Master thesis for the Master of Economic Theory and Econometrics Degree

Educational Wage Returns in Norway

A Sector Selection Model with Endogenous Schooling

Manudeep Singh Bhuller

November 2008

Department of Economics

University of Oslo

Preface

This thesis has been written as part of the Research Program for the Study of Welfare Policies at Statistics Norway funded by the Research Council of Norway.

It has been an interesting, challenging, and certainly instructive process to write this thesis. I would like to thank my supervisor Rolf Aaberge for introducing me to the concepts of selection bias and counterfactual distributions. I certainly value the encouraging and inspiring discussions that I was allowed to be part of. Secondly, I would like to thank Magne Mogstad for helpful comments and invaluable help with the data analysis. This work would not have been possible without their professional and inspirational guidance. I would also like to thank Harald Lunde for giving me access to the necessary data and quick feedback on any enquires.

Apart from the above mentioned, I would also like to thank my colleagues and friends at Statistics Norway for creating a pleasant environment to work in and enabling me to feel as a part of their team, even during my stay abroad as an exchange student. I will also like to thank all the people that have supported and encouraged me through this process. All amendments and advices have been of value.

Oslo, November 11th

Manudeep Singh Bhuller

Contents

1. Introduction

Approximately 30 000 students graduate from universities and colleges in Norway each year. It is commonly believed that graduates have more opportunities in the labour market and are likely to enter high-paid jobs. An individual’s decision to pursue higher education can be seen as an attempt to enhance productive capabilities. In a competitive market, workers are paid according to their productivity. Human capital theory predicts that individuals make systematic investments in higher education by forgoing earnings early in their careers in order to increase expected future lifetime earnings. However, making causal statements on the relationship between education and earnings is difficult. Education is subject to individual choice and those who take part in higher education may differ from those who do not. This study aims at analysing the variation in pecuniary returns to education across different subpopulations using detailed information on individual earnings, education, and a number of other observable characteristics, taking into account the endogeneity of educational choices and sector participation. The human capital earnings framework developed by Becker (1964) and Mincer (1958, 1974) is utilized.

Our approach has the notable features that it aims at estimating well-defined parameters of interest that can under certain conditions be given causal interpretations, and secondly, unlike most other studies on wage differentials, considers a wider set of endogeneity problems. Using microeconometric methods, we formalize a number of selection problems in the earnings regression framework. A question we pose in the empirical analysis is whether and to what extent our estimates of returns to education change when we allow for one or more sources of endogeneity. In addition to the well-know problem of non-random labour force participation, individual choices of schooling and occupational sector are also considered. Greater functional form flexibility is allowed through non-linear returns to education across years of schooling, educational fields and interaction-effects between schooling and work-experience. Unlike earlier studies on educational returns in Norway that are restricted to full-time workers because of insufficient data on individual working hours, we consider the returns to education on hourly wages for a much larger population. The wage returns to education are also likely to be better indicators of the relative increase in worker productivity, because crude measures of full-time work have been used in the earlier studies.

Our primary data source is Statistics Norway's Wage Statistics, comprising a series of cross-sectional micro datasets containing around 1.2 million individuals for 2001. The data on public sector employees is taken from administrative registers covering all employed workers. For employees in privately registered enterprises, the data are based on an annual stratified random sampling of enterprises conducted by Statistics Norway. Depending on the worker's employment industry, between 30 and 90 per cent of private sector employees are covered. The datasets contain information on years of schooling, industry, educational field, age, sex, and municipality of residence, apart from a number of earnings components such as average monthly paid salaries, variable additional allowances, bonuses, commissions, and contracted weekly working hours. On the basis of these measures, we are able to construct gross average hourly wages that are then used in the wage regressions. All estimations are performed in SAS 8.2 and STATA 10.

We find significant variation in the estimated returns to education across gender and sector, with OLS estimates ranging between 4.8 and 7.8 percent. However, there is strong evidence of non-linearity in returns across schooling level and education fields for all population groups we consider. Interaction effects between education and experience are also present, leading to a rejection of the standard log-linear Mincer wage equation. Selection effects are found to be significant, though of a moderate size. In general the OLS estimates are upward biased for the public sector, suggesting positive sorting into the labour market. However, our estimates indicate a negative selection bias for the private sector. The magnitude (and the direction) of the selection bias varies considerably across different educational groups. When taking into account endogeneity of schooling, we find that the OLS estimates are significantly downward biased, and that the bias increases further when we account for sector-participation.

This study comprises of a methodological and an empirical section. Section 2.1 outlines the theoretical framework, followed by Section 2.2 providing an overview of the various theories of investment in human capital. In Section 2.3, various econometric problems related to earnings regressions are outlined. Section 3 contains the empirical analysis. Section 3.1 starts with an introduction to the estimation problem, earlier studies and data sources. OLS estimates are given in Section 3.2, while Section 3.3 considers the problems of endogeneity and selection bias. Section 4 concludes this study. Variable definitions, descriptive statistics and estimation results are given in the appendix.

2. Theory and Method

2.1 Introduction

Economists use the concept of human capital to denote an individual's stock of productive knowledge, acquired through formal education and post-school learning, which enables him to indulge more abundantly in labour market activities.[1] Essentially, the human capital theory tells us that human beings perceive a positive relation between productive capacities and life-time earnings, and thus, being rational income-maximizing individuals, decide on whether to forego current earnings in order to improve their productive capacities and achieve higher future earnings prospects.[2] A credible strength of this approach is that it enables the researcher to calculate the potential stock of productive knowledge for an individual, based on proxy measures such as the costs of schooling, and analyse the incremental increases in earnings following increases in the human capital stock. This has given rise to a vast empirical literature in the field of labour economics documenting ‘rates of return’ to schooling, where the underlying hypothesis is that education is an investment that receives a pecuniary return in the labour market.

The conceptual framework of the theory of investment in human capital is based on the seminal works of Becker (1964, 1967), Ben-Porath (1967) and Mincer (1958, 1974). However, these initial contributions are motivated somewhat differently. The formal life-cycle model of investment in human capital can be traced to Ben-Porath, who analyses the choice of schooling and lifetime earnings in a dynamic optimisation framework. This structural model is further expanded in Rosen (1976) and Heckman (1976). However, these models are difficult to apply in an empirical analysis. Mincer (1958) assumes identical abilities and opportunities for all individuals and employs the principle of equalizing differences to derive a simple relation between earnings and schooling. This allows us to

---

[1] The term *productive knowledge* may here refer to skills, ideas, or earning power that an individual acquires. Rosen (1987) provides an overview of the concept of human capital. Blaug (1976) also includes improvement in health, information retrieval, job search and migration as sources of the overall accumulation of human capital. Unlike an improvement in health, which may directly increase the productivity of the individual, the last three components function indirectly through better matching of worker’s capacities to the demanded job skills.

[2] The signalling theory also postulates a similar relation between education and earnings. It says that education may have a value in the labour market because it may act as a signal of ability or other productivity enhancing factors, independent of the direct effect on productivity. We will ignore such effects in this analysis; see Riley (2001) for an overview.

easily locate the rate of returns to schooling. Meanwhile, Becker (1967) studies schooling choices and individual returns to schooling as the outcome of supply and demand interactions in a simpler structural model.

Despite these apparent conceptual differences between the theories, there is an underlying hedonic interpretation of the labour market. There are costs attached to undertaking schooling: *direct educational costs,* such as tuition fees, expenses on books, etc and *opportunity costs,* restricted mainly to foregone earnings during the schooling period. Thus, the worker would find additional schooling advantageous only if there is a compensation associated with higher education, such as sufficiently higher lifetime earnings. In the long-run, employers would be willing to give higher earnings to schooled workers only if they are more productive than less schooled workers. In a competitive equilibrium, the supply and demand for workers of each schooling level should equate and workers should have no incentives to alter their schooling level. The relationship between lifetime earnings and schooling is then determined in this equilibrium.[3]

In Section 2.2, some alternative approaches to the theory of investment in human capital are presented. A theory-consistent econometric specification of the human capital earnings function deduced by Mincer (1974) is discussed in Section 2.3. We also touch upon some methodological issues recurrent in the literature on schooling returns and present some recent contributions to the methodology of human capital earnings functions.

2.2 Models of human capital investment

2.2.1 The equalizing differences model

One of the earliest contributions to the human capital theory is Mincer (1958), who uses the principle of compensating differences to explain why persons with different levels of schooling receive different earnings over their lives. Occupations differ in the amount of schooling required. Meanwhile, individuals are assumed to be *ex ante* identical in terms of attributes that may affect their earnings and choices of schooling, such as abilities, opportunities, and preferences over non-pecuniary attributes of occupations and educational fields. However, individuals may choose to have different levels of schooling. The model also

[3] For a thorough discussion of hedonic earnings equations, see Rosen (1974) and Willis (1986, 527-529).

assumes that the credit markets are perfect, there are no direct schooling costs (i.e. all investment costs are time costs), the rate of return is exogenously given to the individual, no further human capital investments are undertaken after completion of schooling (i.e. no on-the-job training), the flow of individual earnings is constant throughout the working-life, there is zero depreciation in the human capital stock, and the span of working life is constant across various schooling groups. Changes in macroeconomics conditions that may affect the worker's productivity and earnings during the life cycle are also excluded. The main prediction of the model is that the present value of net lifetime earnings of an individual who has *s* years of education and one who for instance has zero years of education should equate in order to have multiple schooling outcomes as optimising strategies in the population.

The present value of lifetime earnings for an individual with *s* years of education (V_s) can be denoted by

$$V_s = \sum_{t=s}^{T+s} R^t Y_s = R^s \left[\frac{1 - R^{T+1}}{1 - R} \right] Y_s ,$$

where T is the constant span of working life, $R \equiv \frac{1}{1+r}$ is the discount factor with r being the externally determined interest rate, Y_s denotes annual earnings of an individual with *s* years of schooling, while t is the time unit measured in years. Similarly, the present value of lifetime earnings for an individual with zero years of education V_0 is found. Equating it to V_s yields $Y_s / Y_0 = (1+r)^s$ which after taking logarithms and approximating becomes

$$\ln Y_s = \ln Y_0 + s \ln(1+r) \approx \ln Y_0 + sr. \qquad (2.1)$$

Given the assumptions summarized above, equation (2.1) follows almost directly from a tautology of present value of earning streams. Thus, the percentage increase in earnings is proportional to the absolute differences in the time spent at school, with *r* as the coefficient of proportionality. This leads us to the prevalent interpretation of the semi-elasticity of yearly earnings with respect to schooling as the *internal rate of return* that equates lifetime earnings streams for different educational choices and should be equal to the interest rate used by individuals as their discount rate in equilibrium. Apart from the equalization of lifetime earnings, there is little economic substance in this procedure that may enable us to explain the earnings-schooling relationship. For instance, the model does not tell us anything about the circumstances that induce individuals to engage in skills accumulation or *why* they chose their given schooling level.

2.2.2 The accounting identity model

Mincer (1974) develops another model that is motivated by life cycle dynamics of earnings, but yields an algebraically similar empirical specification of the earnings equation. The model emphasizes the relationship between observed earnings, potential earnings, and human capital investment, for both formal schooling and on-the-job investment. Individuals are *ex ante* heterogeneous and there is also variation in the returns to schooling in the population. As earlier, this model also suggests that post-school investment has some alternative costs that need to be accounted for. We may define the observed earnings for an individual with a given education some period *t* after he has joined the labour force as $Y_t \equiv E_t - C_t$, where E_t denotes earning capacity (or what the worker would have earned had there been no post-school investment in the current period), while C_t are the resources devoted to on-the-job training in period *t*. Post-school investment yields a return only in the subsequent period. Thus, potential earnings can be expressed as

$$E_t = E_{t-1} + rC_{t-1} = E_{t-1}(1 + rk_{t-1}) = E_0 \prod_{j=0}^{t-1} \left(1 + rk_j\right),$$

where k_j is defined as the ratio of investment costs C_j to gross earnings E_j in period *j* and can be interpreted as the *time-equivalent* value of investment costs.[4] Introducing the investment ratio in this manner implies that workers allocate all of their time either on work or post-school training.[5] The third equality follows from backwards recursion. Assuming that the rate of return to post-school investment is constant over the life span and investment costs equate earning capacity during schooling (so that $k_j = 1$ for $j \le s$), we can take logarithm of the above expression and express earning capacity in period *t* as

$$\begin{aligned} \ln E_t &= \ln E_0 + s \ln(1 + r_s) + \sum_{j=s}^{t-1} \ln(1 + r_p k_j) \\ &\approx \ln E_0 + s r_s + r_p \sum_{j=s}^{t-1} k_j , \end{aligned}$$

where r_s is the schooling rate of return and r_p can be interpreted as the return to the cumulative time-equivalent post-school investments before period *t*. The last approximation is for small r_s and r_p. To simplify matters ever more, it may be assumed that investment ratio

---

4 Using the definition of observed earnings and the equation for potential earnings and post-school investment, we can easily deduce the relation $Y_j = Y_{j-1} + (1+r)C_{j-1} - C_j$.

5 Problems concerning interpersonal valuation of leisure and allocation of time are not considered. We assume exogenously fixed labour supply in rest of the analysis.

declines linearly with years of experience: $k_{s+p} = k_0\left(1-\frac{p}{P}\right)$ where $p = t - s \geq 0$ is the amount of work experience at year *t*, $k_0 > 0$ is the initial investment ratio and *P* is the total period of positive net investments (assumed to be independent of years of schooling). Under these assumptions, the relationship between potential earnings, schooling and experience is given by

$$\ln E_{s+x} = \ln E_0 + r_s s + \left(r_p k_0 + \frac{r_p k_0}{2P}\right)p - \frac{r_p k_0}{2P}p^2.$$

Recalling that observable earnings are potential earnings less investment costs, the relationship can also be expressed as the following well-known Mincer earnings function

$$\begin{aligned}\ln Y_{s+p} &= \ln E_{s+p} + \ln\left(1 - k_0\left(1-\frac{p}{P}\right)\right) \approx \ln E_{s+p} - k_0\left(1-\frac{p}{P}\right)\\ &= \ln E_0 - k_0 + r_s s + \left(r_p k_0 + \frac{r_p k_0}{2P} + \frac{k_0}{P}\right)p - \frac{r_p k_0}{2P}p^2.\end{aligned} \tag{2.2}$$

While the schooling function in (2.1) follows by equalizing of lifetime earnings, the earnings function in (2.2) enshrines a hypothesis about the optimal post-school investment path. The schooling model given in (2.1) can be regarded as a simplified version of (2.2) under the conditions that there is no post-school investment, the only costs of schooling are foregone earnings, and that each individual faces a constant interest rate. The model (2.2) presents some new insights. Wage dispersion across individuals is now explained by variation in schooling and post-school training. The schooling coefficient should now be interpreted as the *average rate of return* across all schooling investment and not, in general, as a marginal internal rate of return that is appropriate for evaluating the optimality of educational investments. It must also be noted that the intercept and slope coefficients depend on parameters such as (E_0, k_0, r_p, r_s) that may vary across individuals. This motivates the random coefficients specification of the earnings function given in Mincer (1974). It should also be noted that the composite experience coefficients cannot be given a simple intuitive interpretation (such as "returns to years of experience") and would depend on the returns to cumulative time-equivalent post-school investments r_p, initial investment ratio k_0, and the total period of positive net investments *P*.

2.2.3 Market for human capital investments

Becker (1967) studies the individual choices of schooling as outcomes of the interaction between supply and demand for investments in human capital, considering factors such as individual ability, tastes, and family wealth as the determinants of schooling choices. Underlying in the analysis is a concept of *optimal investment* in human capital related to marginal rates of returns and costs of undertaking the investments. The credibility of this approach is that it provides a framework within which it is possible to study the circumstances that induce workers to engage in skills accumulation. This enables the researcher to explicitly model heterogeneity in the individual schooling decisions.

The main assumptions of this model are that the earnings of a person that are unrelated to human capital investment (denoted by E_0 in the previous section) are independent of their schooling choices and can be neglected and, secondly, human capital is homogeneous in the sense that all units are perfect substitutes in production for each other and thus add the same amount to earnings. The model also abstracts from the problems of uncertainty in future outcomes. For a particular person, the demand for human capital is given by the marginal benefit of accumulating human capital, measured by the rate of return on the educational investment. Meanwhile, the effective marginal financing cost of human capital accumulation is measured by the rate of interest and represents the supply curve of capital.

The marginal rate of return depends on the marginal returns and the marginal production costs of investment, and is assumed to decline with the amount of human capital investment. The embodiment of human capital and the limitation on human memory capacity and physical size are the main reason for the declining marginal rate of return. Producing additional capital results in increasing marginal production costs. This result is analogous to the concept of diminishing returns in capital theory. Another peculiar aspect of human capital investment is that an increase in the amount invested usually corresponds to an increase in the time spent investing. This leads to production costs because of foregone earnings during the investment period. Because of finite lifetimes, and therefore shorter lifespan to recover the potential returns, and higher time-valuation with increased capital accumulation, one may expect smaller marginal benefits and higher production costs at later stages of life. This provides a fairly strong case for declining marginal rate of returns.

The marginal cost of financing an additional unit of capital is increasing in the amount of human capital accumulated. The capital markets are assumed segmented with the existence of local and federal subsidies, transaction costs and legal limitations. Although certain sources of funds are cheaper than others, the amounts available to any person from the cheaper sources are usually rationed since the total demand for the funds tend to exceed their supply. The positive inclination of the supply curve illustrates the point that when a person wants to finance more of the human capital investment, the cheapest fund are utilized first, but eventually the individual has to use expensive sources.

The equilibrium outcome of this model is that agents select a path that maximizes the present value of the difference between the benefits and costs of undertaking schooling. However, since both the marginal benefits and financing costs of schooling depend on the path of capital accumulation, the demand and supply curves would neither be uniquely determined nor independent of each other. Becker (1967) assumes that own time and hired inputs are used in fixed proportions to produce human capital, and that a unit of hired inputs is available at a given price (foregone earnings) up to a certain maximum amount, beyond which no time is available at any price. This enables the interpretation of the educational investment at the intersection point between the demand and supply curves as the optimal accumulation path. Corresponding to the optimal accumulation path is the optimal investment period.

Within this general framework, the differences in total human capital investment and rates of return can be explained as the result of variation in individual capacities and/or opportunities. While personal abilities (or IQ) may affect the benefits or costs of producing human capital and thereby alter the demand conditions, differences in family background or other factors determining individual opportunities may lead to variation in financing costs and enter the analysis by changing the supply conditions. Becker’s contribution holds significant importance because it enables us to systematically formulate and analyse the labour market implications of the two prevalent approaches in the policy debate concerning schooling, namely *equality of opportunity* and *equality of ability*.

Rosen (1977) incorporates these basic concepts and presents a reinterpretation of the schooling model under the conditions that there are no post-school investments, constant interest rate, and that costs of schooling are fully captured by forgone earnings.[6] The simple version of Mincer’s schooling function in (2.1) can be regarded as defining a set of iso-wealth curves in $(\ln Y, s)$ with slopes equal to the rate of interest. The iso-wealth curve states that all

[6] See Willis (1986) for a discussion of Rosen’s model. Also, Card (2001) presents a model with utility maximization and consumption choice that yields a similar decomposition of earnings-variation into inequality of abilities and opportunities.

choices of years of schooling along this curve yield the same lifetime earnings. Rosen (1976) assumes that the relationship between schooling and earnings can be given by a continuous function of the form

$$\ln Y(s) = h(s; A), \tag{2.3}$$

where A is some measure of exogenously given economic ability. While the schooling function in (2.1) enshrines the idea that in equilibrium individuals have no incentives to deviate from their schooling choice, the function (2.3) represents a hypothesis about the technology of human capital production. [7] Underlying is a concept of homogenous labour, such that persons with more schooling and/or ability have more efficiency units of labour to rent in the market.

The optimal schooling choice is given by the following problem:

$$max\ V(s) = Y(s)\int_s^{T+s} e^{-rt}dt = \frac{Y(s)e^{-rs}\left(1-e^{-rT}\right)}{r}\ s.t.$$

$$Y(s) = \exp\left\{h(s;A)\right\}.$$

The first-order condition for this problem implies that the individual should continue schooling until the marginal rate of return is equal to the interest rate; $h_s^{'}(s;A) = r.$ Existence of an interior solution to this problem requires diminishing returns of schooling in producing embodied skills (i.e. $h_{ss} < 0$). The optimality condition gives schooling as a function of the individual's ability and the rate of interest, $s^* = s(A,r).$ By inserting this in (2.3) we get the optimal earnings, $\ln Y(s^*) = h(s(A,r);A).$ In general, the opportunities for financing investments in education vary across individuals. In order to capture this variation, we may allow the rate of interest faced by the individual to depend on a set of exogenous variables *Z*, such as family background and non-human wealth, so that optimal earnings are given by $\ln Y(s^*) = h(s(A,r(Z));A)$. Within this model, variation in observed earnings can be fully explained by inequality of abilities and opportunities in the population.[8]

[7] Implicit in the human capital production function is the assumption that ability influences the marginal rate of return to investment. If schooling and ability can be regarded as additively separable in (2.3), the optimal schooling choice would be independent of individual's abilities. In that case, the variation in schooling choices across individuals would be the result of variation in the rates of interest. The rate of return to schooling can then be identified without knowledge of individual abilities. See section 2.3 for problems concerning self-selection and problems of identification in the earnings function.

[8] A generalization of the concept of inequality of ability is *comparative advantage*, which allows for multidimensional variation in the skills. For instance, there is a possibility of occupation-specific skills variation among the workers. A related problem is that of *sorting* or dynamic self-selection, often coexisting in the presence of inter-temporal learning of skills by workers and employees. An important contribution to this literature is Gibbons et al (2005).

2.2.4 Life-cycle determination of earnings

The first formal representation of the human capital accumulation process as an intertemporal optimisation problem is due to Ben-Porath (1967). This structural model of life-time earnings is further expanded in Rosen (1976) and Heckman (1976). Surveys of this literature are presented in Weiss (1987) and Cunha et al (2006). As static models, the life-cycle models of earnings determination state that workers can influence their earnings through various investment activities, such as schooling or on-the-job training. By offering current earnings, they hope to increase future earnings potential. What differentiates life-cycle models from the static framework of Becker-Mincer models is the explicit modelling of the dynamics of human capital accumulation and its relation to the individual earnings profiles, where technological or biological processes such as human capital depreciation or aging of the human body are accounted for. The theory enables us to explain a number of regularities in the earnings literature, such as a life-cycle earnings profile that increases at early ages and declines towards the end of the working period.

The simple model assumes that all firms are identical and compete for workers by offering job opportunities that specify both wage and working hours. Workers have some initial human capital endowment, but are able to attain on-the-job training. Thus, there is a trade-off between current earnings and human capital accumulation and thereby future earning potential. Firms decide the number of workers they should employ, while the workers decide on the hours of work and the intensity of human capital accumulation.[9] Workers are distinguished by their human capital endowment and the time they spend at the firm. However, it is assumed that workers with different skills are perfect substitutes in the production of the composite good.

Equilibrium in this model is attained when the marginal product of hiring a worker is equal to the cost. In order for this to occur, the rental rate on human capital must adjust. When each worker has allocated his time among jobs with varying training intensity levels given his current human capital endowment, an efficient earnings frontier $Y = Y(K, \dot{K}, h)$ is generated, where Y is current earnings net of explicit training costs, h is time spent at work or at school, K is the current human capital endowment and $\dot{K}$ is the change in human capital following

[9] Schooling investment can be considered a special case of this model when the human capital accumulation intensity is equal to one and current earnings are equal to zero.

the current investment. The human capital accumulation process can be captured in a production function, $\dot{K} = G(K, h, x),$ where x represents the training intensity level. Hence a worker with human capital endowment K, who works h hours in a firm with training intensity level x, will earn $(1-x)RKh,$ where R is the rental rate of capital. In order to simplify the model, we may regard labour supply as exogenous. Thus, we focus only on the allocation of time in the market, taking the total amount of non-leisure time as predetermined. The worker has a fixed lifetime of length T and operates in an economy with a perfect capital market facing a fixed rate of interest, r. The worker's choice of the periodic training intensity profile, $x(t),$ generates an optimal path for human capital accumulation.

