Oxford Economic Papers 46 (1994), 658–675

COSTS AND REVENUES OF INVESTMENT IN ENTERPRISE-RELATED SCHOOLING

By WIM GROOT*, JOOP HARTOG†, *and* HESSEL OOSTERBEEK†

* *Department of Economics, Leiden University, P.O. Box 9521 RA Leiden, The Netherlands*

† *Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*

1. Introduction

EMPIRICAL research on returns to schooling is usually concerned with returns to pre-work education or formal education only. Research on the returns to enterprise-related schooling and worker training has been initiated only recently. Mincer (1988) compares the year-to-year wage growth of workers and firms with job training with the wage growth in firms and periods without training. He finds that the wage growth of a year's training is approximately 4.4%. Barron *et al.* (1989) and Brown (1989) find that job training for new employees who have had no previous job training yields wage growth of 7.5–9%. Holzer (1988) finds a smaller effect: 4.4%. Lynch (1991) concludes that the wages of young male workers who received one year of company training grow by 11%, while the wages of those who did not receive training grow by only 4%. Lillard and Tan (1986) find for older cohorts that job training initially leads to wage growth of 10.8%, but that this effect gradually diminishes. Booth (1991) estimates wage equations for male and female workers, in which she includes a dummy for the incidence of job-related training and a variable for the number of training days for the past two years. She finds that the estimated impact of training incidence is to increase earnings for men (women) by 11.2% (16.3%).[1]

Based on a sample of these findings Mincer (1989) calculates that—with a rate of depreciation of company training of 4% (12%)—the average rate of return on job training for different sub-groups ranges from 8.7% (18.5%) to 26% (37.5%). Similarly, Vaughan (1989) (referred to in OECD 1991) in a summary of training costs and average rates of return for men, finds that the average return ranges from 4% to just over 25% depending on the data set used, the sub-group of males workers examined, and the depreciation rate chosen.

In OECD (1991) some of the research on enterprise-related training is surveyed. This survey concludes that enterprise training has a positive impact on wages, 'though the extent varies by type of training, and the degree to which

[1] As noted by Booth, the training variables are treated as exogenous in the wage equations. Because of self-selection bias and simultaneity bias such training estimates may be an over- or under-estimate of the true training effect.

the relation reflects only productivity increases associated with training investments is not resolved' (OECD 1991).

With respect to the amount of the investment in enterprise-related schooling, it is frequently found that more highly educated workers, male workers, and workers with a lower propensity to quit participate more in enterprise-related schooling (ERS in the sequel) than female workers with lower education, and workers with shorter expected tenures (see OECD 1991 for a review). Why is it that some workers participate more in ERS than others? In general two answers can be given to this question. The first is that for some workers the revenues from investing in ERS are higher than for others. For example, the expected revenues increase with the expected tenure of the job. The second explanation is that for some workers the costs of schooling are lower than for others. For example, it is sometimes argued that higher educated workers have capabilities which make them more trainable, and that therefore they have to incur lower costs to acquire a given amount of human capital. In a simple regression type of analysis in which the amount of ERS is regressed on variables like education, gender, and expected tenure it is impossible to determine whether differences in the amount of training are due to differences in the costs of training or differences in the (expected) revenues. To address the question why some workers participate more in ERS than others we, therefore, have to formulate a structural model.

In Becker's (1975) theoretical framework of investments in general and specific human capital both post-schooling earnings and investments in schooling are endogenous. Post-schooling wage rates are determined by the amount of investment in human capital; investments in human capital are determined by expected post-schooling wages. Almost an infinite number of studies have calculated the rate of return to investments in general education. In view of this large amount of research it is somewhat remarkable that only a few studies have acknowledged this simultaneity between wages and investment in schooling. In most studies on the rate of return to formal education, the amount of the investment in human capital is taken to be exogenous.[2] Only a few studies have taken account of the fact that the investment in education is endogenous (see Kenny *et al.* 1979; Garen 1984; and Oosterbeek 1992). However, these studies usually do not account for the full simultaneity between the investment in human capital and wages. An exception is the study by Willis and Rosen (1979). However, in this study the investment in post-compulsory schooling is treated as a discrete binary decision variable instead of a continuous variable. This implies that the amount of the investment in schooling is known in advance; individuals only have to decide whether or not to invest in schooling.[3]

In the analysis of the returns to enterprise-related schooling the assumption of a fixed investment is untenable, as there is usually no general curriculum for

[2] The most notorious example is calculating the rate of return to education by specifying a wage equation with years of formal education as an exogenous variable.

[3] The same holds for the model for the analysis of training programs presented in Björklund and Moffitt (1987).

all workers and enterprise-related courses differ in length. Some workers invest in enterprise-related schooling and some do not, and for those who do there is a large variance in the amount of the investment.

In this paper, we apply a structural human capital model, where wages and training are estimated simultaneously. Our model aims at exploiting a rather special dataset, in which not only the incidence of training is known but also the length and intensity of the training period. Another special feature is the availability of IQ-scores measured when the individual was about 12 years old, so we can better control for ability than is usually the case in training models. A shortcoming of the data is lack of information on the age at which the individual participated in the training program. Also, the individuals were all about the same age when interviewed (about 43), but this is a less serious restriction, as the position at age 43 may give a fair indication of the lifetime opportunities of an individual.

The outline of this paper is as follows. In Section 2 the investment in schooling-cum-wages model is described. We define enterprise-related schooling as all schooling organized by the firm, and provided either by the firm itself or by some outside institution hired by the firm. Section 3 contains the econometric specification. In Section 4 we describe how the conditional marginal and total revenues and the conditional marginal costs can be calculated from the parameter estimates. The data are described in Section 5. The estimation results are in Section 6. The last section contains the conclusions.

