s$ (with $\mu$ a preference parameter), which can be rewritten as $U = \ln w - (r - \mu)s$. Maximizing this function subject to equation (3), gives optimum amount of schooling:

[8] Substitution of (3) in (2) and of (2) in (1) gives:

$$U = \beta_0 + \beta_1 s + \beta_2 s^2 + \beta_3 a + \beta_4 a^2 + \beta_5 sa - rs + \alpha \ln s + C,$$

where $C$ is a constant term including $\ln(r)$ and the integral term in (2). Setting the derivative with respect to $s$ equal to zero, gives:

$$\beta_1 + 2\beta_2 s + \beta_5 a - r + \frac{\alpha}{s} = 0.$$

This equality has two solutions:

$$s_{1,2} = \frac{r - \beta_1 - \beta_5 a \pm \sqrt{(\beta_1 + \beta_5 a - r)^2 - 8\beta_2 a}}{4\beta_2}. \tag{$*$}$$

When we set $\alpha$ equal to 0 in equation (1) and then maximize with respect to $s$ given equations (2) and (3) we obtain as solution for the optimum amount of schooling

$$s^* = \frac{r - \beta_1 - \beta_5 a}{2\beta_2} \tag{$**$}$$

Obviously expression ($**$) should also result from the general expression when $\alpha$ is set equal to zero in ($*$). When we do so, it turns out that we only retrieve ($**$) when the term with the square root is subtracted. If this term is added, $s$ equals zero.

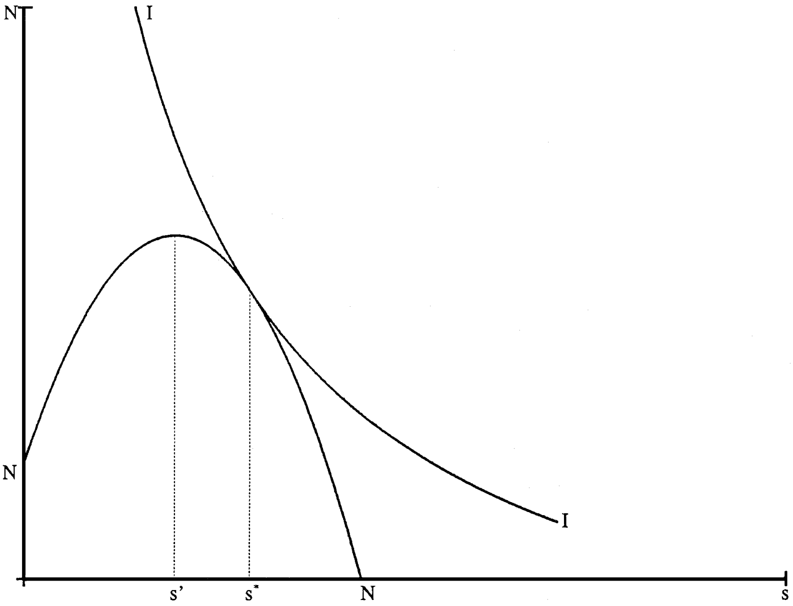


**Fig. 1.** Illustration of the model

$$s^* = \frac{(r - \mu) - \beta_1 - \beta_5 a}{2\beta_2}. \tag{4'}$$

This is similar to the expression that we obtain with the Cobb-Douglas form if $\alpha = 0$ (the pure investment model) except that $r$ is now replaced by $(r - \mu)$. Hence, interpreting the rate of return that follows from a wealth maximization framework as a rate of return corrected for the direct utility of schooling, (implicitly) assumes a quasi-linear utility function. We have three reasons to prefer the Cobb-Douglas specification over the quasi-linear form. First, with equation (4′) it is not possible to disentangle "access to funds" ($r$) and "taste for schooling" ($\mu$); not even by identifying $r$ and $\mu$ by different sets of variables, because these will still have a constant term in common. Note that it is a special feature of the quasi-linear function that $r$ and $\mu$ are lumped together, and not a special feature of the Cobb-Douglas that $r$ and $\mu$ are separated. With more involved utility functions (CES for instance), the two factors will also appear in separate terms; it is in that case, however, not possible to write the schooling equation in an explicit form. Second, the quasi-linear utility function implies that in case of extra income, all extra income will be spent on schooling. This is a rather unrealistic implication, and it is not what people do when they win a lottery, or find money on the doorsteps (cf. Shea 1996). Finally, we did estimate the model based on equation (4′), and comparing the fit of our preferred model and that model shows that the preferred model performs better in terms of absolute or squared differences between predicted and realized amounts of schooling, and predicted amounts of schooling are also for more cases closer to the realized amounts. In sum, the Cobb-Douglas function performs better and has less stringent implications than the quasi-linear function. And more important, it enables us to disentangle investment and consumption motives in schooling choices.[9]

