Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation: Dissertation Summary

Pedro Carneiro
University of Chicago
Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation

Pedro Carneiro

The economic return to schooling is a fundamental parameter of interest in many different areas of economics and public policy. It is one of the most frequently estimated parameters in empirical economics. Economists interested in growth are concerned about the role of schooling in productivity and growth. Economists studying inequality and poverty seek to learn how schooling increases the incomes of the poor. Therefore, the evaluation of policies that promote school attendance is a central research question. The increase in earnings due to additional schooling (what is usually called the return to schooling) is a main component of the benefits of proposed policies.

The specification most often used to estimate the return to schooling comes from Mincer (1974):

\[ \ln Y = \alpha + \beta S + \varepsilon , \]

where \( \ln Y \) is log earnings, \( S \) is years of schooling, and \( \beta \) is the return to schooling.\(^1\) There are two important sources of heterogeneity to consider in this equation. The first source is \( \varepsilon \), and it influences the potential earnings of an individual uniformly, no matter what level of schooling he chooses. In this paper selection in levels arises if \( \varepsilon \) is correlated with schooling. This could happen if those who have more years of schooling also have higher earnings ability as measured by \( \varepsilon \). The other source of heterogeneity is \( \beta \), the individual gain in wages from additional schooling. We say there is selection in returns if \( \beta \) is correlated with schooling. This could happen if those who choose to get more schooling are also the ones who benefit the most from schooling. Both types of selection are sources of econometric problems.

Most of the earlier work on this topic considered a representative agent model where \( \beta \) is the return to schooling for the representative agent. Assume, for instance, that we want to evaluate a policy that increases college enrollment by amount \( K \). \( \beta \) is the return to college for an individual, so \( B = \beta^* K \) is the total benefit of the policy. In the traditional approach \( \beta \) is either assumed to be the same for everyone (possibly conditioning on a set of individual characteristics) or assumed to be a random variable uncorrelated with \( S \). The main econometric problem in estimating \( \beta \) (or the average \( \beta \), in the case where \( \beta \) is random) is that \( S \) can be correlated with \( \varepsilon \) because of unobserved ability. Therefore it is assumed that there is only selection in levels but no selection in returns, and the usual way to correct for this is to use linear instrumental variables (IV).

If there is heterogeneity in \( \beta \) in the population and individuals respond to these differences when choosing their level of schooling (if, for example, individuals decide to enroll in school based on the returns they face), two additional problems can emerge. The first is an economic problem: No single number summarizes the distribution of returns. It becomes necessary to define what the parameter is that one is interested in to answer a specific economic question. Separate policies target groups of individuals that are located in different sections of the distribution of \( \beta \). To compute \( B \) we need to know where in the distribution of \( \beta \) the new entrants are coming from. Even though many economists focus on the average return to schooling in the population, this is only one of many parameters that can be defined, and, in general, is not the relevant answer to most policy questions.\(^2\) A more interesting pair of questions is, “What is the return to the marginal entrant?” and “Is it above or below the return to the average student?”

The second problem is an econometric problem. Once we have defined the parameter of interest, how can we estimate it? The usual intuitions about how instrumental variables work break down in a model where \( \beta \) is heterogeneous in the population and individuals respond to these differences when making their schooling decisions (Heckman and Vytlacil 2003 and Carneiro, Heckman, and Vytlacil 2001). In general, applications of the linear IV method (the solution to the selection problem in the common return model) will not generate estimates of policy-relevant parameters.