2. Optimal investment

The starting point of the analysis is the specification of the wage equation. For those who have invested in enterprise-related schooling, wages are dependent on human capital and personal characteristics as well as on the amount of the investment in enterprise-related schooling. Enterprise-related schooling may not only have a direct wage effect, it may also have a wage augmenting effect by enhancing the productivity of other human capital variables. Let X_1 and X_2 be vectors of human capital variables and personal characteristics, S the actual amount of ERS, and β and γ associated parameter vectors. We assume that the wage rate is a linear function of personal chacteristics (including human capital variables), the amount of the investment in ERS, ERS squared, and the interaction between personal characteristics and ERS

$$w_1 = \beta_0 + X'_1\beta_1 + X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 \quad (1)$$

Thus, if the worker does not participate in ERS, the wage equation becomes

$$w_0 = \beta_0 + X'_1\beta_1 \quad (2)$$

The difference between w_1 and w_0 gives the per hour total revenue (TR) of the investment in ERS. However, optimality conditions are derived from marginal rather than total considerations. The marginal revue (MR) of the investment is found by differentiating with respect to S, which gives the per hour marginal revenue, times the number of working hours that remain after the investment period is finished, times an appropriate discount rate. We assume that the product of the number of hours and the discount rate equals the constant δ.[4] Hence, we assume that the total discounted marginal revenues are proportional to the marginal revenues per hour of work

$$\text{MR} = \delta\, \partial w_1/\partial S = \delta(\gamma_1 + X'_2\beta_2 + \gamma_2 S) \tag{3}$$

The marginal revenues are linear in the amount of the investment in enterprise schooling. The coefficient of the investment variable S, γ_2, gives the slope of the marginal revenue curve, while γ_1 gives the intercept of the MR curve. The intercept of the marginal revenue curve is further determined by personal characteristics X_2 with associated coefficients β_2.

The marginal costs (MC) of the investment in ERS are assumed to be a linear function of a constant term α_0, the amount of the investment S with associated coefficient α_1 and of personal characteristics Y with associated coefficients α_2. The MC equation is

$$\text{MC} = \alpha_0 + \alpha_1 S + Y'\alpha_2 \tag{4}$$

The worker invests in ERS until the marginal costs equal the marginal revenues, i.e. the optimal amount of ERS (S^*) is found by equalizing MC and MR and solving for S

$$\text{MR} = \text{MC}$$

$$\delta(\gamma_1 + X'_2\beta_2 + \gamma_2 S) = \alpha_0 + \alpha_1 S + Y'\alpha_2 \tag{5}$$

If we solve eq. (5) for S we get an expression for the optimal amount of investment in ERS (S^*)

$$S^* = (\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 Y'\alpha_2)/(\alpha_1 - \delta\gamma_2) \tag{6}$$

For the net revenues to be maximum at this point the second order condition requires $\delta\gamma_2 - \alpha_1 < 0$.

Our assumption that the observed amount of training is equal to the amount where the worker equalizes his marginal costs and marginal revenues deserves some discussion. First, we note that sufficient conditions of this assumption to hold are:

(i) the worker, and not the firm, decides upon the amount of ERS;
(ii) the worker faces no constraint regarding for instance the availability of training slots;
(iii) the worker is risk neutral.

If condition (i) is not fulfilled because firm and worker jointly decide on the ERS amount, it suffices if the firm's share in the marginal costs equals its share

[4] Remember that all individuals in the sample are the same age. Hence, if they all would retire at the same age and have identical discount rates, this would be an innocent assumption. The problem of differences in time elapsed since training is discussed below.

in the marginal revenues. According to Hashimoto (1981) this condition reflects long-run competitive equilibrium which requires that the present value of the returns exactly equals costs.

An important objection against condition (ii) is that involuntary unemployment is a clear manifestation of rationing of training slots. Remember however, that our dataset was collected in 1983 among workers employed in 1983 who entered the labour force in the 1960s. This was a decade in which unemployment in the Netherlands was very low. Of course, there may also be rationing for employed workers. In the absence of any empirical information, we can only assume that workers can always obtain the training that they optimally want to take. It would certainly be valuable to search for information on which the assumption can be tested. We return to the issue in the concluding section.

Finally, given the relatively small amounts of investment involved in ERS programs, the condition of risk neutrality seems not too severe.

3. Econometric specification

We observe the wage rate of workers who have participated in ERS ($S > 0$), the wage rate of workers who have not ($S = 0$), and the amount of the investment in ERS. For the amount of ERS we assume that we observe $S^* = S$ if $S^* > 0$, and $S = 0$ if $S^* \leqslant 0$. From the wage equations and the equation for the optimal amount of the investment we can identify the marginal cost and marginal revenues of ERS.

For estimation purposes we add random disturbances to the model; μ for the training equation and ε for the wage equation. To allow for non-captured differences between workers who have and workers who have not been in training, we specify different errors for wages of trained and untrained workers.

If $S^* > 0$

$$w_1 = \beta_0 + X'_1\beta_1 + X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 + \varepsilon_1 \tag{7}$$

and

$$S = S^* = (\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 - Y'\alpha_2)/(\alpha_1 - \delta\gamma_2) + \mu \tag{8}$$

If $S^* \leqslant 0$

$$w_0 = \beta_0 + X'_1\beta_1 + \varepsilon_0 \tag{9}$$

$$S = 0$$

Let $f(\varepsilon_1, \mu, \rho_1)$ be the bivariate density function ε_1 and μ, i.e. the joint density of the errors terms of the optimal training equation and the post-training wage equation. Similarly, let $f(\varepsilon_0, \mu, \rho_0)$ be the bivariate density function of ε_0 and μ, the error terms of the investment equation and the non-training wage equation. The correlation coefficient of ε_1 and μ, and ε_0 and μ are denoted by ρ_1 and ρ_0 respectively. The likelihood function of this model is (cf. Kenny *et al.* 1979)

$$L = \Pi_{S>0} f(\varepsilon_1, \mu, \rho_1) \Pi_{S=0} \int_{-\infty}^{-J} f(\varepsilon_0, \mu, \rho_0)\, \mathrm{d}\mu \tag{10}$$