The problem facing the worker can thus be outlined as maximizing the lifetime earnings given the human capital production function:

$$max \int_0^T RhK(1-x)e^{-rt}dt \quad s.t.$$

$$\dot{K} = G(K, h, x), \quad K(0) = K_0, \quad 0 \le x \le 1,$$

where t is the worker's age, h is a predetermined function of t and the function $x(t)$ is the object of choice. The solution to this control problem is characterised by the maximization of the discounted earnings of the worker at each age. There is no closed form solution to this problem, unless the human capital production function is given a specific functional form. Presenting the whole solution procedure is outside the scope of this study, however some general results can be stated: Initially the worker invests at full intensity ($x = 1$), however over the years the training intensity declines and eventually reaches zero. Earning capacity increases at the range with high investment intensity and declines in the range with low investment intensity, due to a positive rate of depreciation. Observed earnings increase during the first phase because of increased human capital accumulation and reduced investment intensity.[10]

[10] Weiss (1987) compares a number of different specifications of the human capital production function. The model can be extended to allow endogenous labor supply, uncertainty and heterogeneity in the returns to schooling, as discussed in Cunha et al (2006).

2.3 Estimating Returns to Education

2.3.1 A theory-consistent econometric specification

The primary objective of this study is to estimate the pecuniary returns to education on individual earnings. A major concern is whether these estimates can be given causal interpretations. Making causal statements based on observed data requires prior assumptions. The researcher needs to (i) explicitly formulate theoretical models providing economic relations between well-defined entities, (ii) specify empirical functions reflecting the postulated theoretical relations, usually in terms of certain parameters of interest, and (iii) highlight the assumptions that need to be placed on the underlying data generating process to allow identification of economic relations from the available data. In this section, we will be concerned with points (ii) and (iii), and discuss how these issues influence our estimation strategy. Section 2.3.2 introduces the concept of individual level causal effects in a potential outcomes model, and defines some conventional parameters of interest. In Section 2.3.3, we discuss the importance of functional form restrictions and the problems related to misspecification. Section 2.3.4 presents a thorough discussion of selection bias, with an introduction to the multiple-selection problem. Estimation strategies used to correct selection bias are discussed in Section 2.3.5.

Econometric specification

Using the theoretical relationship (2.2), we specify the econometric earnings-schooling equation as

$$y_i = \alpha + \beta_s s_i + \gamma_0 p_i + \gamma_1 p_i^2 + \varepsilon_i, \tag{2.4}$$

where y_i are log-earnings for an individual *i* with schooling level *s* and potential experience *p*, and β_s can be interpreted as the rate of return to schooling.[11] (2.4) is the famous Mincer earnings equation and has been widely used in the empirical literature.[12] The residual is assumed to fulfill the restriction $E\left[\varepsilon_i \middle| s_i, p_i\right] = 0,$ implying that no omitted variables are correlated with schooling and/or experience.

[11] The reduced-form parameters (γ_0, γ_1) capture the structural parameters (k_0, r_p, P) in (2.2) and cannot be given a simple interpretation. Estimates of experience coefficients can be used to estimate the return to the cumulative time-equivalent post-school investments r_p if one of the structural parameters, for instance the whole period of net positive post-school investments *P*, is fixed.

[12] Recent surveys of this literature are found in Card (1999), Harmon el al (2003) and Blundell et al (2005).

Specifying an econometric earnings-schooling relationship is not unproblematic, since a number of theoretical frameworks can be used to model the same relations. We focus exclusively on the Mincer-Becker-Rosen setup, assuming that individuals maximize life-time earnings by choosing years of schooling and follow a given post-school training intensity profile. This generates an earnings profile with the wage rate being equal to the marginal productivity of labour. The parameter of interest is the internal rate of return on schooling, earlier defined as the capital interest rate that equates the present values of the earning streams for different choices of schooling levels by the agents. Two sets of assumptions need to be fulfilled before earnings regressions can be said to provide consistent estimates of the internal rate of return. As outlined in the previous sections, the earnings function (2.2) is deduced from strong behavioural assumptions on the agents in the economy. Secondly, parameters in the empirical counterpart of the earnings-schooling relation can be identified from the available data only under certain conditional independence restrictions. Here, we repeat the theoretical assumptions made in earlier sections, while the more econometric issues are discussed in the remaining sections.

Internal rate of return

When individuals chose between different schooling levels, they take account of direct costs, including both monetary and effort costs as well as indirect costs. In addition, income taxes and length of working life may depend on the schooling level. These factors were not accounted for while deducing the earning function (2.2). Heckman et al (2006) use a "direct" solution method to illustrate the importance of such assumptions. For an individual with earnings $Y(s,x)$ having schooling s and experience x, the present-value of the lifetime earnings is given by

$$V(s)=\int_0^{T(s)-s}(1-\tau)e^{-(1-\tau)r(x+s)}Y(s,x)dx-\int_0^s ve^{-(1-\tau)rz}dz,$$

where v denotes private tuition and other non-pecuniary costs of schooling, τ is a proportional income tax rate, r is before-tax interest rate and $T(s)$ is the schooling-dependent last age of earnings. [13] It can be shown that maximizing the above term with respect to s and rearranging the resulting terms gives

[13] Apart from changes in the opportunity costs of schooling, the variable *v* also captures a number of non-pecuniary costs or benefits associated with schooling (mental pressure or stress) or having a high-skilled job (status).

$$\tilde{r} = \underbrace{\frac{[T'(s)-1]e^{-\tilde{r}(T(s)-s)}Y(s,T(s)-s)}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 1)}} + \underbrace{\frac{\int_0^{T(s)-s} e^{-\tilde{r}x}\left[\frac{\partial \log Y(s,x)}{\partial s}\right]Y(s,x)dx}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 2)}} - \underbrace{\frac{v/(1-\tau)}{\int_0^{T(s)-s} e^{-\tilde{r}x}Y(s,x)dx}}_{\text{(Term 3)}},$$

where $\tilde{r} = (1-\tau)r$ is the after-tax interest rate.[14] Term 1 is the effect of the change in working-life associated with additional schooling. Term 2 is the weighted average effect of schooling on log-earnings, while term 3 is the cost of tuition and effort costs as a fraction of lifetime income measured at age *s*. The parameter $\tilde{r}$ can be interpreted as a marginal internal rate of return to schooling that takes into account tuition costs, changes in the retirement age and the structure of schooling returns over the life cycle. This rate of return should equal the after-tax interest rate with credit markets, once all costs and benefits from schooling are considered. Now, if we assume that v = 0 (i.e., no tuition or effort costs), $T'(s) = 1$ (no loss of work span from schooling), and that the schooling and experience components of earnings are multiplicatively separable,[15] so that $Y(s,x) = \mu(s)\varphi(x)$ (i.e., parallel log-earnings experience profiles across schooling levels), then $\tilde{r} = \mu'(s)/\mu(s)$. If all these assumptions hold for all *s*, then earnings must be log linear in schooling and $\mu(s) = \mu(0)e^{\rho_s s}$, so that $\tilde{r} = \rho_s$, i.e., a linear fit of log-earnings on schooling provides an estimate of the internal rate of return to schooling. In general, the difference between the internal rate of return and the schooling coefficient can be decomposed into three parts: a life-earnings effect (Term 1), the deviation of present schooling returns from the overall life-time average (Term 2), and existence of schooling costs that are not offset by earnings during schooling (Term 3).

To summarize, we are assuming that schooling precedes work, individuals predict their future earnings with certainty, that our earnings measure captures the full benefits of the schooling investment (i.e. no differences in non-pecuniary advantages across jobs requiring different educations), the only costs of schooling are foregone earnings (i.e. no direct costs), no externalities or general equilibrium effects of taking higher education, the economy is in a steady state without any wage or productivity growth, and that the log-earnings function is linear in schooling and additively separable in schooling and experience.

---

[14] We assume that there is no uncertainty about future earnings. Alternatively, we may assume that individuals look at the mean of the log earnings distribution when forecasting their earnings and ignore any potential person-specific deviations from that profile, and thereby base their schooling decisions on *ex ante* rates of return. See Heckman et al (2006) for an extensive discussion of uncertainty and expectations formation in this context.

[15] The earnings function derived in (2.2) embeds this property because a specific human capital accumulation profile is assumed. The investment ratio is equal to one during the initial schooling period and declines linearly during the fixed period of working life, with the post-school investment ratio being independent of the level of schooling.

2.3.2 *Potential outcomes and causal effects*

In order to fully explore the importance of functional form restrictions and endogeneity of choice variables, we can formulate the individual schooling-earnings relation in a potential outcomes model.[16] For brevity, we focus only on the role of schooling in earnings determination and overlook work experience and other covariates in the potential outcomes model. The underlying motivation for this modelling framework is that schooling choices are likely to be the result of systematic decisions, so that the sample of individuals who chose a given level of schooling is not random. Simply comparing differences in outcomes for graduates and high school dropouts will therefore not give the causal effect of high school on earnings.

Consider a situation where individuals choose between J schooling levels. The potential log-earnings of an individual *i* with schooling level *j* is denoted by y_i^j. The actual schooling choice of individual *i* is given by S_i. When an ordinal structure can be placed on the schooling levels, it is helpful to introduce the dummy variables S_{ji} and $\tilde{S}_{ji}$ defined as

$$S_{ji} = 1\left[S_i = j\right],$$

$$\tilde{S}_{ji} = 1\left[S_i \geq j\right].$$[17]

We are interested in the individual-level causal effects $y_i^j - y_i^{j-1}$, being the change in potential log-earnings for an individual *i* by receiving schooling level *j* relative to *j*-1. The relation between observed log-earnings y_i and potential log-earnings y_i^j is given by

$$y_i = \sum_{j=1}^{J} y_i^j S_{ji} = y_i^1 + \sum_{j=2}^{J}\left(y_i^j - y_i^{j-1}\right)\tilde{S}_{ji}. \qquad (2.5)$$

We may let the potential outcomes depend on both observable covariates X_i and unobserved factors u_i^j, such that $y_i^j = f_j(X_i, u_i^j)$ for all *j*. This specification is fairly general, allowing the log-earnings to differ not only across schooling levels for the same individual, but also across individuals having the same schooling level.[18] To simplify, we assume additive separability

---

[16] The potential outcomes model presented here is a modified version of Blundell et al (2005).

[17] The ordinal structure enables us to derive the marginal returns of an additional year of schooling in the potential outcome model. Otherwise, we may place a non-ordinal categorical structure on schooling and interpret the coefficients as the "cumulative" returns relative to some benchmark category, such as no education. An application of this is in the empirical part, where we construct educational groups by combining years of schooling and educational fields.

[18] However, we rule out spill-over or general equilibrium effects and effects of schooling-specific observable characteristics, so that $X_{ji} = X_i$ for all *j*. Thus, Rubin's (1986) stable unit-treatment value assumption is satisfied.

between observable and unobservable factors, so that $y_i^j = m_j(X_i) + u_i^j$. By inserting this expression in (2.5) we get

$$y_i = m_1(X_i) + \sum_{j=2}^{J} \left\{ m_j(X_i) - m_{j-1}(X_i) \right\} \tilde{S}_{ji} + \sum_{j=2}^{J} \left\{ u_i^j - u_i^{j-1} \right\} \tilde{S}_{ji} + u_i^1 .$$

The state-specific unobservable components of earnings can be written as $u_i^j = \nu_i + \upsilon_i + b_{ji}$, for all *j*, with ν_i representing unobservable individual traits that affect log-earnings equally at all schooling levels, level-specific unobservable marginal returns given by b_{ji}, and υ_i being the standard residual (capturing, say measurement errors). Thus, the above equation becomes

$$y_i = m_1(X_i) + b_{1i} + \sum_{j=2}^{J} \beta_{ji}(X_i) \tilde{S}_{ji} + \nu_i + \upsilon_i , \qquad (2.6)$$

where $\beta_{ji}(X_i) \equiv m_j(X_i) - m_{j-1}(X_i) + b_{ji} - b_{j-1i}$ are individual *i*'s private returns from having received schooling level *j* relative to *j*-1. Following Angrist and Imbens (1995), the equation (2.6) is a heterogeneous returns model with variable treatment intensity. The set of parameter given by $\beta_{ji}(X_i)$ captures variation in marginal returns to schooling across *N* individuals and *J* schooling levels, through both observable and unobservable characteristics. We may consider some simplifications: If we place the restriction $\beta_{ji}(X_i) = \beta_j$, the model (2.6) reduces to a homogeneous returns model with non-linear marginal returns to schooling. By further imposing $\beta_{ji}(X_i) = \beta$ and $\sum_{j=2}^{J} \tilde{S}_{ji} = s_i$, we get the standard homogeneous returns one-factor human capital model. When post-school investments are excluded, this corresponds to the model (2.4).

Population-level parameters of interest

The above modelling framework has similarities with the treatment-effect literature. An individual's choice of schooling level *j* can be understood as a "treatment", while the causal change in log-earnings brought about through this choice is the effect or treatment response. Similarly, the causal change $y_i^j - y_i^{j-1}$ is the individual-level treatment effect when we are comparing the outcome from treatment *j* with the outcome from treatment *j*-1. Constructing (2.6) for a given individual is impossible as we never observe the same individual in different states at the same point in time. To identify the causal effects comparisons across counterfactual states are needed. Usually, we must limit the analysis to averages of these individual-level effects over some populations of interest. A parameter of interest is the (conditional) average treatment effect or ATE, defined as

$$ATE(j, j-1 | x) \equiv E\left(y_i^j - y_i^{j-1} | X_i = x\right) = E\left(\beta_{ji}(X_i) | X_i = x\right),$$

where expectations are taken with respect to the distribution of individuals, conditioning on covariates *X* associated with the observed individual characteristics. This is the effect of assigning an individual with schooling *j*-1 from a subpopulation conditional on *X* to schooling *j*, averaged over the factors that determine log-earnings but are not captured by *X*. Similarly, the unconditional expectation provides the average treatment effect for the overall population. Another parameter in this literature is the conditional average treatment effect on the treated or TT, defined as

$$TT(j, j-1 | x) \equiv E\left(y_i^j - y_i^{j-1} | S_{ji} = 1, X_i = x\right) = E\left(\beta_{ji}(X_i) | S_{ji} = 1, X_i = x\right).$$

These parameters are the mean impact of assigning individuals having *j*-1 levels of schooling to level *j* for those who get treatment, conditional on *X*. We may also define a parameter for nonparticipants ($S_{ji} = 0$), i.e. the treatment effect on the untreated

$$TUT(j, j-1 | x) \equiv E\left(y_i^j - y_i^{j-1} | S_{ji} = 0, X_i = x\right) = E\left(\beta_{ji}(X_i) | S_{ji} = 0, X_i = x\right).$$

These parameters tell us how assigning more schooling to those who have not received it would affect their outcomes. The essential difference here is that while the TT and TUT parameters condition on individual choices, the ATE does not and is policy-invariant under weaker conditions (see Heckman and Vytlacil (2007)).

Bias associated with unconditional means

We may start by considering the simple estimator β_j^M that computes the observed difference in unconditional means of log-earnings for individuals with schooling level *j* and *j*-1.

$$\begin{aligned}
\beta_j^M &\equiv E\left[y_i | S_{ji} = 1\right] - E\left[y_i | S_{j-1i} = 1\right] \\
&= E_X\left[E\left[y_i | X_i, S_{ji} = 1\right] - E\left[y_i | X_i, S_{j-1i} = 1\right]\right] \\
&= E_X\left[E\left[y_i^j - y_i^{j-1} | X_i, S_{ji} = 1\right] + \left(E\left[y_i^{j-1} | X_i, S_{ji} = 1\right] - E\left[y_i^{j-1} | X_i, S_{j-1i} = 1\right]\right)\right] \\
&= E_X\left[E\left[\beta_{ji}(X_i) | X_i, S_{ji} = 1\right] + \left(E\left[y_i^{j-1} | X_i, S_{ji} = 1\right] - E\left[y_i^{j-1} | X_i, S_{j-1i} = 1\right]\right)\right] \\
&= E_X\left[TT(j, j-1 | X_i)\right] + \text{bias term.}
\end{aligned}$$

The first equality follows from the law of iterated expectations, while the second equality follows from the definition of observed earnings (2.5). As pointed out earlier, non-random schooling choices make it difficult to draw causal statements based on direct comparisons of average log-earnings. This is captured in the bias term above. Generally, the bias from an OLS fit of log-earnings on years of schooling can be decomposed into two components: Specification bias and selection bias due to ‘selection on unobservables’ (i.e. through ν_i, υ_i, or b_{ji}). In the next sections, we will focus on problems of misspecification and selection in this general modelling framework and extend the discussion on the parameters of interest under various modifications of the human capital model.

2.3.3 *Specification bias*

For now, we disregard selection bias by simply assuming $E\left[\nu_i \mid X_{i,}\tilde{S}_{ji}\right] = E\left[\upsilon_i \mid X_i, \tilde{S}_{ji}\right] = 0$ in (2.6), and focus on the problem pertaining to specification bias. We may consider the estimator $\beta_j^M(X_i) \equiv E\left[y_i \mid X_i, S_{ji} = 1\right] - E\left[y_i \mid X_i, S_{j-1i} = 1\right]$, that computes the observed difference in the conditional means of log-earnings for individuals with schooling level *j* and *j*-1. Under the conditional independence restrictions (‘selection on observables’), it follows that $\beta_j^M(X_i) = E\left[\beta_{ji}(X_i) \mid X_i, S_{ji} = 1\right]$. Angrist and Krueger (1999) show that the OLS regression coefficient β_s from a regression of log earnings on schooling gives a weighted average of the marginal effects $\beta_j^M(X_i)$, expressed as

$$\beta_s = E_X\left[\sum_{j=2}^{J}\beta_j^M(X_i)\mu_j(X_i)\right]E_X\left[\sum_{j=2}^{J}\mu_j(X_i)\right]^{-1},$$
$$\mu_j(X_i) = \left(E\left[S_i \mid X_i, S_i \geq j\right] - E\left[S_i \mid X_i, S_i < j\right]\right)\left(\Pr\left[S_i \geq j \mid X_i\right]\right)\left(1 - \Pr\left[S_i \geq j \mid X_i\right]\right).$$

With heterogeneous marginal returns, the OLS estimator for the linear model depends on the distribution of schooling levels and other covariates in the population. The weighting formula above has a sum and an expectation. The inner sum provides a weighted average of $\beta_j^M(X_i)$ across different schooling levels, given a particular value of *X*. This averaging matters if the marginal returns are nonlinear across schooling levels. The marginal returns are weighted using the functions $\mu_j(X_i)$, that weight $\beta_j^M(X_i)$ for each *j* in proportion to the change in the conditional mean of schooling level and give proportionally more weight to schooling levels that are closer to the conditional sample median schooling level. The expectation then averages this sum according to the distribution of *X*. This averaging matters if the marginal returns vary across different subpopulations. It should be noted that the OLS estimates exclude observations having $\Pr\left[S_i \geq j \mid X_i\right]$ equal to 0 or 1. These include values of X where schooling does not vary across observations. Thus, there is the problem of non-overlapping support of the observables.

In order to explore the extent of heterogeneity in returns to schooling, one may estimate a model similar to (2.4) separately for different subpopulations with non-linear returns to schooling. A parametric estimation of (2.6) would require knowledge of the functional form of $m_j(X_i)$ at each *j*. The OLS fit would generally control linearly for the observable variables, and can potentially suffer from a specification bias if the true model contains higher order terms of the *X*s, or interactions between *X*s and schooling. It may seem obvious to specify an arbitrarily flexible estimation model to solve any specification bias, such as a fully saturated model of schooling levels and conditioning variables. Unless large datasets are available, one is likely to run into problems of dimensionality. However, there are some other alternative estimation methods available in the econometrics literatures, such as matching on conditioning variables and matching on propensity score.

Matching

The matching approach enables the researcher to construct comparison groups by carefully reweighting the observables on the common support. As the flexible regression approach above, matching does not solve the selection on unobservable characteristics. In the case of schooling returns, the counterfactual log-earnings for individual *i* with schooling level *j* are assumed to be independent of the schooling variable S_{ji}, such that $y_i^{j'} \perp S_{ji} \left| X_i \right.$, where *j´* denotes non-treatment outcomes. This assumption of selection on observables requires that conditional on the observed attributes, the distribution of the counterfactual outcomes y^j in the treated group is the same as the observed distribution of y^j in the non-treated group. This implies $(\nu_i, \upsilon_i) \perp S_{ji} \left| X_i \right.$ in the case of (2.6) above. However, heterogeneous returns to schooling b_{ji} can be correlated with the schooling decision S_{ji}, if $(\nu_i, \upsilon_i) \perp b_{ji} \left| X_i \right.$ also holds. Thus, individuals may decide to acquire schooling on the basis of their unobserved individual-specific returns, as long as this is uncorrelated with their conditional potential outcome for an alternative schooling choice.

General nonparametric matching methods allow very flexible specifications of $m_j(X_i)$ and can thereby avoid the potential specification bias associated with an OLS fit. Matching requires that for each treated observation $(y_i : i \in \{S_{ji} = 1\})$, we can look for one or more non-treated observations $(y_i : i \in \{S_{ji} = 0\})$ such that the non-treated observations correspond to the required counterfactual group.[19] This assumes that there is a common support region for *X* where the treated and non-treated groups overlap. Usually, this is satisfied by the requirement $P(S_{ji} = 1 | X_i) < 1$, which prevents *X* from being a perfect predictor of treatment status. Meanwhile, finding exact matches can be extremely difficult if *X* includes a wide range of variables. However, Rosenbaum and Rubin (1983) show that the problem of high-dimensional matching can be simplified to the problem of finding treated and non-treated observations with the same one-dimensional *propensity score* for attaining treatment *j*, $p_j(X_i)$.[20] The conditional independence assumption remains valid, as $0 < p_j(X_i) < 1$ and $y_i^{j'} \perp S_{ji} | X_i$ imply $y_i^{j'} \perp S_{ji} | p_j(X_i)$. Thus, causal effects on earnings from different schooling levels can be drawn from comparisons within subpopulations defined by $p_j(X_i)$. After averaging across the relevant subpopulations, the average effect of having schooling level *j* relative to having schooling *j*-1 is given by $E_X\left[E\left[y_i^j - y_i^{j-1} | p_j(X_i), p_{j-1}(X_i)\right]\right]$.

Although the matching method provides a more flexible estimation procedure compared to OLS, strong identifying assumptions are still called for. Firstly, depending on the set of conditioning variables, the identifying independence assumption can be controversial. In some situations, it may not be plausible to rule out selection on unobservables. For instance, using experimental data Smith and Todd (2005) show that propensity score matching estimates are highly sensitive to the set of variables included in the scores.[21] Secondly, there is a possibility that there are no non-treatment observations with a similar propensity score for all the participants. The estimated treatment effect must then be redefined as the mean treatment effect for those treated falling within the common support. And finally, if the effect

[19] When there are multiple treatment and non-treatment observation on the common support of *X*, some sort of weighting between the different observations can be done, such as nearest neighbour, calliper, or kernal-based matching. For instance, comparisons of twins or siblings earnings outcomes have been used to estimate returns to schooling. This can be seen as an application of calliper matching.

[20] The 'generalized' propensity score used in the multi-treatment case above is the probability of attending schooling level *j* given the observed characteristics (see Imbens (2000)).

[21] Other studies evaluating the matching methods are Heckman, Inchimura, Smith and Todd (1998), Dehejia and Wahba (2002) and Heckman and Navarro-Lozano (2004).

of schooling differs across individuals who have received it, restricting the comparison to observations found on the common support may not give the mean treatment effect on the treated group. In comparison, the parametric models "solve" the common support problem by means of functional form assumptions to extrapolate outside the common support.