*Estimation*

Estimation of the model set out above should take into account the relevant rules and institutions that affect the individual's choice. With regard to schooling choices, the major rule that restricts free choice is compulsory schooling up to a certain age. All persons who have an optimum below the amount that is taken at that age, are not permitted to choose their preferred solution. Assuming monotonicity of preferences and of the opportunity set, all these persons will prefer to leave school at the age that schooling is no longer compulsory. In empirical data, this produces a spike in the frequency distribution of schooling.[10] The proper way to deal with this, is to specify the schooling equation in a tobit-like way. For the individuals in the dataset that we employ for our empirical analysis, the end of compulsory schooling coincides with 6 years of primary schooling. With regard to the remainder of the schooling distribution we follow Garen (1984) in assuming that schooling is a continuous variable. Individuals can choose exactly the amount that they prefer. The description of the Dutch schooling system in Groot and Oosterbeek (1994) shows that for actual years of schooling attained, this is a fairly adequate assumption.

Estimation of the model captured by equations (3) and (4) now proceeds as follows. An individual's discount rate $r$ is unobserved. We follow Willis and Rosen (1979) by assuming that $r$ depends upon a vector of social background characteristics $Z$ (including a constant) and a parameter vector $\gamma$; hence $r = Z\gamma$. Further-more we allow for individual differences in the taste for schooling by making $\alpha$ a function of social background characteristics and of ability; $\alpha = W\delta$. Including an ability measure reflects the notion that someone's preference for schooling depends on how easy or difficult school is to him. In our choice of variables (discussed in the next section) we allot background variables that relate to the family's income to the vector $Z$, whereas variables related to attitudes towards schooling are included in $W$.

For all persons who have optimal schooling amounts above six years, actual ($s$) and optimal ($s^*$) amount are assumed to be equal. For all persons with optimal amounts equal to or less than six years, we observe actual amounts equal to six. After adding error terms to equations (3) and (4) we have:

If $s^* > 6$:

$$s = \frac{Z\gamma - \beta_1 - \beta_5 a - \sqrt{(\beta_1 + \beta_5 a - Z\gamma)^2 - 8\beta_2 W\delta}}{4\beta_2} + u_1 = h(Y) + u_1, \quad (5)$$

where $Y$ includes the terms in $W$, $Z$ and $a$.

---

[9] Card (1999) specifies the utility function $U(w, s) = \ln w - h(s)$, with $h$ an increasing convex function. Although this function seems rather general, it does not include the Cobb-Douglas function for values where $\alpha/s$ exceeds $r$, neither does it include the quasi-linear model if $\mu$ exceeds $r$. Moreover, Card does not attempt to distinguish between forgone earnings and the psychic benefits from schooling.

[10] See Table 1 in the next section.

If $s^* \leq 6$

$$s = 6, \tag{6}$$

and the common wage equation:

$$\ln w = \beta_0 + \beta_1 s + \beta_2 s^2 + \beta_3 a + \beta_4 a^2 + \beta_5 sa + \beta_6 t + \beta_7 t^2 + u_2, \tag{7}$$

where $t$ is work experience. Assuming that $u_1$ and $u_2$ follow a joint normal distribution, the logarithmic likelihood function of this model is (cf. Kenny et al. 1979):

$$\ln L = \Sigma_{s>6} \ln f(u_1, u_2) + \Sigma_{s=6} \ln \int_{-\infty}^{-h(Y)} f(u_1, u_2)\, du_1, \tag{8}$$

where $f$ is the density function of the normal. Note that even if the vectors $Z$ and $W$ have all their elements in common, $\gamma$ and $\delta$ are still identified due to the nonlinearity of the square root. To make identification not solely dependent on functional form, we also assign different elements to $Z$ and $W$.

## 3. Data and choice of variables

The dependent variables in our model are the log wage rate and the amount of schooling. Moreover, the specification of the equations in the previous section reveals that we need adequate measures of ability and social background. A dataset that contains such information is the so-called Brabant survey. We start with a brief description of the creation and special features of this dataset and then discuss the variables that will be employed in the empirical analysis.

In 1952, one quarter of the sixth-grade pupils (approx. age 12 years) in the Dutch province of Noord-Brabant were sampled. Thirty years later, the observations (on school, intelligence and family background) proved to be still available. In 1983, the same individuals were contacted to collect data on education, labor market status, earnings, etc. To obtain the 1983 information, the individual's changes of address since 1952 were traced through the services of municipal administrations of population. This is possible as city population administrations register the city of destination if somebody leaves town, and as individuals are legally compelled to register. A questionnaire was sent to all individuals for whom a valid address was found. After mailing two reminders, the remaining male non-respondents were approached by an interviewer. It was decided (for budgetary reasons) to approach only men because of their higher labor force participation rate. The response on the basis of valid addresses was 58%.

