



ÉCOLE POLYTECHNIQUE
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

**ESTIMATING COMPLEMENTARITY BETWEEN
EDUCATION AND TRAINING**

Christian BELZIL
Jørgen HANSEN
Nicolai KRISTENSEN

November 2008

Cahier n° 2008-38

DEPARTEMENT D'ECONOMIE

Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033

<http://www.enseignement.polytechnique.fr/economie/>
<mailto:chantal.poujouly@polytechnique.edu>

ESTIMATING COMPLEMENTARITY BETWEEN EDUCATION AND TRAINING*

Christian BELZIL¹
Jørgen HANSEN²
Nicolai KRISTENSEN³

November 2008

Cahier n° 2008-38

Abstract: In this paper, we formulate and estimate a structural model of post-schooling training that explicitly allows for possible complementarity between initial schooling levels and returns to training. Precisely, the wage outcome equation depends on accumulated schooling and on the incidence of training. The effect of training on wage growth depends on individual permanent endowments as well as on education. We find evidence of statistically significant complementarity, i.e. the higher educated receive the highest return to the MBA-type training considered here.

Key Words : Skill Complementarity, Dynamic Treatment Effects, Dynamic Programming, Random Coefficients.

Classification JEL: I2, J2, J3

* We thank Paul Bingley, Hessel Oosterbeek, Bertel Schjerning, Ian Walker, Frédéric Warzynski and seminar participants at AKF (Copenhagen) for useful comments on previous versions of the paper.

¹ Department of Economics Ecole Polytechnique, ENSAE, IZA and CIRANO

² Concordia University, CIRANO, CIREQ, IZA

³ Danish Institute of Governmental Research and Aarhus School of Business, University of Aarhus

1 Introduction

In most schooling systems, education is accumulated through successive steps conditional on completion of a lower schooling level. There is therefore rarely any difference in educational attainment among a cross-section of individuals currently attending the same education (grade) level. However, this is not the case in post-schooling training participation, where educational background may differ widely across participants. For this reason, it is natural to investigate whether the incidence of training and its effect on wages depends on schooling.

This question motivates the current paper. More precisely, we estimate the effect of training on post-training wage outcomes in a context where individual returns depend on permanent individual-specific heterogeneity and on the level of schooling achieved by the individual.

The issue of possible correlation between individual returns to training and accumulated schooling is far from new. In a series of papers (*i.a.* Heckman 2000, Carneiro and Heckman 2003, Knudsen et al. 2006, and Cunha and Heckman 2007), Heckman and co-authors argue that complementarity in learning exists. Skill formation is a dynamic process with strong synergies that lead early investments to promote later investments: skill begets skill.

In a dynamic skill accumulation model, complementarity may take at least two distinct forms. First, accumulated skills may decrease the psychic cost of investing in further skill accumulation. In such a case, individual age earnings profiles would tend to display positive correlation between schooling and post-schooling wage growth, even if the wage regression is separable in education and experience (or any post-schooling training activities). However, with data on earnings, schooling and training decisions, the econometrician would be able to estimate the returns to education and training within a classical (separable) Mincer wage equation.

A second type of dynamic skill accumulation model would arise if, for a given level of individual permanent endowments, the wage gain associated with training participation depends on accumulated schooling. In such a case, schooling would also magnify the effect of post-schooling human capital investments. However, for a given level of skill endowment, the wage equation would no longer be linearly separable in schooling and training. As a result, the returns to additional training will be higher for individuals with a relatively high formal schooling, say a college degree, *vis-à-vis* individuals with a lower level of education, say high school.

In both cases, the policy implications of complementarity are important and far-reaching: Investment in early childhood +development is more cost-

effective all else equal, vis-à-vis seeking to upgrade qualifications of unskilled adult workers. As there exists a widespread belief that OECD countries have experienced a rapid shift in demand toward more skilled workers, the implications of the studies cited above are very important for how to effectively design policies that deal with this shift in demand of skilled labor.¹

At a pure econometric level, estimating the effects of training on wages is relatively complicated. Because both training and schooling decisions are potentially affected by individual permanent endowments, IV estimation, typically based on the availability of policy shocks (or natural experiments), is ill-equipped to tackle this issue. Indeed, IV estimation is usually associated to the notion of Local Average Treatment Effect (LATE), which was introduced by Imbens and Angrist (1994). If the dispersion of individual wage gains reflects differences in schooling as well as differences in skills (which would be the case if there is complementarity), classical IV would hardly be able to disentangle the causal effect of schooling from the heterogeneity component. Because the LATE is an average return for a sub-population affected by a policy change (the instrument), it is not even clear that such an estimate would be of any help in determining the incidence (or the degree) of complementarity between schooling and training.²

In this paper, we focus on dynamic skill accumulation that takes the form of a causal effect of education level on individual private returns (wage gains) to training participation. We investigate the degree of complementarity between formal schooling level and MBA-type of training (described in detail in section 2.3). To this end, we use data from Danish administrative employer-employee records spanning from 1980 to 2003 and condition the sample on participants with highest completed formal schooling in one of two levels. Level 1 includes high school degree or some (partly) vocational education at an equivalent level (12-13 years) while Level 2 consists of people with at least a bachelor degree or its equivalence (15 years or more). This provides us with a unique opportunity to analyze heterogeneity and complementarity in returns to training between different skills groups. We do this by estimating the returns to training using a structural model of the

¹Heckman and co-authors are not the only scholars finding results in support of complementarity, see e.g. Shaw (1989) and Meghir and Palme (2005). However, complementarity in learning has not gone unchallenged. For instance, Krueger (2003) argues for the opposite view that returns to training are at least as big at the bottom of the income distribution as at the top.

²The empirical relevance of the LATE parameter is not unanimously recognized. For instance, several economists have claimed that the LATE parameter is not “policy-relevant” for a variety of reasons, see e.g. Heckman and Vytlačil (2007).

training decision, in which the wage equation is formulated in a correlated random coefficient wage regression model (CRCWRM) in the vein of Belzil and Hansen (2007).

The main finding of the paper is the existence of complementarity between schooling and training. More precisely, we find that individual returns (wage gains) to training are magnified by individual schooling level. The results therefore indicate a dynamic causal effect of schooling and imply that the standard representation of the Mincer wage equation, which is built on an assumption of separability between schooling and training, may be erroneous.³ While it is not clear that the result regarding complementarity may be extrapolated to all possible post-schooling investment decisions, we believe that one of the major contribution of this paper is the development of a methodology that address possible non-separability between the effects of schooling and post-schooling human capital investments on wages.

The paper is organized as follows. In the next section, we present the Danish data that have been used to implement our model. The setup of the behavioral model is found in section 3. The estimation strategy is discussed in section 4 while the parameter estimates are presented in section 5. Section 6 provides a discussion of the main findings while section 7 concludes the paper.

2 Data

2.1 Danish Register Data

Data come from administrative records maintained by Statistics Denmark. The records constitute a longitudinal database that contains information about *all* individuals aged 15 to 74 (demographic characteristics, education, labor market experience, tenure and earnings) and employees in *all* workplaces in Denmark during the period 1980-2003. Persons and workplaces are matched at the end of November each year. Consequently, only changes of employment November-to-November are accounted for, not intervening changes.

The database has been merged with a course register database that contains information on all government co-sponsored formal training courses, vocational as well as classroom training. This register includes detailed information on the business training program described below.