2.3.4 Selection bias

In most data generating processes, sampling rules other than random sampling are used to draw from some underlying population of interest. Such sampling rules may be outcomes of choice behaviour of agents being studied and/or decisions made by sample survey designers. In such cases, the researcher has a non-random sample of the overall population, i.e. there is selection bias. One type of identification problem is to recover the conditional probability distribution of the underlying population from an observed sample, in which the realizations of the conditioning variables are always observed but the realizations of outcomes are censored.[22] This would usually require strong prior assumptions. For instance, we observe wages only for individuals participating in the labour market but we might be interested in the joint distribution of (potential) wages and schooling years for all persons whether or not they work. However, selection problems are more general and can occur in any context where different population groups are being compared and individuals pertaining to the different groups are not subject to random assignment. Job-search strategies, occupational choices, schooling attainment, migration, etc., are variants of selection problems that have been studied within the Mincer earnings regression framework.[23]

As shown in the preceding section, unless strong conditional independence restrictions are placed on the residuals we would encounter problems of selection bias in estimating log-earnings. In the econometric literature, the selection problem has been characterized in two ways: The traditional structural approach estimates decision rules derived from economic models in an attempt to recover the primitives of economic theory (technology and preferences). Given that the decision rules are correctly specified, the estimates of the primitives can be used to construct counterfactuals, and thereby identify causal effects. Commonly, a joint distribution of potential outcome and explanatory variables is specified

[22] See chapter 2 in Manski (1995) for an introduction to the selection problem encountered in identifying the probability distributions of partially observed outcome variables.

[23] Widely cited applications of selection models are Lee (1978) on unionism, Willis and Rosen (1979) on schooling, Björklund and Moffitt (1987) on job training, Heckman and Sedlacek (1990) on sectoral choice and Pessino (1991) on migration.

and the observed distribution is deduced by introducing sampling rules through *weighting functions*. However, the recent program-evaluation literature treats the selection problem as a *missing data problem* and attempts to use observable data to impute the relevant counterfactual outcomes.[24] The current popularity of the experimental approach stems mainly precisely from the concern that the economic theory does not provide us the functional form, the distributional assumptions or what to control for in a regression analysis. Thereby, it is impossible to recover the primitives of economic theory.[25] As emphasised by Angrist and Krueger (1999), the experimental approach seeks situations where one has a reasonable presumption that the omitted variables are uncorrelated with the variable of interest. Such situations may arise if the researcher can use random assignment, or if the forces of nature or human institutions provide something close to random assignment. By controlling for confounding variables, using before and after comparisons of the same agents, and by using instrumental variables as a source of exogenous variation, the experimental approach tries to mimic random assignment. Accordingly, a regression analysis is given a causal interpretation when the estimated coefficient approximates the causal effect that would have been revealed in a true experiment. The potential outcomes model and preceding discussion on the population-level parameters of interest is largely derived from the experimental literature.

Ability bias and comparative advantage

In Section 2.2.3, the concept of individual earnings ability was introduced within a human capital production function in the Rosen-Becker model. We also showed that optimal schooling and earnings depend on individual ability. Before we proceed, it may be important to restate what our concept of ability constitutes. One way of understanding ability is as an operational factor indicating intrinsic or natural talents that enable the individual to succeed. The main problem with this view is that there is no objective measure of an individual's intelligence. The closest we have are IQ-test scores or residuals from a linear fit of test scores on schooling. However, using such proxy measures that do not fully capture the unobservable ability may not solve the problem and may even increase the bias, as shown by Frost (1979).

[24] Heckman (2001) provides a discussion of the distinction between structural modelling and treatment literature in the context of selection models.

[25] Thurow (1983, pp. 106-107) argues: "Economic theory almost never specifies what secondary variables (other than the primary ones under investigation) should be held constant in order to isolate the primary effects. ... When we look at the impact of education on individual earnings, what else should be held constant: IQ, work effort, occupational choice, and family background? Economic theory does not say. Yet the coefficients of the primary variables almost always depend on precisely what other variables are entered in the equation 'holding everything else constant'."

Another view is that ability is "excess earnings capacity" that is not captured by formal schooling or observable factors.[26] This concept of ability may include factors such as motivation, attitude or other unobservable forces driving earnings. It is extremely hard to motivate a one-dimensional empirical counterpart for such concepts. Few observational studies have attempted to include factors other than test scores. In most cases the researcher does not have adequate measures of IQ and is forced to exclude the ability variable from the earnings regressions. Thus, individual earnings ability is one of the unobservable factors that are likely to explain schooling choices and earnings dispersion across individuals. Estimation is further complicated if we allow for comparative advantage or multi-dimensional skills variation. In that case, the schooling coefficient would not only be ridden by omitted ability bias, but the bias would vary across different subgroups.

To illustrate this problem, let's again consider the model (2.6). By separating the observable and unobservable factors we can write this as

$$y_i = m_1(X_i) + \sum_{j=2}^{J} \bar{\beta}_j(X_i)\tilde{S}_{ji} + \underbrace{\sum_{j=2}^{J}\left\{b_{ji} - b_{j-1i} - \left(\bar{b}_j - \bar{b}_{j-1}\right)\right\}\tilde{S}_{ji} + \nu_i + \upsilon_i + b_{1i}}_{\omega_i}, \qquad (2.7)$$

where $\bar{\beta}_j(X_i) \equiv m_j(X_i) - m_{j-1}(X_i) + \bar{b}_j - \bar{b}_{j-1}$ are the average private returns to schooling level j relative to j-1 (ATE), and $\bar{b}_j = E\left[b_{ji}\right]$ is the population mean of the unobservables. We assume that the functional forms of $m_j(X_i)$ are known for each j, so that it is possible to perform a correctly specified OLS regression. However, as long as there is correlation between $\tilde{S}_{ji}$ and the composite error term ω_i, i.e. $E\left[\omega_i \mid X_{i,}\tilde{S}_{ji}\right]$ is non-zero, estimates from the OLS regression will be biased. As pointed out by Blundell et al (2005), such correlation may arise from different sources. The residual term ν_i captures unobserved individual traits that affect log-earnings equally across all schooling levels (the absolute advantage). The standard ability bias arises from the likely correlation between ν_i and $\tilde{S}_{ji}$. If workers with higher earnings ability also acquire more schooling, this correlation will be positive and there will be an upward bias in the estimated average return. Meanwhile, another source of bias is through correlation between the unobserved individual returns b_{ji} (the comparative advantage) and the

[26] In Griliches's (1977) words: "[Ability] is an unobserved latent variable that both drives people to get relatively more schooling and earn more income, given schooling, and perhaps also enables and motivates people to score better on various tests."

schooling decision given by $\tilde{S}_{ji}$. Individuals may choose to pursue an education based on the knowledge of their own capabilities or interests specific to the chosen schooling level or field. These are unobserved by the researcher and hence cannot be controlled for. The direction of this bias is not clear and will depend on the average returns among the subpopulations with the different schooling levels. Another issue is measurement errors in the schooling variables $\tilde{S}_{ji}$, captured by υ_i, that lead to downwards bias and could partially offset the upward ability bias, as emphasized by Griliches (1977).

A multiple-selection model

The data used by the researcher has been subject to a number of sample selection processes and could generate biased estimates even if schooling choices are random. For instance, individuals face the choice of participating in the labour force. The choice of working can be made arbitrarily flexible, by allowing the individual to work in different occupations and industries, with non-participation being one of the alternatives in this choice-set. While we observe schooling, location, and other demographic characteristics (X_i, S_{ji}) for each individual, irrespective of their participation in the labour force, an individual's earnings are available only for his own choice of occupation. Thus, we are facing a situation with partial observation of the dependent variable. This becomes a problem when such censoring is non-random. Individuals work when they choose to do so. The hourly wage rate a worker is offered influences the decision to work. By neglecting this dependence, we are assuming that information on individuals found in the data can be generalized to those outside the labour force.

Considering multiple selection mechanisms simultaneously in the schooling returns model (2.7) is not straightforward. Maddala (1983) characterises this as a *polychotomous-choice selection* model. To give an illustration of the problem, we may consider the labour force participation decision of an individual in a latent variable model.[27] Each individual is allowed to either work in private sector, public sector, or remain outside the labour force. For brevity, all outside options, such as self-employment, unemployment, or disability, are considered as non-participation. Our parameter of interest is still the average private returns to schooling $\overline{\beta}_j^l(X_i)$, now conditional both on choices of schooling *j* and sector *l*. By incorporating these

[27] Mincer (1962), Gronau (1974) and Heckman and Willis (1977) are some of the early contributions to this literature, analysing labour force participation decisions of women.

choice mechanisms in the human capital model (2.7) we arrive at the system of equations given by (2.9).

(2.9a) are the latent wage offers, expressing the potential wages that an individual is offered in the public sector W_i^1 and in the private sector W_i^2, assumed to depend upon observable characteristics (X_i, S_{ji}) and some unobservable factors $(\varepsilon_{1i}, \varepsilon_{2i})$. (2.9b) gives an expression for the latent reservation wage W_i^R, that depends on variables Z_i^R excluded from the earnings equations (2.9i), in addition to the characteristics (X_i, S_{ji}). Equations (2.9c)-(2.9e) specify the selection rules. The latent sectoral preferences (2.9c) are dependent on the wage gap between the offered wage rates in public and private sector (W_i^1, W_i^2) and the reservation wage W_i^R and the non-pecuniary benefits related to each option (Z_i^1, Z_i^2, Z_i^R). It is assumed that the observable characteristics (Z_i^l, X_i, S_{ji}) and the unobservables $(\varepsilon_{Di}^1, \varepsilon_{Di}^2, \varepsilon_{Di}^R)$ affecting the decision variables are additively separable. (2.9d)-(2.9e) are choice indicators. The actual sector-dependent wages are given by (2.9f), with the observed wage rate W_i being equal to the offered wage rate W_i^l if sector *l* is chosen. (2.9g) defines observed earnings Y_i as the product of observed wage rate W_i and observed working hours H_i for an individual *i*, assuming that the individual can only hold a single job at a wage rate invariant to hours worked. The equations (2.9i) states that the sector-specific log-earnings are given by the earnings relation (2.7) only if the individual accepts the wage offer. For non-participants the observed hours of work H_i and the observed labour market earnings Y_i are both zero, as given in (2.9j).

$$W_i^l = f_l(X_i, S_{ji}, \varepsilon_{li}), \qquad \text{for } l = \{1,2\}, \tag{2.9a}$$

$$W_i^R = f_R(Z_i^R, X_i, S_{ji}, \varepsilon_{Ri}), \tag{2.9b}$$

$$D_i^{l*} = m_l\left(W_i^{k=1,2,R}, Z_i^{k=1,2,R}\right) = g_l(Z_i^1, Z_i^2, Z_i^R, X_i, S_{ji}) + \varepsilon_{Di}^l, \quad \text{for } l = \{1,2,R\}, \tag{2.9c}$$

$$D_i = k \qquad \text{if } \varepsilon_i^k \equiv \max_{l=(1,2,R), l\neq k} D_i^{l*} - \varepsilon_{Di}^k < g_k(Z_i^1, Z_i^2, Z_i^R, X_i, S_{ji}) \tag{2.9d}$$

$$D_i^l = 1 \qquad \text{if } D_i = l, \quad \text{for } l = \{1,2,R\}, \tag{2.9e}$$

$$W_i = W_i^l \qquad \text{if } D_i^l = 1, \quad \text{for } l = \{1,2\}, \tag{2.9f}$$

$$Y_i \equiv W_i \cdot H_i, \tag{2.9g}$$

$$y_i = \log\left[W_i^l \cdot H_i\right] = m_1^l(X_i) + \sum_{j=2}^{J} \bar{\beta}_j^l(X_i)\tilde{S}_{ji} + \omega_i^l \qquad \text{if } D_i^l = 1, \quad \text{for } l = \{1,2\}, \qquad (2.9h)$$

$$Y_i = H_i = 0 \qquad \text{if } D_i^R = 1, \qquad (2.9i)$$

where

$$\bar{\beta}_j^l(X_i) = m_j^l(X_i) - m_{j-1}^l(X_i) + \bar{b}_j^l - \bar{b}_{j-1}^l,$$

$$\omega_i^l = \sum_{j=2}^{J}\left\{b_{ji}^l - b_{j-1i}^l - \left(\bar{b}_j^l - \bar{b}_{j-1}^l\right)\right\}\tilde{S}_{ji} + \nu_i^l + \upsilon_i^l + b_{1i}^l, \qquad \text{for } l = \{1,2\}.$$

This system of equations is motivated by a reservation-wage model, predicting that the individual chooses to work only if offered earnings are greater than some lowest acceptable wage level (a reservation wage), and chooses not to work otherwise. The sector choices are based on valuation of various non-pecuniary benefits and the wage gaps. Similar selection problems arise in estimation of wage differentials, as wages are observed only for those who work. The above problem is posed without explicitly specifying the relation between earnings and wages or counterfactual sector-specific wages, such that only equations (2.9a)-(2.9e) are of interest. We specify (2.9h) in terms of earnings, as these relationships follow from the human capital theory of lifetime earnings maximization. However, in the returns to schooling literature there is an ambiguity as to what constitutes a proper measure of the dependent variable (Willis, 1986) and a number of studies report returns to schooling both on hourly wages and monthly/annual earnings. This has obvious implications for interpretation of the returns to schooling parameters.[28]

We are interested in estimating the sector-specific earnings relation (2.9h), especially computing the average private returns to schooling $\bar{\beta}_j^l(X_i)$. An OLS fit on (2.9h) would suffer from a large number of problems. Apart from the biases arising through a likely correlation between $\tilde{S}_{ji}$ and the composite error terms ω_i^l, i.e. the *ability bias* due to correlation between

[28] By assuming exogenously fixed labour supply we have simplified matters. Log(earnings) can normally be decomposed into log(wages) + log(hours worked). Unless workers decide on wages and working hours simultaneously, the effect of schooling on log(earnings) could be similarly decomposed into a wage (or marginal productivity) component and an hours of work component. However, simultaneity in the choice of working hours and wage rates is likely, as suggested by Moffitt (1984), Tummers and Woittiez (1991), and Vella (1993). Perhaps distinguishing between potential and actual hours of worker in the above model could provide a solution to this problem. Hardoy and Schøne (2006) allow choice of part-time and full-time employment in a switching regression model with endogenous switching for Norwegian data. They find "no evidence of systematic selection bias" with small wage gaps between part-time and full-time workers when using capital income as an instrument for choice of employment interval. However, to our knowledge there is no satisfactory solution to this problem in the empirical literature when we allow for multiple sources of selection.

ν_i^l and $\tilde{S}_{ji}$, the *returns bias* due to correlation between b_{ji}^l and $\tilde{S}_{ji}$, and the *measurement error bias* due to correlation between υ_i^l and $\tilde{S}_{ji}$, there would be an additional *participation bias* when there is correlation between any of the random terms in the composite errors ω_i^l and the unobserved determinates of sector choice ε_{Di}^l. To see this, we take the conditional expectations of (2.9h), given observable characteristics, schooling and sector choices, such that

$$E\left[y_i \mid X_i, \tilde{S}_{ji}, D_i^l = 1\right] = m_1^l(X_i) + \sum_{j=2}^{J} \bar{\beta}_j^l(X_i)\tilde{S}_{ji} + E\left[\omega_i^l \mid X_i, \tilde{S}_{ji}, D_i^l = 1\right] \text{ for } l = \{1,2\}.$$

By expanding the composite errors ω_i^l according to (2.7) and inserting the decision rules (2.9d), we can express the last terms as

$$E\left[\sum_{j=2}^{J}\left(b_{ji}^l - b_{j-1i}^l\right)\tilde{S}_{ji} + \nu_i^l + \upsilon_i^l + b_{1i}^l \middle| \varepsilon_i^l < g_l(Z_i^1, Z_i^2, Z_i^R, X_i, S_{ji}), X_i, \tilde{S}_{ji}\right] \neq 0 \text{ for } l = \{1,2\}.$$

In general, this conditional truncated mean would be non-zero, irrespective of various correlations between $\tilde{S}_{ji}$ and ω_i^l. Without restricting the joint distribution of random terms $(b_{ji}^l, \nu_i^l, \upsilon_i^l, \varepsilon_i^l)$ with $j = (1,...,J)$ over each sector *l*, the conditional mean of the errors terms cannot be given a specific functional form. Nor is it easy to make qualitative predictions on the direction of the selection bias. In a simple selection model with exogenous schooling assignment, no heterogeneity in returns and a binary choice between labour force participation and staying unemployed, the error terms in the earnings or wage equation and the participation equation are often assumed to follow a joint normal distribution. This yields a closed form for the conditional means. A discussion on the multinomial logit model with general selection is given in McFadden (1980) and Lee (1983).

2.3.5 *Estimation methods*

This section compares two non-experimental estimation methods, namely control function and instrumental variables, which have been commonly used in response to the selection problems discussed earlier.[29] The focus here is on the methodological distinctions between the estimation methods, especially the identifying assumptions and their validity in different contexts. We focus exclusively on the selection problem, abstracting from biases related to misspecification and measurement errors.

[29] Comprehensive surveys of various non-experimental estimation methods are provided by Blundell et al (2005), Card (1999), and Heckman and Navarro-Lozeno (2004).

Instrumental variables

Instrumental variables (IV) techniques are applied in situations where there are reasons to believe that there is a correlation between observable schooling measures and the unobservables in the earnings equation (2.6). IV methods seek to establish exogenous variation in schooling measures and identify the parameters of interest using this variation. In order for this to be possible, such variables must be correlated with the true measure of schooling and uncorrelated with the unobservables in the wage equation. For simplicity, we restrict the schooling coefficients to be the same across different schooling levels in (2.6) such that $m_j(X_i) - m_{j-1}(X_i) = b(X_i),\ b_{ji} - b_{j-1i} = b_i, b_0 = E[b_i]$, and $\sum_{j=2}^{J} \tilde{S}_{ji} = s_i$, so that

$$y_i = m_1(X_i) + b(X_i)s_i + b_0 s_i + \underbrace{(b_i - b_0)s_i + \upsilon_i + \nu_i}_{e_i}, \qquad (2.10)$$

where e_i is the composite residual. An IV Z_i would satisfy the conditions $E\left[\nu_i \middle| Z_i, X_i\right] = 0$ and $E\left[\upsilon_i \middle| Z_i, X_i\right] = 0.$ By taking the conditional expectation of (2.10) we get the following:

$$E\left[y_i \middle| Z_i, X_i\right] = m_1(X_i) + \left(b(X_i) + b_0\right) E\left[S_i \middle| Z_i, X_i\right] + E\left[(b_i - b_0)S_i \middle| Z_i, X_i\right].$$

The IV method can be applied by replacing the conditional expectation of schooling with its prediction in both its linear and interaction terms. Since the last term in the equation above contains the interaction between the schooling variable and the unobservable individual returns, the IV method would produce inconsistent estimates of the ATE and the TT unless stronger assumptions are made. One such assumption is $b_i = b_0$, implying that given the conditioning variables *X*, the returns are homogeneous across individuals. Following the IV assumptions, the last term is zero. In addition, the IV estimator requires a specification of $m_1(X_i)$ and $b(X_i)$, and would in general be subject to the potential misspecification bias associated with OLS. However, in cases where the above assumptions are satisfied and the true functional form is known, the IV method provides benchmark estimates that account for endogeneity of schooling.[30]

In general heterogeneous returns models where $b_i \neq b_0$, the situation is more complicated. An important parameter of interest in the literature is the average effect of treatment on the treated (ATT). Under some conditions, it is possible to consistently estimate ATT in the

[30] In the multi-treatment framework with each schooling level having its level-specific coefficient, at least as many instruments are needed as the number of levels in order to identify all the parameters.

presence of heterogeneous returns. Assume that we consider two groups: college graduates $(S_{1i}=1)$ and high school graduates $(S_{1i}=0)$. The treated individuals are those who have received college education. We assume that the condition $E[b_i|Z_i, X_i, S_{1i}=1]=E[b_i|X_i, S_{1i}=1]$ is satisfied, implying that conditional on X the instrument Z is not correlated with the individual-specific component of the return b_i for those who have received the treatment. Thus, the schooling decision is unrelated to the individual gains. Suppose that there is a single binary instrument $Z_i \in \{0,1\}$. Conditional expectation of (2.7) gives:

$$E[y_i|Z_i, X_i]=m_0(X_i)+\{b(X_i)+E[b_i|X_i, S_{1i}=1]\}E[S_{1i}|Z_i, X_i].$$

Under the above condition, the IV estimation of the schooling coefficient provides a consistent estimate of the conditional effect of treatment on the treated:

$$\begin{aligned}\hat{\beta}_{IV}(X) &\equiv \frac{E[y_i|X_i, Z_i=1]-E[y_i|X_i, Z_i=0]}{E[S_{1i}|X_i, Z_i=1]-E[S_{1i}|X_i, Z_i=0]}\\ &= b(X_i)+E[b_i|X_i, S_{1i}=1]\\ &= E[y_i^1-y_i^0|X_i, S_{1i}=1]\equiv \hat{\beta}_{ATT}(X).\end{aligned}$$

However, if there is comparative advantage, the IV estimator will not estimate the ATT. Following Imbens and Angrist (1994), there is a literature that aims at interpreting IV estimates in such cases. Under some weaker conditions, the IV estimates the local average treatment effect (LATE). Angrist and Imbens (1995) expand this approach to the models with variable treatment intensity. For now, we may consider the same schooling levels as above. It is helpful to introduce four subgroups of individuals: those who attend only high school irrespective of the value of the instrument (the 'never takers'), those who always choose to acquire college education (the 'always takers') and those who are induced by the instrument to change their behaviour, either in a perverse way (the 'defiers') or in line with the instrument (the 'compliers'). The last group is made up of individuals who would have chosen to pursue college education if the instrument were in place, but not otherwise. It is useful to make the restriction that the instrument has the same directional effect on all those whose behaviour it changes, such that either one of the two last groups can be excluded (usually the defiers). In situations where this *monotonicity condition* is satisfied, the IV estimator can be interpreted as the average returns among those individuals who are induced to change behaviour because of a change in the instrument.[31] This may also provide some intuition for why the IV estimate usually varies depending on the instrument that is used. There may be some systematic variation in the returns across different subpopulations that are affected by different

[31] In interpreting the IV estimator as LATE, the schooling decision does not need to be unrelated to the individual gains.

instruments. Using IV estimates to make predictions for the whole population may not be right if the instrument affects the population disproportionately. In addition, as in the homogeneous returns model, the credibility of the IV approach hinges on the untestable assumptions concerning correlation between the error terms in the outcome equation and the instrument. Reflecting over these issues is critical in assessing the validity of IV estimates.

