Swedish Institute for Social Research (SOFI)

Stockholm University

WORKING PAPER 2/2014

THE DISTRIBUTION OF LIFETIME EARNINGS RETURNS TO COLLEGE
by

Martin Nybom

The Distribution of Lifetime Earnings Returns to College[*]

Martin Nybom[†]

Abstract

I use Swedish registry data to estimate the distribution of lifetime earnings returns to college, and to what extent returns vary with observed and unobserved characteristics. Access to high-quality ability measures also allows me to examine heterogeneity with respect to cognitive and noncognitive ability. A local IV technique is applied to recover marginal and average treatment effects in the presence of selection on gains. The findings support the notion of self-selection, but mainly on observable characteristics. Moreover, returns vary substantially with respect to both cognitive and noncognitive ability, thus supporting the idea of important complementarities between formal schooling and informal skills.

[†]Swedish Institute for Social Research (SOFI), Stockholm University (email: martin.nybom@sofi.su.se)

Introduction

Research on the returns to education is a classical part of economics. With a primary focus on standard omitted-variable problems, a large amount of estimates have been produced using instrumental variable (IV) methods (see Card, 1999). In a recent paper, however, Carneiro et al. (2011) document that returns vary and that individuals select into schooling based on their idiosyncratic gains. With such selection on gains, standard IV identifies local average treatment effects (LATE) of potentially low external relevance (Angrist and Imbens, 1995).[1] But a potential resolution is provided by the marginal treatment effect (MTE) approach, a method that falls somewhere in the middle of the two extremes of reduced form and structural methods (Heckman and Vytlacil, 1999, 2001, 2005). Using the structure of the Roy model, this approach allows for heterogeneity in the response to treatment, and estimation of the distribution of treatment effects using semiparametric methods. Knowledge of the distribution of effects in turn allows for estimation of a range of summary treatment effects of interest.

While there is growing interest in this approach, existing applications are few and have despite heavy data requirements been limited to survey data. This paper contributes to the MTE literature by estimating lifetime returns to college in Sweden using a large registry-based data set. The paper has three objectives: (i) to explore the applicability of the MTE approach and compare with standard methods; (ii) to provide the first estimates of lifetime returns to college in Sweden while taking self selection into account; and (iii) to assess the degree of heterogeneity in returns and the relative importance of heterogeneity that is observable and unobservable to the researcher. In particular, I will make use of high-quality data on cognitive and noncognitive abilities to analyze their influence on returns.[2]

Such evidence may shed new light on several broader questions. To evaluate the effects of educational policy, for instance, it is essential to know both *who* gains from schooling and *how much* they gain. An example is the debate on the

---

[1]The external relevance of the LATE depends on context and instruments: in some cases it can be very hard to interpret, in other cases it is more straightforward (e.g., if a certain reform constitutes the instrument and the effect of the reform is of primary interest).

[2]There is an important conceptual difference between "abilities", "skills" and, e.g., "test scores". While I use the terms abilities and skills interchangeably, I recognize that my measures are only assessments (or proxies) of actual ability (or skills). When I describe the data below I will, however, argue that these proxies are of unusually high reliability.

optimal size of the college sector, for which it is necessary to concentrate on the distribution of returns. Exposing such heterogeneity may also deepen the understanding of earnings inequality in general. A related question is if the rise in the college premium since the 1980s (Autor et al., 2008) primarily reflects a general shift in the demand for college-educated workers, or rather changes in the degree of "school-skill complementarity", i.e., an increased demand for college-educated workers equipped with certain skills that are produced independent of college (Blackburn and Neumark, 1993; Taber, 2001).

Whereas recent work (e.g. Carneiro et al., 2011) has focused heavily on unobservable heterogeneity, I also address the role of heterogeneity with respect to observable characteristics. I devote special attention to two high-quality measures of cognitive and noncognitive ability that are based on high-stake tests from the mandatory enlistment to the Swedish military. I examine if these abilities influence returns and assess their relative importance, thus casting new light on the complementarity of formal education and different skill types.

The study thus relates to the ongoing debate on the role of cognitive and noncognitive ability for educational and labor market outcomes. The notion of ability has recently shifted from a simple concept primarily related to IQ, such as in the single-skill signaling model (Arrow, 1973) and the g factor (Herrnstein and Murray, 1994), to a multidimensional view that especially recognizes the importance of noncognitive or "soft" skills. A large literature concerns the reduced-form earnings returns to various skill measures: some focus on cognitive skills or IQ (e.g., Murnane et al., 1995; Zax and Rees, 2002); some on noncognitive and personality traits (e.g., Nyhus and Pons, 2005; Mueller and Plug, 2006); and others consider both types of skills jointly (e.g., Heineck and Anger, 2010; Lindqvist and Vestman, 2011).

But these studies rarely address how such skills transmit into earnings and wages – in particular via endogenous schooling choices. For example, do those with high IQ earn more because of their IQ as such, or because high IQ makes educational investments less costly or more beneficial, thus indirectly causing increased earnings? Some evidence suggests that returns to education vary with respect to cognitive ability. In two recent Swedish studies, Nordin (2008) and Öckert (2012) estimate heterogeneous returns relying on selection-on-observables assumptions. The former finds that the return to a year of schooling increases at a diminishing rate by level of cognitive ability, the latter that the return to college is increasing with respect to secondary school GPA. Carneiro

and Lee (2009) estimate returns to college within the MTE framework using the National Longitudinal Survey of Youth (NLSY) and find that cognitive skills are positively related to the return, and at an increasing rate. Heckman et al. (2006a) is a rare exception that considers multiple skills jointly. Using a factor structure model, they find that both cognitive and noncognitive skills are important in explaining several economic and non-economic outcomes within different schooling groups.

A key difference in this paper is that I apply a semiparametric approach, with less restrictive parametric assumptions. The main merit, however, is the data. While most related studies rely on NLSY data, this paper is based on rich administrative data that cover a large and representative sample of the Swedish male population. The size of the data set provides more flexibility compared to previous studies. I use measures of close to full lifetime earnings, minimizing the influence of life-cycle effects in my estimates (Bhuller et al., 2011). Apart from enabling me to estimate actual lifetime returns, this aspect of the data also allows me to explore how estimates vary with the age at which earnings are observed by the researcher. My ability measures are collected at a uniform age prior to college, thus offsetting concerns about endogeneity in malleable ability measures (Hansen et al., 2004). Moreover, my measure of noncognitive ability is unusual in that it is an overall judgement of psychological capability that stems from interviews with certified psychologists. This is in contrast with most related work, which typically relies on combinations of different self-reported answers about one's personality.

As instruments for college attendance, I use distance to the closest university and short-run fluctuations in unemployment and average earnings in the municipality of residence at the end of high school. I estimate the average treatment effect of a year of college to be a 4.5 percent increase in lifetime earnings, which is not very different from what is provided by OLS and standard IV. My findings are consistent with Carneiro et al. (2011) in that agents select into college based on individual gains. The positive selection is manifested in that the difference between the average treatment effect on the treated and the untreated is consistently positive and statistically significant across specifications. While evidence on unobserved heterogeneity is weak, observed characteristics are shown to capture a large part of the total heterogeneity that drives self-selection. Finally, the heterogeneity with respect to both cognitive and noncognitive ability is substantial, and of comparable magnitude. Those in the bottom of the respective

ability distributions have strongly negative returns, while those in the top earn returns that are about twice as high as the average return. My findings thus corroborate the idea of school-skill complementarities on the labor market.

The rest of this paper is structured as follows. Section 1 provides a brief theoretical illustration of the college decision. In Section 2, I show how the generalized Roy model is used to define the MTE, and briefly discuss identification and estimation. I describe the data in Section 3 and present the results in Section 4. I conclude by discussing some implications of my findings.

1   Theoretical Illustration

The decision rule in the Roy model can be seen as a reduced form of a more elaborate theoretical model. To fix ideas, I will illustrate the college decision with a discrete-choice model that builds on those in Keane and Wolpin (2001) and Keane (2002). I extend the model by introducing heterogeneity so that distinct ability or skill types are allowed to affect the costs and benefits of acquiring college education. In addition to the assumptions in Keane (2002), I therefore assume that agents are endowed with a set of abilities $\mathbf{A}$ that can impact on both the indirect time costs and the direct utility (or consumption value) of going to college, as well as college and non-college earnings.[3] The wage rate in period 1 is $w_1(\mathbf{A})$. In the following periods, the wage rate is $w_2(\mathbf{A}) + \beta(\mathbf{A})$ if the agent attended school, and $w_2(\mathbf{A})$ otherwise.

For a given ability realization $\mathbf{A} = \mathbf{a}$, the value function conditional on college attendance is

$$V_S \mid \mathbf{a} = \max_{\{h,b\}} u\left[y_1 + b + hw_1(\mathbf{a}) - t, L - h - s(\mathbf{a})\right] + \varphi(\mathbf{a}) + \rho^{-1} u\left[w_2(\mathbf{a}) + \beta(\mathbf{a}) - rb, 1\right], \tag{1}$$

and the value function for not attending is

---

[3]The assumptions imply that agents: are infinitely lived in discrete time; decide whether to attend college in period 1 to a direct cost (e.g., tuition, transaction or moving costs) denoted t; face a discount factor ρ and interest rate r; can borrow or save b in period 1 with annuity payments rb from period 2 and onwards; devote time to work h and can work while in college; receive an exogenous parental transfer y_1 in period 1; receive non-monetary utility from college denoted $\varphi(\textbf{.})$; receive utility from consumption c and leisure l through a concave utility function denoted $u(c,l)$ with $L \geq l \geq 0$; and inelastically supply one unit of labor after period 1.

$$V_0 \mid \mathbf{a} = \max_{\{h,b\}} u\left[y_1 + b + w_1(\mathbf{a}), L - h\right] + \rho^{-1} u\left[w_2(\mathbf{a}) - rb, 1\right]. \qquad (2)$$

Utility maximization gives

$$u_1(c_1, l_1) = r\rho^{-1} u_1(c_2, l_2) \qquad (3)$$

$$w_1(\mathbf{a}) u_1(c_1, l_1) = u_2(c_1, l_1), \qquad (4)$$

as first-order conditions for *inter-* and *intra*temporal optimality, respectively, given interior solutions. A first-order Taylor expansion of V_S around V_0 at the point of indifference, combined with the two first-order conditions, gives the (approximate) decision rule to attend college if and only if

$$\lambda_1^{-1} \varphi(\mathbf{a}) + r^{-1}\beta(\mathbf{a}) \geq w_1(\mathbf{a}) s(\mathbf{a}) + t, \qquad (5)$$

where $\lambda_1 = u_1(c_1, l_1)$. This has several implications. First, parental transfers y_1 affect the decision only through the marginal utility of consumption. If there is no non-monetary utility from schooling, parental transfers do not affect the schooling decision.[4] If $\varphi(\mathbf{a}) > 0$, then larger parental transfers increase attendance rates by decreasing the marginal utility of consumption and thereby increasing $\varphi(\mathbf{a})/\lambda_1$.[5] Second, the higher the interest rate as well as the direct and time costs of getting a degree, the lower is attendance.