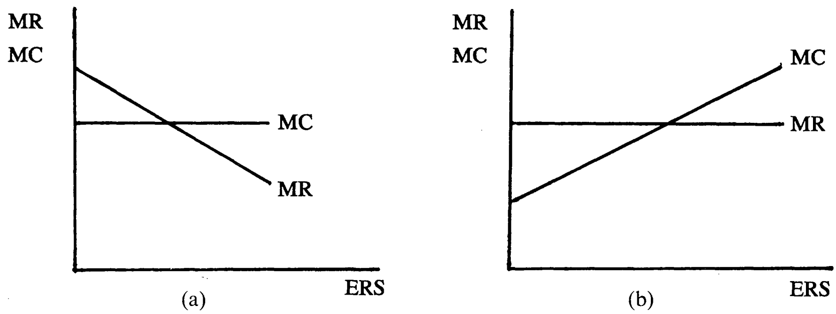

FIG. 1. Optimal investment. (a) $\alpha_1 = 0$; (b) $\delta\gamma_2 = 0$

This joint distribution is characterized by

$$E(\varepsilon_1)^2 = \sigma_1^2, \quad E(\mu)^2 = \sigma_2^2, \quad E(\varepsilon_0)^2 = \sigma_0^2, \quad E(\varepsilon_1\mu) = \rho_1\sigma_1\sigma_2, \quad E(\varepsilon_0\mu) = \rho_0\sigma_0\sigma_2$$

and

$$E(\varepsilon_1\varepsilon_0) = 0$$

$$J = \{(\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 - Y'\alpha_2)/(\alpha_1 - \delta\gamma_2)\}/\sigma_2$$

As both the slope of the MR equation, $\delta\gamma_2$, and the slope of the MC equation, α_1, have to be estimated from the amount of investment in ERS, we can not identify both α_1 and $\delta\gamma_2$. To solve this under-identification problem we have to impose restrictions on either α_1 or $\delta\gamma_2$. We present two sets of estimation results. In the first α_1 is set equal to zero. This amounts to assuming that the MC curve is constant (see Fig. 1a). In the other $\delta\gamma_2$ is set to zero, implying that the MR curve is horizontal (see Fig. 1b). The figures illustrate the inferences about MC and MR. In Fig. 1a, α_1 is set equal to zero, which implies by the second order condition that the MR-curve has a negative slope. The slope ($\delta\gamma_2$) and intercept ($\delta[\gamma_1 + X_2\beta_2]$) of the MR-curve are (up to the factor δ) derived from the estimation of the wage equations. Given the MR-curve and the observed amount of ERS, which is assumed to be optimal, we can infer the individual-specific constant marginal cost ($\alpha_0 + Y\alpha_2$). Similarly, in Fig. 1b, the slope of the MR-curve is assumed to be zero, the intercept is $\delta[\gamma_1 + X_2\beta_2]$. Given the observation of the optimal amount of ERS, a point on the positively sloped MC-curve is determined. Admittedly, we deal with the identification problem in a very crude way. But by cutting the Gordian knot, we can evaluate both ends of the rope and see what they imply.

A drawback of the model is that, instead of the more conventional Mincerian semi-logarithmic wage equation, the wage equation is specified in wage levels. The reasons for this is that we need levels of marginal costs and marginal benefits, not relative marginal costs and benefits. Starting from a log-specification, it is awkward to transform back to levels of costs and benefits to determine optimal investment levels.

4. The costs and benefits of enterprise-related schooling

The total wage gain of ERS can be calculated by the difference between the expected wage rate with ERS and the expected value of the wage rate without ERS. Because participants and non-participants differ in both their observed and unobserved characteristics the average wage gains for both groups need not be equal; on the contrary, it seems likely that those who participated earned higher gains than those who did not participate would have gained had they participated. For ease of notation, let $X = \{X_1, X_2\}$, $\alpha = \{\alpha_0, \alpha_1, \alpha_2\}$, $\beta = \{\beta_0, \beta_1, \beta_2\}$, and $\gamma = \{\gamma_1, \gamma_2\}$. First, we discuss the wage gain for participants. Let $E(w_1 \mid S > 0, X, \gamma, \beta)$ be the expected wage rate after a year of ERS for a participant in ERS with characteristics X. $E(w_0 \mid S > 0, X, \gamma, \beta)$ is the expected wage rate a participant in ERS would have received if he/she had not invested in ERS. The total wage gain to investment in ERS for participants in ERS is calculated by

$$E(w_1 \mid S > 0, X, \gamma, \beta) - E(w_0 \mid S > 0, X, \gamma, \beta)$$
$$= X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 + (\rho_1\sigma_1 - \rho_0\sigma_0)\lambda_1 \quad (11)$$

where

$$\lambda_1 = \varphi(u)/(1 - \Phi(u)) \quad \text{and} \quad u = [-(\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 - Y'\alpha_2)/(\alpha_1 - \delta\gamma_2)]/\sigma_2$$

The last term in eq. (11), $(\rho_1\sigma_1 - \rho_0\sigma_0)\lambda$, is the self-selection effect.

In a similar way we can determine the wage gain non-participants would have received, had they participated in ERS. Let $E(w_1 \mid S = 0, X, \gamma, \beta)$ be the expected value of the wage rate non-participants ($S = 0$) would have received if they had participated in ERS, and let $E(w_0 \mid S = 0, X, \gamma, \beta)$ be the expected wage rate a non-participant receives without ERS. The expected average wage gain of ERS for non-participants is now

$$E(w_1 \mid S = 0, X, \gamma, \beta) - E(w_0 \mid S = 0, X, \gamma, \beta)$$
$$\times X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 - (\rho_1\sigma_1 - \rho_0\sigma_0)\lambda_2 \quad (12)$$

where $\lambda_2 = \varphi(u)/\Phi(u)$.