Control function estimation

While the matching method completely excludes selection on unobservables and the IV method generally provides the LATE (under the monotonicity condition), there is another approach known as control function estimation that aims at recovering the average treatment effect (ATE) in the presence of selection on unobservables by placing additional restrictions. An explicit modelling of the schooling selection process is usually required. In the multi-treatment framework of equation (2.6) above, we can assume that an individual *i*'s choice of schooling level *j* is subject to the following binary response model

$$\tilde{S}_{ji} = 1\left\{ f_j\left(Z_i, X_i\right) + \eta_i \geq 0 \right\} \quad \text{with} \quad \eta_i \perp Z_i, X_i .$$

Conditional on the function f_j, the unobservables in the outcome equation (2.6), ν_i and b_{ji}, are assumed to be distributed independently of the schooling variable. This implies that the structure placed on the selection process completely removes the problem of endogeneity in the schooling measures. Another way of stating this requirement is by the following conditions:

$$\nu_i - \bar{\nu} = r_{\nu\eta}\eta_i + \varepsilon_{\nu i}$$

$$b_{ji} - \bar{b}_j = r_{b_j\eta}\eta_i + \varepsilon_{b_j i} \quad \text{with} \quad \eta_i \perp \varepsilon_{\nu i}, \varepsilon_{b_j i} .$$[32]

Given the above assumptions, the conditional expectations of the unobservables is given by

$$E\left[(\nu_i - \bar{\nu}) \middle| X_i, Z_i, \tilde{S}_{ji} = 1\right] = r_{\nu\eta} K_{ji}(X_i, Z_i) ,$$

$$E\left[(b_{ji} - \bar{b}_j) \middle| X_i, Z_i, \tilde{S}_{ji} = 1\right] = r_{b_j\eta} K_{ji}(X_i, Z_i) \qquad \text{for } j = 1,\ldots, J.$$

The K_{ji} terms are the control function that account for the dependence of the unobservables determinants of log-earnings on the choice of schooling.[33] By taking the conditional expectation of (2.7) and inserting the above terms, we get

---

[32] Assuming joint normality between the unobservables in the outcome equation and the assignment equation, as is common in the literature, would also lead to similar relations between the error terms. For instance, the two-step estimator of Heckman (1979) that is widely applied in self-selection problems is a special case of the above model with bivariate normal distribution of the unobservables, no heterogeneity in returns, and a binary selection rule.

$$E\left[y_i \mid X_i, Z_i, \tilde{S}_{ji} \right] = m_1(X_i) + \overline{b}_1 + \overline{v} + \sum_{j=2}^{J} \overline{\beta}_j(X_i)\tilde{S}_{ji} + r_{v\eta}K_{1i} + \sum_{j=2}^{J} r_j K_{ji}\tilde{S}_{ji},$$

where $\overline{\beta}_j(X_i) = m_j(X_i) - m_{j-1}(X_i) + \overline{b}_j - \overline{b}_{j-1}$ is the conditional ATE, $r_j \equiv r_{b_j\eta} - r_{b_{j-1}\eta} + r_{v\eta}$, and the conditional independence condition $E\left[\upsilon_i \middle| X_i, \tilde{S}_{ji}, K_{ji} \right] = 0$ holds, i.e. no measurement errors etc. Replacing K_{ji} terms with their estimated counterparts, specifying the functional forms of $m_j(X_i)$ and performing a linear fit on the resulting equation will provide estimates of $\overline{\beta}_j(X_i)$ for different schooling levels *j*. Under common distributional assumptions on the unobservables, the functional form of K_{ji} may alone provide identification. However, without excluded 'instruments' *Z* problems of multicollinearity can arise between the K_{ji} terms and the *X* covariates in the outcome equation. In practice the estimator is found to perform poorly if there are no excluded variables, as shown by Puhani (2000).

This procedure provides some interesting insights. By invoking control function assumptions, not only are we able to estimate the ATE in a multi-level model of heterogeneous returns, but also other parameters of interest such as the TT or the TNT. Secondly, under the structure that is imposed on the selection process, the *r*-coefficients are informative on the presence and direction of the selection process. Selection on unobserved ability ($r_{v\eta}$) and selection on unobserved returns ($r_{b_j\eta}$) can be tested separately. Such tests can provide some guidance on the reliability of the conditional independence assumption in the matching specification. However, the control function approach rests on strong exclusion restrictions. The explicit modelling of the schooling selection process can be viewed as equivalent to assuming knowledge of the exogenous variation in educational choices for each subpopulation. Thus, this knowledge enables the researcher to not only locate the LATE, but the ATE for the overall population. While the presence of selection can be tested through the estimated *r*-coefficients, the credibility of the underlying full-information criteria placed on the selection process is difficult to test. By placing more structure, we are able to predict more. However, the validity of this approach rests on the strong assumptions made on the selection process.

[33] Unless distributional assumptions are placed on the unobservables, it is not possible to derive a functional form for K_{ji}. However, the control functions can usually be written as functions of the participation propensities $p_j(X_i, Z_i)$, which can be easily calculated using a multinomial estimation. In such cases, the control functions can be approximated by polynomials in the participation propensities.

3. Empirical Analysis

3.1 Introduction

3.1.1 Overview

In this part we estimate the returns to education based on Norwegian wage data for paid workers for the year 2001. Section 3.1.2 provides a review of earlier studies on educational returns that use similar methods and comparable data. The data sources are described in Section 3.1.3. We restrict our attention to the three main problems discussed earlier, namely specification, selection and endogeneity. Cross-sectional OLS estimates are given in Section 3.2. Focusing on the specification issues, we allow for non-linearity in the schooling measure and interactions between schooling and experience in Section 3.2.2. Effects of controlling for labour market characteristics, such as industry and region, are discussed in Section 3.2.3. All wage equations are estimated separately for gender and sector subgroups. Section 3.3 considers the problems related to endogeneity of schooling and selection bias due to non-random choices of labour force participation and sector choice. Descriptive statistics, tables and estimation results are given in the appendix.

3.1.2 Earlier studies

Given the peculiar nature of an institutionalised wage setting and the availability of rich register datasets on earnings, there have been a large number of studies analysing earnings dispersion in Norway. Apart from descriptive studies on earnings/income inequalities, economists have also been concerned with issues relating to bargaining regimes, local labour markets (regions and industries), skills returns (education, formal or informal training), establishment effects, performance pay, and certainly, gender wage differentials. An updated overview of this literature is not available, however, fairly lengthy surveys of studies on earnings dispersion and educational pay are given in Dale-Olsen (1997), Barth and Røed (1999) and Raaum (1999).

Hægeland, Klette, and Salvanes (1999) and Aakvik, Salvanes, and Vaage (2003) are the two prominent studies on educational returns in Norway. Both of these studies touch upon problems related to endogeneity and selection bias. Hægeland et al compare the estimates of returns to education across different cohorts, sectors and over time and find stability in educational returns. Self-selection into education is corrected by using an IV technique. County during childhood is used as an identifying instrument for choices of schooling levels. It is assumed that the region where a person grew up does not by itself influence earnings when conditioning on the level of education and other variables. The educational choice equations also include individual's gender and a number of other family background variables as explanatory variables. The earnings equations are estimated only for full-time workers (those working more than 30 hours) in 1980 and 1990. The number of children, their age-composition, and the disposable income of the rest of the household are used as determinants of full-time employment to account for selection bias. These variables are excluded from the earnings regressions. After correcting for education and employment selection, the results indicate non-linear returns across schooling levels, while the returns are stable across cohorts.

Aakvik et al (2003) estimate a comparative advantage model for schooling, in which the returns to education vary at different levels of education. Problems relating to endogeneity of schooling choices and heterogeneity in the returns to education are discussed. In order to correct for endogeneity of schooling in this variable treatment intensity framework, instruments need to be specified at each level of education. Aakvik et al pursue an experimental IV approach, exploiting the staged implementation of increase in years of compulsory education through a major reform in the school system in the 1960s. The reform extended the minimum years of schooling from seven to nine years. In particular, the researchers look at annual earnings in 1995 across individuals with different schooling levels born in the period 1948-57. The earnings equations are estimated only for full-time workers. These cohorts were exposed to the reform that took more than ten years to implement. Comparisons across schooling levels are possible because the same birth cohorts are observed through both types of compulsory school systems, since the reform was implemented in different years across municipalities. Also after controlling for selection effects, the results indicate non-linearity in returns to education.

3.1.3 Data sources

Our primary source is the Wage Statistics for 2001 provided by Statistics Norway.[34] This comprises a set of cross-sectional micro datasets containing about 1.2 million individuals and contains information on both public and private sector employees. The data on public sector

[34] Documentation provided in Statistics Norway (2005).

employees is taken from administrative registers covering all employed workers. For employees in privately registered enterprises, the data is based on an annual stratified random sampling of enterprises conducted by Statistics Norway.[35] Apart from the primary sector, all other industries are represented. The enterprises are characterised as small, medium-sized or large, depending on the number of employees.[36] For small and medium-sized enterprises, the sampling rates vary between 10-20 per cent and 40-50 per cent respectively, while all large enterprises are sampled. Within each sampled enterprise, all workers are counted. Depending on the worker's employment industry, between 30 and 90 per cent of private sector employees are covered.

The datasets contain information on a number of earnings components such as average monthly paid salaries, variable additional allowances, bonuses, commissions, and contracted weekly working hours. Our constructed measures of gross average hourly wages are based on these variables. See Section 6.1 for detailed variable definitions. In this respect, this dataset is unique. Few studies build on detailed information on earnings components and working hours for such a large population. The datasets also provide several other measures such as years of schooling, educational field, industry, age, sex, and municipality of residence. We remove immigrants, students, and individuals having multiple jobs from the sample. Only paid workers in the age group 20-64, with contracted working hours between 1 and 75 hours per week and an hourly wage between NOK 50 and NOK 2000 are included. Table A.1 provides a description of the data selection process.

3.2 OLS Estimation

3.2.1 Standard specification

While other studies on educational returns in Norway use data on annual earnings for full-time workers, we use wage data for a much larger population covering both part-time and full-time employees. Although the theoretical models in Section 2.2 are formulated in terms of lifetime earnings and, under the assumptions made clear earlier, can be reduced to the annual earnings equation (2.4), there is ambiguity in the empirical literature as to what

[35] The variables used for stratification are industry and size of the enterprise. The survey is conducted in September-October once every year. Data on monthly earnings is drawn from the sample population and seasonal corrections are made to make the data representative for any month. See Grini (2007) for details on the stratification method.

[36] Enterprises having employees under a certain minimum ('cut-off') are excluded from the sample. The cut-off limits used in WS97-07 vary between <3 and <5 across different industries and sampling years. Removing the very small enterprises is a way of excluding self-employed workers.

constitutes a proper measure of the dependent variable. Both returns to education on hourly wages and annual earnings are found in the literature. Our simple theoretical framework is silent on issues relating to labour supply. One way to circumvent this ambiguity is to impose that the agents work full-time once they are employed with a fixed hours of work supply. However, the empirical literature on *wage differentials* based on the Mincer earnings equations implicitly assumes that the log-earnings function is additively separable into a log-working hours component and a log-wages function. The latter is assumed to have exactly the same functional form as the log-earnings function. Unless schooling has an effect on choice of working hours or that the wage rates depends upon working hours, the returns to education can be estimated either on earnings or wage data. We do not disregard that schooling has an effect through working hours. However, for simplicity we assume that wages are invariant to the choice of working hours. This assumption can be relaxed by allowing endogenous choice of working hours. By restricting our attention to the returns to education on hourly wages, we may interpret the schooling coefficients as the relative increase in worker productivity. Estimates of educational returns from annual earnings data for the subpopulation of full-time workers can be given a similar interpretation. However, as working hours are not precisely observed in earlier studies and crude measures are used for full-time work, there may be considerable measurement errors. Thus, we view the effect of schooling on average hourly wages as a better indicator of productivity increase.

We consider the following specification of the empirical wage equation

$$\ln w_i = \alpha + \beta_s s_i + \gamma_0 p_i + \gamma_1 p_i^2 + \varepsilon_i, \qquad (3.1)$$

where w_i is the wage rate for individual *i* with schooling level *s* and potential experience *p*, and β_s is the rate of return to schooling.[37] For now, we place the conditional independence restriction $E\left[\varepsilon_i \middle| s_i, p_i\right] = 0,$ deferring the problems of endogeneity and selection bias to Section 3.3. The data is segregated across gender and sectors and the wage equations are estimated separately for each group. Table A.11 provides OLS estimates of (3.1). We find that the estimates of education returns range between 4.8 and 7.8 percent, and are moderate compared to estimates for other countries (see Psacharopoulos and Patrinos (2004)). There is considerable variation in the educational returns between the public and the private sector,

[37] Using potential experience as a proxy for post-school human capital investments may be problematic, especially since it generally overestimates working experience for females. There is also an upper bound on experience. In addition, identification problems may arise when analyzing cohort effects because of perfect co-linearity between birth years, years of schooling and potential experience.

and between male and female workers employed in private sector. The simple model is capable of explaining a strikingly large portion of the wage dispersion in the public sector with R-squared values higher than 0.40. This is likely to be an artefact of a rigid and centralized wage structure based on education and seniority pay in the public sector. This is in contrast to the private sector where we are more likely to find incidence of performance pay and a decentralized wage bargaining. Table A.7 also documents higher variance in log-earnings for private sector industries. When gender and sector dummies are excluded from the pooled regression, the R-squared value falls to 0.2467. The schooling coefficient estimates are almost unchanged.

3.2.2 Functional form flexibility

In this section we relax two simplifying assumptions implicit in the standard specification (3.1) above. Firstly, interaction effects between education and experience are allowed. In other words, we do not restrict the log-wage experience profiles to be independent of schooling levels. Table A.12 reports the results from this estimation. We find significant interaction effects between schooling and experience. However, the schooling coefficient can no longer be interpreted as the marginal return to education. Figure A.3 shows how the returns to education vary with years of experience in different subpopulation. The estimates provide evidence for concavity in the returns profile across years of experience for each of the four gender-sector groups. However, the interaction effects are relatively small in the public sector. In the private sector, returns to education increase to the points where male workers have 20 years of experience and female workers have 15 years of experience, and fall sharply after that. The overall estimates suggest falling returns to education. This finding is in contrast with the assumption of parallel experience profiles across schooling levels, implicit in the specification (3.1).[38]

Secondly, we allow non-linear returns across years of schooling. The descriptive statistics provided in the appendix highlight some interesting facts in this context. Figure A.1 shows a non-linear relationship between years of schooling and average log wages. We observe spikes in the wage profile around 14/15, 18, and 20 years of schooling, irrespective of gender or sector. Such patterns may indicate differences in schooling returns across educational fields and/or irregularly high returns for completed degrees. We explore the possibility of non-linear

[38] Heckman et al (2006) also find evidence of converging experience profiles in the US census data for 1970-1990. An alternative way of testing this assumption is to estimate the experience profiles separately at each educational level and see if the experience profiles are parallel across schooling levels.

returns by (i) controlling for educational fields, and by (ii) replacing the continuous schooling variable by indicators for educational groups (see Tables A.3 and A.4 for definitions of the different categories). Table A.13 reports the estimation results once we have controlled for differences in educational fields.

Controlling for educational fields enables us to capture some of the wage dispersion within each schooling year, however maintaining linear schooling returns, we assume that the systematic non-linearity observed in the earnings-schooling schedule is caused solely by differences in the composition of educational fields across schooling years. This procedure does not allow us to test whether non-linearity is an intrinsic component of the wage schedule.[39] By combining educational fields and years of schooling, we construct new educational groups that enable us to detect non-linearity easily. We start by specifying the non-linear schooling model with dummy regressors for educational groups eg_j

$$\ln w_i = \alpha + \sum_{j=2}^{21} \beta_j eg_{ji} + \gamma_0 p_i + \gamma_1 p_i^2 + v_i, \tag{3.2}$$

where each of the *j* educational groups are allowed to have different coefficient. We continue to assume that the residual restriction $E[v_i \mid eg_i, p_i] = 0$ holds. Estimation results are displayed in Table A.14. We also consider the case of multilevel interaction effect, allowing each educational group to have a group-specific experience profile, such that

$$\ln w_i = \alpha + \sum_{j=2}^{21} \beta_j eg_{ji} + \gamma_0 p_i + \gamma_1 p_i^2 + \sum_{j=2}^{21} \delta_{0j} eg_{ji} p_i + \sum_{i=2}^{21} \delta_{1j} eg_{ji} p_i^2 + u_i. \tag{3.3}$$

We estimate equation (3.3) for the whole population, and find significant deviations in the *cumulative* educational returns. We also find significant variation in the group-specific returns across years of experience. Estimation results are presented in Table A.15 and Figure A.4.

3.2.3 Labour market characteristics

The human capital model assumes that the labour market is competitive, i.e. there are no wage rigidities, mobility costs, wage bargaining etc, so that wages equal the marginal productivity of labour. This theory does not predict any relationship between wages and employer characteristics. By contrast, the empirical evidence shows consistent variation in earnings across different industries, occupations, labour market regions, bargaining regimes, and other

[39] It is uncertain if variation in educational fields fully explains the three spike-points. Individuals having 14/15 years of schooling are highly likely to be enrolled in a vocational college, while those having undergraduate tertiary education are evenly distributed across educational fields.

firm characteristics. Meanwhile, some alternative theories of wage determination, such as the efficiency wage theory and agency theories, suggest that employer characteristics do matter.[40] Therefore, it is a common practise to control for employer and other labour market characteristics in wage regressions. However, estimates of returns to education are sensitive to inclusion of additional covariates. The interpretation of the schooling coefficient is no longer straightforward. Disregarding the wage-productivity relationship embedded in the human capital production function (2.3), we may consider the set of options that emerge once an individual takes more education, such as the possibility to work in highly paid industries, firms, or living in richer neighbourhoods, as a part of the returns to education. If the parameter of interest is the composite effect of education on wages, adding covariates that depend on education can bias our estimates unless proper corrections are made to handle the endogenous character of such variables.

At present, we disregard the problems discussed above and estimate the model (3.2) controlling for employer characteristics, with dummies for industry and labour market region.[41] Table A.16 gives the estimation results. It is likely to believe that individuals having different schooling are not randomly distributed across regions or industries. The composition of various educational groups is also likely to be different across regions. By adding dummies for labour market region and industry, we are able to control for effects of any unmeasured characteristics that are constant across individuals within each region-industry subgroup. Overlooking the endogeneity of schooling choices, we may interpret the education coefficients as the average returns to a given education when the individual is randomly assigned to a region-industry subgroup. Without these controls the education coefficients may capture the effects of being hired in a high-paid sector or living in a region with higher wages. As noted by Pereira and Martins (2004), the estimates for educational returns fall when controls for region and industry are added. We also find significant region-industry effects for all population groups. When controls for labour market characteristics are added, the model is able to explain more of the wage dispersion. The increase in R-squared is especially significant for the private sector. This indicates that employer characteristics are relatively more important determinants of the wage setting outcomes in the private sector, compared to the public sector.

---

[40] See Stiglitz (1986) for a formal overview of efficiency wage theories. Krueger and Summers (1988) provide some evidence for inter-industry wage differentials, while Molho (1992) presents a theory of local pay determination, considering regional wage differentials.

[41] The data material does not allow estimation of the more general model (3.3) when we add regional and industry dummies.

3.3 IV Estimation

3.3.1 Introduction

In the previous sections, we have ignored the problems relating to endogeneity of educational choices and selection bias arising through non-random participation in the labour market. In this section, we restrict our sample to non-immigrants who are currently not enrolled as students and are between 23 and 43 years of age in 2001. The primary reason for considering this particular age group is the availability of large scale demographic data for parents, such as their education, work status, occupation, industry, and housing location during the individual's childhood. Using the decennial National Population and Housing Censuses for the period 1960-2001 (see Vassenden (1987) and Statistics Norway (2006)), we are able to collect data on a large number of demographic variables, such as the number of children, parental characteristics, and marital status, in addition to non-labour earnings for a large portion of the population in this age group. An overview of the data selection process for merging the census data with the earnings data is given in Table A.2.

We follow two estimation strategies: Selection biases are corrected using the standard Heckman two-step and maximum likelihood (ML) procedures, while the two-stage least squares (2SLS) estimation method is used to handle the endogeneity of schooling. We consider both labour market participation and sector choices in a simultaneous ML estimation of the wage and participation equations, assuming that error terms follow a joint-normal distribution. Finally, we consider sector choices and endogenous schooling together. This is done in a two-step procedure, where we first estimate the inverse Mills ratio (IMR) for sector choice and impute this in the wage equation. Next, we estimate the wage equations by 2SLS to correct for the endogeneity of schooling. Section 3.3.2 provides estimates of the wage equations correcting for selection biases, while the presence of both endogenous schooling choices and selective sector-participation is considered in Section 3.3.3.

3.3.2 Selection corrections

Wages are observed only for those individuals who have a positive labour supply. Individuals outside the labour market are supposed to have an offered wage below their reservation wage. As schooling has a positive influence on wages, people with little schooling will on average have a lower offered wage and are less likely to participate. We observe the wages for only those individuals with little schooling who receive comparatively high wage offers. Thus, the OLS estimates of wages on years of schooling will be downward biased.

Heckman (1979) characterizes this as a latent-variable problem. We can illustrate this in the following model with categorical educational variables

$$\ln w_i^* = \alpha_1 + \sum_{j=2}^{21} \beta_{1j} eg_{ji} + \gamma_{10} p_i + \gamma_{11} p_i^2 + X_{1i}^{'} \delta_{10} + u_{1i}, \tag{3.4a}$$

$$D_i^* = \alpha_2 + \sum_{j=2}^{21} \beta_{2j} eg_{ji} + \gamma_{20} p_i + \gamma_{21} p_i^2 + X_{2i}^{'} \delta_{20} + Z_i^{'} \delta_{21} + u_{2i}, \tag{3.4b}$$

$$\ln w_i = \ln w_i^*, \qquad \text{if } D_i^* > 0, \tag{3.4c}$$

Equation (3.4a) gives the latent log-wage offered to an individual i with education j, experience p, in a labour market defined by the set of variables $X_1^{'}$. (3.4b) is the participation equation, with the decision to work being influenced by the individual's education, experience, labour market characteristics and a set of variables Z excluded from the wage equation (3.4a). We are interested in the wage returns for having a specific education. However, wages are observed only for those who work, as expressed in (3.4c). According to economic theory, people who are offered a relatively lower wage than what corresponds to their education are less likely to work, as their reservation wage is more likely to be higher than their offered wage. Thus, the error terms u_1 and u_2 are expected to be positively correlated. The error terms are commonly assumed to have a bivariate normal distribution, such that

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}\right] \tag{3.5}$$

Under assumption (3.5), (3.4c) becomes a probit selection equation with the outcome variable $D_i = 1$ if $D_i^* > 0$ and zero otherwise. There are two prominent ways of estimating the model (3.4). The standard two-step estimator proposed by Heckman considers the conditional expectation of log-wages for those inside the labour market

$$E\left[\ln w_i^* \middle| D_i = 1\right] = \alpha_1 + \sum_{j=2}^{21} \beta_{1j} eg_{ji} + \gamma_{10} p_i + \gamma_{11} p_i^2 + X_{1i}^{'} \delta_{10} + E\left[u_{1i} \middle| D_i = 1\right].$$

Under the joint normality of the errors terms, the last term can be expressed as

$$E\left[u_{1i}\middle|D_i=1\right]=\frac{\sigma_{12}}{\sigma_2}\frac{\phi\left(-\left(\alpha_2+\sum_{j=2}^{21}\beta_{2j}eg_{ji}+\gamma_{20}p_i+\gamma_{21}p_i^2+X_{2i}^{'}\delta_{20}+Z^{'}\delta_{21}\right)\Big/\sigma_2\right)}{1-\Phi\left(-\left(\alpha_2+\sum_{j=2}^{21}\beta_{2j}eg_{ji}+\gamma_{20}p_i+\gamma_{21}p_i^2+X_{2i}^{'}\delta_{20}+Z^{'}\delta_{21}\right)\Big/\sigma_2\right)}.$$

The so-called inverse Mills ratio (IMR) in the expression above can be estimated in a probit model. In the second step, the estimated IMR $\hat{\lambda}(.)$ can be imputed in the wage equation and the resulting wage equation

$$\ln w_i=\alpha_1+\sum_{j=2}^{21}\beta_{1j}eg_{ji}+\gamma_{10}p_i+\gamma_{11}p_i^2+X_{1i}^{'}\delta_{10}+\frac{\sigma_{12}}{\sigma_2}\hat{\lambda}\left(eg_{ji},p_i,X_{2i}^{'},Z_i^{'}\right)+v_i, \qquad (3.6)$$

can be estimated for those observations where the wages are observed. This is the Heckman two-step method that corrects for selection bias pertaining to non-random labour force participation. This characterization of the selection problem is similar to the omitted variable problem with IMR being the omitted variable if OLS was used on the sub-sample with observed wages. The two-step method is a limited-information maximum likelihood (LIML) method, where the second step uses information on the participating individuals only. However, a more efficient full-information maximum likelihood (FIML) is also possible. Under the assumption (3.5), the likelihood function for the model (3.4) is

$$L=\prod_{D_i=0}\left[1-\Phi\left(\frac{\alpha_2+\sum_{j=2}^{21}\beta_{2j}eg_{ji}+\gamma_{20}p_i+\gamma_{21}p_i^2+X_{2i}^{'}\delta_{20}+Z^{'}\delta_{21}}{\sigma_2}\right)\right]\times$$

$$\prod_{D_i=1}\left[\begin{array}{l}\Phi\left\{\left(\begin{array}{l}\alpha_2+\sum_{j=2}^{21}\beta_{2j}eg_{ji}+\gamma_{20}p_i+\gamma_{21}p_i^2+X_{2i}^{'}\delta_{20}+Z^{'}\delta_{21}\\+\frac{\sigma_{12}}{\sigma_1^2}\left(\ln w_i-\alpha_1-\sum_{j=2}^{21}\beta_{1j}eg_{ji}-\gamma_{10}p_i-\gamma_{11}p_i^2-X_{1i}^{'}\delta_{10}\right)\end{array}\right)\Bigg/\sqrt{\sigma_2^2-\frac{\sigma_{12}^2}{\sigma_1^2}}\right\}\times\\\frac{1}{\sigma_1}\phi\left(\frac{\ln w_i-\alpha_1-\sum_{j=2}^{21}\beta_{1j}eg_{ji}-\gamma_{10}p_i-\gamma_{11}p_i^2-X_{1i}^{'}\delta_{10}}{\sigma_1}\right)\end{array}\right].$$

Some exclusion restrictions are necessary for the selection corrections to be empirically tenable (Puhani, 2000). This is because of the almost linearity of the IMR. Therefore, we need to find some “instruments” that appear in the participation equation, but are absent from the wage equation. If not, there could be collinearity between the explanatory variables in the wage equation and the IMR. Traditionally, the number and age-composition of children have been used to identify female labour force participation, as they are likely to affect work status but not the offered wages. This is controversial if women spend time outside the labour force because of pregnancy and child-care. This would result in lower labour market experience and thereby a lower wage offer. However, we neglect this type of child-wage discrimination.