The focus of this paper is on the role of individual abilities. For simplicity, consider $\mathbf{a} = a$ as a unidimensional ability realization such as a standard notion of cognitive ability. Differentiating the decision rule with respect to a gives:

$$\varphi'(a)/\lambda_1 + \varphi(a)\lambda_1'\lambda_1^{-2} + r^{-1}\beta'(a) \geq w_1'(a)s(a) + w_1(a)s'(a). \qquad (6)$$

Ability thus affects the decision rule through several mechanisms: (i) through the non-monetary utility (or consumption value) of college $\varphi'(a)$; (ii) indirectly through the effect of $w_1(a)$ on the marginal utility of consumption; (iii) through the monetary return $\beta(a)$; (iv) through the time cost of acquiring schooling $s(a)$; and (v), since $s(a)$ implies foregone earnings, through the opportunity cost $w_1(a)$. For example, assume that ability increases the non-monetary utility

---

[4]This changes if agents are credit constrained (Keane, 2002). This dimension has been frequently studied in the US perspective, but is not explicitly addressed here as absence of tuition fees suggests it is less important in Sweden.

[5]This is a well-know result and also pointed out in Keane (2002).

of college, as well as both first- and second-period earnings (i.e., absolute advantage), and that it lowers the time cost. In this case the only mechanism that works against attending college is $w_1'(a)s(a)$, i.e., first-period earnings at given time cost. If non-college earnings differ little by ability so that $w_1'(a)$ is small or even negative (i.e., comparative advantage), then the effect of ability on attendance is unambiguously positive. Given positive partial ability effects, the first two terms in (6) imply that increased ability both increases $\varphi(a)$, thus inducing more consumption of schooling, and lowers λ_1, which indirectly encourages the individual to consume even more schooling through $\varphi(a)$. The term $r^{-1}\beta'(a)$ is the effect of ability on the long-run earnings return to college, which I will examine in detail in the empirical part below.

2   Econometric Model

Empirical work on the returns to schooling traditionally seeks to estimate variations of the equation

$$Y = \alpha + \beta S + X'\delta + \varepsilon, \tag{7}$$

where Y denotes earnings, X a vector of covariates, ε the error term, and S can be years of schooling, a vector of schooling levels, or an indicator variable of, e.g., college education. While standard omitted-variable bias (S correlated with ε) has typically received most attention, the issue of heterogeneous "sorting on gains" (S correlated with β) has recenly been gaining attention. Heckman et al. (2006b) distinguish between non-essential and essential sources of heterogeneity, where the former implies sorting by observable and the latter by unobservable characteristics. Although standard IV may in principle take the former into account, the latter will typically cause the LATE to diverge from average and marginal treatment effects.[6]

To move beyond the LATE, Heckman and Vytlacil (1999, 2001, 2005) general-ize the MTE first introduced by Björklund and Moffitt (1987). Their approach has two cornerstones: (i) a choice-theoretic structure, based on the Roy model, that defines each individual's margin of indifference towards selecting into

---

[6]Under observable heterogeneity, IV may recover average and marginal treatment effects provided that the functional forms of the regression equations are sufficiently flexible so that the LATE coincides with these parameters. This would, for example, include interactions between "sorting variables" and the endogenous treatment variable.

treatment, and (ii) local IV (LIV) estimation of marginal treatment effects.

The Generalized Roy Model

The generalized Roy model offers a choice structure for policy analysis in which agents self select into treatment based on their expected gains.[7] The decision rule in the binary version of the model can be seen as the reduced form of a model of college choice such as the one in Section 1.

Let S be a binary choice indicator with $S = 1$ if the agent selects into treatment and $S = 0$ if not. Moreover, let the potential outcomes in the two states be

$$Y_j = \mu_j(\mathbf{X}, \mathbf{A}) + U_j, \quad \text{for } j = 0, 1 \tag{8}$$

where $\mathbf{X}$ is a set of observed regressors, $\mathbf{A}$ is a set of observed ability measures, μ_j are unknown functions, and U_j are unobserved random variables that need not be orthogonal to $\mathbf{X}$ and $\mathbf{A}$. By inserting (8) for both states into a standard switching regression, the observed outcome can be written

$$Y = \mu_0(\mathbf{X}, \mathbf{A}) + S\left[\mu_1(\mathbf{X}, \mathbf{A}) - \mu_0(\mathbf{X}, \mathbf{A}) + U_1 - U_0\right] + U_0. \tag{9}$$

The individual benefit of treatment is the difference between potential outcomes $Y_1 - Y_0 = \mu_1(\mathbf{X}, \mathbf{A}) - \mu_0(\mathbf{X}, \mathbf{A}) + U_1 - U_0$. Thus, the average treatment effect conditional on $\mathbf{X} = \mathbf{x}$ is given by $\text{ATE}(\mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$, and the average ability-specific treatment effect conditional on $\mathbf{X} = \mathbf{x}$ and $\mathbf{A} = \mathbf{a}$ is $\text{ATE}(\mathbf{x}, \mathbf{a}) = \mu_1(\mathbf{x}, \mathbf{a}) - \mu_0(\mathbf{x}, \mathbf{a})$. Moreover, conditioning on $S = 1$ or $S = 0$ defines the average treatment effect of the treated (ATT) and of the untreated (ATU), respectively.[8]

Let I_S denote the net benefit of selecting into college. An individual's decision rule can then be written as a standard latent-variable model of observed and unobserved components:

$$I_S = \mu_S(\mathbf{Z}) - V, \tag{10}$$

where the individual selects into college ($S = 1$) if $I_S \geq 0$, and otherwise not ($S = 0$). $\mathbf{Z}$ is observed and may include some or all of $(\mathbf{X}, \mathbf{A})$, but also components

---

[7]The original version of the model is due to Roy (1951). Although different in style and notation, the essence of the model is similar to the one in Willis and Rosen (1979).

[8]We have $\text{ATT}(\mathbf{x}) = \text{ATE}(\mathbf{x}) + E(U_1 - U_0 \mid S = 1, \mathbf{X} = \mathbf{x})$ and $\text{ATU}(\mathbf{x}) = \text{ATE}(\mathbf{x}) + E(U_1 - U_0 \mid S = 0, \mathbf{X} = \mathbf{x})$, and, for the ability-specific effects, $\text{ATT}(\mathbf{x}, \mathbf{a}) = \text{ATE}(\mathbf{x}, \mathbf{a}) + E(U_1 - U_0 \mid S = 1, \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a})$ and $\text{ATU}(\mathbf{x}, \mathbf{a}) = \text{ATE}(\mathbf{x}, \mathbf{a}) + E(U_1 - U_0 \mid S = 0, \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a})$.

$\mathbf{Z} \setminus (\mathbf{X}, \mathbf{A})$ that are excluded from $(\mathbf{X}, \mathbf{A})$. V is unobserved and represents a latent resistance to select into college. Moreover, assume that V is continuous with a strictly increasing cumulative distribution F_V, and that (U_0, U_1, V) are conditionally independent of $\mathbf{Z}$. $\mathbf{Z} \setminus (\mathbf{X}, \mathbf{A})$ thus work as exogenous cost-shifters that affect the outcome only through the college decision.

Finally, let the propensity score $P(\mathbf{z}) \equiv \Pr(S = 1 \mid \mathbf{Z} = \mathbf{z}) = F_V [\mu_S(\mathbf{z})]$ denote the conditional probability of attending college. Define $U_S = F_V(V)$ such that U_S represents the quantiles of V and is uniformly distributed. The latent index can be rewritten using $F_V(\mu_S(\mathbf{Z})) = P(\mathbf{Z})$ so that $S = 1$ if $P(\mathbf{Z}) > U_S$. Thus, $P(\mathbf{Z})$ and U_S represent the observed and unobserved inducement to college: the higher is $P(\mathbf{Z})$, the more inducement from the observables $\mathbf{Z}$; the lower is U_S, the more unobserved inducement. For a high U_S it thus takes a high inducement from $\mathbf{Z}$ to attend college. If $P(\mathbf{Z}) = U_S$, the individual is indifferent to attending.

Estimating the Marginal Treatment Effect

The marginal treatment effect (MTE) is defined by

$$
\begin{aligned}
\text{MTE}(\mathbf{x}, \mathbf{a}, u_S) &\equiv E(Y_1 - Y_0 \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}, U_S = u_S) \\
&= \mu_1(\mathbf{x}, \mathbf{a}) - \mu_0(\mathbf{x}, \mathbf{a}) + E[U_1 - U_0 \mid U_S = u_S],
\end{aligned} \quad (11)
$$

and can be identified across the support of U_S. Identification is provided by local perturbations of the propensity score induced by $\mathbf{Z} \setminus (\mathbf{X}, \mathbf{A})$ at quantile u_S. Persons with high $P(\mathbf{Z})$ identify the return for those with high U_S, and vice versa. Local perturbations at high (low) $P(\mathbf{Z})$ induce persons with high (low) U_S to change treatment status. If the treatment effect is homogeneous with respect to U_S, then the MTE as a function of U_S is flat. If the MTE correlates with U_S conditional on $(\mathbf{X}, \mathbf{A})$, then there is unobserved heterogeneity.

A key virtue of the MTE approach is that summary parameters – e.g., the ATE, ATT, ATU and LATE – can be recovered using estimates of the MTE and appropriate weights (Heckman et al., 2006b). An average MTE at each level of U_S can be obtained by integrating over $(\mathbf{X}, \mathbf{A})$ conditional on $U_S = u_S$. Integrating over U_S yields the unconditional ATE. The same procedure, conditioning on $S = 1$ or $S = 0$, gives the unconditional ATT and ATU, respectively.

The ability-specific ATE can be obtained by integrating over $\mathbf{X}$ and U_S while conditioning on $\mathbf{A} = \mathbf{a}$, such that $\text{ATE}(\mathbf{a}) = E_{\mathbf{X}, U_S \mid \mathbf{A} = \mathbf{a}} [\text{MTE}(\mathbf{x}, \mathbf{a}, u_S)]$ traces out

the ATE at a given value of $\mathbf{A}$. However, the interpretation of ATE($\mathbf{a}$) as the contribution of ability to the treatment effect will be confounded if there are heterogeneous treatment effects with respect to variables in $\mathbf{X}$ that correlate with $\mathbf{A}$. An alternative is instead to impose a separable version of equation (9) with $\mu_0(\mathbf{X}, \mathbf{A}) = \mathbf{X}\delta_0 + \mathbf{A}\gamma_0$ and $\mu_1(\mathbf{X}, \mathbf{A}) = \mathbf{X}\delta_1 + \mathbf{A}\gamma_1$. The ability-specific treatment effect is then given by $\gamma_1 - \gamma_0$.