Next we derive expressions for the marginal revenues for participants and non-participants. From the presentation of the model in Section 2 it follows that expected marginal revenues are found by differentiating the wage equation with respect to schooling. Let $E(\mathrm{MR} \mid S > 0, X, \gamma, \beta)$ be the marginal revenue of ERS for participants. The expression for $E(\mathrm{MR} \mid S > 0, X, \gamma, \beta)$ is found by differentiating $E(w \mid S > 0, X, \gamma, \beta)$ with respect to S. The expected wage rate after participation in ERS for participants is given by

$$E(w_1 \mid S > 0, X, \gamma, \beta) = \beta_0 + X'_1\beta_1 + X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 + \rho_1\sigma_1\lambda_1 \quad (13)$$

Recall that

$$\lambda_1 = \varphi(u)/(1 - \Phi(u)) \quad \text{and} \quad u = [-(\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 - Y'\alpha_2)/(\alpha_1 - \delta\gamma_2)]/\sigma_2$$

Differentiating eq. (13) with respect to S also involves differentiating λ_1 with respect to u. If u increases the individual becomes more likely to participate in ERS.[5] The expected marginal revenues of ERS for participants in ERS is found by

$$\begin{aligned} E(\mathrm{MR} \mid S > 0, X, \gamma, \beta) &= \delta\, \partial E(w_1 \mid S > 0, X, \gamma, \beta)/\partial S \\ &= \delta[\gamma_1 + X'_2\beta_2 + \gamma_2 S - \rho_1\sigma_1(\lambda_2^2 - \lambda_1 u)] \end{aligned} \quad (14)$$

In general the sign of $\lambda_1^2 - \lambda_1 u$ is ambiguous. However, for participants the marginal revenues will be greater than the marginal costs. So for participants $u < 0$ and the sample selection effect $\lambda_1^2 - \lambda_1 u$ is positive. With negative selection, i.e. if $\rho_1 < 0$, the interpretation of this effect is that the marginal revenues for participants are higher than the marginal revenues for an average individual in the sample. With positive selection ($\rho_1 > 0$) the reverse holds. For non-participants the marginal revenues will be less than the marginal costs and $u > 0$, For non-participants the sign of the selection effect cannot be determined *a priori*.

In a similar way the marginal revenues of participation for non-participants can be calculated. The expected wage rate upon participation for non participants is

$$E(w_1 \mid S = 0, X, \gamma, \beta) = \beta_0 + X'_1\beta_1 + X'_2\beta_2 S + \gamma_1 S + 0.5\gamma_2 S^2 - \rho_0\sigma_0\lambda_2 \quad (15)$$

Differentiating with respect to S yields the marginal revenue for a non-participant

$$\begin{aligned} E(\mathrm{MR} \mid S = 0, X, \gamma, \beta) &= \delta\, \partial E(w_1 \mid S = 0, X, \gamma, \beta)/\partial S \\ &= \delta[\gamma_1 + X'_2\beta_2 + \gamma_2 S + \rho_0\sigma_0(\lambda_2^2 + \lambda_2 u)] \end{aligned} \quad (16)$$

Again, for non-participants $u > 0$ as the marginal costs exceed the marginal revenues. Then $(\lambda_2^2 + \lambda_2 u) > 0$. With negative selection ($\rho_0 < 0$) this implies that for a non-participant the marginal revenues of the first unit of schooling are lower than average. Again, the reverse holds if the marginal revenues are greater than the marginal costs or if $\rho_0 > 0$.

Expressions for the marginal costs of the investment can also be derived for participants and non-participants separately. Let $E(\mathrm{MC} \mid S > 0, Y, \alpha, \beta)$ be the marginal costs of ERS for participants. The marginal costs for a participant are calculated by

$$E(\mathrm{MC} \mid S > 0, Y, \alpha, \beta) = \alpha_0 + \alpha_1 S + Y'\alpha_2 + \rho_1\sigma_1\lambda_1 \quad (17)$$

For a non-participant the marginal costs are

$$E(\mathrm{MC} \mid S = 0, Y, \alpha, \beta) = \alpha_0 + \alpha_1 S + Y'\alpha_2 - \rho_0\sigma_0\lambda_2 \quad (18)$$

[5] As shown by Dolton and Makepeace (1987) $\mathrm{d}\lambda_1/\mathrm{d}u = (\lambda_1^2 - \lambda_1 u)$. Also $\partial u/\partial[(\delta\gamma_1 + \delta X'_2\beta_2 - \alpha_0 - Y'\alpha_2)/(\alpha_1 - \delta\gamma_2)]/\sigma_2 = -1$.

5. The data

The data are taken from the Brabant survey 1983. This data set contains information on 2,587 individuals who were in the sixth grade of primary school in the Dutch province of Brabant in 1952. These individuals were interviewed in 1952 and 1983. The 1952 records include information on IQ and social background. The 1983-questionnaire includes questions about post-primary schooling careers, ERS, present job status, earnings, etc.[6]

From this data set we have taken a sub-sample of wage-earners. This reduces the sample to 1,559 observations. After further elimination of observations with missing values on essential variables (wages and participation in ERS), 1,057 observations could be used in the analysis. So, the data are a cohort of employees who are approximately 43 years old in 1983.

The amount of ERS is defined by the answers on two survey questions. For a maximum of three enterprise-related courses it was asked:

(i) how long did the enterprise-related course take (in weeks);
(ii) how large a part of the week was spent on the course.