The sample has some distinct features that are worth pointing out. It has already been mentioned that all observations came originally from the same region, the province of Noord-Brabant. Earlier analysis (Hartog and Pfann 1985)[11] indicated that this does not necessarily detract from the survey's representativeness for the Netherlands as a whole, as many variables relating to

[11] This report contains a detailed description of the dataset, in Dutch.

labor force participation, industrial structure and earnings in Noord-Brabant do not differ greatly from values observed for the whole country. Another feature is that the dataset pertains to a cohort of men born in 1940 and for whom wage data have been collected in 1983. It can be argued that the dataset is somewhat outdated thereby reducing the relevance of the empirical findings. For several reasons, we think this is not the case. First, it is advantageous to have data from one single age cohort of individuals who have reached a stable labor market position. This precludes that the findings are affected by cohort or age effects and early career labor market transitions. A cohort of people in their early 40s is very suitable for this purpose. Such cohorts are not easily available. Of course it would be preferable to have access to data from a cohort of individuals being in their early 40s in 1998 rather than in 1983. It is important to note that both cohorts would have been confronted with basically the same educational structure. The main difference is that a cohort born in 1940 was confronted with a compulsory school leaving age of 12, while those born in 1955 had to stay in school until they reached age 15. From the perspective of an analysis of optimum schooling choices it is preferable to have data from a situation with the lower minimum school leaving age, because this raises the (expected) number of unrestricted observations. Finally, we believe that the kind of model we develop in this paper should be able to find out behavioral regularities that are independent of the specific period in which they are observed. Economists typically assume that the kind of preferences analyzed in our model reveal a large degree of stability.[12]

We restrict our analysis to the subsample of males. The reason for this is that we know from earlier analysis that females in this sample earn far less than males, and also that many women who work are employed in part-time jobs. In order not to mix up schooling decisions and the returns to human capital with participation decisions and possible discrimination against females, we think it better to exclude women from the analysis.[13] An additional reason is the different sampling procedures applied for males and females.

As the dependent variable in the wage equation, we have chosen the log of the net hourly wage rate. From a theoretical point of view this seems to be the best measure. The dependent variable in the schooling equation is the actual number of years of post-compulsory education. In the sample that we use, 162 out of 839 cases have zero years of post-compulsory education; for these persons we assume that their optimal level of schooling is $s^* \leq 6$.[14] Table 1 shows the distribution of actual years of schooling in our sample.

The dataset contains several variables that relate to social background. In our theoretical structure we distinguish between two effects of social background on schooling choices. The first effect is that children from poorer

[12] Evidence for this assumption in the context of schooling choices is presented in Oosterbeek and Webbink (1995), who estimate logit models for the choice to enroll in higher education in the Netherlands in 1982 and 1993. The find that changes in the enrollment rate between these two years are entirely attributable to changes in people's characteristics while behavior (parameter estimates) have remained constant.

[13] Fishe *et al.* (1981) present evidence that for females in the US choices concerning schooling and labour force participation are interdependent.

[14] Notice that it is immaterial for the results whether we measure years of schooling or years of post-compulsory schooling. In the first case those who leave school as soon as possible have 6 years of schooling, in the latter case they have 0 years of schooling. This scaling only affects the estimates of the constant terms.

**Table 1.** Counts of actual years of schooling

| years of schooling | count | years of schooling | count |
|---|---|---|---|
| 6 | 162 | 18 | 15 |
| 7 | 23 | 19 | 14 |
| 8 | 129 | 20 | 10 |
| 9 | 93 | 21 | 9 |
| 10 | 141 | 22 | 7 |
| 11 | 61 | 23 | 3 |
| 12 | 51 | 24 | 3 |
| 13 | 24 | 25 | 0 |
| 14 | 28 | 26 | 1 |
| 15 | 22 | 27 | 1 |
| 16 | 27 | 28 | 0 |
| 17 | 14 | 29 | 1 |

families have less access to financial funds and may even face liquidity constraints. Therefore they are assumed to have higher discount rates, and this reduces their propensity to invest. The second effect is that children from families with an academic tradition may have different preferences with regard to schooling than children from families that lack such a tradition, i.e. the value of $\alpha$ in the utility function depends on social background. We have to assign different social background variables to the two effects (to the vectors $Z$ and $W$). In our view the best choice here is to allot the variables that relate to family's income to the first effect ($Z$), and variables related to the attitudes towards education to the second effect ($W$).