Heckman and Vytlacil (1999, 2001, 2005) propose a semiparametric estimator for the MTE. Their approach relies on that the expected value of Y depends on $P(\mathbf{Z})$, so that $P(\mathbf{Z})$ serves as a local IV (LIV):

$$\Delta_{LIV}(\mathbf{x}, \mathbf{a}, u_S) = \frac{\partial E(Y \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}, P(\mathbf{Z}) = p)}{\partial p} \Big|_{p = u_S} = MTE(\mathbf{x}, \mathbf{a}, u_S). \quad (12)$$

The LIV approach thus involves the estimation of the partial derivative of the conditional expectation of Y with respect to p. In the empirical analysis I will impose separability, so that the expected value in (12) can be written

$$E(Y \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}, P(\mathbf{Z}) = p) = \mathbf{x}\delta_0 + \mathbf{a}\gamma_0 + p\left[\mathbf{x}(\delta_1 - \delta_0) + \mathbf{a}(\gamma_1 - \gamma_0)\right] + K(p),$$
$$(13)$$

where $K(p) = E(U_1 - U_0 \mid S = 1, P(\mathbf{Z}) = p)$. Thus, the expected outcome depends on three components: non-college earnings, a treatment effect attributed to observed characteristics, and $K(p)$, representing an effect attributed to unobserved characteristics. Using equations (12) and (13), the estimator becomes

$$\text{MTE}(\mathbf{x}, \mathbf{a}, u_S) = \mathbf{x}'(\delta_1 - \delta_0) + \mathbf{a}'(\gamma_1 - \gamma_0) + \frac{\partial K(p)}{\partial p} \Big|_{p = u_S}. \quad (14)$$

In order to compute the MTE, I thus need to estimate $(\delta_1 - \delta_0)$, $(\gamma_1 - \gamma_0)$ and $\partial K(p)/\partial p$.[9] I first estimate the college choice using a probit to obtain estimates of $P(\mathbf{Z})$. I then estimate the coefficients in equation (13) using a semiparametric version of double residual regression. Specifically, I estimate separate local linear regressions of each of the regressors and the outcome on the predicted propensity score, and retrieve their respective residuals.[10]

---

[9]The implementation follows the guidelines for LIV estimation ("Semiparametric Method 1") presented in Heckman et al. (2006b), and in more detail, at: http://jenni.uchicago.edu/underiv.

[10]Notice that, if $n_X + n_A$ denotes the total number of variables in $(\mathbf{X}, \mathbf{A})$, this step involves the estimation of in total $2 \times (n_X + n_A) + 1$ regressions. This is since equation (13) also contains interaction terms between each of the variables in $(\mathbf{X}, \mathbf{A})$ and the propensity score. The local

Estimates of $\delta_0, \gamma_0, (\delta_1 - \delta_0)$, and $(\gamma_1 - \gamma_0)$ are then obtained by regressing the residual associated with the outcome on the residuals associated with the variables in $(\mathbf{X}, \mathbf{A})$. Given these estimates, $\partial K(p)/\partial p$ can be estimated using standard nonparametric techniques.[11] Finally, the LIV estimate of $\mathrm{MTE}(\mathbf{x}, \mathbf{a}, u_S)$ is computed by plugging in these estimates into equation (14). I obtain estimates of the summary treatment effects by applying weights computed from the data (see Appendix A.1).

An alternative to the semiparametric approach is to impose parametric assumptions on the unobservables and derive the expression for the MTE. This approach, relying on joint estimation of the choice and outcome equations as an endogenous switching regression, is more in line with the work of Willis and Rosen (1979) and Björklund and Moffitt (1987). Similar to above, the parametric MTE estimates can be combined with weights to compute summary treatment effects. As a comparison to the semiparametric approach, I will also report estimates from a parametric version of the model, assuming joint normality of (U_0, U_1, V). In this case,

$$\mathrm{MTE}(\mathbf{x}, \mathbf{a}, u_S) = \mathbf{x}'(\delta_1 - \delta_0) + \mathbf{a}'(\gamma_1 - \gamma_0) - (\sigma_{1V} - \sigma_{2V})\Phi^{-1}(u_S), \qquad (15)$$

where $E(U_1 - U_0 \mid U_S = u_S) = -(\sigma_{1V} - \sigma_{2V})\Phi^{-1}(u_S)$ and has a variance that is normalized to one.[12] I estimate the parameters $\delta_1, \delta_0, \gamma_1, \gamma_0, \sigma_{1V}, \sigma_{2V}$ and their standard errors by maximum likelihood and plug them into equation (15) to obtain estimates of $\mathrm{MTE}(\mathbf{x}, \mathbf{a}, u_S)$.

3 Data and Sample Restrictions

My data are based on a random draw of 35 percent of all Swedish men born 1951-57. The estimation sample is obtained by merging several registers, including data on earnings, ability assessments, educational attainment, personal and

---

linear regressions are estimated for the set of values of p that is contained in the support of $P(\mathbf{Z})$. I use gaussian kernels and select the bandwidth that minimizes the mean leave-out square prediction error when excluding one percent of the sample at a time. This provides an estimated optimal bandwidth of 0.16.

[11]This term is also estimated using local linear regression, gaussian kernels, and an estimated optimal bandwidth of 0.15.

[12]Moreover, $\sigma_{1V} = Cov(U_1, V)$, $\sigma_{0V} = Cov(U_0, V)$, and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.

local labor market characteristics, and data on family members. The analysis is restricted to men since the ability data come from the military draft.

Measure of Lifetime Earnings. I use individual data on annual labor earnings from tax-declaration files for the years 1968-2007. These data come with a number of advantages: they are almost entirely free from attrition and reporting errors; pertain to all jobs; and are not censored. I approximate log lifetime earnings by the log of mean (non-missing) earnings over ages 25-50.[13]

In estimation of the returns to education in general, and when based on an explicit decision model in particular, the relevant outcome (or maximand in the model) is the stream of earnings across the lifetime. In contrast, it has been common in the literature to use single-year or short-run measures, often from around age 30. Yet, if the age-earnings relationship is systematically related to components in the earnings equation, the interpretation of such estimates will be restricted to the age-specific earnings distribution.[14] Since I use earnings data that span over 26 years, such issues will not be of concern here.

Measures of Cognitive and Noncognitive Ability. An attractive feature of the data set is that it includes information from the mandatory military enlistment's tests of cognitive and noncognitive ability. The enlistment typically takes place at age 18 and includes two days of physical, intellectual, and psychological tests and evaluations.[15]

The measure of cognitive ability is based on scores on a test of general intelligence that has been conducted since the 1940s. The test consists of four

---

[13]The actual measure is "Inkomst av tjänst" in Swedish, which includes labor earnings and labor-related benefits but not income from self employment. Using the slightly different measure "Arbetsinkomst", which includes income from self employment, yields very similar results. I trim annual earnings below 10 000 SEK (just over $1000) and discount to present values at age 20 using an annual rate of 0.02.

[14]Empirically, heterogeneous earnings profiles will cause non-classical measurement error when short-run earnings measures are used as proxies for lifetime earnings, giving rise to a so called "life-cycle bias" (Haider and Solon, 2006). That such life-cycle effects can be quantitatively important has been shown in recent studies (e.g., Nybom and Stuhler, 2011; Bhuller et al., 2011). Using measures from early age appears especially precarious when estimating the returns to college. College graduates may then only recently have entered the labor market, reflected in relatively low earnings as of higher rates of on-the-job investments and job switching.

[15]For the cohorts born 1951-57, only a tiny fraction were exempted from the enlistment, mainly due to disability. Although the test scores are drawn from a similar age for all, concerns about joint causality of schooling, latent skills, and test performance (see Hansen et al., 2004) leads me to restrict the sample to individuals with similar educational attainment at the age of enlistment by excluding high school dropouts. I examine the sensitivity of the results by dropping this restriction in Section 4.

subtests of logical, verbal, and spatial ability, as well as technical comprehension, each graded on a discrete scale from 1 to 9. These scores are then transformed to a discrete general variable between 1 and 9 that follows a Stanine scale.[16]

The measure of noncognitive ability stems from interview-based evaluations made by certified psychologists. In the interview, the enlistee's psychological profile and capacity to fulfill the requirements of military duty are evaluated. Central to this is the ability to cope with stress and contribute to group cohesion. Other valued traits include willingness to assume responsibility, independence, emotional stability, outgoing character, persistence, and ability to take initiatives. Motivation for joining the military is not considered. The interview is semi-structured in the sense that the psychologist follows a manual that states topics to discuss and how to grade answers. Scores are given on four subscales and as an overall assessment that follows a Stanine scale between 1 and 9.

This variable is valuable in two ways: it provides a general "omnibus" measure of noncognitive ability and it is based on a psychologist's experience from a personal encounter with the individual, which is likely to capture more aspects of a personality than what can be deduced from survey questionnaires.

Data on Education and Background Characteristics. The education register provides data on highest level of education, in what year it was achieved, and from what type of study program.[17] I observe educational attainment at ages 33-39 and let the college dummy equal one if one has a minimum of three years of college studies.[18] I also obtain various data on personal and family characteristics. The censuses provide data on birth date, country of birth, and geographical residence. I use a multigenerational register to identify (biological) family members and obtain data on number of siblings, father's and mother's years of schooling, and father's earnings.

---

[16]Carlstedt (2000) discusses this test as well as the Swedish military's history of psychometric testing in detail. Carlstedt shows that the test is a good measure of general intelligence, in contrast with tests measuring the more malleable concept of crystallized intelligence. For a good summary of the military enlistment data, see also Lindqvist and Vestman (2011).

[17]In the 1970s, when most sample subjects made their college choices, admission to higher education was largely unrestricted. Most programs were open to anyone with a high school degree and many faculties had no formal application procedure. There were also no tuition fees and the system of student aid was generous (see Erikson and Jonsson, 1993).

[18]In Section 4, I analyze the sensitivity of the results by using a less strict definition of the college variable (at least one semester). Throughout I also use a measure of years of schooling to rescale estimates to annualized values.

Instrumental Variables.   As instruments for college attendance, I use distance to the closest university and short-run fluctuations in unemployment and average earnings in the municipality of residence at the end of high school. The distance instrument was first used by Card (1995) and has later been frequently applied, for example in LIV applications such as Carneiro et al. (2011).[19] I construct a continuous measure of typical travel distance by car between the local municipality in which the individual resided in 1965 and the closest university. Earlier work has instead typically used either indicators for whether a college is located in one's county or, in the continuous case, more crude measures such as "as the crow flies" generated from geographical coordinates.[20]

The exogeneity of distance instruments has been questioned in studies based on both US (Cameron and Taber, 2004) and Swedish data (Kjellstrom and Regner, 1999). As recommended in these studies, it is thus crucial that I can condition on measures of ability and family background.

I also include as instruments short-run fluctuations in the local labor market at the end of high school, conditioned on permanent local labor market conditions.[21] While both current and permanent labor market conditions are in the individual's information set at the time of the college decision, current conditions do not contain any additional information about the future *conditioned on* the permanent component. If this argument holds, then such innovations in the local labor markets can be excluded from the outcome equation. As measures of current conditions, I use the unemployment rate and average earnings in the municipality of residence at age 20. As permanent measures, I use municipal averages of unemployment and earnings over the years 1968-1988.[22]

Model Specification and Sample Statistics.   The linear-in-parameter representations of $(\mathbf{X}, \mathbf{A})$ include linear and quadratic terms of cognitive and

---

[19]Other papers that have used variations of this instrument include Kling (2001), Currie and Moretti (2003), Cameron and Taber (2004), and Carneiro and Lee (2009).