³The misspecification of the Mincer equation has already been noted in Heckman et al. (2003), as well as in Belzil (2007)

2.2 The Training Program

We analyze the effects of a government co-sponsored training program in business administration⁴, which has a long tradition in Denmark dating back more than 50 years and until recently this training program was the only MBA-type training program available in Denmark. The curriculum consists of studies in business economics with classes taking place in the evening (or in a few instances on weekends). The designed course load is equivalent to two years' full-time studies, divided into two parts of equal size. It is possible to stop after Part I but a diploma is obtained after completion of the second part only⁵. The program is considered as training although it clearly has some academic component to it and as such is close in nature to education and (most) participants work full-time both during Part I as well as Part II of the program.⁶

In Part I of the program, students are introduced to fundamental theories and methods in social science. The minimum educational background required in order to be eligible to enter Part I is a high school degree.

Part II of the program includes more in-depth studies in some field of business economics.⁷ Students who have completed Part I successfully and who have a minimum of two years (full-time equivalents) of relevant work experience are eligible for Part II.

In this paper, the sample is restricted to individuals who have completed Part I of the program and we model their decision to either stop after Part I or continue (perhaps after one or more sabbatical years) with Part II of the program. As mentioned in the introduction, we select individuals who belong to one of two levels of education; High School or College and above.⁸

2.3 Descriptive Statistics

We limit our sample to males, born 1960-1970 who completed Part I of the training program before 1989. We follow these individuals from the time they complete Part I and as long as we observe them in the data.

⁴In Danish called "Handelsdiploma" or simply "HD".

⁵So the Handelsdiploma (HD) requires Completion of both Part I and Part II.

⁶In the nomenclature applied by Statistics Denmark this course is defined as post-schooling training. Furthermore, it takes place in the evening time and is attended by full-time workers as many other types of training.

⁷Possible fields include: Finance, Organization, Accounting, Logistics & Supply Chain Management, Financial Counselling and Retail & Key Account Management.

⁸As the focus here is on complementarity it is not necessary to model selection into Part I as well. Such an extension would not be trivial since the duration of part I also can be of various length.

By conditioning on cohorts born 1960 or later, keeping in mind that the register database starts in 1980, we ensure that we observe individuals in our sample from around the time their labor market experience potentially starts accumulating (around age 19-20).

This leaves us with a sample of about 3,500 individuals (54,000 observations) of which 2,654 individuals have no schooling beyond high school while 857 individuals have completed college or more, cf. Table 1. About 75 percent choose to continue with Part II.

[Table 1 about here]

Looking at the raw wage differences between individuals who select into Part II of the training versus those who only take Part I reveals that the median wage among those who complete Part II is markedly higher than those who do not, cf. Figure 1. Of course, this does not convey any information about whether this is an effect of training or purely a sign of positive selection of the most able and motivated.

[Figure 1 about here]

In formulating our structural model, we need to know the approximate time between completion of Part I and start of Part II (if ever entering). Not surprisingly, Figure 2 shows that about 90 percent of those who continue do so immediately, whereas the remaining participants enroll within the following two years. Likewise, the model takes into account the fact that duration of Part II of the training program, although it typically takes two years, may stretch over 4-5 years, cf. Figure 3.⁹

[Figures 2 and 3 about here]

From the data we can observe that the reason for prolonged duration for many training participants is *not* that they fail to pass exams but rather that they deliberately choose a less intensive course-load and for this reason stretch the duration over more than two years. This observation is taken into account when we estimate the probability that it takes more than two years to complete Part II.

⁹As can be seen from Figures 2 and 3 there is virtually no difference between the two education groups in their "time to enrol" and "time to complete".

3 A Structural Dynamic Model

As mentioned above, we consider individuals who have completed Part I of the training program, and we model their decision to continue in the training program (to enter Part II) or not. Every decision is made at the beginning of each period.

3.1 The Model

3.1.1 Choice Set

Individuals are assumed to be rational and forward-looking. In every period, they make choices in order to maximize their expected net present value of utility until they retire.

In our model, the choice set is very simple. Upon completion of Part I of the training program, individuals have to decide whether to continue with the second part of the program or not. All choices are made within the first three years after completion of Part I, but workers may be active in training up to and including period 5 in the model. Differences in training duration mainly reflect individual choices of various intensity levels (and not failure to pass exams).

3.1.2 Utility of Work

The instantaneous utility from work, denoted $U^W(\cdot)$, is set equal to the real gross log wages, $\ln w_{it}$. As the individuals in our sample experience only very low degrees of unemployment, we model their career paths as if they are full-time employed in every period. The log wage received by individual i in period t is given by

$$U_{it}^W(\cdot) = \ln w_{it} = \varphi_{1i}TR_{it} + \varphi_2Exper_{it} + \varphi_3Exper_{it}^2 + \delta_i\mathbf{I}_{i,college} + \nu_i^W + \varepsilon_{it}^W, \quad (1)$$

where $\mathbf{I}_{college}$ is an indicator which equals 1 if the individual has a college degree (including bachelors and masters) and zero otherwise, δ_i is an individual-specific parameter, TR_{it} is an indicator for having completed Part II of the training, φ_{1i} is the individual-specific wage return to training, and φ_2 and φ_3 measure the effects of experience and experience squared. Unobserved heterogeneity is accommodated both through the individual specific

intercept term, ν_i^w , and through the returns to education, training and education and training interacted (the complementarity term). Finally, ε_{it}^W represents idiosyncratic stochastic wage shocks.

The individual-specific returns to training can be written as

$$\varphi_{1i} = \mathbf{I}_{i,college} \cdot \varphi_{ci} + \omega_{1i}$$

where the term φ_{ci} is an unrestricted type i specific parameter that represents the difference in returns to training for those with a college degree relative to those with a high school diploma. A positive φ_{ci} would imply that schooling has a causal effect on the return to training and it could be interpreted as evidence in favor of complementarity in learning. At this stage, it is important to note that, in general, ω_{1i} is correlated with $\mathbf{I}_{i,college}$.

Note also that because college attendance is affected by individual heterogeneity, the difference between actual training returns of high school graduates and college graduates would generally be different from φ_{ci} .

3.1.3 Utility of Training Participation

Non-pecuniary utility is measured in monetary equivalents and enters the model as utility of learning and utility of leisure (disutility of loss of leisure), and these utility components are allowed to depend on a series of observable covariates which, following Belzil and Hansen (2002, 2003 and 2007), enter the model exogenously.

The per-period utility of selecting into the second part of the training is therefore given by

$$U_{it}^{TR} = X_{it}^{TR} \lambda + \nu_i^{TR} + \varepsilon_{it}^{TR} \quad (2)$$

The matrix of covariates, X_{it}^{TR} , includes age and age squared, marital status, number of children and local labor market conditions, i.e. the local unemployment rate (by education group, gender, and age) across time. It also includes an indicator for whether college is completed or not and tuition fee. Lastly, the utility of training participation depends on unobserved individual heterogeneity, ν_i^{TR} , orthogonal to observed covariates while ε_{it}^{TR} represents idiosyncratic stochastic shocks.

The unobserved utility component, ν_i^{TR} , captures unobserved heterogeneity which arises if individuals are endowed with different unobserved pre-training skills or differ in terms of their intrinsic motivation. The term may also capture idiosyncratic knowledge about the time and effort required for completion of the training. Each of these utility components can not be

separately identified so we think of ν_i^{TR} as the combined net outcome of unobserved factors.