The amount of the investment was found by multiplying (i) and (ii). The total investment in ERS was calculated by summing over the three courses and is measured in months. ERS is defined as organized by the company and accessible to employees of that company. These courses can either be followed within the company itself or at some outside training institution. Also, they can either be organized for employees of a specific company or for employees of other companies as well.

Unfortunately it cannot be deduced from the data how long ago the individual received the training. We consider this the most serious shortcoming of our dataset. The marginal cost equation is best thought of as reflecting the psychic cost of training, or the required monetary compensation for training. While this no doubt may vary with age and experience, and hence depend on the timing of training, we don't feel too uncomfortable about not knowing this timing. We are far less at ease with the effect on the marginal returns. Essentially, we have assumed that the wage mark-up from training does not depend on age or experience. Therefore, we have not accounted for possible depreciation of ERS. If the value of human capital has depreciated, the observed marginal revenues underestimate the true marginal revenues of ERS. Neither can we account for variations in the discount factor δ due to differences in remaining working life. Hence, at best we can think of the results as an average effect for different timing of training.

The variables that we use refer to experience, education, IQ, gender, and job level. IQ is a test score at age 12. Job level is a grading of jobs (by job analysts) by required level of mental ability, from 1 ('very simple labour') to 7 ('work on scientific basis'). There is no obvious way of assigning the variables to X_1,

[6] Details of the survey can be found in Hartog (1989), and more extensively in Hartog (1992).

TABLE 1
Sample averages (standard deviations)

	Total sample	*Participants*	*Non-participants*
No. of observations	1,057	268	789
Participate in ERS	0.25	1	0
ERS in months	2.15 (8.42)	8.47 (15.05)	0
Net wage rate (guilders per hour)	16.77 (18.41)	17.61 (11.29)	16.48 (20.27)
Years of education after primary school	4.61 (3.79)	4.37 (2.96)	4.69 (4.03)
Years of experience	25.28 (4.42)	26.11 (3.60)	25.00 (4.63)
IQ	103.38 (13.38)	105.35 (12.52)	102.70 (13.84)
Female	0.18	0.05	0.23
Education drop-out	0.49	0.63	0.45
Education low	0.16	0.16	0.17
Education intermediate	0.15	0.12	0.16
Education high	0.05	0.01	0.06
General education	0.61	0.57	0.62
Job level	4.60 (1.69)	4.82 (1.44)	4.52 (1.77)

X_2, and Y. For X_1 (the linear part of wages) we chose experience, experience squared, IQ, gender, job level, and years of education. We allowed training effects to interact (in X_2) with IQ, gender, and years of education. Y, determining marginal cost of training, includes IQ, education by type (five levels and general education[7] as opposed to vocational education) and gender. Gender and IQ appear in all three sets (X_1, X_2, and Y). Education also appears in each set, but is sometimes measured in years, sometimes in interval dummies, to facilitate identification. In terms of variables rather than specification, we only excluded job level from marginal training cost and the general education dummy from the wage equations.

Table 1 contains some sample characteristics. For the entire sample the following conclusions can be drawn from the descriptive statistics:

(i) About a quarter of the workers have invested in within-company training.
(ii) The average amount of ERS for the total sample is 2.15 months. If we only take participants the average amount is 8.47 months.
(iii) An average worker in our sample has about four-and-a-half years of post-compulsory education. About 60% have a general education.
(iv) Almost 82% of the workers in the sample are male.

From a comparison of participants and non-participants the following conclusions can be drawn:

[7] The educational dummies are education dropout, education low, education intermediate, education high, and general education. The omitted category is primary education only.

(i) On average, participants earn about 1.13 guilders or 11% more per hour than non-participants.
(ii) Participants have invested less in formal education than non-participants. Workers with general education are under-represented among the participants.
(iii) The average IQ of participants is higher than for non-participants.
(iv) Women are under-represented among the participants.
(v) On average, participants have a higher job level than non-participants.

As mentioned before, the data contain no information on the training, so we do not know how long ago training actually took place. And with individuals of the same age, there is no variable that could help us predict at what age training takes place for different individuals. Hence we can only attempt an interpretation in terms of an average age at which training takes place. In a proper discounting framework, marginal cost, and benefits should be evaluated at the same moment, i.e. that of the actual decision. If we assume that both marginal cost and marginal benefit are independent of age and experience, we may claim that our estimated functions are a proper indication of cost and benefits at the moment of decision. The discount factor δ is then equal to the discount factor $\int_D^{T-D} e^{-rt}\, dt$, where D is the time of decision and T is the time of retirement, used to calculate the present value of a benefit over the rest of the working life. Now, indeed, we do not know at what time the individual actually made his decision (his D), but as a very crude approximation we might say that individuals with 25 years of experience on average made their decision 12.5 years ago, and with present participation rates of the older workers, on average would have another 25 years to go. Hence, the discount factor would compound the marginal gain over 37.5 years. With an interest rate of 5% this would imply a δ of about 17. If either marginal cost or marginal benefit are age/experience dependent and individuals make their investments at different points in their career, the model cannot be estimated without modelling such differences of timing.

6. Estimation results

The appendices present the results of some OLS wage equations and of a Tobit equation on the duration of ERS. These results can serve as a benchmark for the estimation results of the structural model. This comparison will also help to ascertain the weaknesses of our data. In the wage equation for participants in ERS we included the duration of ERS, the duration of ERS squared, and some interaction terms. These variables correspond to the variables in the marginal revenue equation of the structural model. An interesting finding in the wage equation for participants in ERS is that both the coefficient of the duration of ERS and the duration of ERS squared are positive (see Appendix 1). This indicates that wages rise with ERS at an increasing rate. The Tobit results in Appendix 2 show that the amount of ERS decreases with education level. Increasing rewards for ERS is something of a surprise, as we are inclined to expect decreasing returns. There might be a relation with out ignorance on the timing of the training. If longer trainings have been taken more recently, and if there is depreciation on the older, shorter trainings, we would measure an inverse effect of depreciation. If training is a requirement for promotions, the increasing returns may be a combined effect of training *per se* and promotions to better paying jobs. However, we control for experience (crudely) and for job level. It is not quite clear to us yet whether the effect is genuine or not. We will return to it later. Decreasing ERS with increasing level of education is less of a puzzle: most training programs relate to lower level, technical courses.