[20]In 1965 there were 998 municipalities in my sample. I consider six university cities: Stockholm, Uppsala, Gothenburg, Lund, Umeå, and Linköping. Up until the late 1970s, nearly all Swedish college students studied in one of these cities. Distances are calculated using the website eniro.se, which is similar to Google Maps.

[21]Previous papers that have used such instruments include Cameron and Heckman (1998), Cameron and Taber (2004), Carneiro and Lee (2009), and Carneiro et al. (2011). As Cameron and Taber (2004) argue, the impact on schooling is theoretically ambiguous: better labor market conditions increase the opportunity cost of schooling, but also increase the resources of credit constrained households which may promote educational investment.

[22]In effect, I use "non-employment", i.e., one minus the employment rate in the local working-age population, as my measure of unemployment.

noncognitive ability measures (**A**), mother's and father's years of schooling, father's log earnings, number of siblings, permanent local earnings and unemployment, as well as region and cohort dummies (**X**). The exclusion restrictions in $\mathbf{Z}\backslash(\mathbf{X}, \mathbf{A})$ are linear and quadratic terms in local short-run earnings and unemployment, and a cubic polynomial of the distance measure. Following Carneiro et al. (2011), I interact the instruments with linear terms in cognitive and noncognitive skills, mother's years of schooling, and number of siblings, as well as region dummies.[23] The sample statistics are presented in Table 1.

4 Empirical Results

The main objective of this paper is to examine the heterogeneity in returns to college using the MTE approach. However, it is informative to first consider the standard estimation methods.

Results using Conventional Methods

A natural point of departure is to consider standard OLS. I estimate different versions of eq. (7), with controls either only entering independently, or also interacted with the college dummy. To illustrate the role of abilities and selection, the models are estimated both assuming that the ability measures are unobserved and observed (i.e., excluded and included from the set of controls). The results are reported in Table 2.

The lifetime return to a year of college is estimated to be around 6 percent when not controlling for observed abilities (columns 1-2). The estimate falls to about 4.8 percent when these are included as controls, which illustrates the potential (positive) ability bias in OLS estimates. Allowing for interactions between the controls and the college dummy decreases the main estimate further, and the heterogeneity with respect to the two ability measures appears quite substantial and of similar magnitude (columns 4-5). The estimates imply that a one standard deviation increase in cognitive or noncognitive ability

---

[23]In Section 4, I test the sensitivity of the results by using only linear terms of the instruments, as well as no interactions. As a further test, I exclude current labor market conditions from the set of instruments. Moreover, it has been common in previous applications to include variables in the outcome equation that are not in the selection equation, e.g. experience and local earnings and unemployment at prime age. As such variables are likely post-determined, I avoid doing so in the main analysis. However, I will present estimates from such specifications in the sensitivity analysis.

increases the return to a year of college by around 1 percentage point. Even these naive estimates thus suggest that there may be considerable variation in individual returns.

As opposed to OLS, standard IV estimates a causal effect without assuming equal potential outcomes for treated and untreated individuals. I report IV estimates for different sets of instruments in Table 3. Observed heterogeneity is taken into account by including a full set of interaction effects. There is considerable variation in the estimated LATEs across different instruments and first-stage models (linear or probit). Since different instruments identify LATEs for different subpopulations, such variation is expected in the presence of self selection on heterogeneous returns. Nevertheless, when I use $P(\mathbf{Z})$ with the full set of instruments, the estimate is in the lower range and close to both the OLS estimate and the semiparametric estimate of the ATE (see below). Even though this is noteworthy, it is not in itself a rejection of the self-selection hypothesis.

Results using a Normal Selection Model

The traditional approach to estimate the model in Section 2 is to specify a parametric joint distribution for the error terms. Björklund and Moffitt (1987), for example, estimate the MTE assuming that the error terms are jointly normally distributed. Although my main focus is on the semiparametric approach, results based on a normal selection model are useful for purposes of comparison.

Figure 1 shows parametric estimates of the MTE by levels of U_S, conditioned on mean values of $(\mathbf{X}, \mathbf{A})$. The MTE is declining and precisely estimated. A simple test of selection on unobserved gains is a test of whether the slope of the MTE is zero. For the normal selection model this implies testing whether $\sigma_{1V} - \sigma_{2V} = 0$ in eq. (15). I estimate that $\sigma_{1V} - \sigma_{2V} = -0.0171$ with a standard error of 0.0057 (obtained using the delta method). The hypothesis that the slope of the MTE is zero is thus rejected at conventional significance levels, supporting the idea that individuals select into college based on idiosyncratic gains. As a comparison, Carneiro et al. (2011) estimate that $\sigma_{1V} - \sigma_{2V} = -0.2388$ with a standard error of 0.0982. Thus, the degree of unobserved heterogeneity therefore appears to be much larger in the US.

However, Figure 1 only provides an overview picture of unobserved heterogeneity. The graph shows the average return at each quantile of U_S, while, in general, there will be a distribution of returns centered around each quantile. If

I also account for observed heterogeneity, i.e., the variation in $\mathbf{X}$ and $\mathbf{A}$ and their impact on the MTE through $\mathbf{X}'(\delta_1 - \delta_0) + \mathbf{A}'(\gamma_1 - \gamma_0)$, then the magnitude of total heterogeneity is more pronounced. Across all individuals, the heterogeneity associated with observed characteristics gives a variation in returns between -0.0858 and 0.2120.

Table 4 (column 1) reports estimates of summary treatment effects based on the parametric MTE estimates and appropriate weights. The estimated ATE implies a return to one year of college of about 3.4 percent. The corresponding estimates for the ATT and ATU are approximately 4.7 and 3.0 percent, respectively. Table 4 also shows tests of equality between ATT and ATE, ATT and ATU, and ATE and ATU, serving as broad tests for self selection on *total* heterogeneity. All tests reject equality and support the notion that individuals select into college based on idiosyncratic gains. But these results rest on the potentially restrictive normality assumption, and it is thus not clear how reliable they are.

Results Using Local Instrumental Variables

A more flexible approach for estimating the MTE is to use Local IV. LIV estimates $E(Y \mid X = \mathbf{x}, \mathbf{A} = \mathbf{a}, P(\mathbf{Z}) = p)$ semiparametrically and then computes its derivative with respect to p, as in eq. (12). If $(\mathbf{X}, \mathbf{A})$ is not independent of (U_0, U_1, V), a necessary and very demanding condition is that P has full support at each value of $(\mathbf{X}, \mathbf{A})$. But for each combination of $(\mathbf{X}, \mathbf{A})$, variation in P only identifies the MTE across small intervals of V. To reduce the dimensionality of $(\mathbf{X}, \mathbf{A})$, I therefore use an index of $\mathbf{X}'(\delta_1 - \delta_0) + \mathbf{A}'(\gamma_1 - \gamma_0)$ and condition on the partitions of this index.[24] The support of P for each value of the index is nevertheless not large enough. If I instead follow Carneiro et al. (2011) and invoke the assumption that $(\mathbf{X}, \mathbf{A})$ is independent of (U_0, U_1, V), then each of the intervals from the conditional identification can be put together so that the MTE can be identified over almost the entire support of V. It is thus only necessary to examine the marginal support of $P(\mathbf{Z})$ as opposed to the support of $P(\mathbf{Z})$ conditional on $(\mathbf{X}, \mathbf{A})$. This assumption also legitimizes the use of interactions between $\mathbf{Z}$ and components of $(\mathbf{X}, \mathbf{A})$ as instruments in the choice equation.

I estimate $P(\mathbf{Z})$ in a probit model and present estimated average marginal

---

[24]I follow Basu et al. (2007) and condition on demideciles (i.e., 20 uniformly distributed groups) of the scalar index $\mathbf{X}'(\delta_1 - \delta_0) + \mathbf{A}'(\gamma_1 - \gamma_0)$. The results are robust to conditioning on finer partitions of the index (e.g., 50 or 100 uniformly distributed groups).

derivatives in Table 5. The average effect of the distance instrument on college attendance is negative and highly significant. The average effect of local unemployment at age 20 is also negative and significant, whereas local earnings at age 20 is a weaker predictor of college attendance. The instruments are jointly strong predictors of college attendance, as are mother's and father's years of schooling, father's earnings, noncognitive ability, and, in particular, the measure of cognitive ability. In fact, cognitive ability is in terms of average derivatives almost six times stronger than noncognitive ability as a predictor of going to college.

Figure 2 shows the (unconditional) support of the estimated $P(\mathbf{Z})$. There is a lack of support only in the top percentile, whereas the support in the lower part of the interval virtually reaches zero.[25] Given the estimates of $P(\mathbf{Z})$, the next step is to estimate the components of equation (14) and compute the MTEs.

Figure 3 shows how the MTE depends on U_S, with 90 percent confidence intervals computed using the bootstrap.[26] The MTE estimates are evaluated at mean values of $(\mathbf{X}, \mathbf{A})$. Two main results emerge. First, the variation in unobserved heterogeneity is in terms of point estimates small. The difference between the sections of U_S with the highest and the lowest MTEs corresponds only to about 3 percentage points in the returns to a year of college. The overall picture of the MTE across the U_S interval suggests an almost flat MTE and thus little unobserved heterogeneity. Second, the little indicative evidence on self selection on unobserved gains is mixed. For values of U_S up until about 0.5, the MTE declines in U_S (i.e., positive selection). Note that a large majority of observations are located here, thus causing the declining MTE for the parametric model shown in Figure 1. For higher values of U_S the semiparametric MTE is actually even increasing (i.e., negative selection), although the estimates are too imprecise. This shows that the normal selection model can provide an incorrect representation of unobserved heterogeneity. While the overall picture is thus somewhat mixed, the evidence on self selection on unobserved gains is clearly much weaker for Sweden than what is reported for the US in Carneiro et al. (2011).

---

[25]The common support is the intersection of the support of $P(\mathbf{Z} \mid S = 0)$ and the support of $P(\mathbf{Z} \mid S = 1)$. I trim observations for which the estimated $P(\mathbf{Z})$ is either lower than the minimum, or higher than the maximum value of $P(\mathbf{Z})$ for which there is common support.

[26]Estimating a similar model, Heckman et al. (1997) show that the bootstrap provides a better approximation to the true standard errors than asymptotic standard errors. Throughout the paper I use 100 bootstrap replications and, in each iteration of the bootstrap, $P(\mathbf{Z})$ is re-estimated to account for the fact that $P(\mathbf{Z})$ is itself an estimated object.

A simple test of selection on unobserved gains consists of comparing the average MTE across equally spaced adjacent intervals along the support of U_S, i.e., LATEs defined over different subpopulations (see Heckman et al., 2010).[27] Table 6 reports the outcome of the test. Even though the test rejects equality of LATEs for low values of U_S where precision is higher, I cannot reject the joint hypothesis that all adjacent LATEs are equal. While this result is insensitive to choice of bandwidth, the exact shape of the MTE is not. Moving from the estimated optimal bandwidth of 0.15 to lower values strenghtens the (weak) tendency of a U-shaped MTE in Figure 3.