3.1.4 The Stochastic Shocks

We assume that the random shocks $(\varepsilon_{it}^{TR}, \varepsilon_{it}^W)$ are normally distributed and independent of each other as well as independent of training, formal schooling and experience.

3.1.5 The Utility Function

The instantaneous utility is given as the sum of equation 1 and equation 2

$$U_{it} = U_{it}^W + \mathbf{I}_{d=1} \cdot U_{it}^{TR}, \quad (3)$$

where the components of $U^W(\cdot)$ and $U^{TR}(\cdot)$ are specified above. The choice variable, d , equals 1 if the individual chooses to continue with Part II of the training program, and zero otherwise.

3.2 Solving the Problem

3.2.1 The Optimization Problem and Solution Method

For sake of clarity, the state space is partitioned into two components. The first one, \mathbf{s}_{it} , includes period-specific idiosyncratic random shocks to utility $(\varepsilon_{it}^{TR}, \varepsilon_{it}^W)$, accumulated experience ($Exper_{it}$), and schooling ($\mathbf{I}_{college}$) for individual i at time t . The second one, TR_{it} , is equal to 1 if individual i has completed Part II in period t , and 0 if not.

The law of motion implies that the state variable, TR_{it} , equals 0 until the training course is actually completed.

Suppose that the decision to enter Part II is made at the beginning of period t . As individuals face two options, two value functions have to be specified. First, the value function of choosing full-time work and foregoing the training opportunity at time t (denoted $V_{it}^W(\cdot)$), is simply

$$V_{it}^W(\cdot) = U_{it}^W(\mathbf{s}_{it}, TR_{it} = 0) + \beta EV_{i,t+1}(s_{it+1}, TR_{it+1} = 0)$$

where $EV_{i,t+1}$ is the expected maximum utility, cf. below.

Consistent with the data, assume that the second part of the training period always last at least two periods. In order to capture the intrinsic randomness in the time needed to complete the program, we introduce a

continuation probability, θ , which represents the probability that the training program will have a duration beyond two years. This probability is parameterized simply as

$$\theta = \Phi(X_{it}^{dur}\xi + \nu_i^{dur})$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, X_{it}^{dur} includes experience and duration of Part I and ν_i^{dur} is unobserved heterogeneity correlated with other components $(\delta_i, \varphi_{ci}, \omega_{1i}, \nu_i^{TR}, \nu_i^W)$.

The value function associated with the decision to work full-time and continue with Part II of the training program ($d_t = 1$) or not, denoted $V_{it}^{TR}(\mathbf{s}_{it})$, can be expressed by the Bellman equation

$$V_{it}^{TR}(\mathbf{s}_{it}) = U_{it}^{TR}(\mathbf{s}_{it}, TR_{it} = 0) + \beta E\{V_{i,t+1}^{TR}(\mathbf{s}_{it+1}, TR_{it+1} = 0)\} \quad (4)$$

where

$$\begin{aligned} E(V_{it+1}^{TR} | \mathbf{s}_{it+1}, TR_{it+1} = 0) &= U_{it+1}^{TR}(\mathbf{s}_{it+1}, TR_{it+1} = 0) + \\ &\beta\{[(1 - \theta) \times EV_{it+2}^W(\mathbf{s}_{it+2}, TR_{it+2} = 1) + \\ &\theta \times EV_{it+2}^{TR}(\mathbf{s}_{it+2}, TR_{it+2} = 0)]\} \end{aligned}$$

With probability $(1 - \theta)$ he completes Part II in period $t + 2$, while with probability θ the training lasts an additional period.¹⁰

The optimal value function, at date t , is

$$EV_{it}(\cdot) = EMAX\{V_{it}^{TR}(\cdot), V_{it}^W(\cdot)\}$$

The optimal choice is found using backward induction. Once an individual enters into Part II of the training program, he will complete the training in either 2 or 3 periods. In this sense, his outcome is deterministic and the problem can be seen as an optimal stopping problem: once a person enters the training program, there are no more choices to be made. Alternatively, if he does not enter in period t , he has the option of entering in the next period. In Appendix A, we include a more formal explanation of how to compute the *EMAX* used in the backward recursions.

¹⁰From data we observe that Part II sometimes lasts more than three periods. Any duration beyond two periods is lumped together here.

4 Estimation

4.1 The Initial Condition

Although we are not modeling schooling explicitly, it is still important to allow for the distribution of unobserved heterogeneity to be arbitrarily correlated with college completion. This is because individual permanent endowments that may help explain training participation are also likely to have played a role in schooling choices. As a consequence, we formulate the schooling decision as a binary probit equation which also depends on unobserved heterogeneity. That is, the probability of initially having completed college, is simply

$$\Phi(X_i^S \delta + \nu_i^S)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, X_i^S includes mothers and fathers highest completed education and ν_i^S is unobserved heterogeneity correlated with other components ($\delta_i, \varphi_{ci}, \omega_{1i}, \nu_i^{TR}, \nu_i^W, \nu_i^{dur}$).

4.2 Unobserved Heterogeneity

As is customary in this type of models, we introduce unobserved heterogeneity in the form of finite mixtures (Heckman and Singer, 1984). In particular, we assume that there are K types of individuals, each endowed with a value of $(\nu_k^{TR}, \nu_k^W, \delta_k, \varphi_{ck}, \omega_{1k}, \nu_k^S)$, $k = 1, 2, \dots, K$.¹¹

4.3 Identification

As is the case for most structural models set in an intertemporal framework, the estimation of our model requires some explicit functional forms.¹² For instance, we make explicit assumptions concerning the form of the utility function and the distribution of random shocks. Based on these assumptions, we use panel data to identify the probability of actual choices and outcomes. Exogenous variation is obtained from time-varying tuition costs, changes in

¹¹The probability of a given type can then be estimated as $p_k = \frac{\exp(q_k)}{\sum_{j=1}^K \exp(q_j)}$, where one of the q_k 's is normalized to 0. We employ $K = 4$.

¹²Although IV models are typically presented without explicit functional form restrictions, their interpretation also require implicit restrictions, which may take the form of very specific functional form assumptions, cf. Keane (2006).

the distance to nearest institution and local unemployment rates (by gender, region, age and education).¹³

Tuition fees were introduced starting from 1990 (prior to this year the training was free of charge) and have been rising (in real 1980 prices) since then (Figure 5).¹⁴

[Figures 4 and 5 about here]

This introduces time and spatial variation, which helps in the identification of the non-pecuniary part of the model.

4.4 The Likelihood Function

For a given set of parameter values, one can solve the "inner" optimization problem using backward recursions. Then, given a set of actual observed choices, the likelihood of these choices is computed (in an "outer" algorithm).

The likelihood consists of four parts:¹⁵

- (a) The probability of observing a particular sequence of choices. This sequence can be expressed as products of marginal probabilities, conditional on the unobserved heterogeneity components, ν_k^{TR} and ν_k^W , and is denoted $L_1(k)$.
- (b) The probability of completing training in 2 or 3 periods conditional on entering into training, denoted $L_2(k)$.
- (c) The joint densities of observed wages from $\tau = 1$ until $nper_i$ (the number of periods in which individual i is observed), at observed wage w_τ , which can be factored as the product of a normal conditional probability multiplied by the marginal wage density $f(\cdot)$, denoted $L_3(k)$.