We construct a set of variables indicating the number of children in the age groups $a \in \{0,1,...,10,11-15,16-18,18-\}$ for each individual. In addition, we use the annual non-labour income in the household, i.e. capital income and the non-taxable components. These variables are excluded from the wage equations. Apart from the number of children and non-labour income, the individual's education, experience, and labour market region are used as explanatory variables in the participation equation. Summary measures on the variation in the excluded variables between the participating and the non-participating populations are given in Table A.8. Especially, we find a significant gap in the share of individuals with a self-employed father for the employed and the non-participating male populations. Using these variables, we estimate probit participation equations. Results are given in Table A.17. We find the number of children in different age groups is a particularly strong predictor of labour force participation solely for females. Meanwhile, non-labour income has a negative effect for all groups. We also find that the likelihood of an individual being a paid-worker reduces significantly if the parents were self-employed during the individual's childhood.

The wage equations are estimated using both the simultaneous ML and the two-step Heckman procedures. Results are given in Table A.18. We find small differences in the estimated coefficients, though the two-step estimates are slightly smaller for all groups and the standard errors are smaller for the ML estimates. Selection effects are found to be significant, though relatively small. Coefficients of the participation-correcting IMRs are significant. We find the OLS estimates to be upward biased for each group. The difference in the cumulative returns is usually on the scale of 1-3 percent points. However, there is greater disparity between sub-sample OLS and ML estimates for males (> 0.01) than females (<0.01), with the positive selection bias being large for males. These findings are consistent with the sorting hypothesis stating that individuals with higher returns are more likely to enter the labour market. Though, the magnitude varies across different educational groups. However, our estimates show the correlation between the errors terms (*rho*) to be negative in each case. This contradicts the view that there is a positive selection bias and is puzzling given the change in the educational returns. The residual variation in log-wages (*sigma*) is estimated to be greater for the male population.

Sector choices

Empirical evidence shows that there are persistent wage gaps for observably equal individuals across sectors. OLS estimates show that the returns to education are considerably higher in the private sector (see Tables A.19 and A.20). However, it is difficult to say if the higher returns to education are caused solely by sector-specific attributes (such as working conditions, flexible schedules, job-stability, bargaining regimes etc) or differences in the unobserved worker skills composition across sectors. It is likely that sector choices are results of optimising behaviour by the agents. An individual would choose to work in a sector only if the net wage offer and the non-pecuniary benefits are higher than in the outside option, that is working in another sector, being unemployed, or self-employment. In order to make causal statements on sector-specific wage differentials, we need to have knowledge of the potential wages of an individual in each sector. Since potential wages are unobserved for the outside option, this problem is paramount to the latent-variable problem considered earlier. The wages are observed only for an individual who participates in the labour force *and* only for the sector where he/she is employed.

We consider this problem in the following latent-variable model

$$\ln w_i^{l*} = \alpha_1 + \beta_1 s_i + \gamma_{10} p_i + \gamma_{11} p_i^2 + X_{1i}^{'} \delta_{10} + u_{1i}^l, \qquad \text{for } l = \{1,2\}, \tag{3.7a}$$

$$D_i^{l*} = \alpha_2 + \beta_2 s_i + \gamma_{20} p_i + \gamma_{21} p_i^2 + X_{2i}^{'} \delta_{20} + Z_{1i}^{'} \delta_{21} + Z_{2i}^{'} \delta_{22} + u_{2i}^l, \tag{3.7b}$$

$$\ln w_i = \ln w_i^{l*}, \qquad \text{if } D_i^{l*} > 0, \tag{3.7c}$$

Equation (3.7a) is the latent sector wage equation, with (3.7c) indicating that the observed wages for an individual are equal to the sector-specific wage offers if the individual chooses to work in that sector. As in the previous section, we assume that the error terms u_1^l and u_2^l are assumed to have a bivariate normal distribution. Thus we can estimate the sector-participation equation (3.7b) in a probit model. It is important to note that there is an additional set of excluded variables Z_2 in (3.7b), such as parental work status, occupation and industry, identifying the individual's sector choices. Individuals choose between either working in a given sector or to follow the outside option, i.e. work in another sector or remain outside the work force. By reducing this polychotomous problem to a binary choice, we encounter some efficiency loss. However, this considerable simplifies our analysis. Sector-specific wage equations are estimated simultaneously with the sector-participation equations for both males and females. In order to avoid collinearity problems between the explanatory variables and the correction terms, exclusion restrictions on sector choices are needed. For this purpose, we exploit the information on parental work status (self-employment or paid-workers) and combinations of parental occupation and industries during the individual's childhood. As before, the number of children at different age groups and non-labour income are also included.

The first-stage probit estimates are given in Table A.17, while Tables A.19 and A.20 provide the estimates of the wage equations for different subpopulations. The results suggest some interesting patterns. Sector-specific wage differentials persist even after correcting for participation effects. For public sector employees the OLS estimates are upward biased. The bias is considerably larger for males. These results are consistent with the sorting hypothesis with the most able workers entering the labour force, and sorting behaviour being more widespread for males. However we find the opposite to be the case in the private sector. For most educational groups, the selection-corrected estimates exceed the OLS estimates with a few percentage points. The sector-specific IMRs have significant negative coefficients for all groups. The correlation parameters are also negative. These results cast doubt on our previous conclusions on the existence of positive sorting.

3.3.3 Endogenous schooling

Earlier studies on the returns to education in Norway by Hægeland et al (1999) and Aakvik et al (2003) focus on the endogeneity of schooling choices. Both studies emphasize non-linear returns to education. Hægeland et al consider this in an explicit ordered multinomial probit choice model of the years of schooling, while Aakvik et al construct a comparative advantage model with sequential choices of seven schooling levels. Both studies utilise instruments to find exogenous variation in schooling choices. While the former approach exploits the ordered structure on schooling and uses the same set of instruments for every choice of years of schooling, the latter requires a valid instrument at each level of schooling. This is done in an experimental framework, exploiting the staged implementation of increase in years of compulsory education from seven to nine years through a major reform in the school system in the 1960s. The estimated returns to education can be given a LATE interpretation (see

Section 2.3.5), being the returns "for a person acquiring an extra year of education just because of the educational reform and who would have dropped out of education after seven years otherwise" (Aakvik et al (2003), p. 28).

Our preliminary estimates suggest significant non-linearity in the wage returns to education across gender and sector, also after correcting for participation bias. However, further research is needed to uncover the importance of non-linear returns of education on hourly wages in presence of endogenous schooling choices. For simplicity, we disregard non-linear returns to schooling. Meanwhile, we allow heterogeneous returns to schooling but assume that schooling decisions are unrelated to the individual gains. Thus, there is only an ability bias and no returns bias. In this case, the IV estimates provide consistent estimates of the ATT (see section 2.3.2), as shown by Blundell et al (2005).

We estimate the probit participation equations (3.7b) and calculate the sector-specific IMRs λ_i^l. After inserting λ_i^l in the wage equations (3.7a) we get

$$\ln w_i^l = \alpha_1 + \beta_1 s_i + \gamma_{10} p_i + \gamma_{11} p_i^2 + X_{1i}^{'} \delta_{10} + \frac{\sigma_{12}^l}{\sigma_2^l} \hat{\lambda}^l + v_i^l, \qquad \text{for } l = \{1, 2\}. \tag{3.8}$$

However, we may still have an endogeneity problem because of unobservable characteristics, for instance worker ability, affecting both wages and schooling choices and thereby creating a correlation between the error term v_i^l and s_i. In the schooling returns literature, it is common to use supply side variables, such as proximity to educational institutions, and family background variables such as parental education, income and work status, as IVs for schooling. A comprehensive survey of studies based on IV methods is given in Card (1999). However, the validity of most IVs used in the empirical literature is questionable. For now, we follow the earlier studies and use parental education and county/region where the individual grew up (as a proxy for geographic differences in accessibility of schools).

The equations (3.8) are estimated by 2SLS taking into account the endogeneity of schooling. Wage equations both with and without sector-specific IMRs are estimated, to see how endogeneity and selection interact. Results are given in Tables A.21-A.25. As most previous studies based on IV methods, we find that OLS estimates to be significantly downward biased. In fact, our IV estimates rise even further when we take into account sector choices. These results are puzzling since omitted worker ability, which is likely to correlate positively with schooling, would suggest that more able workers have higher education and thereby create an upward bias the OLS estimates. Secondly, while our preliminary selection corrections support the positive selection hypothesis, the direction of the selection bias is opposite when we consider endogenous schooling. This is the case for all subgroups we consider. On average, we find that the OLS estimates have a negative endogeneity bias of about 30 percent and a negative selection bias of about 4 percent relative to the selection corrected IV estimates.

One reason for the higher IV estimates could be that we are using parental education as an IV without controlling directly for family background in the wage equations. Bound and Jaeger (1996) suggest that this may enhance the unobserved differences between the characteristics of the treatment and comparison groups implicit in the IV approach. This is likely if we can expect intergenerational persistence in unobserved earnings potential. If so, the validity of parental education as an IV is highly questionable. Griliches (1977) points to measurement errors as another source of bias in the OLS estimates. If ability biases are relatively small, then some of the disparity in the IV and OLS estimates may reflect downward bias in the OLS estimates due to measurement errors in the schooling variable. This can certainly be the case for the register data that we use, especially for older cohorts.

Yet another case is that there is some underlying heterogeneity in the returns to education, with availability of schools being comparably more important than individual ability in the choice of schooling. In that case, the IV estimates based on birthplace or childhood-county are most likely to affect the schooling choices of individuals with higher-than-average marginal returns to education. With low accessibility of schooling, these individuals would have taken relatively low schooling. Thus, the IV estimates may give us the LATE for high-return individual who pursue higher education because of a higher accessibility to schools. This may give us a positive difference between the IV and OLS estimates. When we include childhood-region as the only instrument, we do find lower returns to education than the returns we get by also including parental education. However, the IV estimates are still higher than the OLS estimates. These results may indicate intergenerational persistence in unobserved earnings potential, while we cannot rule out the LATE interpretation.

4. Conclusion

In this study we estimate the returns to education on hourly wages in Norway, taking into account the endogeneity of schooling and occupational sector selection. Our results show significant variation in the estimated returns to education across different population groups, with the wage return to an additional year of schooling ranging between 4.8 and 7.8 percent. However, we find strong evidence of non-linearity in the returns across different schooling levels and education fields for all population groups we consider. Interaction effects between education and experience are also present, leading to a rejection of the standard log-linear Mincer wage equation.

Selection effects are found to be significant, though of a moderate size. The magnitude (and direction) of the selection bias varies across different educational groups. In general the OLS estimates are upward biased for the public sector, with the bias being considerably larger for males. These results are consistent with the positive sorting hypothesis suggesting that the more able workers are more likely to participate in the labour force. However, we find the opposite to be the case for private sector workers. These results cast doubt on the existence of positive sorting. Sector-specific differences in the returns to education persist even after correcting for selection effects. We find significantly higher educational returns in the private sector, especially for male workers.

As most previous studies based on IV methods, we find that OLS estimates to be significantly downward biased. In fact, our IV estimates rise even further when we take into account sector choices. These results are puzzling since omitted worker ability, which is likely to correlate positively with schooling such that that more able workers also have higher education, is supposed to give an upward bias in the OLS estimates. Secondly, while our preliminary selection corrections support the positive selection hypothesis, the direction of the selection bias is opposite when we consider endogenous schooling. This is the case for all subgroups we consider. On average, we find that the OLS estimates to have a negative endogeneity bias of about 30 percent and a negative selection bias of about 4 percent relative to the selection corrected IV estimates.

5. References

Aakvik, A., K.G. Salvanes and K. Vaage (2003): "Measuring Heterogeneity in the Returns to Education in Norway Using Educational Reforms", CEPR, Discussion Papers No. 4088.

Angirst, J.D. and G.W. Imbens (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity", Journal of the American Statistical Association 90, No. 430, 431-442

Angrist, J. and A. Krueger (1999): "Empirical Strategies in Labor Economics", Chapter 23 in O. Ashenfelter and D. Card, eds., Handbook of Labor Economics, Volume 3, North Holland, New York.

Barth, E. and Røed, M. (1999): "The Return to Human Capital in Norway", in R. Asplund and P.T. Pereira, eds., Returns to Human Capital in Europe. ETLA Research Institute of the Finnish Economy, Helsinki.

Becker, G. (1964): Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. National Bureau of Economic Research, New York.

Becker, G. (1967): "Human Capital and the Personal Distribution of Income: An Analytical Approach", Woytinsky Lecture No. 1. University of Michigan, Institute of Public Administration, Ann Arbor.

Ben-Porath, Y. (1967): "The production of human capital and the life cycle of earnings", Journal of Political Economy 75 No. 4, 352-365

Bhuller, M.S. (2008): "Inndeling av Norge i arbeidsmarkedsregioner (Labour market regions in Norway)", Statistisk sentralbyrå, Mimeo.

Björklund, A. and R. Moffitt (1987): "The estimation of wage gains and welfare gains in self-selection", Review of Economics and Statistics 69 No. 1, 42-49

Blundell, R., L. Dearden and B. Sianesi (2005): "Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey", Journal of Royal Statistical Society 168 Part 2, 473-512

Bound, J. and D. A. Jaeger (1996): "On the validity of season of birth as an instrument in wage equations: a comment on Angrist and Krueger's 'Does compulsory school attendance affect schooling and earnings?'", NBER Working paper no. 5835, Cambridge.

Card, D. (1999): "The causal effect of education on earnings", in O. Ashenfelter and D. Card eds., Handbook of Labor Economics, Volume 5. North-Holland, New York.

Card, D. (2001): "Estimating the return to schooling: Progress on some persistent econometric problems", Econometrica 69 No. 5, 1127-1160

Cunha, F., J. J. Heckman, L. J. Lochner, and D. V.Masterov (2006): "Interpreting the evidence on life cycle skill formation", in E. A. Hanushek and F. Welch eds., Handbook of the Economics of Education, Volume 1. North-Holland, New York.

Dale-Olsen, H. (1997): Lønn for arbeid: En gjennomgang av empiriske studier av lønnsforskjeller i Norge på 80- og 90-tallet. Institutt for samfunnsforskning, Oslo.

Dehejia, R. and S. Wahba (2002): "Propensity Score-Matching Methods For Nonexperimental Causal Studies", The Review of Economics and Statistics 84, No. 1, 151-161

Frost, P.A. (1979): "Proxy Variables and Specification Bias", The Review of Economics and Statistics 61, No. 2, 323-325

Gibbons, R., L.F. Katz, T. Lemieux, and D. Parent (2005): "Comparative Advantage, Learning, and Sectoral Wage Determination", Journal of Labor Economics 23, No. 4, 681-723

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems", Econometrica 45, No. 1, 1-22.

Gronau, R. (1974): "Wage Comparisons - A Selectivity Bias", Journal of Political Economy 82, No. 6, 1119-1143

Grini, H. (2007): "Quality Report on the Norwegian Structure of Earnings Survey 2002", Report 2007/14. Statistics Norway, Oslo.

Hardoy, I. and P. Schøne (2006): "The Part-Time Wage Gap in Norway: How Large is It Really?", British Journal of Industrial Relations 44, No. 2, 263-282

Harmon, C. P., H. Oosterbeek, and I. Walker (2003): "The Returns to Education: Microeconomics", Journal of Economic Surveys 17, 115-156

Heckman, J.J. (1976): "A Life-Cycle Model of Earnings, Learning, and Consumption", Journal of Political Economy 84, No. 4, 11-44

Heckman, J.J. (1979): "Sample Selection Bias as a Specification Error", Econometrica 47, No. 1, 153-61

Heckman, J.J. (2001): "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture", Journal of Political Economy 109, No. 4, 673-748

Heckman, J.J., H. Ichimura, J. Smith, and P. E. Todd (1998): "Characterizing selection bias using experimental data", Econometrica 66, No. 5, 1017-1098

Heckman, J.J. and X. Li (2004): "Selection bias, comparative advantage and heterogeneous returns to education: evidence from China in 2000", Pacific Economic Review 9, No. 3, 155-171

Heckman, J.J., L.J. Lochner, J. Lance and P.E. Todd (2006) "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond", in E. A. Hanushek and F. Welch eds., Handbook of the Economics of Education, Volume 1. North-Holland, New York.

Heckman, J.J. and S. Navarro-Lozano (2007) "Dynamic discrete choice and dynamic treatment effects", Journal of Econometrics, Elsevier 136, No. 2, 341-396

Heckman, J.J. and G. Sedlacek (1990): "Self-Selection and the Distribution of Hourly Wages", Journal of Labor Economics 8, No. 1, 329-363

Heckman, J.J. and E.J. Vytlacil (2007): "Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation", in J.J. Heckman and E. Leamer eds., Handbook of Econometrics, Volume 6. Elsevier, Amsterdam.

Heckman, J.J. and Willis, R.J. (1977): "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women", Journal of Political Economy 85, No. 1, 27-58

Hægeland, T., T.J. Klette and K.G. Salvanes (1999): "Declining Returns to Education in Norway? Comparing Estimates across Cohorts, Sectors and over Time", Scandinavia Journal of Economics 101, No. 4, 555-576

Imbens, G. W. (2000): "The role of the propensity score in estimating dose-response functions", Biometrika 83, 706-710.

Krueger, A.B. and L.H. Summers (1988): "Efficiency wages and the inter-industry wage structure", Econometrica 56, 259–293

Lee, L. (1978): "Unionism and Wage Rates: Simultaneous Equations Model with Qualitative and Limited Dependent Variables", International Economic Review 19, No. 2, 415-433

Lee, L. (1983): "Generalized Econometric Models with Selectivity", Econometrica 51, No. 2, 507-12

Maddala, G.S. (1983): Limited-dependent and qualitative variables in econometrics. Econometric Society Monographs No. 3. Cambridge University Press, Cambridge.

Manski, C.F. (1995): Identification Problems in the Social Sciences. Harvard University Press: Cambridge.

McFadden, D. (1980): "Econometric Models of Probabilistic Choice", in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data. MIT Press, Cambridge.

Mincer, J. (1958): "Investment in human capital and personal income distribution", Journal of Political Economy 66, No. 4, 281-302

Mincer, J. (1962): "Labor-Force Participation of Married Women: A study of Labor Supply" in H. G. Lewis, ed., Aspects of Labor Economics. Princeton University Press, New Jersey.

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

Molho, I. (1992): "Local Pay Determination", Journal of Economic Surveys 6, No. 2, 155-94

Moffitt, R. (1984): "The Estimation of a Joint Wage-Hours Labor Supply Model", Journal of Labor Economics 2, No. 4, 550-66

Pereira, P.T. and P.S. Martins (2004): "Returns to education and wage equations", Applied Economics 36, No. 6, 525-531

Pessino, C. (1991): "Sequential Migration: Theory and Evidence from Peru", Journal of Development Economics 36, 55–87

Psacharopoulos, G. and H. A. Patrinos (2004): "Returns to investment in education: a further update", Education Economics 12, No. 2, 111-134

Puhani, P.A. (2000): "The Heckman Correction for Sample Selection and Its Critique", Journal of Economic Surveys 14, No. 1, 53-68

Raaum, O. (1999): "Inntektseffekter av utdanning i Norge - en litteraturoversikt", Arbeidsnotat. Frischsenteret, Oslo.

Rosen, S. (1976): "A Theory of Life Earnings", Journal of Political Economy 84, No. 4, 45-67

Rosenbaum, P.R. and D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika 70, 41–55.

Rubin, D.B. (1986): "Statistics and Causal Inference: Comment: Which Ifs Have Causal Answers", Journal of the American Statistical Association 81, No. 396, 961-962

Statistics Norway (2005): NOS D 362 Wage Statistics 2005, Oslo.

Statistics Norway (2006): NOS D 353 Population- and Housing Census 2001, Oslo.

Stiglitz, J.E. (1986): "Theories of Wage Rigidity", NBER Working Papers 1442. National Bureau of Economic Research, New York.

Thurow, L.C. (1983): Dangerous Currents: The State of Economics. Oxford University Press, Oxford.

Tummers, M. and I. Woittiez (1991): "A simultaneous wage and labor supply model with hours restrictions", Journal of Human Resources 26, 393–423

Vassenden, K. (1987): "Folke- og Boligtellingene 1960, 1970 og 1980 (The Population- and Housing Census 1960, 1970 and 1980)", Rapporter fra Statistisk Sentralbyrå 87/2, Oslo.

Vella, F. (1993): "Non-Wage Benefits in a Simultaneous Model of Hours and Wages: Labor Supply Functions for Young Females", Journal of Labor Economics 11, 704-723

Weiss, Y. (1987): "The Determination of Life-Cycle Earnings", in O. Ashenfelter and R. Layard eds., Handbook of Labor Economics, Volume 1. North Holland, Amsterdam.

Willis, R. and S. Rosen (1979): "Education and Self-Selection", Journal of Political Economy 87, 7-36

6. Appendix

6.1 Variable definitions

Basic paid monthly salaries (BPS). This variable captures the actual cash payments from the employer to the employee for work rendered, and can be denoted as the regular basic monthly earnings in NOK. Taxes, national insurance contributions or other payments deducted by the employer are included in this measure.

Variable additional allowances (VAA). This variable includes allowances for inconvenient working conditions and work during off-hours, call-out allowance, shift allowance, and other irregularities. The measure is a calculated monthly average based on the aggregate allowances given to the employee between 1 January and the time of the census.

Bonuses and commissions (BC). This variable includes allowances that are usually not connected with specific duties and where the payments occur irregularly with respect to the period in which they are earned or to which they apply. Profit sharing, production allowance and gratuities figure in this category. The measure is an adjusted monthly average based on the aggregate allowances between 1 October of the previous year and the time of the census.

Gross average monthly earnings. This variable adds all the three previous measures (BPS + VAA + BC), and provides a measure of the total monthly earnings. By adjusting the aggregate additional allowances and bonus measures by a monthly average, we assume that the irregular payments are distributed smoothly across the year. Such payments are an intrinsic part of many job contracts and should therefore feature in a measure of overall earnings. Meanwhile, overtime payment not included as they are directly related to the additional time/effort made by the individual worker.

Contracted weekly working hours (H). This variable provides a measure of the average number of hours per week for the year or for the last month given in the job contract. Meal breaks are excluded, however, no deductions are made for absences due to holiday, illness, leave of absence or the like. Due to overtime hours or other irregularities in the work schedule, this measure may not be equal to actual working hours.

Gross hourly wage (W). Using the above measure of gross average monthly earnings and contracted weekly working hours, we are able to construct an estimate of gross hourly wages. The formula used for this calculation is: (BPS + VAA + BC)/(4.3 x H), when using the average number of weeks in a month equal to 4.3.

Average monthly overtime payment. This variable covers the sum of cash compensation for work done beyond contractual working hours. Overtime compensation is a calculated average per month over the period 1 January to the time of census.