Lastly, I turn to my estimates of the ATE, ATT, and ATU. Since I do not have strictly full support on P, these parameters cannot be estimated in exact accordance with their definitions. I can, however, compute approximations of these parameters, denoted $\widetilde{\text{ATE}}$, $\widetilde{\text{ATT}}$, and $\widetilde{\text{ATU}}$, for which I rescale the weights to integrate to one over the common support.

Table 4 (column 2) reports the estimates together with a set of simple tests for self selection on total heterogeneity. The semiparametric estimate of the $\widetilde{\text{ATE}}$ suggests a return to one year of college of about 4.5 percent. As expected, the estimated $\widetilde{\text{ATT}}$ is larger and $\widetilde{\text{ATU}}$ smaller, although the differences are relatively small. Nevertheless, these differences are all statistically significant, thus indicating sorting into college based on total heterogeneity. It is those who actually have selected into college that, on average, also have the highest returns. What is more surprising is that the returns for those who have chosen not to go to college are still significantly positive. This is in contrast with Carneiro et al. (2011), who report an $\widetilde{\text{ATU}}$ that is close to zero. Finally, it is worthwhile to compare with the estimates from the normal selection model (column 1). The normal model captures the full support of U_S and the resluting estimates are on average somewhat lower. But the pattern in terms of the differences across ATE, ATT, and ATU is very similar. As unobserved heterogeneity seems to be of relatively little importance in my data, I now turn to examine the role of observed heterogeneity, and in particular ability-specific heterogeneity.

---

[27]The test is based on 100 bootstrap replications of the MTE, evaluated at mean values of **X** and **A**. I take the average of the MTE in equally spaced intervals along the support of U_S and compute the statistics $T = | \text{LATE}^j - \text{LATE}^{j+1} |$ (i.e., the absolute value of the difference between two adjacent LATEs j and $j+1$) and $T_b = | (\text{LATE}_b^j - \text{LATE}_b^{j+1}) - (\text{LATE}^j - \text{LATE}^{j+1}) |$, where LATE_b^j is the bth bootstrap replication of LATE^j. The statistics for the joint test are $C = \sum_{j=1}^{J-1} (\text{LATE}^j - \text{LATE}^{j+1})^2$ and $C_b = \sum_{j=1}^{J-1} \left[(\text{LATE}_b^j - \text{LATE}_b^{j+1}) - (\text{LATE}^j - \text{LATE}^{j+1}) \right]^2$. The p-value is the proportion of bootstrap replications for which $T_b > T$ (or $C_b > C$ for the joint test).

Evidence on Observable and Ability Heterogeneity

Figure 4 shows the component of the MTE that is attributable to total observed heterogeneity along the scalar index $\mathbf{X}'(\delta_1 - \delta_0) + \mathbf{A}'(\gamma_1 - \gamma_0)$. First, note that this curve is not comparable to Figure 3, which plotted the MTE across the distribution of U_S. U_S is an unobserved variable, while the scalar index on the x-axis in Figure 4 is itself estimated. However, both are components of expected returns and may therefore affect selection into college. Figure 4 implies that the variation in observed heterogeneity is quite substantial, and the slope of the curve indicates that observed characteristics impact on returns across the entire distribution. The point estimates suggest that those with the most favorable observed characteristics (i.e., that complement formal college education the most) on average have a return that is about 20 percentage points higher than those with the least favorable characteristics. In reality, the heterogeneity is even larger since there is also considerable variation within the groups on the x-axis.

An important question is whether observed heterogeneity matters more than unobserved heterogeneity. My results suggest that it does. A simple illustration can be provided by comparing the variation in returns attributed to the scalar index $\mathbf{X}'(\delta_1 - \delta_0) + \mathbf{A}'(\gamma_1 - \gamma_0)$ on the one hand and the unobserved factor U_S on the other hand. My estimates imply that the variance of the former is about 50 times larger than the latter. Intuitively, this is sensitive to choice of bandwidths. When moving from the estimated optimal bandwidth of 0.15 to 0.10 or even 0.05, the relative importance of unobserved heterogeneity increases. Yet, even for a bandwidth of 0.05 the variance in returns attributed to observed characteristics is about four times larger than that attributed to the unobserved factor.

Figures 5a and 5b show the ATE conditional on the measures of cognitive and noncognitive ability, respectively. There is a strong relationship between both measures and the estimated ATE. Moreover, both the pattern and the magnitude of the heterogeneity are roughly similar for the two measures, although the negative effects at the low end are more pronounced for cognitive ability. In contrast, the positive complementarity with college education at the top end, as suggested by the point estimates, is larger for noncognitive ability. Belonging to the top category in the cognitive (noncognitive) ability-distribution implies a return to a year of college that is around 2.5 percentage points (around 5.5 percentage points) higher than the average.

A potential explanation to the high resemblance across the two measures

would be that they are highly correlated. This correlation is about 0.19 in my sample, suggesting that this could only be a partial explanation. Moreover, it is possible that the ability measures are correlated with other control variables that impact on observed heterogeneity. In Table 7 (column 2), I therefore report the (semiparametric) estimates of observed heterogeneity with respect to the ability measures from the outcome equation. Despite the fact that these estimates are thus conditional on observed heterogeneity from other covariates, they imply a pattern similar to the comparison of the conditional ATEs.[28] The estimates, presented as average derivatives, imply that an increase of one standard deviation in either cognitive or noncognitive ability on average increases the return to one year of college by around 2.2 percentage points. The return to a typical college period of 4-5 years would thus on average increase by around 10 percentage points as an individual is shifted up one standard deviation in either of the ability distributions. These estimates are larger than the comparable OLS estimates in Table 2, but similar in terms of the relative importance of the two measures. In total this adds up to quite robust evidence on positive complementarities between informal skills and formal education, and that noncognitive skills appear to matter as much for the payoff from going to college as cognitive skills.

Moreover, the estimates in Table 7 (column 1) imply relatively modest direct effects from the ability measures in the outcome equation (note that the estimates in this column are not annualized). While the positive direct effects suggest that both high cognitive and noncognitive ability entail an absolute advantage on the labor market, the degree of absolute advantage in terms of the ratio between the direct and indirect effects seems somewhat weaker for noncognitive ability. Instead, the indirect effect via schooling is *in relative terms* more important for noncognitive ability, although the difference is not large.

The resemblance between the estimates in Table 7 (column 2) and the conditional ATEs is not surprising. First, the impact on observed heterogeneity from other covariates is small compared to the impact from the measures of cognitive and noncognitive ability.[29] Most observed heterogeneity thus seems

---

[28]Note that these estimates are not conditional on unobserved heterogeneity as they rely on the auxiliary assumption that the observed and unobserved heterogeneity components are uncorrelated. The previously discussed estimates of unobserved heterogeneity are derived conditional on estimated observed heterogeneity, and I can thus not control for unobserved heterogeneity when I estimate observed heterogeneity.

[29]The estimates of observed heterogeneity with respect to other control variables are not shown here, but are available upon request.

to be captured by these two variables. Second, the previous evidence did not suggest any large effects related to unobserved heterogeneity. What is maybe more surprising is that the heterogeneity with respect to noncognitive ability is so large, and roughly comparable to the one with respect to cognitive ability. The choice-equation estimates implied that cognitive ability is a much stronger predictor of selection into college than noncognitive ability. If selection were purely driven by expected monetary benefits, then the two ability measures should have a more equal predictive power of college attendance. However, the model of college choice in Section 1 illustrates some potential explanations for why this need not be the case. For example, cognitive ability might impact on the cost of going to college more than noncognitive ability, either in terms of time costs (yielding more leisure) or psychic costs (less headache) for a given achievement. It could also be due to heterogeneity in the valuation of college as a consumption good; the level of cognitive ability might influence the direct utility derived from going to college more positively than the level of noncognitive ability. Such explanations could all contribute to the result that cognitive ability seems to trigger selection into college much more strongly than noncognitive ability, despite having roughly similar effects on monetary returns.[30]

Sensitivity Analyses and Life-Cycle Effects

A simple way to analyze the robustness of my estimates is to examine how $\widetilde{\text{ATE}}$, $\widetilde{\text{ATT}}$, and $\widetilde{\text{ATU}}$ vary across specifications. In addition, I report a straightforward test of selection on returns: a test of the null that $\widetilde{\text{ATT}} = \widetilde{\text{ATU}}$, i.e., whether the average person attending college has the same return as the average person not attending. Results are reported in panels A, B and C of Table 8.

First, the results in panel A concern choice of sample and specification of the outcome equation. I excluded from my baseline sample everyone without a high school degree. This is in line with Willis and Rosen (1979), whereas Carneiro et al. (2011) include dropouts. Column 2 in panel A indicates that the main estimates as well as estimated heterogeneity ($\widetilde{\text{ATT}} - \widetilde{\text{ATU}}$) stay virtually unchanged when including dropouts. The model that I use has the limitation that it restricts the college variable to be binary. An intuitive critique of estimates of heterogeneous returns is that different people might choose different types of

---

[30]There are of course other potential explanations that go beyond this simple model, e.g., differences in preferences such as discount factors (as emphasized by Willis and Rosen, 1979) or risk attitudes, and differential forecasting errors.

college in terms of quality or field of study. One could, in principle, extend the method used in this paper to multiple schooling types, but that would require distinct instruments for each schooling transition (Heckman et al., 2006b). To explore this limitation, I report estimates with indicators for a set of prestigeous fields of study (STEM, law, business, IT) included in the outcome equation. The resulting estimates in column 3 are clearly lower, which might be because these indicators are pre-determined and not exogenous. However, the estimate of heterogeneity is similar.

A way of improving precision is to include other types of additional controls in the outcome equation, again assuming they are exogenous. Column 4 shows estimates from using local unemployment and earnings at about age 35 as such additional controls (i.e., similar to Carneiro et al., 2011). The estimates in column 4 are somewhat smaller than the baseline, while estimated heterogeneity is marginally larger.

Second, the results in panel B concern the specification of the choice equation. Column 1 shows that the estimates are relatively unaffected by using "any college" (minimum one semester) as the college indicator. Columns 2-4 show alternative specifications of the instruments. Using only linear terms in $\mathbf{Z}$ (column 2) produces larger estimates, while excluding interactions with $\mathbf{Z}$ (column 3) produces somewhat smaller estimates and larger standard errors. One might worry that the local labor market instruments affect selection into college by shifting expected returns, despite the fact that I only use the innovations in these variables in $\mathbf{Z}$. The estimates in column 4, for which only the distance variable is included in $\mathbf{Z}$, are somewhat smaller than the baseline. However, estimated heterogeneity is remarkably similar to the baseline across all the different specifications of the choice equation.