The following variables are included into the vector $Z$. First, the occupation of the respondent's father. In the original questionnaire this variable was coded into 15 categories. In the present analysis this information is compressed into 4 classes.[15] Low level occupations are lower clerical personnel, farm laborers, industrial laborers and a few categories with a very small number of observations (disabled, retired and temporarily unemployed). High level occupations are secondary and university teachers and managers. Intermediate level occupations are made up of primary school teachers and middle-ranking employees. Self-employed fathers are a separate category. Furthermore, as an important determinant of a family's income we include the father's level of schooling. The mother's level of schooling is less important as a predictor of family income since the dataset refers to a province and a period (the 1950s) where rather traditional family values prevailed, implying that

[15] The same compression has been used before in Hartog (1988) and Hartog and Oosterbeek (1993).

most mothers worked full-time in the household and only very few of them worked outside the household.

In the vector $W$ we include a dummy variable that takes value 1 if the respondent is from an antisocial family. The score for this variable was filled in by the respondent's teacher in the 6th grade. The teacher could choose between: strongly antisocial, antisocial and normal. The first two possibilities constitute the dummy value of 1. The implicit assumption that we make is that the teacher bases his choice for an important part on the family's attitudes towards schooling. This seems reasonable since this is the main area where a teacher and parents have common interests.

The child's ability enters our model in two ways: through the human capital production function and through the taste for schooling. In our dataset we have two different measures for ability available. The first is taken from a standard IQ test. The test consisted of six subsets, relating to numbers, words, analogies and spatial orientation. The second is the average school mark in the final year of primary education. The average mark is composed of separate marks for six different fields: history, science, applied figures, reading, expression and exercises with blanks. Given the character of the two ability measures we think it natural to use IQ as the ability measure that enters the structural earnings equation and to use the average mark as the ability measure that affects the taste for schooling. IQ relates to the more general type of ability that is important for a person's performance in a job, while the average mark is closely connected to how easy or difficult school is for a person, and therefore to the joy of schooling. An advantage of the ability measures in this dataset is that they refer to scores that were measured before any further schooling could be attended. Therefore both measures are exogenous with respect to the schooling variable. This is a feature not present in many other datasets that contain ability measures.

The wage equations also include the actual amount of work experience (squared). Although all respondents were born in the same year experience is not simply the difference between present age and school-leaving age. Occurrences which cause a difference between actual and potential experience are military service, temporary disability and temporary unemployment. According to the definition of potential experience, the correlation between experience and schooling is minus one; in our dataset, the correlation is $-0.63$.

Table 2 presents the mean values and standard deviations of the variables that we use in the analysis. The figures are presented separately for those who have pursued post-compulsory education and those who have not.

We conclude this section with a brief comment on the most outstanding figures in Table 2. There are a number of differences between the two subgroups. The social background indicators show that, as expected, those who have no post-compulsory education come from lower socio-economic backgrounds than those with (some) additional schooling. In particular, the proportion coming from antisocial families is larger in the $s = 6$ group and the proportion of children with a father working in the intermediate occupation group is larger in the $s > 6$ group. Further-more the two groups differ in the average level of ability, with the average levels of the two groups almost one standard deviation apart.

The high proportion of children with self-employed fathers may seem peculiar. The reason for this is that Noord-Brabant is a rural area and in the 1950s many people were running their own farms.

**Table 2.** Descriptive statistics

| variable | $s > 6$ | $s = 6$ |
|---|---|---|
| log wage rate | 2.73 (0.52) | 2.47 (0.46) |
| years of schooling | 11.42 (3.80) | 6 |
| antisocial = 1 | 0.06 | 0.28 |
| education father (1–6) | 2.39 (0.77) | 2.20 (0.22) |
| occupation father high = 1 | 0.04 | 0.01 |
| occupation father intermediate = 1 | 0.13 | 0.03 |
| occupation father independent = 1 | 0.32 | 0.30 |
| average mark in final year | 5.85 (1.36) | 4.70 (1.28) |
| IQ | 103.9 (12.8) | 95.1 (13.4) |
| experience in years | 24.9 (4.14) | 28.1 (3.41) |
| number of cases | 677 | 162 |

## 4. Estimation results

This section starts with presenting the results from an instrumental variable approach. Then we continue with reporting and discussing the estimation results of the model outlined in Section 2. These results are presented separately for the discount rate, the schooling preference parameter and the wage equation.

### *IV results*

The discussion in the previous section motivated the choice of identifying variables, and thereby also the choice of instrumental variables. These instruments are: father's education, father's occupation, the family's social status and the average mark in final year. Estimation results are reported in Table 3.

The first column in Table 3 reports the conventional OLS earnings equation including as regressors also IQ. These results point to a return to schooling equal to 4.3 per cent. The second column shows the results from a reduced form schooling equation; only the coefficients for the excluded instruments are reported. As expected, the number of years of schooling a person attained, increases with the level of education of the father, the family's social status and the average final mark in sixth grade. Using the predicted amount of schooling as a regressor in the earnings equation reported in the third column, we find a sharp increase of the returns to schooling. The point estimate now equals 9.4 per cent, being significantly different from the 4.3 per cent estimated woth OLS. This finding is consistent with many of the recent papers applying IV and concluding that OLS estimates are biased downward.