¹³In particular note that only about half of the institutions that offer Part I of the training program also offer Part II (Figure 4). This change in distance to Part II institutions provides variation, which should be helpful for identification. However, empirically it turned out not to be significant and it was therefore excluded from the estimation.

¹⁴Note that prices differ between institutions. This could raise a concern for quality differences in the training also. However, a recent report from the Ministry of Education concludes that despite the substantial differences in tuition costs there appears to be no quality difference across institutions (MoE, 2000).

¹⁵Details of how the model is solved and how the likelihood is constructed can be found in Appendix A.

- (d) The probability of having a college degree (i.e. endogenizing the initial conditions), denoted $L_4(k)$.

The k -index represents the discrete type mixture for the unobserved heterogeneity components. Hence, the log-likelihood to be maximized is the product over types

$$\ln L = \ln \sum_{k=1}^4 p_k \cdot L_1(k) \cdot L_2(k) \cdot L_3(k) \cdot L_4(k). \quad (5)$$

5 Estimates of the Structural Model

The presentation of the structural parameter estimates is split into 4 different sets; those pertaining to the non-pecuniary costs of training (Table 2), the wage equation (Table 3), the continuation probability and initial conditions (Table 4) and the distribution of unobserved heterogeneity components (Table 5).

[Tables 2, 3, 4 and 5 about here].

5.1 The Cost of Training

Given the additive form of the Utility function, negative parameter estimates for the non-pecuniary utility components, equation 2, indicate that an increase in a specific variable tends to decrease the utility of training (increase the psychic cost). We note that the results indicate that the incidence of training increases with the local unemployment rate and decreases with tuition. Other things being equal, those who were older when they terminated Part I are less likely to enrol in Part II.

Interestingly, the effect of accumulated education (the college indicator) on the net utility of training is significantly negative for alle four types (the weighted average is equal to -0.81). This does not necessarily imply that education increase the psychic cost of training. To see the argument, one has to re-interpret the negative effect of education as a composite effect on distinct components of the per-period utility of training. For instance, the utility of training may incorporate (i) a pure consumption value of the training cost (prestige associated with training, etc.), (ii) a component that

reflects psychic learning costs, or (iii) a component that measure the effect of education on time availability for training (for instance, reduction in leisure may be higher for college educated).

Finally, the utility (or the cost) of training decreases (increases) with the number of young children.

5.2 The Wage Equation

Estimates of the wage equation are found in Table 3. The individual-specific return to college completion (ranging from 0.01 to 0.17) indicates a large premium to college completion. The population average return is equal to 0.15. When reported on a per-annum basis, these estimates imply wage returns to education (ignoring the effect of education on the return to training) ranging roughly between 4 percent and 5 percent per year.

Not surprisingly, training completion is also associated with a relatively high wage return. For those who have completed high school, the returns range between 0.09 and 0.28. As Part I does not result in a diploma part of these estimated returns to Part II (vis-à-vis Part I) may also be returns to credentials.

For those who have completed college, the return to training is magnified by 0.10 (weighted average across types), after conditioning on unobserved heterogeneity. The population average of the individual-specific return to training is 0.14. Evidently, individual differences in returns to training are dominated by individual permanent endowments.

Returning to the issue of complementarity, our estimates indicate a statistically significant degree of complementarity between schooling and training. Interestingly, complementarity differs markedly in magnitude across the four types ranging from zero (type 3) to 0.22 (type 2).

5.3 The Continuation Probability and Initial Conditions

The estimates in Table 4 show that there is considerable heterogeneity in the time required to complete Part II of the training program. As expected, older and more experienced workers appear to complete the program faster than younger and less experienced workers. Moreover, the duration of Part II is longer for those who took longer time to complete the first part of the training program.

The initial condition equation indicates, as is known in most countries, that the probability of graduating from college is highly correlated with parents' education. Unobserved heterogeneity is also found important.

5.4 The Heterogeneity Distribution

A thorough review of the main heterogeneity components (utility of training, utility from college and training, return to college, return to training and complementarity in training) and their standard deviations reveals a fair degree of cross-sectional dispersion in individual endowments. This is particularly true for the individual returns to training (the ω_{1i}) with a standard deviation equal to 0.09 and the complementarity parameter (φ_{ci}) with a standard deviation equal to 0.12.

Individuals of types 1 and 4 are endowed with the highest utility of training (lower learning cost) as well as the highest returns to college (these types also constitute the majority of the population, about 85%). This may be interpreted as evidence that both the wage return to college and the psychic costs of training are affected by skills that are academic in nature. However, ignoring the effect that schooling may have on returns to training, the wage return to training appears to be higher for those who are endowed with lower utility of training (types 2 and 3). This result is not surprising within a coherent behavioral model. A high positive correlation between utilities and returns (to training) would most likely imply a very large incidence of post-schooling training. Obviously, the frequency of post-schooling training predicted by the model is naturally limited by the observed frequencies in the data, since structural parameters must maximize the likelihood of observing the data (including training decisions frequencies).

6 Discussion

We now discuss a few issues that are naturally raised in the present context. In particular, we consider (i) the degree of heterogeneity bias that would prevail if we estimated the wage equation by OLS, (ii) the dynamic effects of schooling on post-schooling wage growth (the return to training), and (iii) the effect of an increase in training costs (tuition).

6.1 Heterogeneity Bias

One of the advantages of estimating a structural dynamic model with unobserved heterogeneity is that the estimated parameters of the Mincer wage regression are interpreted as causal effects net of heterogeneity bias. For instance, in the context of a model with endogenous schooling decisions and exogenous post-schooling wage growth, a structural model offers an explicit opportunity to quantify the classical "ability bias".

In the present model, the OLS bias is a complicated quantity. The difference between OLS estimates and their true population counterpart depends on multi-variate heterogeneity (individual wage intercept, individual returns to college, and individual returns to training). It is therefore interesting to compare our estimates of the return to training and education (and possibly interactions) with parameters from an OLS regression.

Overall, the degree of heterogeneity bias appears to be important. The OLS estimate of the effect of college on earnings - equal to 0.91 - is about 6 times as high as the average type-specific estimates. On the other hand, the OLS estimate of the return to training, around 0.13, appears to be only slightly smaller than the true average return to training (about 0.14).

Interestingly, the OLS estimate of the effect of schooling on the return to training is negative and therefore does not suggest any complementarity. The estimate, which is equal to -0.33, when compared to the average type-specific structural estimate (0.10), may be interpreted as evidence in favor of important heterogeneity bias.

6.2 The Explanatory Power of Schooling on Individual Returns to Training

As noted earlier, differences between individual returns to training of college graduates and high school graduates should be affected by the causal effect of schooling on training returns as well as a composition effect. To verify this claim and in order to quantify the importance of complementarity, we regress individual-specific returns to training on a college completion indicator. To compute individual returns, we needed to simulate schooling outcomes by type so as to recover the relationship between individual endowments and schooling.

Perhaps surprisingly, this regression indicates that the difference in returns between college graduates and high school graduates is around -0.24. In other words, high school graduates have higher returns to training, even though the true causal effect is positive 0.10 (on average). This indicates the importance of self-selection. If many individuals endowed with a high level of labor market skills (including a high return to business training) stop at high school, the results may display a large differential between college and high school graduates in terms of returns to business training. Yet, this differential is not inconsistent with a positive causal effect of education and training.