Finally, in panel C I exploit the nearly career-long earnings data to examine how the estimates vary across different definitions of the outcome variable. I first use an imputed wage measure as the outcome in order to examine the role of labor supply (column 1).[31] The estimates are somewhat larger in size, while standard errors and estimated heterogeneity are clearly lower. This may suggest that labor supply plays a role for evidence on selection on returns, as measured by annual earnings. Columns 2-4 show evidence on life-cycle effects

---

[31]I follow the procedure of Antelius and Björklund (2000) who show that left truncating these data, so that low earnings observations and likely part-time workers are excluded, gives similar estimates of the returns to schooling as when using wage measures. I thus use average annual earnings conditional on having annual earnings above 75 000 SEK (about $10000).

in the estimates. In column 2, average earnings across ages 26-30 are used as the outcome and the resulting estimates are now negative. This is not surprising as many at this age might either still be in school or just recent entrants on the labor market. If earnings are instead observed in their late 30s (column 3) the estimates are similar to the baseline. In contrast, if earnings are observed in their late 40s (column 4), the estimated effects are considerably larger than the baseline. Estimated heterogeneity is larger than the baseline for earnings observed at the earlier ages and smaller for older ages, thus indicating that heterogeneity is decreasing in age of observed earnings. This analysis highlights that heterogeneity in earnings profiles makes it very easy to under- or overstate lifetime returns depending on at what age the outcome variable is observed. One remedy is to use long-run averages of earnings whenever possible, and to avoid using short-run measures from ages when current earnings approximate lifetime earnings poorly.

5 Conclusions

I applied the MTE approach of Heckman and Vytlacil (1999, 2001, 2005) to a large registry-based data set of Swedish males. My analysis of the lifetime returns to college revealed a relatively modest role for heterogeneity in general, and for unobserved heterogeneity in particular, at least in comparison to previous evidence (e.g., Carneiro et al., 2011). Yet, total heterogeneity (mainly via observed characteristics) seems to be an important phenomenon, and this holds across various specifications and sample definitions. It is unclear, however, whether the divergence from previous evidence is due to differences in data quality or contextual setting (Sweden vs. the US). A possible explanation for both smaller returns and less heterogeneity could be a lower degree of selection into college in Sweden and most notably a more compressed wage structure. Recent quasi-experimental evidence also lends support to the finding of low returns to college in Sweden (see Öckert, 2010).

Moreover, I provided new evidence on ability heterogeneity using measures of cognitive and noncognitive ability from military enlistment tests. The results implied that both cognitive and noncognitive ability have a large influence on lifetime returns, thus indicating that "school-skill complementarities" (i.e., between formal schooling investments and independently produced abilities) are important features of the labor market. Since the effect of noncognitive

ability is about as large as that of cognitive ability, it is puzzling that the former has much less influence on the probability of selecting into college. If selection were purely driven by expected monetary returns, then the two ability types should have a much more equal predictive power of college attendence. A potential explanation is that cognitive ability also has a much stronger impact on either the costs (e.g., time or psychic costs) or the consumption value of going to school. An intriguing avenue for future research is thus to enable a more causal interpretation of the cost side of such ability heterogeneity, for example by introducing exogenous return-shifters in the Roy model.

Some lessons regarding the applicability of Local IV estimation of the MTE can also be learned from my analysis. While a large sample size is clearly helpful in the LIV setting, the existence of strong continuous instruments remains crucial. Even though my large sample clearly produces more precise estimates of treatment effect parameters compared to previous applications based on survey data, the standard auxiliary assumptions are still needed. Most MTE applications have to date been concerned with the effects of education. Applications to other research questions and contexts are therefore desirable, as is further development of estimation and identification. A recent example of both is Brinch et al. (2012) who show that a linear MTE-curve can be estimated using binary instruments, and apply these insights to the quantity-quality tradeoff of children.

A main conclusion from my results is that there is only limited overall evidence on unobserved heterogeneity. Still, there is some tendency of a U-shaped, or at least L-shaped, pattern of the MTE, which is notable as it differs from the monotonically decreasing curve reported in Carneiro et al. (2011). One explanation to this pattern could simply be that the effect of unobserved heterogeneity is more complex than what is commonly assumed. Several sources could generate a non-monotonic MTE, including heterogeneity in time or risk preferences, assymetric information about costs and benefits of college, and differences in economic resources or access to credit. As Brinch et al. (2012) demonstrate, a non-monotonic MTE can be derived from such underlying data, for example represented by a mixture of multiple normal distributions. A methodological implication is that a typical univariate normal selection model will impose an incorrect representation of treatment effect heterogeneity, which is also strongly confirmed in my analysis.

References

ANGRIST, J., AND G. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity," *Journal of the American Statistical Association*, 90(430), 431–442.

ANTELIUS, J., AND A. BJÖRKLUND (2000): "How Reliable are Register Data for Studies of the Return on Schooling? An examination of Swedish data," *Scandinavian Journal of Educational Research*, 44(4), 341–355.

ARROW, K. (1973): "Higher Education as a Filter," *Journal of Public Economics*, 2(3), 193–216.

AUTOR, D., L. KATZ, AND M. KEARNEY (2008): "Trends in US Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 90(2), 300–323.

BASU, A., J. HECKMAN, S. NAVARRO-LOZANO, AND S. URZUA (2007): "Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: an Application to Treatments of Breast Cancer Patients," *Health Economics*, 16(11), 1133–1157.

BHULLER, M., M. MOGSTAD, AND K. SALVANES (2011): "Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings," *IZA Discussion Paper No. 5788*.

BJÖRKLUND, A., AND R. MOFFITT (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection," *The Review of Economics and Statistics*, 69(1), 42–49.

BLACKBURN, M., AND D. NEUMARK (1993): "Omitted-Ability Bias and the Increase in the Return to Schooling," *Journal of Labor Economics*, pp. 521–544.

BRINCH, C., M. MOGSTAD, AND M. WISWALL (2012): "Beyond LATE with a Discrete Instrument," *Statistics Norway Discussion Paper No. 703*, (703).

CAMERON, S., AND J. HECKMAN (1998): "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy*, 106(2), 262–333.

CAMERON, S., AND C. TABER (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling," *Journal of Political Economy*, 112(1), 132–182.

CARD, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. Christofides, K. Grant, and R. Swidinsky, pp. 201–222. University of Toronto Press, Toronto.

——— (1999): *The Causal Effect of Education on Earnings*vol. 3A of *Handbook of Labor Economics*, chap. 30, pp. 1801–1863. Elsevier, Amsterdam.

CARLSTEDT, B. (2000): "Cognitive Abilities: Aspects of Structure, Process and Measurement," Ph.D. thesis, University of Gothenburg.

CARNEIRO, P., J. HECKMAN, AND E. VYTLACIL (2011): "Estimating Marginal Returns to Education," *The American Economic Review*, 101(6), 2754–2781.

CARNEIRO, P., AND S. LEE (2009): "Estimating Distributions of Potential Outcomes Using Local Instrumental Variables with an Application to Changes in College Enrollment and Wage Inequality," *Journal of Econometrics*, 149(2), 191–208.

CURRIE, J., AND E. MORETTI (2003): "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *The Quarterly Journal of Economics*, 118(4), 1495–1532.

ERIKSON, R., AND J. JONSSON (1993): *Ursprung och utbildning*, SOU 1993:85. Utbildningsdepartementet.

HAIDER, S., AND G. SOLON (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *The American Economic Review*, 96(4), 1308–1320.

HANSEN, K., J. HECKMAN, AND K. MULLEN (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1-2), 39–98.

HECKMAN, J., D. SCHMIERER, AND S. URZUA (2010): "Testing the Correlated Random Coefficient Model," *Journal of Econometrics*, 158(2), 177–203.

HECKMAN, J., J. STIXRUD, AND S. URZUA (2006a): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24(3), 411–482.

HECKMAN, J., S. URZUA, AND E. VYTLACIL (2006b): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, 88(3), 389–432.

HECKMAN, J., AND E. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96(8).

——— (2001): "Local Instrumental Variables," in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, pp. 1–46, New York, NY. Cambridge University Press.

HECKMAN, J., AND E. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64(4), 605–654.

HEINECK, G., AND S. ANGER (2010): "The Returns to Cognitive Abilities and Personality Traits in Germany," *Labour Economics*, 17(3), 535–546.

HERRNSTEIN, R., AND C. MURRAY (1994): *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.

KEANE, M. (2002): "Financial Aid, Borrowing Constraints, and College Attendance: Evidence from Structural Estimates," *The American Economic Review (Papers and Proceedings)*, 92(2), 293–297.

KEANE, M., AND K. WOLPIN (2001): "The Effect of Parental Transfers and Borrowing Constraints on Educational Ettainment," *International Economic Review*, 42(4), 1051–1103.

KJELLSTROM, C., AND H. REGNER (1999): "The Effects of Geographical Distance on the Decision to Enrol in University Education," *Scandinavian Journal of Educational Research*, 43(4), 335–348.

KLING, J. (2001): "Interpreting Instrumental Variables Estimates of the Returns to Schooling," *Journal of Business and Economic Statistics*, 19(3), 358–364.

LINDQVIST, E., AND R. VESTMAN (2011): "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3(1), 101–128.

MUELLER, G., AND E. PLUG (2006): "Estimating the Effect of Personality on Male and Female Earnings," *Industrial and Labor Relations Review*, 60(1), 3–22.

MURNANE, R., J. WILLETT, AND F. LEVY (1995): "The Growing Importance of Cognitive Skills in Wage Determination," *The Review of Economics and Statistics*, 77(2), 251–66.

NORDIN, M. (2008): "Ability and Rates of Return to Schooling - Making Use of the Swedish Enlistment Battery Test," *Journal of Population Economics*, 21(3), 703–717.

NYBOM, M., AND J. STUHLER (2011): "Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation," *IZA Discussion Paper No. 5697*.

NYHUS, E., AND E. PONS (2005): "The Effects of Personality on Earnings," *Journal of Economic Psychology*, 26(3), 363–384.

ÖCKERT, B. (2010): "What's the Value of an Acceptance Letter? Using Admissions Data to Estimate the Return to College," *Economics of Education Review*, 29(4), 504–516.

——— (2012): "On the Margin of Success? Effects of Expanding Higher Education for Marginal Students," *Nordic Economic Policy Review*, (1), 111–158.

ROY, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *The Review of Economic Studies*, 68(3), 665.

WILLIS, R., AND S. ROSEN (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5), 7–36.

Zax, J., and D. Rees (2002): "IQ, Academic Performance, Environment, and Earnings," *Review of Economics and Statistics*, 84(4), 600–616.