The estimation results are in Table 2. In the model in which $\alpha_1 = 0$, education high and IQ have a significant effect on the marginal cost of ERS. In the model with $\delta\gamma_2 = 0$ neither of these coefficients differs significantly from zero. All significant variables in the cost function have the expected sign. High educated workers and workers with higher ability (as measured by the IQ test score) have lower marginal costs. A high education decreases the marginal costs by 40 cents; a one point higher IQ score reduces marginal cost by less than 1 cent in the equation with a constant marginal cost function, and by 4 cents in the equation in which the marginal revenue function is constant.

Of the personal characteristics in the marginal revenue function only IQ is significant. The sign is negative, indicating that the marginal revenues decrease with ability. If we interpret ability as a component of a person's stock of human capital, this finding is consistent with the notion that there are decreasing returns-to-scale in human capital investments.

The coefficients of the wage equation are fairly similar in both specifications. Wages increase with years of education, IQ, and job level. Males earn more than females. The estimated variances and correlation coefficients are fairly similar in both estimations as well. The value of the likihood function in the estimations in which the marginal cost curve is constant is higher than the value of the likelihood function of the estimations in which the marginal revenue curve is assumed to be horizontal. This implies that a model with constant marginal costs of investing in ERS would be preferred over a model with constant marginal revenues if we let a goodness-of-fit criterion—such as for example the Leamer (1978)–Schwarz (1978) criterion—decide between the two models.

However, before relying on statistical measures, we should first consider whether conformity of the outcomes with theoretical restrictions (first principles) is decisive for the choice between the two specifications. Both the marginal cost equation and the marginal revenue equation are upward sloping in the amount of ERS. The result that the marginal revenue equation is increasing in the amount of ERS correspond to the finding in the OLS wage equations that for participants in ERS an increase in the investment in ERS raises wages at an increasing rate (see Appendix 1). In the model with constant marginal revenues ($\delta\gamma_2 = 0$), an upward sloping marginal cost curve ($\alpha_1 = 0$) satisfies the second order condition for the optimum to be a maximum. With constant cost

TABLE 2
Parameter estimates (t-values in brackets)

	Model with $\alpha_1 = 0$ (marginal cost curve is constant)	*Model with $\gamma_2 = 0$ (marginal revenue curve is constant)*
Marginal costs: $\alpha_0 + \alpha_1 S + Y'\alpha_2$		
Intercept (α_0)	−0.027 (0.06)	6.716 (1.15)
Education drop-out	0.123 (2.37)	−0.629 (0.56)
Education low	0.003 (0.05)	−0.192 (0.35)
Education intermediate	−0.090 (1.05)	0.330 (0.89)
Education high	−0.398 (2.75)*	1.532 (0.78)
IQ	−0.006 (1.68)†	−0.047 (1.38)
Female	−0.025 (0.08)	1.536 (0.66)
General education	0.040 (1.36)	−0.251 (0.57)
α_1	0	0.076 (0.65)
Marginal revenues: $\delta[\gamma_1 + X'_2\beta_2 + \gamma_2 S]$		
Intercept ($\delta\gamma_1$)	0.713 (1.85)†	3.020 (7.29)*
Years of education	−0.011 (1.47)	0.028 (1.51)
IQ	−0.010 (3.13)*	−0.026 (6.94)*
Female	0.267 (0.84)	0.084 (0.21)
$\delta\gamma_2$	0.015 (12.1)*	0 0
Wage equation: $\beta_0 + X'_1\beta_1$		
Intercept (β_0)	5.128 (0.60)	7.107 (0.79)
Years of education	0.576 (3.83)*	0.577 (3.31)*
Experience	−0.047 (0.07)	−0.289 (0.39)
Experience^2	−0.005 (0.30)	−0.000 (0.02)
IQ	0.097 (2.74)*	0.109 (2.75)*
Female	−4.302 (3.11)*	−2.429 (1.66)†
Job level	0.944 (3.23)*	0.775 (2.33)*
Variances and correlation coefficients		
σ_0	19.538 (39.47)*	19.508 (39.60)*
σ_1	7.360 (20.76)*	9.348 (17.72)*
σ_2	19.593 (21.16)*	19.635 (21.15)*
ρ_0	−0.093 (1.66)†	−0.263 (1.94)†
ρ_1	0.153 (0.99)	−0.059 (0.88)
Loglikelihood	−5,834.55	−5,892.40

* Significant at 5% level.
† Significant at 10% level.

($\alpha_1 = 0$), an upward sloping marginal revenue curve ($\delta\gamma_2 > 0$) however, doesn't satisfy the second order condition.[8] On economic grounds, we therefore prefer the results in the second column of Table 2, and we will henceforth limit the discussion to those based on this result.

The results have a fairly regular structure. Marginal cost increase with the

[8] In addition to what we suggested above about the unknown timing of the training, this might be caused by a linearization of an inverse U-shaped marginal revenue curve, in which the marginal revenues first increase with the amount of ERS and decrease afterwards. We have not experimented with more complicated forms.

level of education, indicating that higher educated individuals need more compensation to participate in training. However, it falls with IQ, indicating that more intelligent individuals are more interested in further training. Moreover, their costs fall more than their revenues, which optimally pushes them towards more training. Those with a general education, rather than a vocational education, have more interest in training, as their marginal cost falls. Women have a stronger increase in cost than in revenue, relative to men.