This regression also allows us to decompose the variance of individual returns to training into two components; unobserved skills and accumulated

schooling. The resulting R^2 , equal to 0.26, indicates that differences in schooling explains one fourth of the distribution of individual wage returns to training. The rest is explained by individual specific permanent endowments. This appears to be true whether or not we condition on individual endowments.

This result is consistent with the importance of persistent unobserved heterogeneity displayed in the distribution of individual endowments. As the degree of cross-sectional dispersion in individual skills and tastes is relatively high, increasing the level of schooling for a randomly selected individual will only have a limited effect on the return to training. This type of result is common in structural models that allow for individual heterogeneity.¹⁶

6.3 Increasing Training Tuition

In order to illustrate the role of heterogeneity in training enrolment, we simulated an increase of DKK 10,000 in training tuition.¹⁷ The reactions to such a hypothetical policy depend on a number of individual-specific factors, such as non-pecuniary utility from attending training and the financial benefits of such activities. On average, the fee increase resulted in a drop in (predicted) enrolment by about 25 percent. The decrease in enrolment varies considerably across individual types, with Type 3 showing the largest drop (26.1 percent) and Type 1 individuals showing the smallest drop (23.3 percent). From the structural estimates, we know that Type 2 individuals receive the least non-pecuniary utility from attending Part II of the training program while Type 4 individuals receive the most utility. At the same time, the wage gain from training is lowest for Type 4 individuals and significantly higher for Type 2 individuals. Hence, in this particular example, the difference in reactions to a fee increase is mainly driven by differences in non-monetary utility from attending training.

Obviously, these results are highly dependent on the partial equilibrium nature of our model. In a general equilibrium framework, increasing access to post-schooling training would not only affect the composition of the population of participants, but would also affect wage and training policies adopted by firms. As a result, these effects should really be understood as upper bounds.

¹⁶For instance, Keane and Wolpin (1997) document that increasing college attendance subsidies would be relatively ineffective for the US, precisely because school attendance is primarily explained by individual unobserved factors that are already determined by age 16.

¹⁷In 2007 prices.

7 Conclusion

In this paper, we have investigated the causal effect that schooling may have on subsequent wage gains following training. Specifically, we investigated the complementarity hypothesis and found that training seems to be slightly more beneficial to college graduates than to high school graduates. The result of significant complementarity found here does not apply to all other types of training and as such the results cannot be interpreted as evidence of complementarity in general. The key contribution of this paper is the proposed model, which yields consistent estimation of returns to both education and training without imposing separability between schooling and training.

Obviously, the results reported herein apply to a context where the post-education training course itself has an academic component. In the present context, education and training seem to represent activities that are close in nature.

At this stage, several questions remain unanswered. While education has a small causal effect on the wage gain associated to Part II of the program, it may still be true that it has major effect on the probability of entering, or completing, Part I. If so, it appears important to model training decisions from the beginning of the program.

It should be understood that we have investigated only one type of endogenous post-schooling activities (namely training) by which education can affect post-schooling wage gains. However, within a dynamic skill accumulation framework, accumulated education can affect (positively or negatively) the psychic cost of a large number of other post-schooling activities, or increase its associated wage gain. As the range of post-schooling investment opportunities is wide, it is unlikely that the econometrician can access data on all relevant investment outcomes, cf. Rosenzweig and Wolpin (2000). This conjecture presents major challenges to applied econometricians. In the presence of any type of post-schooling dynamics, the logic behind IV strategies collapse, and structural modeling becomes more complicated.¹⁸ Our plan is to address this fundamental issue in the near future.

¹⁸See Belzil and Hansen (2008).

8 References

- Belzil, C. (2007). Testing the Specification of the Mincer Wage Equation, *IZA Discussion Paper* 2650, forthcoming in *Annals of Economics and Statistics*.
- Belzil, C. and Hansen, J. (2002). Unobserved Ability and the Returns to Schooling. *Econometrica*, **70** (5), 2075-2091.
- Belzil, C. and Hansen, J. (2003). Structural Estimates of the Intergenerational Education Correlation. *Journal of Applied Econometrics*, **18**, 679-696.
- Belzil, C. and Hansen, J. (2007). A Structural Analysis of the Correlated Random Coefficient Wage Regression Model. *Journal of Econometrics*, **140** (2), 827-848.
- Belzil, C. and Hansen, J. (2008) Calibration and IV estimation of a Wage Outcome Equation within a Dynamic Environment, *IZA Discussion Paper* 2994.
- Carneiro, P. and Heckman, J.J. (2002). The Evidence on Credit Constraints in Post-Secondary Schooling. *The Economic Journal*, **112**, 705-734.
- Cunha, F. and Heckman, J.J. (2007). The Technology of Skill Formation. *NBER Working Paper* 12840.
- Heckman, J.J., Lochner, L. and Todd, P (2003). Fifty Years of Mincer Earnings Regressions. *NBER Working Paper* 9732.
- Heckman, J.J.. and Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, **52**, 271-320.
- Heckman, J.J. (2000). Policies to foster human capital. *Research in Economics*, **54**, 3-56.
- Heckman, J.J. and Vytlacil, E. (2007). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. Manuscript.

Imbens, G.W. and Angrist, J.D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica* **62** (2), 467-475.

Keane, M. P. (2006). Structural vs. Atheoretic Approaches to Econometrics. *Journal of Econometrics (In Press)*.

Keane, M.P. and Wolpin, K.I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, **105** (3), 473-522.

Knudsen, E.I., Heckman, J.J., Cameron, J.L. and Shonkoff, J.P. (2006). Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce. *NBER Working Paper* 12298.

Krueger, A.B. (2003). Inequality, Too Much of a Good Thing. Ch. 1 in Heckman, J.J. and Krueger, A.B. (eds.) *Inequality in America - What Role for Human Capital Policies?* The MIT Press Cambridge.

Meghir, C. and Palme, M. (2005). Educational Reform, Ability and Family Background. *American Economic Review* **95** (1), 414-424

Ministry of Education (2000). Takteftersyn - Hovedrapport. *Undervisningsministeriet, Institutionsstyrelsen*. (In Danish).

Rosenzweig, M.R. and Wolpin, K.I. (2000). Natural 'Natural' Experiments in Economics. *Journal of Economic Literature*, **38** (4), 827-874.

Shaw, K. L. (1989). Life-Cycle Labor Supply with Human Capital Accumulation. *International Economic Review*, **30** (2), 431-456.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

A Appendix: The Solution Algorithm and the Log-likelihood

A.1 The Solution Algorithm

The value function of going into training can be written as (omitting individual index i)

$$V_t^{TR}(s_t) = \bar{U}_t^{TR} + \beta E[V_{t+1}(s_{t+1}) \mid s_t, d = 1] + \varepsilon_t^W + \varepsilon_t^{TR}, \quad (\text{A1})$$

and the value function of not going into training is given by

$$V_t^W(s_t) = \bar{U}_t^W + \beta E[V_{t+1}(s_{t+1}) \mid s_t, d = 0] + \varepsilon_t^W, \quad (\text{A2})$$

where \bar{U}^{TR} signifies the deterministic part of the utility function.