6 Tables and Figures

Table 1 Summary Statistics by Treatment Group

	$S=1$		$S=0$	
	Mean	SD	Mean	SD
Log lifetime earnings	12.20	0.50	11.91	0.47
Cognitive test score	6.95	1.54	5.45	1.81
Noncognitive test score	5.93	1.85	5.30	1.76
Mother's years of education	9.99	3.03	8.30	2.13
Number of siblings	1.84	1.17	2.09	1.45
Father's years of education	11.44	3.63	9.01	2.72
Father's log earnings	12.36	1.19	11.90	1.28
Local long-run earnings (SEK/100)	137.95	14.35	134.94	14.47
Local long-run unemployment	0.21	0.04	0.22	0.04
Distance to university (km/100)	0.91	0.99	1.09	1.06
Local short-run earnings (SEK/100)	132.35	19.09	131.85	19.23
Local short-run unemployment	0.26	0.07	0.26	0.07
Non-missing earnings observations	25.72	1.34	25.84	1.04
Years of education	15.90	1.16	11.60	0.90
Number of observations	23146		61043	

Note: Lifetime earnings is computed as the average of all non-missing annual earnings observations for ages 25-50. Father's earnings are computed as the log of average annual earnings for years 1968-1972. Local permanent labor market characteristics are computed as averages across the years 1968-1990 by municipality of residence at age 20. The short-run measures are for age 20. Unemployment is computed as one minus the local working-age employment rate. Distance to university is measured as the closest route by car from the municipality of residence in 1965 to the closest university city. Included in the set of controls are also regional and birth-year dummies (not reported here).

Table 2 OLS Estimates of the Return to a Year of College

	OLS Coefficients				
	(1)	(2)	(3)	(4)	(5)
College dummy (S)	0.0602	0.0601	0.0479	0.0414	0.0417
	(0.0009)	(0.0010)	(0.0009)	(0.0011)	(0.0011)
S*A (Cognitive)	.	.	.	0.0069	0.0064
	.	.	.	(0.0007)	(0.0007)
S*A (Noncognitive)	.	.	.	0.0056	0.0049
	.	.	.	(0.0005)	(0.0005)
Ability controls (A)	.	.	x	x	x
Interactions S*A	.	.	.	x	x
Interactions S*X	.	x	.	.	x

Note: This table reports OLS estimates of the return to college. The controls (**X**) include region and cohort dummies, and linear and quadratic terms of father's and mother's years of schooling, father's log earnings, number of siblings, local long-run unemployment and earnings at age 20. Specifications (3)-(5) add linear and quadratic terms of cognitive and noncognitive ability (i.e. **A**), (2) and (5) include interactions between S and all components of **X**, and (4) and (5) include interactions between S and all components of **A**. The interaction terms (S***A**) are reported as average derivatives. I obtain annualized returns by dividing all estimates by 4.3, which is the average difference in years of schooling for those with $S = 1$ and those with $S = 0$. Standard errors (from 100 bootstrap replications) are in parentheses.

Table 3 IV Estimates of the Return to a Year of College

	IV estimates for different sets of instruments			
	Distance to university	Local earnings	Local unempl.	All
Standard 2SLS	0.0673	0.2003	0.1707	0.0805
	(0.0404)	(0.0425)	(0.0358)	(0.0228)
P(**Z**) as instr.	0.0319	0.0931	0.0773	0.0429
	0.0241	0.0257	0.0235	0.0177

Note: This table reports IV estimates of the return to college for different sets of instruments: distance to university at age 20 in column (1), local short-run earnings and unemployment in (2) and (3), and all these instruments in (4). Row 1 reports two-stage least squares estimates, row 2 estimates using $P(Z)$ as instrument (probit first stage). All specifications include second-stage interactions between predicted college and the components of **X** and **A**. I obtain annualized returns by dividing all estimates by 4.3, which is the average difference in years of schooling for those with $S = 1$ and those with $S = 0$. Bootstrapped standard errors are in parentheses (100 replications).

Table 4 Returns to a Year of College

Model	Normal	Semiparametric
ATE	0.0344	0.0454
	(0.0024)	(0.0172)
ATT	0.0466	0.0643
	(0.0023)	(0.0149)
ATU	0.0298	0.0381
	(0.0025)	(0.0182)
ATT - ATU	0.0169	0.0262
	(0.0010)	(0.0051)
ATT - ATE	0.0122	0.0190
	(0.0007)	(0.0037)
ATE - ATU	0.0046	0.0072
	(0.0003)	(0.0014)

Note: This table reports estimates of the average treatment effect (ATE), treatment effect on the treated (ATT), and treatment effect on the untreated (ATU) from the normal model in equation (15), and the semiparametric model. The latter are thus for $\widehat{ATE}$, $\widehat{ATT}$, and $\widehat{ATU}$, indicating that they are sample specific parameters with weights integrating to one over the empirical support of U_S. Rows 4-6 show the estimated differences between the treatment effect parameters. All estimates are annualized, reflecting the average difference in years of schooling for those with $S = 1$ and those with $S = 0$. Standard errors are obtained using the bootstrap (100 replications).

Table 5 College Decision Model

Controls (**A**, **X**)	Average derivative
Cognitive test score	0.0615
	(0.0008)
Non-cognitive test score	0.0109
	(0.0009)
Mother's years of schooling	0.0154
	(0.0008)
Number of siblings	-0.0112
	(0.0014)
Father's years of schooling	0.0166
	(0.0006)
Father's log earnings	0.0413
	(0.0026)
Local long-run earnings	0.0015
	(0.0010)
Local long-run unempl.	0.5959
	(0.2268)
Instruments (**Z**)	
Distance to university (km/100)	-0.0407
	(0.0116)
Local short-run earnings	-0.0016
	(0.0009)
Local short-run unempl.	-0.6024
	(0.1883)
Joint significance test of **Z**: p-value	0.0000

Note: The table reports average derivatives from the college choice equation (see Section 3 for exact specification) obtained by computing, for each individual, the effect of increasing a variable by one unit on the probability of enrolling in college, and then average across all individuals. The model also includes region- and birth-year dummies (not reported here). Standard errors are obtained using the bootstrap (100 replications).

Table 6 Test for Equality of LATEs over Different Intervals

Range of LATE^j	(.10;.20)	(.25;.35)	(.40;.50)	(.55;.65)
Range of LATE^{j+1}	(.25;.35)	(.40;.50)	(.55;.65)	(.70;.80)
LATE^j-LATE^{j+1}	0.0062	0.0068	0.0015	0.0077
p-value	0.03	0.10	0.80	0.32
Joint p-value		0.29		

Note: This table reports a test of unobserved heterogeneity conducted by testing the equality of LATEs in pairwise adjacent intervals of U_S. I construct intervals of U_S and average the MTE within these intervals by computing $E(Y_1 - Y_0 \mid \mathbf{X} = \bar{\mathbf{x}}, U_S^{L_j} \leq U_S \leq U_S^{U_j})$, where $U_S^{L_j}$ and $U_S^{U_j}$ are the lower and upper bounds of U_S in interval j. This gives the different LATEs and the null of the tests are $H_0 : \text{LATE}^j(U_S^{L_j}, U_S^{U_j}) - \text{LATE}^{j+1}(U_S^{L_{j+1}}, U_S^{U_{j+1}}) = 0$. The bottom row reports the outcome of the test that all adjacent LATEs are jointly equal. All tests take the multiple estimation steps into account by using the bootstrap (100 replications).

Table 7 Average Derivatives for Abilities in the Outcome Equation

	γ_0	$\gamma_1 - \gamma_0$
Cognitive ability	0.0248	0.0128
	(0.0050)	(0.0044)
Noncognitive ability	0.0195	0.0117
	(0.0019)	(0.0012)

Note: This table reports average derivatives of the measures of cognitive and noncognitive ability in the outcome equations for the semiparametric model. The procedure, along with exact specifications of the full set of control variables (not reported here), are described in Section 2. The average derivatives are obtained by computing for each individual the effect of increasing a variable by one unit (keeping all the others constant) on log lifetime earnings and then average across all individuals. One standard deviation of the cognitive or noncognitive measure corresponds to about 1.8 units. Column 1 reports the main effects, whereas column 2 reports the interaction effects (i.e., observed heterogeneity). The estimates in column 2 (not in column 1) are annualized, reflecting the average difference of 4.3 years of schooling for those with $S = 1$ and those with $S = 0$. Standard errors are bootstrapped (100 replications).

Table 8 Sensitivity Analyses and Life-Cycle Effects

(a) Different Samples and Specification of the Outcome Equation

	Baseline	Including high school dropouts	Type of college in $\mathbf{X} \setminus \mathbf{Z}$	$\mathbf{X} \setminus \mathbf{Z}$ as Carneiro et al. (2011)
$\widetilde{ATE}$	0.0454	0.0424	0.0217	0.0290
	(0.0172)	(0.0192)	(0.0163)	(0.0162)
$\widetilde{ATT}$	0.0643	0.0619	0.0431	0.0526
	(0.0149)	(0.0170)	(0.0145)	(0.0142)
$\widetilde{ATU}$	0.0381	0.0351	0.0136	0.0201
	(0.0182)	(0.0202)	(0.0171)	(0.0171)
$\widetilde{ATT} - \widetilde{ATU}$	0.0262	0.0268	0.0295	0.0325
	(0.0051)	(0.0047)	(0.0049)	(0.0048)

(b) Specification of the Choice Equation

	Any college as treatment	Only linear terms in $\mathbf{Z}$	No interactions with $\mathbf{Z}$	Only distance in $\mathbf{Z}$
$\widetilde{ATE}$	0.0550	0.0639	0.0314	0.0292
	(0.0171)	(0.0190)	(0.0249)	(0.0232)
$\widetilde{ATT}$	0.0732	0.0797	0.0517	0.0504
	(0.0154)	(0.0171)	(0.0221)	(0.0202)
$\widetilde{ATU}$	0.0473	0.0579	0.0237	0.0211
	(0.0180)	(0.0199)	(0.0261)	(0.0245)
$\widetilde{ATT} - \widetilde{ATU}$	0.0259	0.0218	0.0280	0.0293
	(0.0045)	(0.0049)	(0.0057)	(0.0060)

(c) Definitions of the Outcome Variable and Life-Cycle Effects

	Lifetime wage ages 25-50	Average earnings ages 26-30	Average earnings ages 36-40	Average earnings ages 46-50
$\widetilde{ATE}$	0.0664	-0.0386	0.0519	0.1195
	(0.0112)	(0.0194)	(0.0214)	(0.0229)
$\widetilde{ATT}$	0.0740	-0.0102	0.0636	0.1285
	(0.0097)	(0.0176)	(0.0189)	(0.0202)
$\widetilde{ATU}$	0.0635	-0.0495	0.0474	0.1160
	(0.0118)	(0.0202)	(0.0225)	(0.0241)
$\widetilde{ATT} - \widetilde{ATU}$	0.0106	0.0393	0.0162	0.0125
	(0.0034)	(0.0050)	(0.0061)	(0.0059)

Note: This table reports estimates of the return to a year of college for the semiparametric model for various samples and specificiations. The estimates of the ATE, ATT, and ATU are computed such that their weights integrate to one over the common support. The table also reports a simple test of self selection: if $\widetilde{ATT} - \widetilde{ATU} = 0$. Panel A reports estimates for different samples and specifications of the outcome equation, Panel B for different specifications of the outcome equation, and Panel C for the definition of the outcome variable and life-cycle effects. Standard errors are obtained using the bootstrap (100 replications).