More important from the perspective of the remainder of this paper, is that the instruments we selected, pass the IV test battery. The test statistics in the bottom half of the second column indicate that the instruments are relevant: the correlation between the excluded instruments and the amount of schooling is not too weak. The test statistics in the bottom half of the third column show that schooling is indeed endogenous (Hausman's $t$-test) and that the instruments are orthogonal to the error of the earnings equation (the over-

**Table 3.** Estimation results; OLS and IV

| | Earnings equation (OLS) | | first stage schooling equation (OLS)[a] | | second stage earnings equation (T-SLS) | |
|---|---|---|---|---|---|---|
| constant | 0.8438 | (2.17)** | 7.1166 | (3.08)*** | 0.4439 | (1.05) |
| schooling | 0.0426 | (7.40)*** | | | 0.0940 | (4.71)*** |
| education father | | | 0.7986 | (5.43)*** | | |
| antisocial | | | −1.1642 | (3.43)*** | | |
| average mark final year | | | 0.5160 | (5.37)*** | | |
| IQ | 0.0040 | (2.89)*** | | | −0.0001 | (0.08) |
| Experience | 0.0825 | (2.62)*** | | | 0.0725 | (2.24)** |
| Experience$^2$ | −0.0017 | (2.43)*** | | | −0.0009 | (1.12) |
| partial $R^2$ excluded instruments | | | 0.043 | | | |
| $F$ test for excluded instruments | | | 26.3*** | | | |
| Hausman $t$-test | | | | | 2.75** | |
| Overidentification test | | | | | 2.445 | |
| adjusted $R^2$ | 0.136 | | 0.520 | | 0.103 | |
| number of cases | 839 | | 839 | | 839 | |

***: indicates significance at the 1%-level
**: indicates significance at the 5%-level
*: indicates significance at the 10%-level
[a]: this equation also includes IQ, father's occupation dummies, experience and experience squared as regressors.

identification test). This certifies the use of these variables as identifiers in our structural model.

The results of the structural model are displayed in Table 4. This table does not include separate information about the effects of exogenous variables on the amount of schooling. This information is provided in Table 5. We discuss each of the structural equations in turn.

## *Discount rate*

The results for the discount rate show that for children of highly educated fathers the discount rate is lower than it is for children of lower educated fathers. Also if the father's occupation belongs to the intermediate level or if the father is self-employed, children have a lower discount rate than children who have fathers with lower level occupations. These results are in line with the notion that within poorer families people attach lower weights to future earnings than people in richer families do. The dummy that takes the value unity if the father's occupation is of high level is not significantly different from zero. This might be due to the fact that only 3.6 per cent of the sample falls within this category. The average discount rate equals 19 per cent; with a maximum value of 21.5 for children with the poorest social background, and 11.6 per cent for children from the richest families. These values seem quite

**Table 4.** Estimation results; Maximum likelihood

| | | |
|---|---|---|
| discount rate: $r = Z\gamma$ | | |
| constant ($\gamma_0$) | 0.2298 | (3.21)*** |
| education father ($\gamma_1$) | −0.0151 | (2.63)** |
| occupation father high level ($\gamma_2$) | −0.0142 | (1.25) |
| occupation father intermediate level ($\gamma_3$) | −0.0232 | (2.15)** |
| occupation father independent ($\gamma_4$) | −0.0141 | (1.92)* |
| school preference parameter: $\alpha = W\delta$ | | |
| constant ($\delta_0$) | 0.3226 | (1.40) |
| antisocial ($\delta_1$) | −0.4611 | (2.19)** |
| average mark final year ($\delta_2$) | 0.1433 | (2.16)** |
| earnings equation | | |
| constant ($\beta_0$) | 1.3035 | (1.57) |
| years of schooling ($\beta_1$) | 0.0261 | (0.87) |
| schooling squared ($\beta_2$) | −0.0027 | (3.32)*** |
| IQ ($\beta_3$) | −0.0020 | (0.13) |
| IQ squared ($\beta_4$) | −0.0000 | (0.41) |
| IQ*schooling ($\beta_5$) | 0.0010 | (2.89)*** |
| experience | 0.0720 | (2.23)** |
| experience squared | −0.0014 | (1.93)* |
| error structure | | |
| $\sigma_1$ | 3.8259 | (35.5)*** |
| $\sigma_2$ | 0.4861 | (37.5)*** |
| $\rho$ | 0.1711 | (2.37)*** |
| loglikelihood | −2582.95 | |
| ***: indicates significance at the 1%-level | | |
| **: indicates significance at the 5%-level | | |
| *: indicates significance at the 10%-level | | |

plausible, although for the children from the poorest families the discount rates are rather high. This suggests that children from these families have no easy access to financial funds.