The probability that an agent enters into the second part of the training program in period t equals the probability that $V_t^{TR}(s_t) > V_t^W(s_t)$ or equivalently that $\varepsilon_t^{TR} > \bar{V}_t^W(s_t) - \bar{V}_t^{TR}(s_t)$, where \bar{V}^{TR} denotes the deterministic part of V^{TR} , i.e. $\bar{V}_t^{TR}(s_t) = \bar{U}_t^{TR}(s_t) + \beta E[V_{t+1}(s_{t+1}) \mid s_t, d = 1]$, and similarly for \bar{V}_t^W .

In order to evaluate the expected value function, we need to integrate out the period-specific shocks ($\varepsilon_t^W, \varepsilon_t^{TR}$). As the shocks are assumed to be *iid* normally distributed, we can use the law of iterated expectations and write (omitting time index t) $E(V^{TR}) = \Pr[V^{TR}(\cdot) \geq V^W(\cdot)] \times E[V^{TR} \mid V^{TR} > V^W]$. These two terms can in turn be written as:

$$\begin{aligned} \Pr[V^{TR} \geq V^W] &= \Pr \left[\frac{\varepsilon^{TR}}{\sigma_{TR}} > \frac{\bar{V}^W - \bar{V}^{TR}}{\sigma_{TR}} \right] \\ &= \Phi \left[\frac{\bar{V}^{TR} - \bar{V}^W}{\sigma_{TR}} \right] \end{aligned} \quad (\text{A3})$$

and

$$E(V^{TR} \mid \varepsilon^{TR} > \bar{V}^W - \bar{V}^{TR}) = \bar{V}^{TR} + \left[\frac{\phi \left(\frac{\bar{V}^W - \bar{V}^{TR}}{\sigma_{TR}} \right)}{1 - \Phi \left(\frac{\bar{V}^W - \bar{V}^{TR}}{\sigma_{TR}} \right)} \right] \times \sigma_{TR}. \quad (\text{A4})$$

see e.g. Wooldridge (2002).

Similar derivations for $E(V^W) = \Pr[V^{TR}(\cdot) < V^W(\cdot)] \times E[V^W | V^{TR} < V^W]$ yields

$$\Pr[V^{TR}(\cdot) < V^W(\cdot)] = 1 - \Phi \left[\frac{\bar{V}^{TR} - \bar{V}^W}{\sigma_{TR}} \right] \quad (\text{A5})$$

and

$$E(V^W | \varepsilon^{TR} > \bar{V}^W - \bar{V}^{TR}) = \bar{V}^W + \left[\frac{\phi \left(\frac{\bar{V}^{TR} - \bar{V}^W}{\sigma_{TR}} \right)}{1 - \Phi \left(\frac{\bar{V}^{TR} - \bar{V}^W}{\sigma_{TR}} \right)} \right] \times \sigma_{TR} \quad (\text{A6})$$

Collecting A3-A6 we can compute the $EMAX[\cdot]$ given by

$$\begin{aligned} EMAX[V^{TR}(\cdot), V^W(\cdot)] &= \Pr[V^{TR}(\cdot) \geq V^W(\cdot)] \times E[V^{TR} | V^{TR} > V^W] + \\ &\Pr[V^{TR}(\cdot) < V^W(\cdot)] \times E[V^W | V^{TR} < V^W]. \end{aligned} \quad (\text{A7})$$

A.2 The Log-Likelihood

As outlined in section 4.2, the log-likelihood can be written as $\ln L = \ln \sum_{k=1}^4 p_k \cdot L_1(k) \cdot L_2(k) \cdot L_3(k) \cdot L_4(k)$.

(a) $L_1(k)$ is the probability of observing the actual choice path for a given individual, conditional on the unobserved heterogeneity components. $L_1(k) = \Pr[(d_1(k) = 1), (d_2(k) = 1), (d_3(k) = 1), (d_4(k) = 1)]$.

(b) $L_2(k)$ is the probability of completing training in 2 or 3 periods conditional on entering into training. We may write this part of the likelihood as

$L_2(k) = \mathbf{I}_{(enter\ part\ II)}(\Phi[\xi_3]^{p_3} \cdot (1 - \Phi[\xi_3])^{(1-p_3)})$, where $\Phi[\xi_3]$ denotes the probability that training lasts 3 periods (conditional on ever entering), and where p_3 denotes an indicator=1 if actual duration was 3 or more periods and zero otherwise. The probability is parameterized as $\xi_3 = \xi_{30} + \xi_{31}(\text{Duration of Part I}) + \xi_{32}(\text{Experience}_j)$.

(c) $L_3(k)$ is the joint probability of being in the labor market in year $\tau + 1$, at observed wage $w_{\tau+1}$. Hence $L_3(k) = f(w_{\tau+1}) \cdot f(w_{\tau+2}) \cdot f(w_{\tau+3}) \cdot f(w_{\tau+4}) \cdots f(w_{nper_i})$, where $nper_i$ denotes the number of periods in which individual i is observed.

(d) $L_4(k)$ is the probability of entering with a college degree as initial value. This part of the likelihood can be written as $(\Phi [\xi_{college}]^{college} \cdot (1 - \Phi [\xi_{college}]^{(1-college)}))$, where $\Phi [\xi_{college}]$ denotes the probability of having a college degree upon entering. This probability is parameterized as $\xi_{college} = \xi_{c0} + \xi_{c1}(\text{mothers education}) + \xi_{c2}(\text{fathers education})$.

Note that when wage densities are simulated, we use all available wage information, i.e. taking into account that the panel is unbalanced.

B Appendix : Figures and Tables

Figure 1 Gross Median Wages 1982-2003, by Education and Training

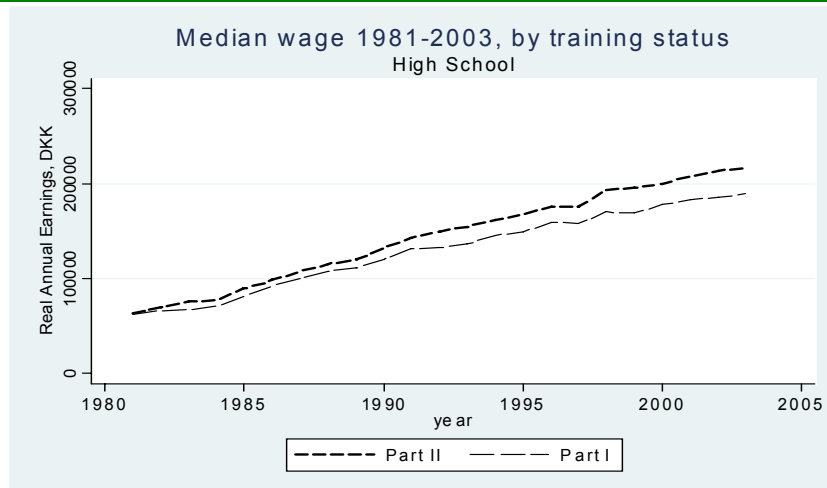
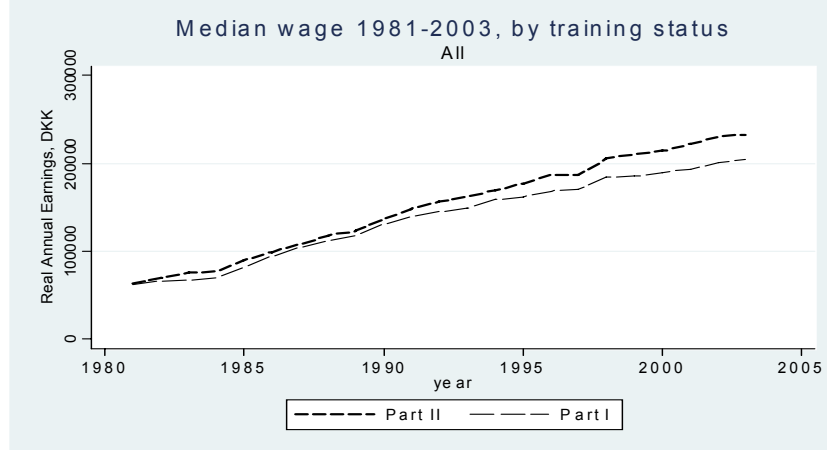