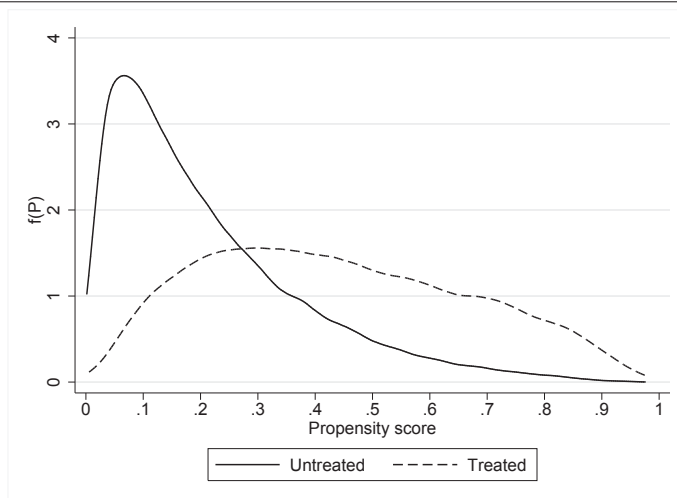**Figure 1** MTE by U_S Estimated from a Normal Selection Model
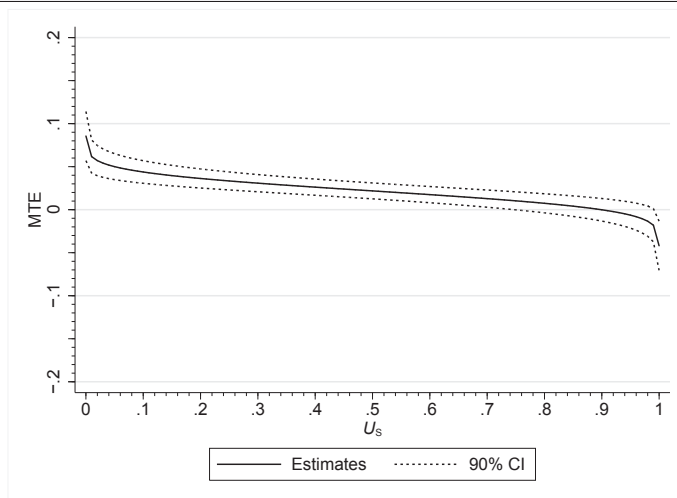
Note: This figure shows point estimates and 90 percent confidence bands of the MTE from the normal selection model in eq. 15 estimated by maximum likelihood. All estimates are conditioned on mean values of **X** and **A**.

Figure 2 Support of $P(\mathbf{Z})$ for untreated ($S = 0$) and treated ($S = 1$)

Note: This figure shows the support of $P(\mathbf{Z})$ for the treated and the untreated. $P(\mathbf{Z})$ is the probability of going to college estimated in a probit regression (see Table 5).

Figure 3 MTE by U_S Estimated by Semiparametric LIV
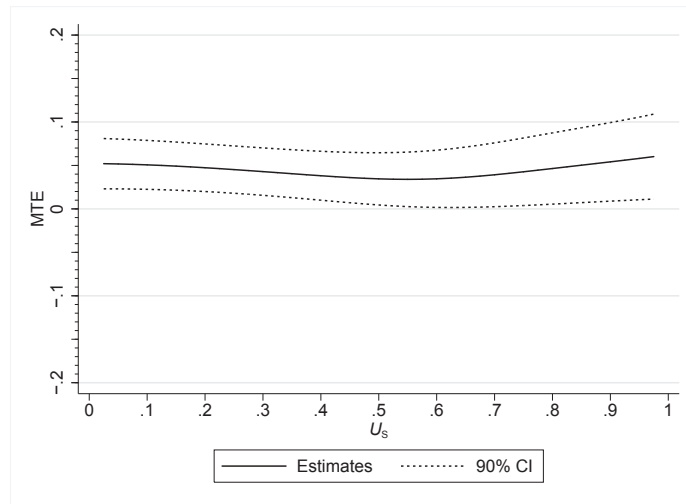
Note: This figure shows point estimates and 90 percent confidence bands of the MTE from the semiparametric model in eq. 11. The model is estimated by the local linear regression procedure described in Section 2. All estimates are conditioned on mean values of **X** and **A**. Standard errors are bootstrapped (100 replications).

**Figure 4** Average MTE by Total Observed Heterogeneity
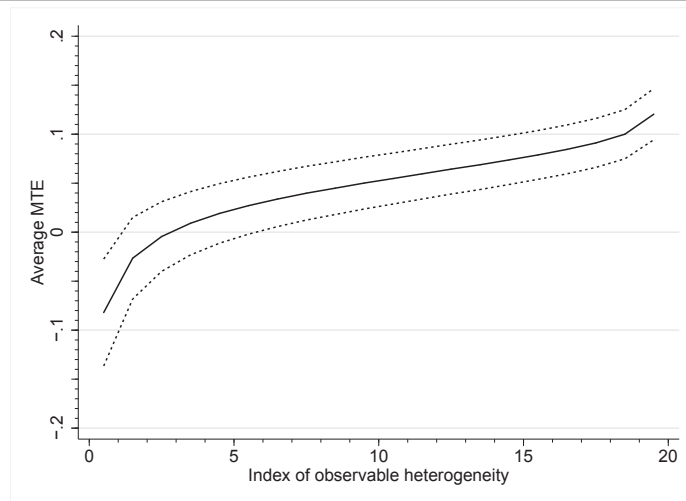
Note: This figure shows the average MTE with 90 percent confidence bands across the index of observed heterogeneity. The index is computed by estimating $\mathbf{X}(\delta_1 - \delta_0) + \mathbf{A}(\gamma_1 - \gamma_0)$ for each individual and splitting the sample into 20 uniformly distributed groups. Standard errors are bootstrapped (100 replications).

37

Figure 5 Observed Ability Heterogeneity in the Return to a Year of College

(a) ATE by Cognitive Ability
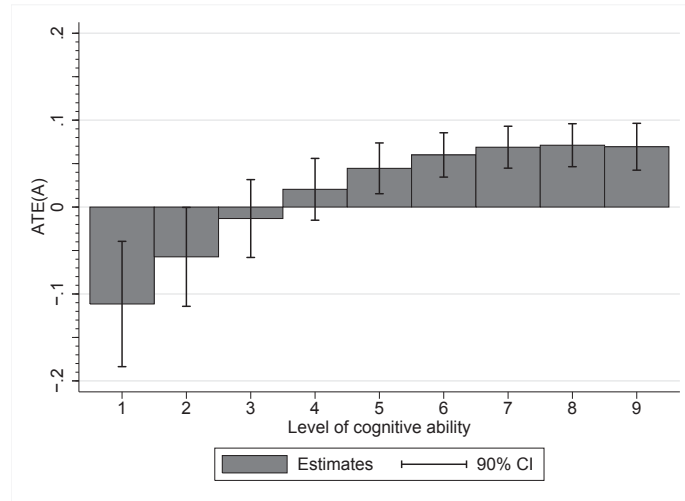
(b) ATE by Noncognitive Ability
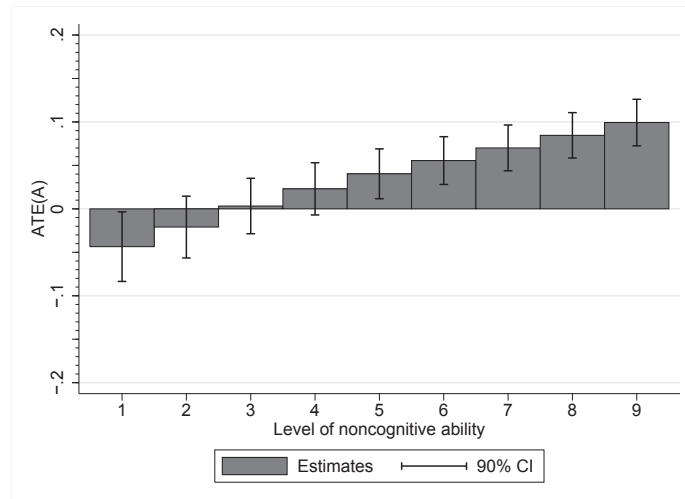
Note: The figures show semiparametric estimates of average treatment effects (ATE) with 90 percent confidence bands conditional on levels of cognitive and noncognitive ability. Standard errors are bootstrapped (100 replications).

Appendix

A.1 Definitions of Weights for ATE, ATT, ATU

Under the LIV approach (and the parametric), all treatment parameters of concern can be identified by using weighted averages of MTE. Heckman et al. (2006b) show that

$$
\begin{aligned}
\text{ATE}(\mathbf{x}, \mathbf{a}) &= E\left[B \mid X = \mathbf{x}, \mathbf{A} = \mathbf{a}\right] = \int_0^1 \text{MTE}(\mathbf{x}, \mathbf{a}, u_S) w_{ATE}(\mathbf{x}, \mathbf{a}, u_S) du_S \\
\text{ATT}(\mathbf{x}, \mathbf{a}) &= E\left[B \mid X = \mathbf{x}, \mathbf{A} = \mathbf{a}, S = 1\right] = \int_0^1 \text{MTE}(\mathbf{x}, \mathbf{a}, u_S) w_{ATT}(\mathbf{x}, \mathbf{a}, u_S) du_S \\
\text{ATU}(\mathbf{x}, \mathbf{a}) &= E\left[B \mid X = \mathbf{x}, \mathbf{A} = \mathbf{a}, S = 0\right] = \int_0^1 \text{MTE}(\mathbf{x}, \mathbf{a}, u_S) w_{ATU}(\mathbf{x}, \mathbf{a}, u_S) du_S,
\end{aligned}
$$

where the weights are

$$
\begin{aligned}
w_{ATE}(\mathbf{x}, \mathbf{a}, u_S) &= 1 \\
w_{ATT}(\mathbf{x}, \mathbf{a}, u_S) &= \frac{\int_{u_S}^1 f(P(\mathbf{Z}) = P(\mathbf{z}) \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}) dP(\mathbf{z})}{E\left[P(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}\right]} \\
w_{ATU}(\mathbf{x}, \mathbf{a}, u_S) &= \frac{\int_0^{u_S} f(P(\mathbf{Z}) = P(\mathbf{z}) \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}) dP(\mathbf{z})}{E\left[1 - P(\mathbf{Z}) \mid \mathbf{X} = \mathbf{x}, \mathbf{A} = \mathbf{a}\right]},
\end{aligned}
$$

and f is the density function of $P(\mathbf{Z})$. By integrating the weighted estimates of $\text{MTE}(\mathbf{x}, \mathbf{a}, u_S)$ over the joint distribution of $(\mathbf{X}, \mathbf{A})$ the estimates of $\text{MTE}(u_S)$ are obtained. In practice, however, I do not condition on $(\mathbf{X}, \mathbf{A})$ nonparameterically. Instead, I follow Basu et al. (2007) and others and condition on, and thus also integrate over, demideciles of the (estimated) scalar index $\mathbf{X}(\delta_1 - \delta_0) + \mathbf{A}(\gamma_1 - \gamma_0)$. Lastly, integrating over $P(\mathbf{z})$ gives the unconditional estimates of ATE, ATT and ATU. The ability-specific ATEs are obtained by evaluating and comparing the treatment parameters at different values of $\mathbf{A} = \mathbf{a}$.