### *Schooling preference parameter*

The value of the schooling preference parameter is lower for children with less educated mothers, for children from families with a low social status and for children who attained lower marks in their final year in primary education. These findings exactly fit our theoretical prediction. The average value of $\alpha$ in the sample equals 1.10 (with a *t*-value of 2.03), suggesting that on average schooling is considered as a good that generates utility. The minimum value of $\alpha$ is 0.24 and also differs significantly from zero (*t*-value 1.27), suggesting that even for the persons with low educated mothers from families with low social status and with low marks, schooling produces utility. The maximum value for $\alpha$ is 1.68 (with $t$-value 2.11), saying that for children with highly educated mothers from normal families and with high marks, the utility weight of schooling is about two and a half times the utility weight of lifetime earnings. Results from a restricted version where $\alpha(= W\delta) = 0$ is clearly rejected by the data. This supports the view that consumption motives do matter for schooling decisions.

### *Wage equation*

Except for the control variables related to experience, the only two variables in the wage equation that have coefficients that differ significantly from zero, are schooling squared and the interaction term of schooling and ability. For a given level of ability the positive value of $\beta_5$ and the negative value of $\beta_2$ imply that the wage-schooling profile exhibits decreasing returns to schooling. For an individual with IQ equal to 100, the rate of return to schooling ($\partial \ln w/\partial s$) equals 10.0 per cent at the end of compulsory education and equals 6.8 per cent after six years of post-compulsory education. The wage-schooling profiles for persons with different ability levels diverge. This is exactly the condition necessary for more able persons to opt for higher levels of schooling than less able persons. This implies that the rate of return to schooling at a given amount of schooling is higher for more able persons. For instance, an individual with IQ equal to 120, has a rate of return equal to 12.0 per cent at the end of compulsory education, and a rate of return equal to 8.8 per cent with six additional years of schooling. The average marginal rate of return at realized amount of schooling and measured ability equals 7.3 per cent. The finding that the average marginal rate of return falls short of the average discount rate, is consistent with the fact that the average schooling preference parameter is positive. Relative to the schooling amount that is optimal from the perspective of pure investment model (with $\alpha = 0$), people invest too much in schooling.

### *Schooling choices*

Table 5 presents the effects of the exogenous variables on the number of schooling years. The effects are evaluated at mean values of IQ and the social background variables.

**Table 5.** Effects of exogenous variables on years of schooling; evaluated at mean values

| exogenous variable ($x$) | $\Delta s$ |
|---|---|
| education father (from 3 to 4) | +0.99 |
| occupation father high level (from 0 to 1) | +0.93 |
| occupation father intermediate level (from 0 to 1) | +1.56 |
| occupation father independent (from 0 to 1) | +0.92 |
| antisocial (from 0 to 1) | −3.08 |
| average mark (from mean value to mean value + 1 s.d.) | +0.88 |
| IQ (from mean value to mean value + 1 s.d.) | +1.17 |

**Table 6.** Average schooling preference parameter, discount rate and marginal rate of return by years of schooling

| years of schooling | $\alpha$ | $r$ | $\partial \ln w/\partial s$ |
|---|---|---|---|
| 6 | 0.869 | 0.191 | 0.090 |
| 7 | 1.000 | 0.190 | 0.083 |
| 8 | 1.003 | 0.193 | 0.082 |
| 9 | 1.068 | 0.191 | 0.080 |
| 10 | 1.130 | 0.189 | 0.075 |
| 11–12 | 1.148 | 0.185 | 0.070 |
| 13–16 | 1.261 | 0.179 | 0.056 |
| $>16$ | 1.285 | 0.169 | 0.033 |

An increase of the father's level of schooling from 3 to 4 (on a scale from 1 to 6), increases the child's optimum amount of schooling with a year. Having a father with a lower level occupation instead of an intermediate level occupation (for instance industrial laborer instead of middle-ranking employee) lowers the optimum amount of schooling by 1.56 years. According to our model these variables affect the schooling decision through the discount rate. Coming from an antisocial instead of a normal family lowers the amount of schooling by 3.08 years. An increase of the average school mark with one standard deviation increases the amount of schooling by 0.88 years. These effects are related to the schooling preference parameter ($\alpha$). Finally, we find that an increase of IQ from the mean to one standard deviation above the mean, increases the schooling optimum by 1.17 years. This latter effect is due to different shapes of the human capital production function for persons with different ability levels.

Another way to display the relation between schooling choices and the underlying structural parameters is included in Table 6. In that table we present the average values of the schooling preference parameter ($\alpha$), the discount rate ($r$) and the marginal returns to schooling ($\partial \ln w/\partial s$) by years of schooling. The patterns in that table are almost monotonic; the average schooling preference parameter rises from 0.87 to 1.29 if years of schooling increase from 6 years to over 16 years. Also, the discount rate decreases from an average value of 19.1 per cent for those with no post-compulsory schooling to 16.9 per cent for the highest schooling class. Finally the average value of the marginal rate of return decreases from 9.0 per cent for the lowest educated to 3.3 per cent for the highly educated.