In order to gain some more insight into the results of Table 2, we performed the following simulations. First we defined a reference individual: a male, with 4.61 years of education (after primary school, i.e. an intermediate level), general education, an IQ of 103 and 25 years of experience. For this individual we predicted the expected duration of training $\hat{S}^*$. Then we calculated the wage differential between training and no training, marginal cost and benefit (evaluated at $\hat{S}^*$). The financial variables are calculated separately for an average individual, an average participant and an average non-participant, to highlight the effect of the selectivity term. The whole exercise is repeated for a female with the same characteristics and for the two other created individuals, a 'dummy' and a 'smarty'. The dummy has an IQ of 90 (one standard deviation below the mean), with only primary education (years educated is zero) and a job level of three (about a standard deviation below the mean). A smarty has IQ of 116, a high education (eight years), and job level 6.2.

The results point to the following conclusions. For an average person from the sample, whether he/she is a reference person, a dummy or a smarty, the optimal investment level is minus 20 months or more. This demonstrates at once the relevance of distinguishing between participants and non-participants on the basis of endogeneous selection. The variations in marginal benefits and marginal costs between our different persons are in this case quite modest. Conditional upon participation the optimal length of ERS varies between 7.43 months for a reference female to 10.40 for a reference male. An interesting finding is that the difference between the participating 'dummy' and the participating 'smarty' is small. The benefits of ERS to the smarty are a bit larger, but so are the costs; as a result the optimal length for a participating dummy exceeds the optimal length for a participating smarty by about one month. The total gain from ERS for participating persons amounts to 100.8 for the reference male, with a discount rate of about 17 (see Section 5); this implies a wage rate increase of 6 guilders for 10.4 full time equivalent months of ERS. The participating reference female has a wage rate increase of about twice as much for an optimal investment that takes only 7.43 months. The results conditional upon non-participation parallel the unconditional findings and the findings conditional upon participation.

7. Evaluation

In this paper we have estimated a structural model for participation in and duration of enterprise-related schooling. We could not fully identify the model,

TABLE 3
Optimal investment, marginal revenues, marginal cost, and wage gains

	Reference male	*Reference female*	*Male dummy*	*Male smarty*
Unconditional				
$\hat{S}^*$	−19.19	−38.41	−20.22	−25.15
$[w_1(\hat{S}^*) - w_0]\delta$	−9.44	−22.11	−14.10	−6.36
MR	0.49	0.58	0.70	0.25
MC	1.94	3.48	2.23	2.16
Conditional on participation				
$\hat{S}^*$	10.40	7.43	10.20	9.30
$[w_1(\hat{S}^*) - w_0]\delta$	100.78	187.01	107.15	123.99
MR	9.69	25.17	10.51	13.33
MC	−0.97	−1.69	−0.81	1.45
Conditional on non-participation				
$\hat{S}^*$	−24.97	−39.58	−25.62	−28.96
$[w_1(\hat{S}^*) - w_0]\delta$	−15.06	−25.28	−20.80	−10.69
MR	0.60	0.64	0.81	0.37
MC	0.22	0.52	0.44	0.07

and choose to estimate two specifications. We preferred the specification with a flat marginal revenue curve, and derived a number of interpretations from it. The dataset and the model suffer from several weaknesses, but we feel we also have unveiled some results worth pondering. The main problem with the dataset is ignorance on the timing of the training. Since we do not know how long ago the training was actually taken, we cannot properly discount the revenues from training accruing over the remainder of the working life. Moreover, we do not know anything about persistence and depreciation of training effects. This is obviously a major shortcoming. The theoretical model imposes some substantial conditions: optimum investment is feasible without rationing; a rather arbitrary solution to the identification problem; no acknowledgement of opportunity cost of training. So, clearly, there is much need for further work; and such further work should be tailormade to a dataset that improves on the shortcomings with which we had to cope.

Yet, there are a number of interesting results. We feel that we certainly should go further down the road of estimating structural models, as they yield decompositions that are not available in reduced forms. Perhaps the most interesting route on which to proceed is to acknowledge the possibility of rationing of training slots. The OLS estimate of the wage function indicates increasing returns to ERS. In the structural model we also find this, if we constrain marginal cost to be constant. However, this is not the specification we prefer: we opted for the restriction where marginal revenue is constant, independent of ERS duration. But even then, the water seeps through the dam. In all our conditional simulations, we find that marginal revenue is greater than marginal cost, in spite of the imposed optimality. This is clearly an undesirable outcome, and one way to deal with it might be to allow for rationing of admission to job slots.

The results for ability are also intriguing. Marginal revenues of ERS falls with childhood IQ, but marginal cost falls even more. In the simulation, for participants in ERS, marginal revenue is higher for a 'dummy' than for a 'smarty', and marginal cost is lower. In the end, a dummy takes more ERS than a smarty. This may reflect a situation of the past. In our dataset, many ERS programs refer to lower level technical training: our data relate to a period prior to the large upswing in company training efforts. It might very well be that at that time, less able individuals rended to end up in jobs where specific human capital was created in formalized ERS programs, rather than left to emerge from informal on-the-job training and learning-by-doing.

Our final evaluation is that we may have found some real effects, but that both better data and more extensive modelling are needed.

ACKNOWLEDGEMENTS

A previous version of this paper has been presented at an *Oxford Economic Papers*' Conference on Vocational Training. We would like to thank two anonymous referees for their helpful comments.

REFERENCES

BARRON, J. M., BLACK, D. A., and LOEWENSTEIN, M. A. (1989). 'Job Matching and On-the-Job Training', *Journal of Labor Economics*, **7**, 1–19.

BECKER, G. S. (1975). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, Second Edition, National Bureau of Economic and Social Research, New York.

BJÖRKLUND, A. and MOFFITT, R. (1987). 'The Estimation of Wage Gains and Welfare Gains in Self-Selection Models', *The Review of Economics and Statistics*, **69**, 43–9.