Years of schooling (S). This variable gives the number of academic years of schooling corresponding to the individual's highest completed level of education. The information is obtained from the register of the Population's Highest Education.

Educational Field (EF). This variable is based on a 6-digit education code from the Norwegian Standard Classification of Education (NUS). We use the second digit to identify educational field and classify eight study fields, such as humanities, social sciences, technical studies etc (see Table A.3). NUS-codes are taken from the National Education Database (NUDB).

Educational Group (EG). By combining years of schooling (in discrete levels) and educational field, we construct a variable indicating educational group. This is helpful in testing non-linear returns to schooling across levels and fields. In all, 21 such schooling categories are defined (see Table A.4).

Potential Labour Market Experience (E). We use the standard measure of potential experience, used since Mincer (1974). Potential experience is calculated as the individual's age at the time of census minus the age at school completion (i.e., E = A - S - 7).

Industry (I). This variable is based on the 5-digit Standard Industrial Classification (NACE) code given in the Wage Statistics. Public sector employees are employed either in public health services, public education, municipal services or central public administration. For private sector workers, the industrial codes are aggregated up to 11 major industries (see Table A.5).

Labour Market Region (R). We use the municipality codes provided in the Central Register of Establishments and Enterprises for firm's location and identify the local labour market region. The regional classification is documented in Bhuller (2008). Table A.6 provides a list of the 46 local labour market regions used in the analysis. Unlike earlier studies, this regional classification is based on worker commutation across municipalities and is not restricted by county-level administrative boundaries.

6.2 Descriptive statistics

Table A.1. Sample selection process I, earnings data

Steps		Number of individuals (N) in 2001
Gross sample (Wage Statistics, 2001)		1 235 241
Age group 20-64		1 159 453
Working hours restricted to 1-75 per week		1 198 173
Hourly wage between NOK 50 and NOK 2000		1 198 190
- Multiple jobs	37 387	
- Immigrants	159 274	
- Students	96 025	
- Missing other covariates	127 493	
Net sample		889 048

Table A.2. Sample selection process II, population

Steps		N in 2001
Residents (Census, 2001)		4 520 947
- Immigrants	522 358	
- Students	424 866	
Age group 23-43		1 371 726
Population of interest		1 040 205
Paid workers		880 841
Earnings data (Wage Statistics 2001)		433 767
Net sample		387 017
Public sector employees		139 570
Private sector employees		147 879
Non-participants		99 568

Table A.3. Description of educational fields

Field	Description	2nd digit in the NUS-code	N in 2001
1	General programmes	0	268 737
2	Humanities and arts	1	39 381
3	Teacher training and pedagogy	2	82 023
4	Social sciences and law	3	22 353
5	Business and administration	4	124 008
6	Natural sciences, vocational and technical subjects	5	187 458
7	Health, welfare and sport	6	130 217
8	Transport, communications, safety and security	8	34 871

Table A.4. Description of educational groups

Group	Description	~ Years of schooling	NUS: 1st digit	NUS: 2nd digit	N in 2001
1	Lower secondary	9	2	-	163 269
2	Upper secondary, basic; general subjects	11	3	0-3,7	46 798
3	Upper secondary, basic; business and administration	11	3	4	44 249
4	Upper secondary, basic; vocational and technical	11	3	5,8	43 169
5	Upper secondary, basic; health, welfare and sport	11	3	6	38 358
6	Upper secondary, final year; general subjects	12	4	0-3,7	64 743
7	Upper secondary, final year; business and administration	12	4	4	33 903
8	Upper secondary, final year; vocational and technical	12	4	5,8	97 388
9	Upper secondary, final year; health, welfare and sports	12	4	6	24 198
10	Post upper secondary, non-tertiary education, all subjects	13+	5	0-8	30 896
11	Tertiary education, undergraduate; arts and pedagogy	16	6	1,2	100 606
12	Tertiary education, undergraduate; social sc., law or business	16	6	3,4	42 816
13	Tertiary education, undergraduate; natural sciences	16	6	5	30 297
14	Tertiary education, undergraduate; health, welfare and sport	16	6	6	58 593
15	Tertiary education, undergraduate; transport and comm.	16	6	8	8 613
16	Tertiary education, graduate; arts and pedagogy	18	7	1,2	10 292
17	Tertiary education, graduate; social science, law or business	18	7	3,4	13 580
18	Tertiary education, graduate; natural sciences	18	7	5	21 843
19	Tertiary education, graduate; health and sports	18	7	6	7 911
20	Tertiary education, graduate; transport and comm..	18	7	8	3 096
21	Postgraduate education; all subjects	21+	8	-	4 430

Table A.5. Description of industries

Industry	Description	Sector	NACE: 2 digits	N in 2001
1	Oil and gas extraction, mining	Private	10 – 14	17 221
2	Manufacturing	Private	15 – 37	116 476
3	Sewage, electricity and water supply	Private	40, 41, 90	3 464
4	Construction	Private	45	32 053
5	Wholesale and retail trade	Private	50 - 52	93 776
6	Hotels and restaurants	Private	55	11 272
7	Transport, storage and communications	Private	60 - 64	61 362
8	Finance, real estate, renting and business activities	Private	65 - 67, 70 - 74	70 615
9	Central government	Public	75	96 392
10	Public educational institutions	Public	76	74 397
11	Municipal services	Public	77	213 777
12	Public health services	Public	78	62 883
13	Private educational services	Private	80	3 257
14	Private health and social work activities	Private	85	17 616
15	Recreational and cultural services	Private	91 - 93	14 487

Table A.6. Description of regions

Number	Local labour market region	Code	N in 2001	Number	Local labour market region	Code	N in 2001
1	Sør-Østfold	11	30 894	24	Sunnfjord (Førde/Florø)	51	9 884
2	Oslo	12	255 342	25	Sognefjord (Sogndal/Årdal)	52	6 370
3	Vestfold	13	35 623	26	Nordfjord	53	4 975
4	Kongsberg	14	6 650	27	Søndre Sunnmøre	54	8 487
5	Hallingdal	15	3 849	28	Ålesund	55	15 163
6	Valdres	21	3 358	29	Molde	56	12 099
7	Gudbrandsdalen	22	5 866	30	Nordmøre	57	4 300
8	Lillehammer	23	7 619	31	Kristiansund	58	1 098
9	Gjøvik	24	13 954	32	Trondheim	61	61 458
10	Hamar	25	17 837	33	Midt-Trøndelag	62	11 888
11	Kongsvinger	26	10 028	34	Namsos	63	6 407
12	Elverum	27	7 458	35	Ytre Helgeland	64	5 484
13	Tynset/Røros	28	4 257	36	Indre Helgeland	65	10 327
14	Nordvest-Telemark	31	3 876	37	Bodø	71	17 917
15	Øst-Telemark	32	4 486	38	Narvik	72	5 986
16	Sør-Telemark	33	25 034	39	Vesterålen	73	5 580
17	Arendal	34	15 164	40	Lofoten	74	3 727
18	Kristiansand	35	25 739	41	Harstad	75	6 404
19	Lister	36	5 596	42	Midt-Troms	76	6 867
20	Stavanger	41	60 134	43	Tromsø	77	17 562
21	Haugesund	42	18 886	44	Alta	81	3 933
22	Sunnhordland	43	9 678	45	Hammerfest	82	4 598
23	Bergen	44	82 537	46	Vadsø	83	4 669

Table A.7. Descriptive statistics for the net sample

Variables	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max	Mean	Std Dev	Min	Max
	Total (N = 889 048)				Male (N = 412 580)				Female (N = 476 468)			
Age	42.77	11.00	20	64	42.52	11.14	20	64	42.99	10.88	20	64
Log-wage	4.99	0.28	3.91	7.56	5.09	0.31	3.91	7.56	4.90	0.21	3.92	7.02
Schooling (years)	13.28	2.74	8	22	13.42	2.75	8	22	13.17	2.73	8	22
Experience (years)	22.49	11.69	0	49	22.10	11.52	0	49	22.82	11.82	0	49
	Public (N = 447 449)				Male, Public (N = 136 500)				Female, Public (N = 310 949)			
Age	44.20	10.67	20	64	45.02	10.76	20	64	43.83	10.61	20	64
Log-wage	4.94	0.21	4.11	7.31	5.04	0.23	4.11	7.31	4.89	0.18	4.45	6.80
Schooling (years)	13.98	2.91	8	22	14.76	2.97	8	22	13.63	2.82	8	22
Experience (years)	23.22	11.50	0	49	23.26	11.19	0	49	23.20	11.64	0	49
	Private (N = 441 599)				Male, Private (N = 276 080)				Female, Private (N = 165 519)			
Age	41.33	11.14	20	64	41.29	11.11	20	64	41.40	11.20	20	64
Log-wage	5.03	0.33	3.91	7.56	5.11	0.34	3.91	7.56	4.91	0.26	3.92	7.02
Schooling (years)	12.58	2.36	8	22	12.75	2.37	8	22	12.30	2.30	8	22
Experience (years)	21.74	11.83	0	49	21.53	11.64	0	49	22.10	12.12	0	49

Table A.8. Means of excluded variables in participation equations across different subgroups

	Total			Males			Females		
Variables	Public	Private	Non-part.	Public	Private	Non-part.	Public	Private	Non-part.
Children_0	0.1137	0.1037	0.1118	0.1378	0.1278	0.0975	0.1060	0.0679	0.1210
Children_1	0.1078	0.1142	0.1472	0.1432	0.1337	0.1072	0.0964	0.0851	0.1730
Children_2	0.1213	0.1254	0.1378	0.1475	0.1386	0.1145	0.1129	0.1058	0.1529
Children_3	0.1222	0.1245	0.1328	0.1407	0.1342	0.1223	0.1162	0.1100	0.1396
Children_4	0.1252	0.1268	0.1331	0.1437	0.1356	0.1249	0.1193	0.1138	0.1383
Children_5	0.1270	0.1294	0.1370	0.1429	0.1374	0.1303	0.1220	0.1174	0.1413
Children_6	0.1243	0.1259	0.1343	0.1361	0.1307	0.1325	0.1205	0.1187	0.1355
Children_7	0.1234	0.1240	0.1298	0.1325	0.1268	0.1293	0.1205	0.1200	0.1300
Children_8	0.1181	0.1214	0.1267	0.1226	0.1219	0.1296	0.1167	0.1206	0.1248
Children_9	0.1173	0.1184	0.1237	0.1186	0.1176	0.1232	0.1168	0.1196	0.1241
Children_ 10	0.1162	0.1130	0.1199	0.1109	0.1105	0.1238	0.1179	0.1168	0.1173
Children_11-15	0.4725	0.4227	0.4593	0.3917	0.3938	0.4786	0.4984	0.4657	0.4468
Children_16-18	0.1600	0.1271	0.1545	0.1020	0.1056	0.1499	0.1787	0.1591	0.1575
Children_18 -	0.1240	0.0892	0.1260	0.0482	0.0554	0.0883	0.1483	0.1393	0.1504
Father_ind Missing	0.0371	0.0367	0.0504	0.0354	0.0367	0.0494	0.0377	0.0368	0.0510
Father_ind1	0.1046	0.0844	0.1474	0.0805	0.0794	0.2164	0.1123	0.0919	0.1028
Father_ind2	0.0100	0.0101	0.0097	0.0089	0.0103	0.0084	0.0103	0.0099	0.0106
Father_ind3	0.2420	0.2834	0.2499	0.2206	0.2904	0.2182	0.2489	0.2730	0.2703
Father_ind4	0.0162	0.0149	0.0113	0.0166	0.0161	0.0097	0.0161	0.0130	0.0123
Father_ind5	0.1307	0.1322	0.1362	0.1190	0.1311	0.1327	0.1345	0.1340	0.1384
Father_ind6	0.1079	0.1284	0.1150	0.1081	0.1261	0.1068	0.1079	0.1319	0.1204
Father_ind7	0.1138	0.1247	0.1184	0.1108	0.1242	0.1114	0.1147	0.1255	0.1230
Father_ind8	0.0334	0.0360	0.0288	0.0371	0.0356	0.0248	0.0322	0.0365	0.0313
Father_ind9	0.2043	0.1491	0.1330	0.2629	0.1502	0.1222	0.1854	0.1476	0.1399
Mother self-employ.	0.0920	0.0801	0.1191	0.0806	0.0770	0.1699	0.0957	0.0848	0.0864
Father self-employ.	0.1627	0.1496	0.2151	0.1369	0.1412	0.2876	0.1710	0.1620	0.1684
Observations	139 570	147 879	99 568	33 935	88 358	39 081	105 635	59 521	60 487

Table A.9. Log hourly wages and years of schooling across different subpopulations

Years	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
	Total (N = 889 048)					Male (N = 420 480)					Female (N = 486 450)				
8	17 655	4.85	0.20	3.92	6.47	8 435	4.93	0.22	3.92	6.47	9 220	4.78	0.15	3.96	5.95
9	36 654	4.83	0.19	3.92	7.31	14 002	4.93	0.21	3.97	7.31	22 652	4.77	0.14	3.92	6.05
10	98 201	4.83	0.22	3.92	6.93	44 935	4.90	0.24	3.92	6.93	53 266	4.77	0.18	3.92	6.67
11	110 129	4.89	0.22	3.93	6.93	34 900	5.01	0.26	3.96	6.93	75 229	4.83	0.16	3.93	6.76
12	66 760	4.91	0.23	3.92	7.29	27 166	5.01	0.27	3.92	7.29	39 594	4.85	0.16	3.93	6.68
13	235 754	4.95	0.27	3.91	7.38	138 620	5.03	0.28	3.91	7.38	97 134	4.84	0.20	3.94	7.02
14	54 675	5.10	0.29	3.95	7.36	25 129	5.22	0.33	3.95	7.36	29 546	5.00	0.21	4.02	6.87
15	45 639	5.13	0.28	3.93	7.56	23 686	5.23	0.31	3.93	7.56	21 953	5.03	0.21	3.95	6.86
16	74 273	5.03	0.21	3.92	7.19	21 436	5.15	0.28	3.97	7.19	52 837	4.97	0.15	3.92	6.50
17	87 196	5.10	0.23	3.95	7.15	32 596	5.19	0.29	4.01	7.15	54 600	5.05	0.16	3.95	6.76
18	29 124	5.37	0.32	3.95	7.08	20 702	5.42	0.32	3.95	7.08	8 422	5.23	0.26	4.10	6.73
19	28 040	5.27	0.27	3.99	7.21	17 173	5.33	0.28	3.99	7.21	10 867	5.19	0.22	4.25	6.69
20	2 026	5.45	0.24	4.63	7.05	1 677	5.47	0.24	4.63	7.05	349	5.36	0.20	4.87	6.09
21	2 397	5.43	0.23	4.57	7.26	1 826	5.45	0.23	4.57	7.26	571	5.35	0.20	4.74	6.04
22	525	5.33	0.21	4.51	6.12	297	5.38	0.23	4.93	6.12	228	5.26	0.17	4.51	5.94
	Public (N = 447 449)					Male, Public (N = 136 500)					Female, Public (N = 310 949)				
8	7 840	4.79	0.14	4.51	6.33	2 001	4.86	0.16	4.51	6.33	5 839	4.76	0.12	4.51	5.58
9	16 594	4.77	0.13	4.48	7.31	3 191	4.85	0.14	4.48	7.31	13 403	4.75	0.12	4.51	6.05
10	32 333	4.76	0.15	4.13	6.75	6 789	4.83	0.18	4.13	6.75	25 544	4.74	0.14	4.45	6.38
11	53 874	4.82	0.14	4.40	6.93	9 087	4.93	0.17	4.40	6.93	44 787	4.80	0.12	4.49	6.21
12	32 373	4.85	0.15	4.45	7.29	6 517	4.92	0.18	4.45	7.29	25 856	4.83	0.13	4.51	6.68
13	80 059	4.84	0.17	4.27	7.29	28 364	4.92	0.19	4.27	7.29	51 695	4.80	0.13	4.45	6.18
14	28 311	4.99	0.17	4.11	6.82	9 608	5.05	0.19	4.11	6.82	18 703	4.96	0.15	4.50	6.14
15	25 668	5.03	0.17	4.48	7.19	10 715	5.09	0.19	4.51	7.19	14 953	4.99	0.14	4.48	6.80
16	56 594	4.97	0.13	4.48	6.20	10 897	5.02	0.15	4.48	6.05	45 697	4.96	0.12	4.51	6.20
17	72 599	5.06	0.13	4.50	6.23	23 723	5.10	0.14	4.50	6.12	48 876	5.03	0.12	4.51	6.23
18	13 817	5.20	0.21	4.37	6.20	8 655	5.24	0.21	4.37	6.20	5 162	5.13	0.18	4.51	6.02
19	23 834	5.25	0.24	4.53	7.21	14 270	5.29	0.25	4.53	7.21	9 564	5.19	0.21	4.53	6.69
20	1 143	5.32	0.13	4.87	6.01	935	5.33	0.13	4.96	6.01	208	5.26	0.11	4.87	5.74
21	1 964	5.41	0.20	4.87	6.77	1 497	5.43	0.20	4.96	6.77	467	5.34	0.18	4.87	6.00
22	446	5.33	0.21	4.93	6.12	251	5.38	0.22	4.93	6.12	195	5.26	0.16	5.03	5.94
	Private (N = 441 599)					Male, Private (N = 276 080)					Female, Private (N = 165 519)				
8	9 815	4.90	0.23	3.92	6.47	6 434	4.95	0.24	3.92	6.47	3 381	4.81	0.18	3.96	5.95
9	20 060	4.88	0.22	3.92	6.69	10 811	4.95	0.22	3.97	6.69	9 249	4.80	0.17	3.92	6.05
10	65 868	4.87	0.24	3.92	6.93	38 146	4.92	0.25	3.92	6.93	27 722	4.79	0.20	3.92	6.67
11	56 255	4.95	0.25	3.93	6.79	25 813	5.05	0.28	3.96	6.79	30 442	4.87	0.20	3.93	6.76
12	34 387	4.98	0.27	3.92	6.80	20 649	5.04	0.29	3.92	6.80	13 738	4.88	0.21	3.93	6.39
13	155 695	5.01	0.29	3.91	7.38	110 256	5.06	0.30	3.91	7.38	45 439	4.89	0.24	3.94	7.02
14	26 364	5.21	0.35	3.95	7.36	15 521	5.32	0.35	3.95	7.36	10 843	5.05	0.27	4.02	6.87
15	19 971	5.26	0.34	3.93	7.56	12 971	5.34	0.34	3.93	7.56	7 000	5.10	0.29	3.95	6.86
16	17 679	5.20	0.32	3.92	7.19	10 539	5.29	0.31	3.97	7.19	7 140	5.06	0.27	3.92	6.50
17	14 597	5.34	0.40	3.95	7.15	8 873	5.45	0.41	4.01	7.15	5 724	5.17	0.33	3.95	6.76
18	15 307	5.52	0.32	3.95	7.08	12 047	5.56	0.32	3.95	7.08	3 260	5.39	0.29	4.10	6.73
19	4 206	5.41	0.37	3.99	7.08	2 903	5.49	0.37	3.99	7.08	1 303	5.24	0.30	4.25	6.21
20	883	5.62	0.24	4.63	7.05	742	5.64	0.24	4.63	7.05	141	5.52	0.21	4.90	6.09
21	433	5.51	0.30	4.57	7.26	329	5.54	0.31	4.57	7.26	104	5.40	0.25	4.74	6.04
22	79	5.34	0.26	4.51	6.09	46	5.41	0.25	4.96	6.09	33	5.25	0.24	4.51	5.82

Table A.10. Log hourly wages across different educational groups and subpopulations

Group	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
	Total (N = 906 930)					Male (N = 420 480)					Female (N = 486 450)				
1	163 269	4.83	0.21	3.92	7.31	72 188	4.91	0.23	3.92	7.31	91 081	4.77	0.17	3.92	6.67
2	46 798	4.89	0.22	3.93	6.76	11 733	5.05	0.27	3.96	6.76	35 065	4.83	0.17	3.93	6.76
3	44 249	4.89	0.22	3.93	7.29	10 582	5.03	0.27	4.06	7.29	33 667	4.84	0.17	3.93	6.68
4	43 169	4.97	0.25	3.92	6.93	35 889	5.00	0.25	3.92	6.93	7 280	4.84	0.18	3.94	5.96
5	38 358	4.83	0.14	3.95	6.73	1 210	4.91	0.20	4.16	6.73	37 148	4.83	0.14	3.95	6.25
6	64 743	4.95	0.30	3.92	7.02	26 360	5.08	0.35	3.92	6.99	38 383	4.86	0.22	3.95	7.02
7	33 903	4.93	0.27	3.93	7.38	12 467	5.05	0.32	3.93	7.38	21 436	4.85	0.20	3.95	6.77
8	97 388	5.00	0.25	3.91	7.29	85 552	5.02	0.25	3.91	7.29	11 836	4.85	0.20	3.95	6.18
9	24 198	4.81	0.14	3.92	6.96	2 168	4.85	0.20	3.92	6.33	22 030	4.80	0.13	3.94	6.96
10	30 896	5.06	0.29	3.95	7.10	19 282	5.14	0.30	3.95	7.10	11 614	4.92	0.22	4.02	6.28
11	100 606	5.04	0.18	3.95	7.11	31 923	5.10	0.20	4.01	7.11	68 683	5.01	0.15	3.95	6.87
12	42 816	5.20	0.35	3.93	7.56	22 789	5.31	0.37	3.93	7.56	20 027	5.08	0.27	3.95	6.76
13	30 297	5.26	0.30	3.92	7.36	25 458	5.29	0.29	3.95	7.36	4 839	5.11	0.25	3.92	6.48
14	58 593	5.00	0.14	4.00	6.43	6 338	5.04	0.19	4.11	6.30	52 255	5.00	0.13	4.00	6.43
15	8 613	5.13	0.28	4.05	6.88	7 311	5.15	0.29	4.05	6.88	1 302	5.02	0.20	4.26	6.06
16	10 292	5.18	0.17	4.10	7.04	5 443	5.21	0.17	4.10	7.04	4 849	5.15	0.16	4.10	6.31
17	13 580	5.26	0.30	4.21	7.08	7 636	5.34	0.32	4.21	7.08	5 944	5.16	0.23	4.25	6.46
18	21 843	5.40	0.30	4.20	7.08	17 214	5.44	0.30	4.20	7.08	4 629	5.26	0.26	4.20	6.73
19	7 911	5.43	0.30	3.99	7.21	4 671	5.48	0.32	3.99	7.21	3 240	5.36	0.26	4.36	6.69
20	3 096	5.37	0.26	4.33	6.85	2 979	5.37	0.26	4.33	6.85	117	5.25	0.21	4.64	5.84
21	4 430	5.44	0.23	4.51	7.26	3 387	5.46	0.24	4.57	7.26	1 043	5.35	0.20	4.51	6.09
						Public (N = 447 449)					Private (N = 441 599)				
1						61 082	4.77	0.14	4.13	7.31	102 187	4.87	0.23	3.92	6.93
2						23 619	4.82	0.14	4.40	6.16	23 179	4.96	0.26	3.93	6.76
3						18 247	4.80	0.13	4.48	7.29	26 002	4.94	0.25	3.93	6.84
4						12 375	4.90	0.18	4.48	6.93	30 794	5.01	0.27	3.92	6.67
5						31 083	4.84	0.13	4.51	6.12	7 275	4.83	0.20	3.95	6.73
6						25 412	4.86	0.19	4.46	6.19	39 331	5.01	0.34	3.92	7.02
7						9 859	4.81	0.15	4.46	6.75	24 044	4.98	0.29	3.93	7.38
8						20 361	4.88	0.17	4.45	7.29	77 027	5.03	0.26	3.91	7.11
9						20 663	4.81	0.12	4.45	6.15	3 535	4.82	0.23	3.92	6.96
10						11 295	4.93	0.18	4.27	6.14	19 601	5.13	0.31	3.95	7.10
11						86 746	5.03	0.14	4.50	6.80	13 860	5.09	0.32	3.95	7.11
12						17 020	5.02	0.18	4.48	6.05	25 796	5.32	0.38	3.93	7.56
13						9 654	5.06	0.16	4.11	7.19	20 643	5.36	0.30	3.92	7.36
14						53 489	5.00	0.12	4.51	6.23	5 104	5.09	0.25	4.00	6.43
15						6 008	5.08	0.22	4.51	6.82	2 605	5.26	0.36	4.05	6.88
16						9 141	5.18	0.14	4.53	6.31	1 151	5.21	0.32	4.10	7.04
17						10 121	5.18	0.22	4.37	6.17	3 459	5.50	0.37	4.21	7.08
18						8 608	5.20	0.16	4.51	6.20	13 235	5.53	0.30	4.20	7.08
19						6 989	5.42	0.30	4.51	7.21	922	5.50	0.33	3.99	6.78
20						2 559	5.31	0.17	4.64	6.21	537	5.65	0.38	4.33	6.85
21						3 118	5.37	0.19	4.87	6.77	1 312	5.59	0.27	4.51	7.26