## *Selection bias and ability bias*

An important topic in recent human capital literature is to what extent the usual Mincerian wage equation produces biased estimates for the rate of return to schooling due to selection bias and ability bias. Selection bias refers to the fact that the amount of schooling included in the wage equation is not exogenous. Our estimation results indicate that this is indeed the fact; the residuals of the wage and schooling equation reveal a positive correlation

**Table 7.** Estimation results; Restricted wage equations

| | Mincerian wage equation | | $\rho = 0$ | | no ability measure | |
|---|---|---|---|---|---|---|
| constant ($\beta_0$) | 1.4809 | (4.07)** | 1.5687 | (1.91)* | 0.8528 | (2.24)** |
| schooling ($\beta_1$) | 0.0483 | (8.87)** | 0.0027 | (0.09) | 0.0952 | (4.11)*** |
| schooling$^2$ ($\beta_2$) | | | −0.0026 | (3.20)*** | −0.0010 | (1.31) |
| IQ ($\beta_3$) | | | 0.0038 | (0.24) | | |
| IQ$^2$ ($\beta_4$) | | | −0.0000 | (0.15) | | |
| IQ*schooling ($\beta_5$) | | | 0.0010 | (2.67)*** | | |
| experience | 0.0828 | (2.62)** | 0.0675 | (2.09)** | 0.0772 | (2.34)** |
| experience$^2$ | −0.0017 | (2.43)** | −0.0011 | (1.85)* | −0.0015 | (2.07)** |
| $\sigma_2$ | | | 0.4808 | (40.9)*** | 0.4919 | (38.5)*** |
| $\rho$ | | | | | 0.1913 | (3.29)*** |

***: indicates significance at the 1%-level
**: indicates significance at the 5%-level
*: indicates significance at the 10%-level

($\rho = 0.1711$). This means that respondents who have attained more schooling than predicted on the basis of their known characteristics, also have higher wages than predicted.

The estimated wage equation also shows that remedying ability bias by just adding an ability measure to the set of regressors, is not correct. This approach at best measures the ‘average’ ability bias, while the estimated wage equation in Table 3 shows that the bias varies with years of schooling and level of ability.

In Table 7 we present results for the familiar Mincerian wage equation along with restricted versions of our model. These restricted versions relate to estimates that take no account of selection bias and to estimates that take no account of ability bias. The first column contains Mincer’s loglinear wage schooling specification and is estimated with OLS. According to these results and the usual interpretation, the rate of return to schooling is equal to 4.8 percent. This finding deviates considerably from the results presented above. It indicates that the rate of return obtained from the standard wage equation underestimates the rate of return found with a more sophisticated model. In this respect the findings in this paper reiterate other recent findings. In the second column we have estimated a wage equation that has a similar specification as the wage equation in Table 4, but where $\rho$ is restricted to be zero, in the third column we re-estimated our full model, but omitted the ability variables. The results in the second column indicate that ignoring selection bias does not harm the estimates to a large extent; the coefficients of the wage equation are hardly affected by this restriction. Ignoring ability differences on the other, provides us with a completely different picture. In that case it seems as if the rate of return is constant over schooling levels and equals 9.5 per cent.

## 5. Conclusion

In this paper we formulated and estimated a structural model of schooling choices and the human capital production function. The schooling choice equation is derived from a utility function that allows for investment and consumption motives with regard to schooling. The human capital production function is quadratic in both schooling and ability and allows for interaction between schooling years and ability. With structural models identification is a crucial issue. Here we achieve identification of the structural parameters by using particular functional forms and by choosing identifying (instrumental) variables. With regard to our choice of a Cobb-Douglas utility function we presented evidence that this form does a better job than the quasi-linear specification implicitely used in models in which individuals are assumed to maximize wealth rather than utility. With regard to the exclusion restrictions we imposed for some variables, we showed that our choices pass the recently developed IV test battery.

The estimation results show that the importance of consumption motives with regard to schooling depend on the mother's level of education, the family's social status and average school marks. For all persons in the sample the schooling preference parameter is positive; ranging from 0.24 to 1.68 as the relative weight of schooling to lifetime income. The discount rate that enters the net value of lifetime income is individual-specific and depends on father's level of schooling and his occupational level. The human capital production function exhibits a decreasing rate of return to schooling for a given level of ability, whereas the profiles for different ability levels diverge if schooling increases.