BOOTH, A. (1991). 'Job-Related Formal Training: Who Receives it and What is it Worth?', *Oxford Bulletin of Economics and Statistics*, **53**, 281–94.

BROWN, J. (1989). 'Why do Wages Increase with Tenure?: On-the-Job Training and Life Cycle Wage Growth Observed within Firms', *American Economic Review*, **79**, 971–91.

DOLTON, P. and MAKEPEACE, G. (1987). 'Interpreting Sample Section Effects', *Economic Letters*, **24**, 373–80.

GAREN, J. (1984). 'The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable', *Econometrica*, **52**, 1199–218.

GROOT, W., HARTOG, J. and OOSTERBEEK, H. (1994). 'Wage and Welfare Gains of Within Company Schooling', in L. Lynch (ed.), *International Comparisons of Private Sector Training*, NBER, University of Chicago Press (forthcoming).

HARTOG, J. (1989). 'Survey Non-Response in Relation to Ability and Family Background: Structure and Effects on Estimated Earnings Functions', *Applied Economics*, **21**, 387–95.

HARTOG, J. (1992). *Capabilities, Allocation and Earnings*, Kluwer, Boston.

HASHIMOTO, M. (1981). 'Firm-Specific Human Capital as a Shared Investment', *American Economic Review*, **71**, 475–82.

HOLZER, H. (1988). 'The Determinants of Employee Productivity and Earnings', NBER Working Paper 2782.

KENNY, L. W., LEE, L.-F., MADDALA, G. S., and TROST, R. P. (1979). 'Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data', *International Economic Review*, **20**, 775–89.

LEAMER, E. (1978). *Specification Searches*, Wiley, New York.

LILLARD, G. and TAN, H. (1986). 'Private Sector Training', Rand Corporation R-3331.

LYNCH, L. (1992). 'Private-Sector Training and the Earnings of Young Workers', *American Economic Review*, **82**, 299–312.

MINCER, J. (1988). 'Job Training, Wage Growth, and Labor Turnover', NBER Working Paper 2680.

MINCER, J. (1989). 'Job Training: Costs, Returns, and Wage Profiles', Columbia University, New York.

MINCER, J. (1989). 'Human Capital and the Labor Market: a Review of Current Research', *Educational Researcher*, **18**, 27–34.

OECD (1991). 'Enterprise-Related Training', *Employment Outlook 1991*, OECD, Paris, 135–75.

OOSTERBEEK, H. (1992). 'Study Duration and Earnings', *Economic Letters*, **40**, 223–8.

RITZEN, J. (1989). 'Government and Job Training', The Eric John Hanson Memorial Lecture Series, **III**.

SCHWARZ, G. (1978). 'Estimating the Dimension of a Model', *Annals of Statistics*, **6**, 461–4.

VAUGHAN, R. (1989). 'Education, Training and Labor Markets: Summary and Policy Implications of Recent Research by Jacob Mincer', Conference Paper Series, The Institute on Education and the Economy, Teachers College, Columbia University, New York.

WILLIS, R. and ROSEN, S. (1979). 'Education and Self-Selection', *Journal of Political Economy*, **87**, S7–36.

APPENDIX 1

Parameter estimates OLS wage equations (t-values in brackets)

	Participants in enterprise-related schooling (ERS)		*Non-participants in enterprise-related schooling (ERS)*	
Dependent variable	*Log of wage rate*	*Wage rate*	*Log of wage rate*	*Wage rate*
Intercept	0.943	−14.542	1.761†	13.896
	(1.441)	(0.903)	(8.761)	(1.345)
Year of education	0.027†	0.370	0.029†	0.801†
	(3.064)	(1.687)	(6.094)	(3.277)
Experience	0.110*	1.739	0.025	−0.687
	(2.006)	(1.292)	(1.480)	(0.795)
Experience^2/100	−0.249*	−4.226	−0.059	0.999
	(2.141)	(1.479)	(1.487)	(0.490)
IQ/100	0.428*	11.028*	0.277*	5.424
	(2.303)	(2.411)	(2.474)	(0.942)
Female	−0.367†	−6.902†	−0.120†	−2.529
	(3.381)	(2.585)	(3.566)	(1,461)
Job level	0.032*	0.817*	0.063†	1.005
	(2.164)	(2.217)	(6.216)	(1.943)
Duration of ERS	0.026	0.785*		
	(1.596)	(1.968)		
Duration of ERS^2/100	0.011†	0.762†		
	(3.967)	(11.442)		
Years of education* duration of ERS/100	−0.001	−0.509		
	(0.017)	(0.284)		
IQ* duration of ERS/100	−0.029*	−1.067†		
	(1.987)	(3.018)		

APPENDIX 1 (*cont.*)

Dependent variable	*Participants in enterprise-related schooling (ERS)*		*Non-participants in enterprise-related schooling (ERS)*	
	Log of wage rate	*Wage rate*	*Log of wage rate*	*Wage rate*
Female* duration of ERS	0.023	0.553		
	(1.588)	(1.545)		
Adj-R^2	0.319	0.569	0.275	0.069
Loglikelihood	−52.348	−910.830	−353.463	−3462.17
No. of observations	268	268	789	789

* Significant at 5% level.
† Significant at 1% level.

APPENDIX 2
Parameter estimates Tobit equations duration of enterprise-related schooling (t-values in brackets)

Intercept	−64.460†	(6.130)
Experience	0.538*	(2.120)
Education drop-out	11.110†	(3.812)
Education low	6.152	(1.812)
Education intermediate	2.916	(0.775)
Education high	−9.448	(1.501)
IQ	0.280†	(4.011)
Female	−19.581†	(6.354)
General education	3.797*	(2.021)
σ	19.583†	(21.170)
No. of observations	1,057	
Loglikelihood	−14,547.98	

* Significant at 5% level.
† Significant at 1% level.