Figure 2 Years between Completion of Part I and Enrollment in Part II, by education

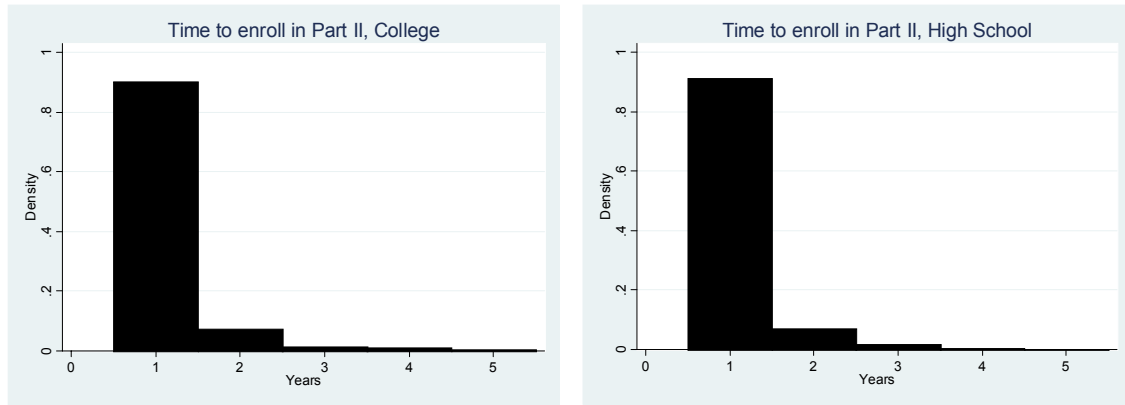


Figure 3 Duration of Training, by education

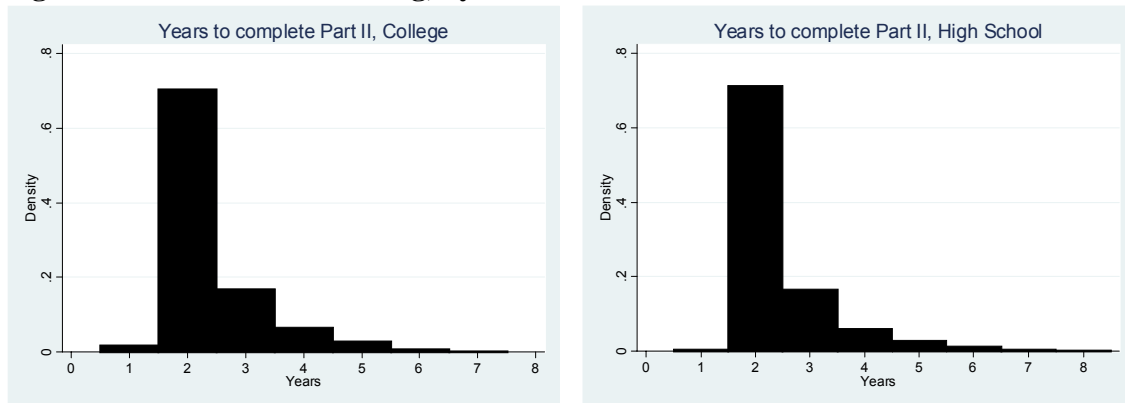
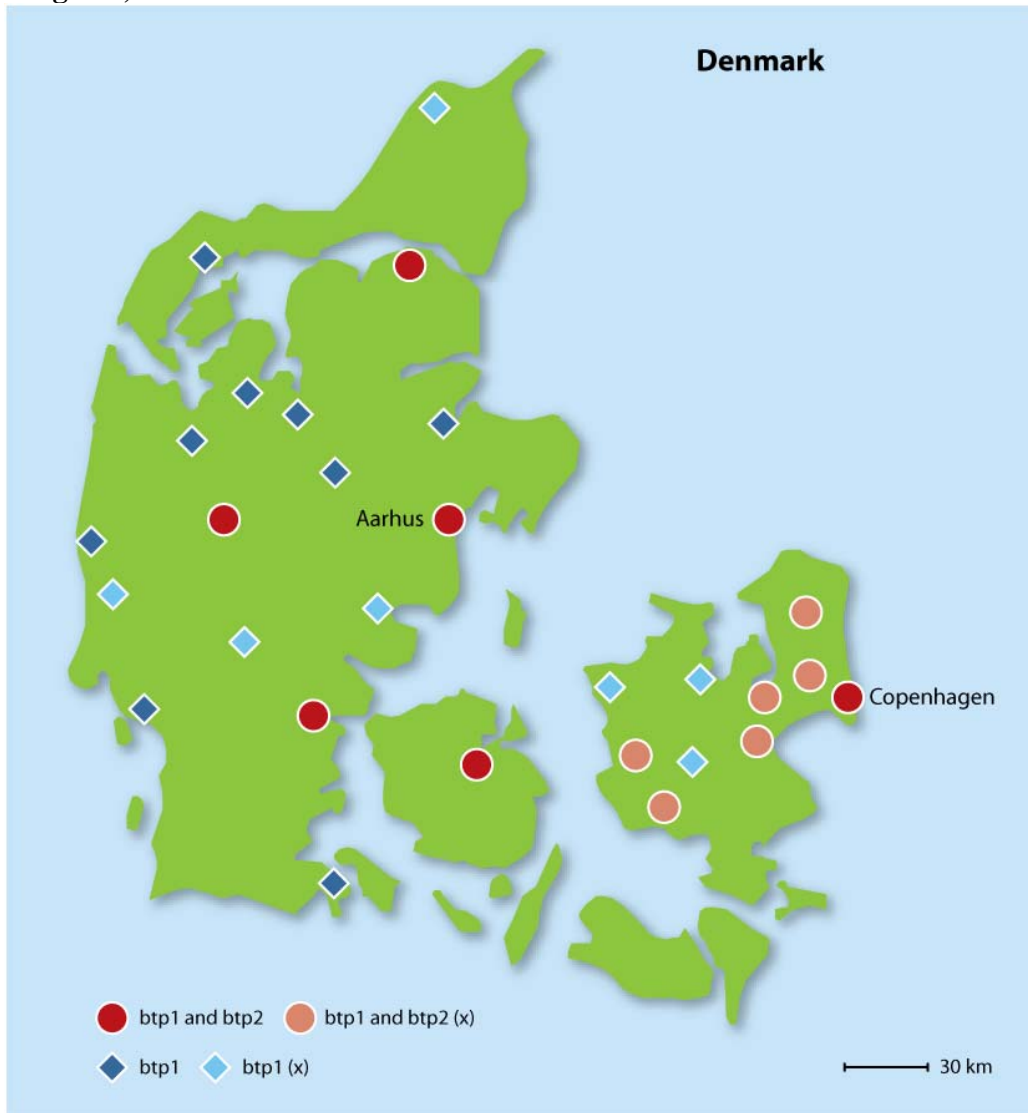
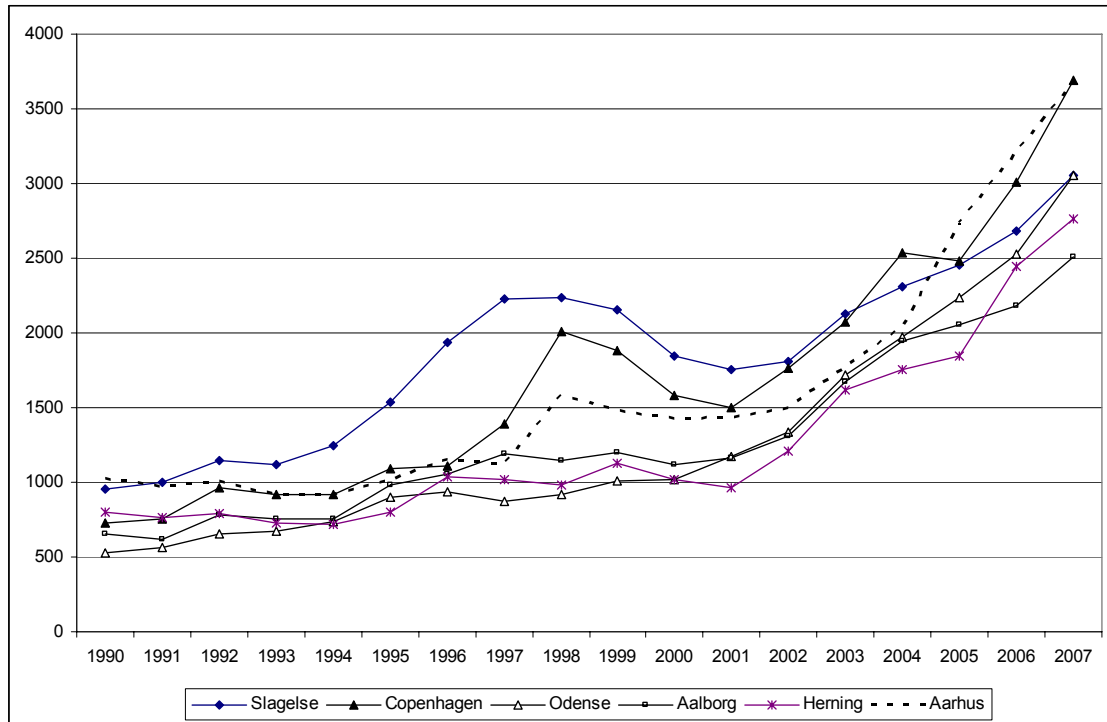