The average value of the marginal rate of return to schooling in our sample is 7.3 percent. The actual value for an individual varies with the amount of schooling and his IQ. For someone with IQ equal to 120 and only 6 years of schooling the rate of return equals 12.0 per cent, whereas someone with IQ equal to 100 and 12 years of schooling faces a rate of return of 6.8 per cent. These results deviate from our estimated Mincerian rate of return which equals 4.8 per cent. This difference indicates that the usual OLS estimate of the rate of return underestimates the true rate of return. This finding is in accordance with the recent findings by, among others, Ashenfelter and Krueger (1994) and Harmon and Walker (1993).

From the point of view of educational policy, our findings have the following implication. Equal participation to education by children with different social backgrounds, is not only hindered by the fact that, due to different discount rates, their opportunity sets are different from those with better social backgrounds, but also by the fact that their attitudes towards schooling are different. These differences in preferences are already apparent when children are 12 years of age. This suggests that early intervention is warranted.

## References

Angrist JD, Krueger AB (1991a) Does compulsory schooling affect schooling and earnings? Quarterly Journal of Economics 106:979–1014

Angrist JD, Krueger AB (1991b) Estimating the payoff to schooling using the Vietnam-era draft lottery, Working Paper #290, Industrial Relations Section, Princeton University

Angrist JD, Newey WK (1991) Over-identification tests in earnings functions with fixed effects. Journal of Business and Economic Statistics 9:317–323

Ashenfelter O, Krueger AB (1994) Estimates of the economic return to schooling from a new sample of twins. American Economic Review 84:1157–1173

Becker GS (1967) Human capital and the personal distribution of income. University of Michigan Press, Ann Arbor
Blackburn ML, Neumark D (1995) Are OLS estimates of the return to schooling biased downward? Another look. Review of Economics and Statistics 77:217–230
Bound J, Jaeger DA, Baker RM (1995) Problems with instrumetal variables estimation when the correlation between the instruments and the enogenous explanatory variable is weak. Journal of the American Statistical Association 90:443–450
Butcher KF, Case A (1994) The effect of sibling sex composition on women's education and earnings. Quarterly Journal of Economics 109:531–563
Card D (1993) Using geographic variation in college proximity to estimate the return to schooling. Working paper 4483, NBER
Card D (1994) Earnings, schooling, and ability revisited. Working paper #331, Industrial Relations Section, Princeton University
Card D (1999) The causal effect of education on earnings. In: Ashenfelter O, Card D (eds.) Handbook of Labor Economics, Elseviers Science Publishers, forthcoming
Davidson R, MacKinnon JG (1993) Estimation and inference in econometrics. Oxford University Press, New York/Oxford
Fishe RPH, Trost PR, Lurie P (1981) Labor force earnings and college choice of young women: an examination of selectivity bias and comparative advantage. Economics of Education Review 1:169–191
Garen J (1984) The returns to schooling: a selectivity bias approach with a continuous choice variable. Econometrica 52:1199–1218
Griliches Z (1977) Estimating the returns to schooling: some econometric problems. Econometrica 45:1–22
Groot W, Oosterbeek H (1994) Earnings effects of different components of schooling; human capital versus screening. Review of Economics and Statistics 76:317–321
Harmon C, Walker I (1995) Estimates of the economic return to schooling for the UK. American Economic Review 85:1278–1286
Hartog J (1988) An ordered response model for allocation and earnings. Kyklos 41:113–141
Hartog J (1994) On human capital and individual abilities. Discussion Paper TI 94–151, Tinbergen Institute
Hartog J, Oosterbeek H (1993) Public and private sector wages in the Netherlands. European Economic Review 37:97–114
Hartog J, Pfann GA (1985) Vervolgonderzoek Noordbrabantse zesdeklassers. University of Amsterdam, Amsterdam
Hartog J, Pfann GA, Ridder G (1989) (Non-)graduation and the earnings function: an inquiry on self-selection. European Economic Review 33:1373–1395
Kenny LW, Lee LF, Maddala GS, Trost RP (1979) Return to college education: an investigation of self-selection bias based on the Project Talent Data. International Economic Review 20:775–789
Kodde DA, Ritzen JJM (1984) Integrating consumption and investment motives in a neoclassical model of demand for education. Kyklos 37:598–606
Lang K (1993) Ability bias, discount rate bias and the returns to education. Boston University, mimeo
Mincer J (1974) Schooling, experience, and earnings. Columbia University Press, New York
Oosterbeek H, Webbink HD (1995) Enrolment in higher education in the Netherlands. De Economist 143:367–380
Schultz TW (1963) The economic value of education. Columbia University Press, New York
Shea J (1997) Does parents money matter? NBER working paper 6026
Uusitalo R (1999) Returns to education in Finland. Labour Economics, forthcoming
Willis RJ (1986) Wage determinants: a survey and reinterpretation of human capital earnings functions. In: Ashenfelter O, Layard R (eds.) Handbook of labor economics, Elseviers Science Publishers
Willis RJ, Rosen S (1979) Education and self-selection. Journal of Political Economy 87:S7–S36