The recent literature reveals substantial concern about heterogeneity in returns to education. Card (1999, 2001) surveys this literature with a special focus on instrumental variables estimates of \( \beta \), since this has been the preferred method of estimation for most of the economists working in this area in recent years (Heckman 1997 and Heckman and Vytlacil 1998). The question he asks is the following: “If \( \beta \) varies in the population, what parameter do we get from instrumental variables estimates of the returns to schooling?” Based on the local average treatment effect (LATE) framework of Imbens and Angrist (1994), he interprets the instrumental variables estimate as a weighted average of returns for individuals induced to change their level of schooling by variation in the instrumental variable (Kling 2001). If the instrument being used is distance to college, IV measures a weighted average of the return for individuals who are induced to increase their schooling by reductions in commuting costs associated with having a college nearer to their place of residence. The IV approach has been advanced as less restrictive and more robust than econometric selection models, which provide an alternative way to
account for selection in the estimation of the returns to schooling (Krueger 2000). However, as suggested by Card’s interpretation, for many policy questions it does not estimate the policy-relevant parameters unless the policy affects the same people that are affected by the instrument. Therefore, the policy evaluation problem is greatly simplified if the assumption of a common \( \beta \) in the population is satisfied, since there is only one parameter of interest, which can be estimated using IV. But, empirically, is this a valid working assumption? This is the main question of this paper.

I focus on the high school–college transition and examine whether selection in the returns to college is an empirically important phenomenon. I present estimates of the returns for the average person in the population, the average person in college, and the average person at the margin between going or not going to college. I analyze different demographic groups in three different datasets: the National Longitudinal Survey of Youth of 1979 (NLSY79), the High School and Beyond (HSB), and the Panel Study of Income Dynamics (PSID). Across these different samples, I find that the average person going to college has a higher return than the marginal person who is indifferent about enrolling in college or not (the only exception is white females in HSB). This suggests that heterogeneity is important and needs to be accounted for in policy analysis. In particular, it suggests that education policies that increase college enrollment in the population attract individuals of low returns to college (relative to the average college student), and that we can estimate exactly the average return to college among the individuals enticed into college by different policies.

I start by presenting a simple model of schooling and earnings, the framework for the analysis performed in this paper. Assume that there are two schooling levels: high school (\( S=0 \)) and college (\( S=1 \)). The log earnings of individual \( i \) if he goes to college are

\[
(2) \quad \ln Y_u - \mu_u(X) + U_{u}.
\]

If he does not go to college his earnings are

\[
(3) \quad \ln Y_w = \mu_w(X) + U_{w},
\]

where \( X \) are observable characteristics for each individual and \( (U_u, U_w) \) are unobservable characteristics such that

\[
E(U_u) = 0, \quad E(U_w) = 0.
\]

Additive separability between \( X \) and \( (U_u, U_w) \) is not necessary for most of what I do in this paper, but it is a convenient assumption that simplifies the analysis and the empirical work (Heckman and Vytlacil 2000, 2003). Individuals can differ in their observables and in their unobservables. This is the potential outcomes model used in studies of unionism, migration, sectorial choice, job training, and in the literature on policy evaluation.

The return to schooling for individual \( i \) is

\[
\beta_i = \ln Y_{u_i} - \ln Y_{o_i} = \mu_i(X) - \mu_i(X) + U_{u_i} - U_{o_i},
\]

and it varies with \( X \) and \( U_{u_i} - U_{o_i} \). We only observe an individual’s earnings for the sector he chooses. Then observed earnings are

\[
\ln Y_i = S_i \ln Y_u + (1 - S_i) \ln Y_w = \mu_u(X_i) + S_i [\mu_i(X_i) - \mu_u(X_i) + U_{u_i} - U_{o_i}] + U_{o_i} = \mu_o(X_i) + S_i \beta_i + U_{o_i},
\]

which is exactly in the same form as (1) (with the inclusion of observable variables \( X \)), where

\[
\alpha_i = \mu_u(X_i), \quad \varepsilon_i = U_{o_i}
\]

(and \( \beta \) is a function of \( X \) and \( U_u - U_w \)). Individual \( i \) chooses to go to college \( (S = 1) \) if

\[
(4) \quad \mu_u(X_i, Z_i) + U_{o_i} > 0,
\]

where \( Z \) are variables that affect the choice of schooling but not the potential earnings in each sector, and \( U_{o_i} \) is the unobservable in the choice of schooling equation. Equation (4) should be interpreted as a reduced form representation of the choice problem, which is typical in applications of selection models. Assume that

\[
(5) \quad \mu_s(X_i, Z_i) \text{ is a nondegenerate function of } Z
\]

(so that \( Z \) is not independent of \( S \)) and that

\[
(6) \quad U_u, U_w, U_o \perp Z \mid X
\]

and

\[
(7) \quad U_u, U_w, U_o \perp X,
\]

where \( \perp \) denotes independence. \( Z \) is a vector of instrumental variables.