Figure 4 Location of Institutions Offering Part I and Part II of the Training Program, 1980-2004



Notes: “btp1 (x)” indicates that Part I of the program was only offered in some of the years. “btp1 and btp2 (x)” indicates that Part I was offered throughout all the entire span of years while Part II was offered only in some of the years.

Figure 5 Tuition Fee for Part II 1990-2007, by main institution (USD 1980 prices)



Notes: Prior to 1990 the training was fully financed by the public. To get 2007 prices, multiply with a factor 2.7, i.e. the tuition fee in 2007 was about USD 7,000-10,000 in real 2007 prices.

Table 1 Number of Individuals, by Education and Training

	Part II Training		Total
	No	Yes	
High School	565	2,089	2,654
College	222	635	857
	787	2,724	3,511

Note: Conditional on experience=7 years. Altogether there are 3,527 individuals in the sample, and about 54,000 observations.

Table 2 Non-pecuniary Utility of Attending Part II of the Training

	Parameter Estimate	Std. Err.	t-value
constant type 1	18.509	0.055	336.356
constant type 2	16.638	0.039	422.603
constant type 3	17.378	0.097	179.569
constant type 4	18.689	0.085	221.249
local unemployment rate	0.021	0.006	3.516
cohort	0.034	0.018	1.945
copenhagen (0/1)	-0.074	0.017	-4.362
age when completing Part I	-0.198	0.004	-46.063
single	-0.054	0.026	-2.037
number of children	-0.186	0.040	-4.659
duration of Part I training	-0.011	0.002	-6.375
tuition (1,000 DKK)	-0.151	0.020	-7.572
college (0/1), type 1	-0.818	0.138	-5.944
college (0/1), type 2	-1.385	0.198	-7.001
college (0/1), type 3	-0.199	0.010	-20.131
college (0/1), type 4	-0.794	0.137	-5.811

Notes: *copenhagen* and *single* are indicator variables. *duration of Part I training* is measured in months and *tuition* is measured in 1,000 DKK (1980 prices).

Table 3 Wage Equation Parameters

	Parameter Estimate	Std. Err.	t-value
wage constant, type 1	11.556	0.010	1157.045
wage constant, type 2	10.830	0.013	856.604
wage constant, type 3	11.930	0.016	746.161
wage constant, type 4	11.275	0.008	1334.856
college, type 1	0.174	0.026	6.617
college, type 2	0.005	0.030	0.182
college, type 3	0.099	0.027	3.643
college, type 4	0.168	0.019	9.094
experience	0.083	0.002	52.452
experience-squared	-0.002	0.000	-19.945
training Part II, type 1	0.156	0.010	15.895
training Part II, type 2	0.243	0.014	17.804
training Part II, type 3	0.282	0.014	20.510
training Part II, type 4	0.091	0.009	10.480
<i>complementarity term</i>			
training Part II x college, type 1	0.087	0.023	3.796
training Part II x college, type 2	0.222	0.027	8.154
training Part II x college, type 3	-0.066	0.071	-0.922
training Part II x college, type 4	0.119	0.016	7.388

Table 4 Remaining Parameter Estimates

	Parameter Estimate	Std. Err.	t-value
$\sigma\text{-}\xi$	0.907	0.020	44.581
$\sigma\text{-}w$	0.425	0.001	294.388
<i>prob(duration > 2 years)</i>			
duration constant, type 1	-1.676	0.060	-28.007
duration constant, type 2	-1.850	0.009	-216.172
duration constant, type 3	-1.303	0.069	-18.870
duration constant, type 4	-1.543	0.064	-24.028
experience	-0.189	0.018	-10.463
duration of Part I training	0.009	0.003	3.322
<i>prob(college degree)</i>			
college constant, type 1	-1.837	0.043	-43.118
college constant, type 2	-1.563	0.058	-26.968
college constant, type 3	-1.382	0.066	-20.912
college constant, type 4	-1.770	0.043	-41.371
mothers education	0.120	0.018	6.879
fathers education	0.063	0.014	4.446

Table 5 Type Probabilities

	q_k	Std. Err.	t-value	p_k
type 1	-0.337	0.069	-4.907	35.49%
type 2	-1.843	0.090	-20.571	7.87%
type 3	-1.968	0.108	-18.152	6.94%
type 4	0.000 (normalized)		---	49.70%

Where

$$p_k = \frac{\exp(q_k)}{\sum_{j=1}^4 \exp(q_j)}$$

C Appendix: Model Validation

Table C1 Observed and Predicted Enrolment, in percent

Period	Observed	Model prediction
1	66.6	57.9
2	4.17	3.6
3	1.02	3.4
Sum	71.8	64.9