Table A.1. Cross sectional returns to schooling across gender and sector

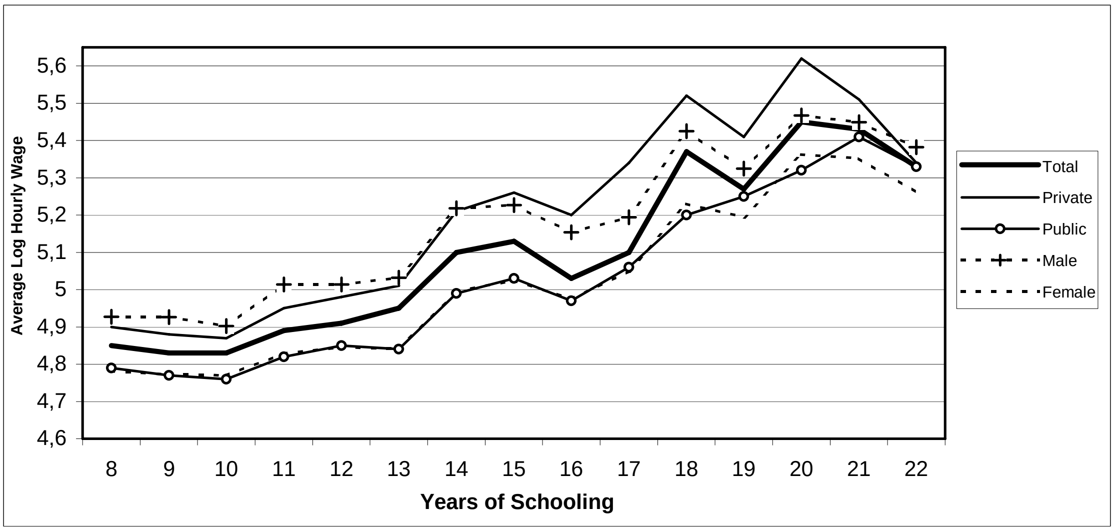


Table A.2. Cross sectional age profiles across gender and sector

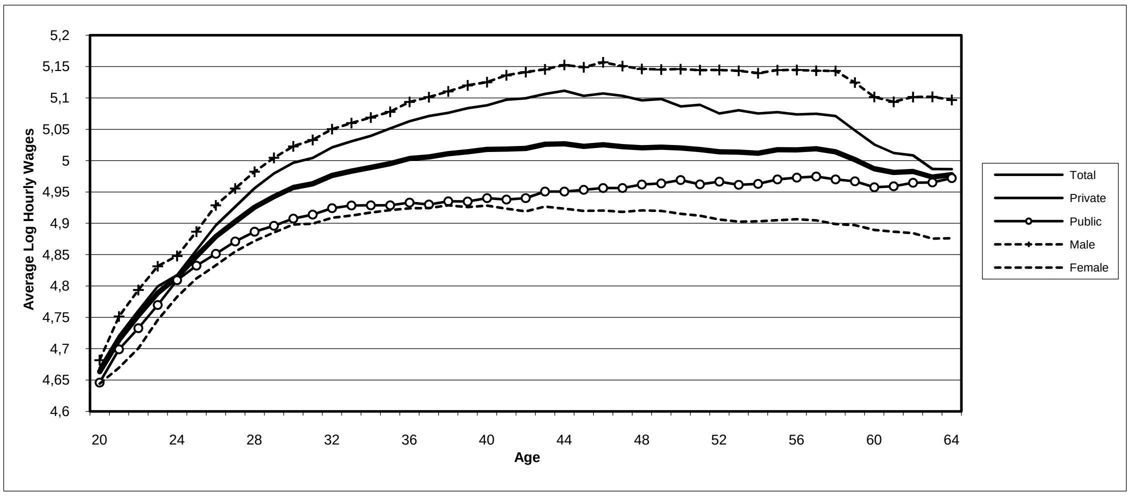
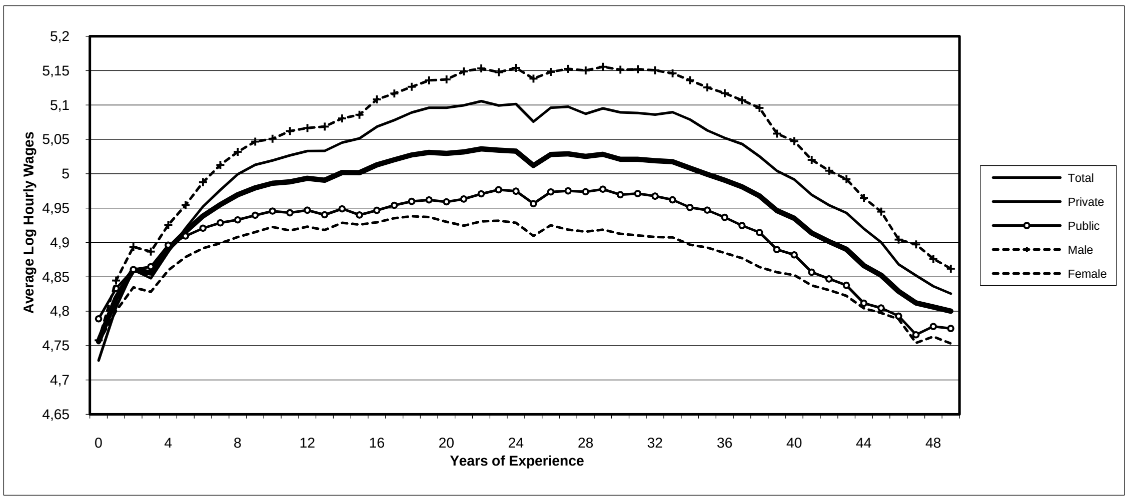


Table A.3. Cross sectional experience profiles across gender and sector



6.3 Estimation results

Table A.11. Wage equations across gender and sectors

	Total		Male, Public		Female, Public		Male, Private		Female, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	3.8502	(0.0017)	4.1106	(0.0031)	4.0936	(0.0017)	3.7818	(0.0037)	3.8932	(0.0043)
Education	0.0583	(0.0001)	0.0493	(0.0002)	0.0483	(0.0001)	0.0781	(0.0002)	0.0636	(0.0003)
Experience	0.0171	(0.0001)	0.0142	(0.0002)	0.0092	(0.0001)	0.0260	(0.0002)	0.0191	(0.0002)
Exp^2 / 100	-0.0241	(0.0002)	-0.0191	(0.0004)	-0.0105	(0.0002)	-0.0381	(0.0004)	-0.0294	(0.0004)
Male	0.1282	(0.0005)	-		-		-		-	
Private	0.1465	(0.0005)	-		-		-		-	
R^2	0.4006		0.4145		0.4748		0.3235		0.2677	
Observations	889 048		136 500		310 949		276 080		165 519	

When gender and sector dummies are removed from the OLS regression on the full sample, the R-squared falls to 0.2467.

Table A.12. Wage equations with interaction effects

	Total		Male, Public		Female, Public		Male, Private		Female, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	3.7314	(0.0046)	4.1188	(0.0107)	4.0886	(0.0054)	3.9862	(0.0105)	3.8879	(0.0106)
Education	0.0651	(0.0003)	0.0478	(0.0007)	0.0477	(0.0004)	0.0611	(0.0008)	0.0615	(0.0008)
Experience	0.0172	(0.0004)	0.0058	(0.0009)	0.0039	(0.0004)	-0.0061	(0.0010)	0.0031	(0.0010)
Exp^2 / 100	-0.0071	(0.0008)	0.0080	(0.0018)	0.0081	(0.0008)	0.0423	(0.0020)	0.0308	(0.0020)
Edu x Exp / 100	0.0229	(0.0029)	0.0708	(0.0060)	0.0507	(0.0030)	0.2706	(0.0075)	0.1644	(0.0077)
Edu x Exp^2 / 100	-0.0019	(0.0001)	-0.0022	(0.0001)	-0.0017	(0.0001)	-0.0069	(0.0002)	-0.0059	(0.0002)
Male	0.1477	(0.0005)	-		-		-		-	
Private	0.1305	(0.0005)	-		-		-		-	
R^2	0.4050		0.4177		0.4778		0.3288		0.2790	
Observations	889 048		136 500		310 949		276 080		165 519	

Table A.13. Wage equations with interaction effects and controls for educational fields

	Total		Male, Public		Female, Public		Male, Private		Female, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	3.6768	(0.0049)	4.1073	(0.0108)	4.0817	(0.0059)	3.8497	(0.0110)	3.8014	(0.0113)
Education	0.0709	(0.0003)	0.0537	(0.0007)	0.0486	(0.0004)	0.0694	(0.0008)	0.0662	(0.0008)
Experience	0.0160	(0.0004)	0.0101	(0.0009)	0.0037	(0.0004)	-0.0086	(0.0010)	0.0017*	(0.0010)
Exp^2 /100	-0.0057	(0.0008)	-0.0047	(0.0017)	0.0080	(0.0008)	0.0449	(0.0020)	0.0317	(0.0020)
Edu x Exp /100	0.0324	(0.0029)	0.0463	(0.0058)	0.0532	(0.0030)	0.2925	(0.0074)	0.1739	(0.0076)
Edu x Exp^2 /100	-0.0020	(0.0001)	-0.0013	(0.0001)	-0.0017	(0.0001)	-0.0071	(0.0002)	-0.0058	(0.0002)
Male	0.1276	(0.0005)	-		-		-		-	
Private	0.1414	(0.0005)	-		-		-		-	
Fields	Yes		Yes		Yes		Yes		Yes	
R^2	0.4136		0.4609		0.4844		0.3448		0.2981	
Observations	889 048		136 500		310 949		276 080		165 519	

* Insignificant at 0.05 significance level.

Table A.14. Wage equations with non-linear returns to education

	Total		Male, Public		Female, Public		Male, Private		Female, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.4692	(0.0010)	4.6117	(0.0022)	4.6160	(0.0010)	4.6101	(0.0021)	4.5591	(0.0022)
EG21	0.6191	(0.0032)	0.5688	(0.0038)	0.5787	(0.0045)	0.7107	(0.0086)	0.7195	(0.0137)
EG20	0.5402	(0.0039)	0.4966	(0.0037)	0.4937	(0.0125)	0.7296	(0.0121)	0.6466	(0.0586)
EG19	0.6710	(0.0025)	0.6565	(0.0030)	0.6316	(0.0025)	0.6351	(0.0136)	0.6539	(0.0098)
EG18	0.5574	(0.0015)	0.3952	(0.0026)	0.4141	(0.0027)	0.6674	(0.0029)	0.6803	(0.0047)
EG17	0.5079	(0.0019)	0.4347	(0.0028)	0.4167	(0.0019)	0.6895	(0.0058)	0.6070	(0.0067)
EG16	0.4047	(0.0022)	0.3592	(0.0029)	0.4141	(0.0020)	0.3225	(0.0109)	0.3994	(0.0099)
EG15	0.3706	(0.0024)	0.3548	(0.0029)	0.3126	(0.0041)	0.4246	(0.0059)	0.3098	(0.0120)
EG14	0.3113	(0.0011)	0.2203	(0.0028)	0.2686	(0.0008)	0.2853	(0.0089)	0.2786	(0.0036)
EG13	0.3856	(0.0013)	0.2520	(0.0025)	0.2654	(0.0027)	0.4514	(0.0024)	0.4684	(0.0045)
EG12	0.3932	(0.0012)	0.2658	(0.0025)	0.2565	(0.0014)	0.5177	(0.0025)	0.4285	(0.0025)
EG11	0.3005	(0.0009)	0.2536	(0.0018)	0.2758	(0.0008)	0.2779	(0.0038)	0.2562	(0.0027)
EG10	0.2072	(0.0013)	0.1735	(0.0027)	0.1085	(0.0018)	0.2683	(0.0026)	0.2030	(0.0030)
EG9	0.1227	(0.0015)	0.0802	(0.0050)	0.0788	(0.0011)	0.0946	(0.0092)	0.0840	(0.0045)
EG8	0.1089	(0.0009)	0.0768	(0.0020)	0.0535	(0.0020)	0.1362	(0.0016)	0.1359	(0.0028)
EG7	0.1241	(0.0013)	0.0821	(0.0039)	0.0613	(0.0016)	0.1827	(0.0029)	0.1299	(0.0022)
EG6	0.1596	(0.0010)	0.1662	(0.0025)	0.1005	(0.0011)	0.2300	(0.0023)	0.1411	(0.0019)
EG5	0.0940	(0.0012)	0.0436	(0.0061)	0.0810	(0.0009)	0.0726	(0.0135)	0.0103	(0.0029)
EG4	0.0639	(0.0012)	0.0730	(0.0023)	0.0307	(0.0024)	0.0653	(0.0020)	0.0680	(0.0035)
EG3	0.0548	(0.0011)	0.0496	(0.0040)	0.0291	(0.0011)	0.0836	(0.0032)	0.0611	(0.0020)
EG2	0.0612	(0.0011)	0.0689	(0.0034)	0.0382	(0.0011)	0.1002	(0.0032)	0.0488	(0.0022)
Experience	0.0176	(0.0001)	0.0159	(0.0002)	0.0094	(0.0001)	0.0265	(0.0002)	0.0199	(0.0002)
$Exp^2 / 100$	-0.0281	(0.0002)	-0.0239	(0.0004)	-0.0142	(0.0002)	-0.0438	(0.0004)	-0.0339	(0.0004)
Male	0.1338	(0.0005)	-		-		-		-	
Private	0.1245	(0.0005)	-		-		-		-	
R^2	0.4265		0.4851		0.5033		0.3529		0.3115	
Observations	889 048		136 500		310 949		276 080		165 519	

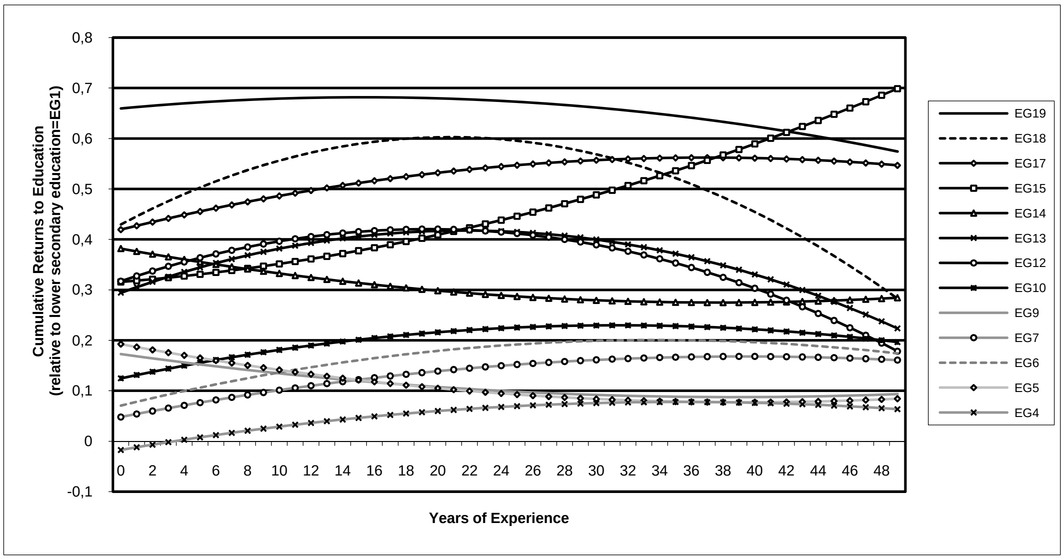
The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-EG21 are the cumulative returns relative to EG1.

Table A.15. Wage equation with non-linearity and multilevel interaction

Variable	Coefficient	SE	Variable	Coefficient	SE	Variable	Coefficient	SE
Intercept	4.4904	(0.0021)						
EG21	0.5897	(0.0150)	EG21 x Exp.	0.0034**	(0.0017)	EG21 x Exp^2 / 100	-0.0079**	(0.0041)
EG20	0.5723	(0.0228)	EG20 x Exp.	-0.0015*	(0.0024)	EG20 x Exp^2 / 100	-0.0007*	(0.0060)
EG19	0.6593	(0.0083)	EG19 x Exp.	0.0029	(0.0010)	EG19 x Exp^2 / 100	-0.0095	(0.0025)
EG18	0.4295	(0.0051)	EG18 x Exp.	0.0167	(0.0006)	EG18 x Exp^2 / 100	-0.0401	(0.0015)
EG17	0.4194	(0.0063)	EG17 x Exp.	0.0077	(0.0008)	EG17 x Exp^2 / 100	-0.0105	(0.0021)
EG16	0.3538	(0.0086)	EG16 x Exp.	0.0025	(0.0009)	EG16 x Exp^2 / 100	-0.0004*	(0.0023)
EG15	0.3156	(0.0067)	EG15 x Exp.	0.0025	(0.0009)	EG15 x Exp^2 / 100	0.0108	(0.0026)
EG14	0.3818	(0.0035)	EG14 x Exp.	-0.0057	(0.0004)	EG14 x Exp^2 / 100	0.0075	(0.0008)
EG13	0.2944	(0.0048)	EG13 x Exp.	0.0113	(0.0005)	EG13 x Exp^2 / 100	-0.0261	(0.0011)
EG12	0.3170	(0.0040)	EG12 x Exp.	0.0107	(0.0004)	EG12 x Exp^2 / 100	-0.0276	(0.0011)
EG11	0.2993	(0.0031)	EG11 x Exp.	-0.0004*	(0.0003)	EG11 x Exp^2 / 100	0.0019	(0.0007)
EG10	0.1247	(0.0055)	EG10 x Exp.	0.0067	(0.0005)	EG10 x Exp^2 / 100	-0.0107	(0.0012)
EG9	0.1726	(0.0039)	EG9 x Exp.	-0.0044	(0.0005)	EG9 x Exp^2 / 100	0.0057	(0.0011)
EG8	0.1056	(0.0031)	EG8 x Exp.	0.0007***	(0.0003)	EG8 x Exp^2 / 100	-0.0022	(0.0007)
EG7	0.0480	(0.0044)	EG7 x Exp.	0.0061	(0.0005)	EG7 x Exp^2 / 100	-0.0077	(0.0011)
EG6	0.0707	(0.0030)	EG6 x Exp.	0.0077	(0.0003)	EG6 x Exp^2 / 100	-0.0114	(0.0007)
EG5	0.1925	(0.0072)	EG5 x Exp.	-0.0059	(0.0006)	EG5 x Exp^2 / 100	0.0075	(0.0011)
EG4	-0.0172	(0.0061)	EG4 x Exp.	0.0054	(0.0005)	EG4 x Exp^2 / 100	-0.0076	(0.0010)
EG3	0.0883	(0.0086)	EG3 x Exp.	-0.0016	(0.0006)	EG3 x Exp^2 / 100	0.0018**	(0.0011)
EG2	0.0518	(0.0145)	EG2 x Exp.	-0.0014*	(0.0009)	EG2 x Exp^2 / 100	0.0048	(0.0014)
Exp.	0.0156	(0.0002)						
Exp^2 / 100	-0.0244	(0.0003)						
Private	0.1341	(0.0005)						
Male	0.1230	(0.0005)						
R-squared	0.4319							
Observations	889 047							

+ The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-21 are the cumulative returns relative to EG1.
* Insignificant for significance level 0.10.
** Insignificant for significance level 0.05.
*** Insignificant for significance level 0.01.

Table A.4. Cumulative returns for different educational groups



Estimation results are given in Table A.15. Returns profiles are drawn only for educational groups having significant estimates.

Table A.16. Wage equations with non-linearity and controls for industry and region

	Total		Male, Public		Female, Public		Male, Private		Female, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.5280	(0.0011)	4.5883	(0.0023)	4.6340	(0.0011)	4.6240	(0.0038)	4.6368	(0.0033)
EG21	0.5709	(0.0031)	0.5159	(0.0037)	0.5341	(0.0044)	0.6354	(0.0081)	0.6292	(0.0130)
EG20	0.5152	(0.0037)	0.4448	(0.0037)	0.4450	(0.0121)	0.6859	(0.0114)	0.5986	(0.0553)
EG19	0.6534	(0.0024)	0.6257	(0.0030)	0.6052	(0.0024)	0.6972	(0.0129)	0.6502	(0.0093)
EG18	0.4879	(0.0015)	0.3607	(0.0026)	0.3620	(0.0027)	0.5766	(0.0028)	0.5687	(0.0045)
EG17	0.4604	(0.0019)	0.3929	(0.0027)	0.3725	(0.0019)	0.6152	(0.0056)	0.5249	(0.0064)
EG16	0.3624	(0.0021)	0.3171	(0.0029)	0.3415	(0.0020)	0.3498	(0.0104)	0.3671	(0.0094)
EG15	0.3415	(0.0023)	0.3071	(0.0029)	0.2658	(0.0040)	0.3853	(0.0056)	0.2581	(0.0114)
EG14	0.2996	(0.0011)	0.2073	(0.0028)	0.2566	(0.0009)	0.3349	(0.0086)	0.3022	(0.0036)
EG13	0.3478	(0.0013)	0.2363	(0.0024)	0.2294	(0.0027)	0.4061	(0.0022)	0.4003	(0.0043)
EG12	0.3562	(0.0011)	0.2438	(0.0025)	0.2273	(0.0014)	0.4702	(0.0024)	0.3670	(0.0024)
EG11	0.2581	(0.0011)	0.2191	(0.0022)	0.2043	(0.0010)	0.2744	(0.0036)	0.2350	(0.0027)
EG10	0.1812	(0.0013)	0.1492	(0.0026)	0.0899	(0.0017)	0.2386	(0.0025)	0.1607	(0.0029)
EG9	0.1260	(0.0014)	0.0739	(0.0049)	0.0805	(0.0011)	0.1173	(0.0086)	0.0934	(0.0043)
EG8	0.1005	(0.0009)	0.0689	(0.0020)	0.0453	(0.0019)	0.1297	(0.0015)	0.1243	(0.0027)
EG7	0.1107	(0.0012)	0.0707	(0.0038)	0.0476	(0.0015)	0.1649	(0.0028)	0.1118	(0.0021)
EG6	0.1432	(0.0010)	0.1407	(0.0025)	0.0848	(0.0011)	0.2077	(0.0022)	0.1164	(0.0018)
EG5	0.1008	(0.0012)	0.0504	(0.0060)	0.0838	(0.0009)	0.0904	(0.0127)	0.0286	(0.0027)
EG4	0.0583	(0.0011)	0.0608	(0.0022)	0.0234	(0.0023)	0.0631	(0.0019)	0.0583	(0.0033)
EG3	0.0472	(0.0011)	0.0435	(0.0039)	0.0226	(0.0011)	0.0824	(0.0030)	0.0485	(0.0019)
EG2	0.0566	(0.0011)	0.0597	(0.0033)	0.0335	(0.0010)	0.0939	(0.0030)	0.0421	(0.0021)
Experience	0.0168	(0.0001)	0.0165	(0.0002)	0.0091	(0.0001)	0.0253	(0.0002)	0.0182	(0.0002)
$Exp^2 / 100$	-0.0268	(0.0002)	-0.0248	(0.0004)	-0.0141	(0.0002)	-0.0415	(0.0004)	-0.0311	(0.0004)
Male	0.1168	(0.0005)	-		-		-		-	
Regions	Yes		Yes		Yes		Yes		Yes	
Industries	Yes		Yes		Yes		Yes		Yes	
R^2	0.4816		0.5179		0.5339		0.4305		0.3871	
Observations	889 048		136 500		310 949		276 080		165 519	

The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-EG21 are the cumulative returns relative to EG1.

Table A.17. Probit estimates for labour force participation and sector choice

	Total		Public, Females		Private, Females		Public, Males		Private, Males	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	-3.0632	(0.0790)	-5.4559	(0.1114)	0.5975	(0.0728)	-5.8070	(0.1306)	0.5702	(0.0793)
Non-lab. inc.	-0.1260	(0.0006)	-0.0708	(0.0007)	-0.0688	(0.0008)	-0.0612	(0.0017)	-0.1240	(0.0013)
Children_0	-0.2989	(0.0086)	-0.0992	(0.0110)	-0.4626	(0.0123)	*0.0115*	(0.0130)	*-0.0193*	(0.0113)
Children_1	-0.3564	(0.0080)	-0.2320	(0.0106)	-0.3044	(0.0114)	*0.0043*	(0.0128)	*-0.0157*	(0.0111)
Children_2	-0.0456	(0.0079)	0.0275	(0.0102)	-0.0223	(0.0108)	*0.0001*	(0.0123)	*-0.0040*	(0.0107)
Children_3	-0.0612	(0.0078)	0.0378	(0.0098)	-0.0626	(0.0103)	*0.0053*	(0.0122)	-0.0298	(0.0106)
Children_4	-0.1125	(0.0076)	*-0.0003*	(0.0094)	-0.1340	(0.0099)	*0.0165*	(0.0119)	-0.0309	(0.0104)
Children_5	-0.1105	(0.0076)	*0.0123*	(0.0093)	-0.1408	(0.0098)	*0.0057*	(0.0119)	*-0.0164*	(0.0103)
Children_6	-0.1049	(0.0076)	*0.0125*	(0.0094)	-0.1188	(0.0098)	*-0.0008*	(0.0120)	-0.0275	(0.0104)
Children_7	-0.0878	(0.0077)	0.0203	(0.0094)	-0.1090	(0.0098)	*0.0176*	(0.0121)	-0.0320	(0.0105)
Children_8	-0.0732	(0.0078)	0.0180	(0.0095)	-0.0788	(0.0099)	*-0.0013*	(0.0124)	-0.0240	(0.0107)
Children_9	-0.0582	(0.0079)	0.0233	(0.0095)	-0.0793	(0.0099)	*-0.0021*	(0.0126)	*-0.0007*	(0.0109)
Children_ 10	-0.0486	(0.0080)	0.0514	(0.0096)	-0.0778	(0.0100)	*-0.0196*	(0.0130)	*-0.0078*	(0.0112)
Child. 11-15	*0.0049*	(0.0046)	0.1161	(0.0055)	-0.0795	(0.0058)	*-0.0148*	(0.0074)	*-0.0066*	(0.0064)
Child. 16-18	0.0269	(0.0075)	0.1405	(0.0085)	-0.1140	(0.0089)	*-0.0016*	(0.0131)	*-0.0039*	(0.0112)
Child. 18 -	-0.0725	(0.0079)	0.0257	(0.0086)	-0.1520	(0.0090)	-0.0489	(0.0164)	*-0.0139*	(0.0137)
Father_indM	-		-0.0886	(0.0290)	*0.0001*	(0.0306)	-0.1018	(0.0371)	-0.0687	(0.0330)
Father_ind1	-		*-0.0159*	(0.0275)	*-0.0032*	(0.0291)	-0.1592	(0.0352)	-0.3024	(0.0313)
Father_ind2	-		*-0.0551*	(0.0386)	0.0883	(0.0405)	-0.1606	(0.0518)	0.0962	(0.0455)
Father_ind3	-		-0.0815	(0.0255)	0.0773	(0.0270)	-0.1431	(0.0319)	0.0870	(0.0289)
Father_ind5	-		*-0.0271*	(0.0262)	*0.0419*	(0.0277)	*-0.0264*	(0.0329)	-0.0559	(0.0298)
Father_ind6	-		-0.1517	(0.0264)	0.1213	(0.0278)	-0.1542	(0.0332)	*0.0488*	(0.0299)
Father_ind7	-		-0.0906	(0.0263)	0.1024	(0.0278)	-0.0873	(0.0331)	*0.0083*	(0.0299)
Father_ind8	-		-0.1960	(0.0298)	0.1346	(0.0312)	-0.1726	(0.0376)	*0.0476*	(0.0339)
Father_ind9	-		-0.0684	(0.0259)	0.0453	(0.0275)	0.0598	(0.0322)	-0.1266	(0.0294)
Mother_self	-0.2379	(0.0071)	-0.0479	(0.0119)	*-0.0117*	(0.0125)	-0.0540	(0.0160)	-0.1127	(0.0133)
Father_self	-0.1930	(0.0091)	-0.0561	(0.0099)	*0.0106*	(0.0102)	-0.1223	(0.0136)	-0.1489	(0.0113)
Pseudo R^2	0.2632		0.2419		0.1585		0.2330		0.1907	
- ln L	162 588		118 219		109 562		63 664		89 932	
Participants	287 449		105 635		59 521		33 935		88 358	
Population	387 017		225 643		225 643		161 374		161 374	

Apart from annual non-labour income and number of children at different age groups, dummies for gender, age, educational group and labour market region were also added to the regressions. Non-labour income includes various transfers and capital income and is scaled by 10000. Combinations of industry and occupational codes for both parental are also used. Father_self = Father self-employed during the individual's childhood, Mother_self = Mother self-employed during the individual's childhood. Cursive coefficient estimates are insignificant at a significance level of 0.1.

Table A.18. Wage equations with selection correction for employment

	OLS, Total		Two-step, Total		MLE, Total		MLE, Males		MLE, Females	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.7548	(0.0041)	4.7910	(0.0043)	4.7738	(0.0042)	4.6835	(0.0076)	4.6841	(0.0044)
EG21	0.5660	(0.0056)	0.5335	(0.0058)	0.5489	(0.0057)	0.5655	(0.0081)	0.5655	(0.0082)
EG20	0.5187	(0.0057)	0.4859	(0.0058)	0.5014	(0.0057)	0.5312	(0.0071)	0.4783	(0.0222)
EG19	0.6941	(0.0046)	0.6866	(0.0047)	0.6901	(0.0046)	0.7537	(0.0084)	0.6590	(0.0051)
EG18	0.4967	(0.0027)	0.4669	(0.0029)	0.4811	(0.0028)	0.5005	(0.0042)	0.4737	(0.0039)
EG17	0.4461	(0.0031)	0.4245	(0.0032)	0.4347	(0.0032)	0.4732	(0.0055)	0.4052	(0.0036)
EG16	0.3444	(0.0049)	0.3213	(0.0051)	0.3323	(0.0050)	0.3321	(0.0090)	0.3338	(0.0054)
EG15	0.3364	(0.0035)	0.3057	(0.0037)	0.3202	(0.0036)	0.3559	(0.0049)	0.3093	(0.0070)
EG14	0.3318	(0.0019)	0.3029	(0.0021)	0.3166	(0.0019)	0.2874	(0.0057)	0.2922	(0.0018)
EG13	0.3652	(0.0024)	0.3389	(0.0025)	0.3515	(0.0024)	0.3714	(0.0035)	0.3271	(0.0037)
EG12	0.3687	(0.0019)	0.3453	(0.0021)	0.3564	(0.0020)	0.4211	(0.0034)	0.2974	(0.0022)
EG11	0.2698	(0.0019)	0.2462	(0.0021)	0.2574	(0.0020)	0.2899	(0.0043)	0.2254	(0.0020)
EG10	0.1711	(0.0022)	0.1548	(0.0022)	0.1625	(0.0022)	0.1993	(0.0035)	0.1204	(0.0027)
EG9	0.1374	(0.0024)	0.1140	(0.0025)	0.1251	(0.0024)	0.1106	(0.0089)	0.0992	(0.0022)
EG8	0.0975	(0.0015)	0.0840	(0.0015)	0.0904	(0.0015)	0.1075	(0.0022)	0.0882	(0.0024)
EG7	0.1041	(0.0019)	0.0894	(0.0020)	0.0964	(0.0019)	0.1464	(0.0042)	0.0711	(0.0019)
EG6	0.1405	(0.0016)	0.1318	(0.0016)	0.1360	(0.0016)	0.2012	(0.0031)	0.0941	(0.0017)
EG5	0.1353	(0.0022)	0.1185	(0.0022)	0.1264	(0.0022)	0.0987	(0.0106)	0.1050	(0.0019)
EG4	0.0436	(0.0025)	0.0368	(0.0025)	0.0401	(0.0025)	0.0541	(0.0035)	0.0339	(0.0039)
EG3	0.0564	(0.0025)	0.0439	(0.0026)	0.0498	(0.0025)	0.0768	(0.0066)	0.0419	(0.0023)
EG2	0.0461	(0.0052)	0.0836	(0.0053)	0.0659	(0.0053)	0.1528	(0.0183)	0.0366	(0.0046)
Experience	0.0186	(0.0003)	0.0189	(0.0003)	0.0188	(0.0003)	0.0266	(0.0007)	0.0157	(0.0004)
Exp^2 / 100	-0.0322	(0.0011)	-0.0345	(0.0011)	-0.0334	(0.0011)	-0.0477	(0.0022)	-0.0313	(0.0012)
Male	0.1359	(0.0009)	0.1350	(0.0009)	0.1355	(0.0009)	-		-	
IMR	-		-0.0644	(0.0020)	-0.0338	(0.0012)	-0.0540	(0.0023)	-0.0237	(0.0012)
Rho	-		-0.3153		-0.1673	(0.0060)	-0.2287	(0.0095)	-0.1431	(0.0074)
Sigma	-		0.2042		0.2018	(0.0003)	0.2361	(0.0005)	0.1655	(0.0003)
R^2	0.4653		-		-		-		-	
- ln L	-		-		108 910		61 032		29 639	
Participants	287 449		287 449		287 449		122 293		165 156	
Population	-		387 017		387 017		161 374		225 643	

Region and industry controls are added to each regression. The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-EG21 are the cumulative returns relative to EG1.

Table A.19. Wage equations with selection correction for female sector choices

	OLS, Public		MLE, Public		OLS, Private		MLE, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.6528	(0.0028)	4.4707	(0.0033)	4.5467	(0.0079)	4.6261	(0.0087)
EG21	0.5281	(0.0078)	0.5174	(0.0078)	0.6409	(0.0172)	0.6420	(0.0175)
EG20	0.4218	(0.0179)	0.4065	(0.0180)	0.6879	(0.0802)	0.7299	(0.0805)
EG19	0.6350	(0.0042)	0.6240	(0.0043)	0.6437	(0.0178)	0.6957	(0.0180)
EG18	0.3561	(0.0043)	0.3515	(0.0043)	0.5985	(0.0066)	0.5788	(0.0068)
EG17	0.3591	(0.0031)	0.3456	(0.0032)	0.5258	(0.0097)	0.5578	(0.0098)
EG16	0.3003	(0.0044)	0.2874	(0.0045)	0.4031	(0.0179)	0.4455	(0.0180)
EG15	0.2885	(0.0060)	0.2759	(0.0061)	0.2888	(0.0188)	0.3127	(0.0190)
EG14	0.2733	(0.0016)	0.2573	(0.0020)	0.3244	(0.0065)	0.3897	(0.0072)
EG13	0.2313	(0.0042)	0.2274	(0.0042)	0.4203	(0.0064)	0.4021	(0.0065)
EG12	0.2250	(0.0024)	0.2204	(0.0024)	0.3755	(0.0039)	0.3620	(0.0040)
EG11	0.2020	(0.0018)	0.1886	(0.0020)	0.2543	(0.0047)	0.2995	(0.0052)
EG10	0.0855	(0.0030)	0.0837	(0.0030)	0.1550	(0.0047)	0.1431	(0.0048)
EG9	0.0883	(0.0018)	0.0751	(0.0021)	0.0968	(0.0076)	0.1473	(0.0079)
EG8	0.0470	(0.0029)	0.0489	(0.0029)	0.1272	(0.0040)	0.1159	(0.0041)
EG7	0.0464	(0.0022)	0.0457	(0.0022)	0.1066	(0.0032)	0.0867	(0.0033)
EG6	0.0781	(0.0018)	0.0762	(0.0018)	0.1137	(0.0030)	0.1083	(0.0030)
EG5	0.1070	(0.0017)	0.0970	(0.0018)	0.0547	(0.0053)	0.0895	(0.0056)
EG4	0.0175	(0.0045)	0.0177	(0.0045)	0.0559	(0.0068)	0.0510	(0.0069)
EG3	0.0252	(0.0027)	0.0251	(0.0027)	0.0647	(0.0040)	0.0492	(0.0041)
EG2	0.0276	(0.0048)	0.0290	(0.0048)	0.0407	(0.0086)	0.0493	(0.0087)
Experience	0.0122	(0.0003)	0.0124	(0.0003)	0.0289	(0.0008)	0.0274	(0.0008)
$Exp^2 / 100$	-0.0257	(0.0011)	-0.0271	(0.0011)	-0.0617	(0.0027)	-0.0573	(0.0027)
IMR	-		-0.0166	(0.0012)	-		-0.0709	(0.0033)
Rho	-		-0.1322	(0.0097)	-		-0.2981	(0.0138)
Sigma	-		0.1251	(0.0003)	-		0.2196	(0.0001)
R^2	0.4835		-		0.3991		-	
- ln L	-		47 859		-		101 363	
Participants	105 635		105 635		59 521		59 521	
Population	-		225 643		-		225 643	

Region and industry controls are added to each regression. The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-EG21 are the cumulative returns relative to EG1.

Table A.20. Wage equations with selection correction for male sector choices

	OLS, Public		MLE, Public		OLS, Private		MLE, Private	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.7000	(0.0069)	4.7680	(0.0088)	4.5955	(0.0090)	4.6230	(0.0092)
EG21	0.4887	(0.0081)	0.4372	(0.0088)	0.6068	(0.0125)	0.6246	(0.0126)
EG20	0.4178	(0.0063)	0.3530	(0.0076)	0.6573	(0.0160)	0.7036	(0.0162)
EG19	0.6498	(0.0067)	0.5922	(0.0077)	0.7246	(0.0296)	0.8039	(0.0298)
EG18	0.3336	(0.0061)	0.3086	(0.0063)	0.5722	(0.0050)	0.5658	(0.0050)
EG17	0.3219	(0.0056)	0.2702	(0.0066)	0.6148	(0.0092)	0.6472	(0.0094)
EG16	0.2530	(0.0076)	0.1902	(0.0087)	0.3337	(0.0207)	0.3811	(0.0208)
EG15	0.2584	(0.0050)	0.1967	(0.0064)	0.3871	(0.0089)	0.4258	(0.0092)
EG14	0.1967	(0.0051)	0.1311	(0.0067)	0.3689	(0.0142)	0.4246	(0.0145)
EG13	0.2399	(0.0054)	0.2183	(0.0055)	0.4232	(0.0041)	0.4146	(0.0042)
EG12	0.2297	(0.0049)	0.2051	(0.0052)	0.5006	(0.0041)	0.4973	(0.0041)
EG11	0.2030	(0.0048)	0.1436	(0.0062)	0.3120	(0.0066)	0.3529	(0.0070)
EG10	0.1062	(0.0052)	0.0861	(0.0054)	0.2334	(0.0041)	0.2280	(0.0041)
EG9	0.0580	(0.0081)	0.0086*	(0.0088)	0.1328	(0.0150)	0.1560	(0.0151)
EG8	0.0649	(0.0038)	0.0557	(0.0039)	0.1264	(0.0025)	0.1179	(0.0026)
EG7	0.0707	(0.0073)	0.0627	(0.0073)	0.1715	(0.0048)	0.1662	(0.0049)
EG6	0.1178	(0.0047)	0.1000	(0.0049)	0.2227	(0.0037)	0.2238	(0.0038)
EG5	0.0500	(0.0099)	0.0069*	(0.0104)	0.1115	(0.0169)	0.1281	(0.0170)
EG4	0.0274	(0.0058)	0.0204	(0.0059)	0.0663	(0.0041)	0.0614	(0.0041)
EG3	0.0567	(0.0124)	0.0541	(0.0124)	0.0966	(0.0076)	0.0864	(0.0076)
EG2	0.0664	(0.0235)	0.0959	(0.0234)	0.0655	(0.0231)	0.1316	(0.0233)
Experience	0.0178	(0.0008)	0.0187	(0.0008)	0.0329	(0.0009)	0.0329	(0.0009)
Exp^2 / 100	-0.0335	(0.0028)	-0.0381	(0.0028)	-0.0652	(0.0028)	-0.0633	(0.0028)
IMR	-		-0.0446	(0.0029)	-		-0.0548	(0.0031)
Rho	-		-0.2704	(0.0168)	-		-0.2148	(0.0119)
Sigma	-		0.1649	(0.0008)	-		0.2552	(0.0007)
R^2	0.4450		-		0.4100		-	
- ln L	-		49 703		-		93 410	
Participants	33 935		33 935		88 358		88 358	
Population	-		161 374		-		161 374	

Region and industry controls are added to each regression. The reference group is lower secondary schooling (EG1). Coefficient estimates for EG2-EG21 are the cumulative returns relative to EG1.
* Insignificant for significance level 0.10.

Table A.21. Wage equations with endogenous schooling and selection correction, whole sample

	OLS		Two-step, Selection		MLE, Selection		2SLS, Education		2SLS w/ IMR	
Variable	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	4.1213	(0.0050)	4.1647	(0.0055)	4.1524	(0.0053)	3.6773	(0.0110)	3.5902	(0.0158)
Education	0.0594	(0.0002)	0.0573	(0.0002)	0.0579	(0.0002)	0.0884	(0.0007)	0.0922	(0.0009)
Experience	0.0181	(0.0003)	0.0180	(0.0003)	0.0181	(0.0003)	0.0166	(0.0004)	0.0169	(0.0003)
Exp^2 / 100	-0.0269	(0.0011)	-0.0274	(0.0011)	-0.0272	(0.0011)	-0.0034	(0.0013)	-0.0036	(0.0013)
Male	0.1336	(0.0009)	0.1332	(0.0009)	0.1334	(0.0009)	0.1218	(0.0010)	0.1235	(0.0010)
IMR	-		-0.0373	(0.0019)	-0.0265	(0.0014)	-		0.0988	(0.0039)
Rho	-		-0.1814		-0.1294	(0.0066)	-		-	
Sigma	-		0.2057		0.2052	(0.0003)	-		-	
Adj. R^2	0.4454						0.4038		0.3992	
- ln L					114 329					
Participants	287 449		287 449		287 449		287 449		287 449	
Population			387 017		387 017				387 017	

Region and industry controls are added to each regression. Standard errors of the two-step estimates are corrected for the first stage uncertainty.

Table A.22. Wage equations with endogenous schooling and selection correction, public sector females

	OLS		Two-step, Employment		Two-step, Sector		2SLS, Education		2SLS w/ IMR	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	4.1624	(0.0042)	4.1690	(0.0052)	4.2204	(0.0055)	3.9239	(0.0110)	3.8617	(0.0227)
Education	0.0462	(0.0002)	0.0452	(0.0003)	0.0427	(0.0003)	0.0605	(0.0007)	0.0635	(0.0012)
Experience	0.0098	(0.0003)	0.0099	(0.0003)	0.0103	(0.0003)	0.0082	(0.0003)	0.0078	(0.0004)
Exp^2 / 100	-0.0142	(0.0011)	-0.0150	(0.0011)	-0.0171	(0.0011)	0.0015	(0.0014)	0.0036*	(0.0016)
IMR	-		-0.0103	(0.0015)	-0.0283	(0.0014)	-		0.0377	(0.0038)
Rho	-		-0.0790		-0.2150		-		-	
Sigma	-		0.1305		0.1317		-		-	
Adj. R^2	0.4346						0.4095		0.4049	
Participants	105 635		105 635		105 635		105 635		105 635	
Population			225 643		225 643				225 643	

Region and industry controls are added to each regression. Standard errors of the two-step estimates are corrected for the first stage uncertainty.
* Insignificant for significance level 0.01.

Table A.23. Wage equations with endogenous schooling and selection correction, private sector females

	OLS		Two-step, Employment		Two-step, Sector		2SLS, Education		2SLS w/ IMR	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	3.8676	(0.0106)	3.9455	(0.0114)	3.9203	(0.0110)	3.4111	(0.0228)	3.4649	(0.0250)
Education	0.0642	(0.0005)	0.0615	(0.0005)	0.0651	(0.0005)	0.0938	(0.0014)	0.0960	(0.0015)
Experience	0.0259	(0.0008)	0.0247	(0.0008)	0.0241	(0.0008)	0.0271	(0.0009)	0.0248	(0.0009)
Exp^2 / 100	-0.0475	(0.0027)	-0.0460	(0.0027)	-0.0416	(0.0027)	-0.0338	(0.0028)	-0.0255	(0.0028)
IMR	-		-0.0497	(0.0028)	-0.0519	(0.0027)	-		-0.0685	(0.0031)
Rho	-		-0.2280		-0.2364		-		-	
Sigma	-		0.2181		0.2194		-		-	
Adj. R^2	0.3772						0.3376		0.3389	
Participants	59 521		59 521		59 521		59 521		59 521	
Population			225 643		225 643				225 643	

Region and industry controls are added to each regression. Standard errors of the two-step estimates are corrected for the first stage uncertainty.

Table A.24. Wage equations with endogenous schooling and selection correction, public sector males

	OLS		Two-step, Employment		Two-step, Sector		2SLS, Education		2SLS w/ IMR	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	4.1964	(0.0097)	4.1798	(0.0128)	4.2265	(0.0127)	3.7315	(0.0251)	3.5603	(0.0519)
Education	0.0489	(0.0004)	0.0497	(0.0006)	0.0474	(0.0006)	0.0717	(0.0016)	0.0835	(0.0028)
Experience	0.0165	(0.0008)	0.0165	(0.0008)	0.0168	(0.0008)	0.0133	(0.0009)	0.0112	(0.0009)
Exp^2 / 100	-0.0261	(0.0028)	-0.0260	(0.0028)	-0.0270	(0.0028)	0.0040**	(0.0036)	0.0097*	(0.0039)
IMR	-		0.0073*	(0.0037)	-0.0098	(0.0026)	-		0.0954	(0.0084)
Rho	-		0.0428		-0.0573		-		-	
Sigma	-		0.1702		0.1703		-		-	
Adj. R^2	0.3799						0.3319		0.3130	
Participants	33 935		33 935		33 935		33 935		33 935	
Population			161 374		161 374				161 374	

Region and industry controls are added to each regression. Standard errors of the two-step estimates are corrected for the first stage uncertainty.
* Insignificant for significance level 0.01.
** Insignificant for significance level 0.1.

Table A.25. Wage equations with endogenous schooling and selection correction, private sector males

	OLS		Two-step, Employment		Two-step, Sector		2SLS, Education		2SLS w/ IMR	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Intercept	3.8025	(0.0112)	3.8676	(0.0122)	3.8200	(0.0114)	3.0473	(0.0258)	3.0532	(0.0284)
Education	0.0749	(0.0005)	0.0718	(0.0005)	0.0753	(0.0005)	0.1244	(0.0016)	0.1276	(0.0018)
Experience	0.0331	(0.0009)	0.0333	(0.0009)	0.0327	(0.0009)	0.0313	(0.0009)	0.0303	(0.0009)
Exp^2 / 100	-0.0608	(0.0028)	-0.0599	(0.0028)	-0.0585	(0.0028)	-0.0296	(0.0031)	-0.0227	(0.0031)
IMR	-		-0.0668	(0.0048)	-0.0301	(0.0035)	-		-0.0691	(0.0043)
Rho	-		-0.2573		-0.1171		-		-	
Sigma	-		0.2596		0.2573		-		-	
Adj. R^2	0.3896						0.3082		0.3086	
Participants	88 358		88 358		88 358		88 358		88 358	
Population			161 374		161 374				161 374	

Region and industry controls are added to each regression. Standard errors of the two-step estimates are corrected for the first stage uncertainty.