Theoretically, heterogeneity is an important problem; however, dealing with heterogeneity is difficult. The fundamental question I ask in this paper is, “Empirically, how important is it to account for heterogeneity in the evaluation of education policy?”

In the next section I assess the empirical importance of heterogeneity and selection for educational policy evaluation. The central parameter of my analysis is the marginal treatment effect (MTE), which is defined as

\[
\text{MTE}(x, u) = E(\beta \mid X = x, U_{w} = u) = \mu_o(x) - \mu_u(x) + E[U_{u} - U_{o} \mid X = x, U_{w} = u]
\]

(Heckman and Vytlacil 1999, 2000, 2001b, 2003; Carneiro, Heckman, and Vytlacil 2001; and Bjorklund and Moffitt 1987). MTE(\( x, u \)) measures the average return to schooling for individuals with \( X = x \) and \( U_{w} = u \). \( X \) and \( U_{w} \) are variables that determine whether or not an individual
goes to school (see equation (4)). As an example, suppose $X$ is a scalar and $\mu(x,Z)$ is increasing in $X$, so that people are more likely to go to school the higher their $X$ and the higher their $U_y$. If the MTE is increasing in $X$ and $U_y$, then the average return for those going to college (individuals with high $X$ and high $U_y$) is higher than the average return for those not going to college (individuals with low $S$ and low $U_y$). Dependence between $\beta$ and $(X,U_y)$ leads to dependence between $\beta$ and $S$ and to a model where there is selection on returns. However, if $\beta$ is independent of $X$ and $U_y$, then there is no selection. In that case, MTE$(x,u)$ = $\bar{\beta}$, where $\bar{\beta}$ is a constant independent of $X$ and $U_y$.

Notice that the marginal treatment effect is a function of an unobservable variable: $U_y$. There is a large literature in econometrics that develops methods to deal with selection of unobservables. In this paper I use a nonparametric selection model and apply the method of local instrumental variables (LIV) introduced by Heckman and Vytlacil (2000) and applied in Carneiro, Heckman, and Vytlacil (2001). To focus on unobservables, and also for easier exposition, I keep the conditioning on $X$ implicit. The goal is to estimate $E(\ln Y_j | X=x, U_y=u) = \tilde{\beta}(x) + E[U_j - U_y | X=x, U_y=u]$ $\tilde{\beta}(x) = \mu(x) - \mu(x)$.

The method of local instrumental variables requires that there exist a continuous instrument $Z$ satisfying (5) and (6). The intuition of the method is best explained with an example. Suppose the model is the following:

\[ S = I[-Z \gamma + U_y > 0], \]

where $Z$ (the instrumental variable) is tuition in county of residence $\gamma > 0$ and $U_y$ is "ability" and it is unobserved.

Assume we start by using only two counties: A and B. In county A, $Z=100$, and in county B, $Z=200$. The two counties are equal in every aspect except tuition. We can estimate $\beta$ by standard IV using tuition as the instrument and data only from counties A and B:

\[ \hat{\beta}_{IV}^{100,200} = \frac{E(\ln Y_j | Z=100) - E(\ln Y_j | Z=200)}{E(S | Z=100) - E(S | Z=200)} = E[\beta | S(100) = 1, S(200) = 0] = E(\beta | 100 < U_y < 200). \]

This is the local average treatment effect for the case where the instrument takes values $Z=100$ and $Z=200$. It is the average return for individuals who go to college if $Z=100$ but do not go if $Z=200$. Therefore these individuals are at the margin between going to college and not going if $Z$ is between 100 and 200. The fact that they are at the margin at such a low level of tuition means that they have low ability, $U_y$.

Suppose now we take two different counties. County C has $Z=2100$ and County D has $Z=2200$. Using C and D only,

\[ \hat{\beta}_{IV}^{2100,2200} = \frac{E(\ln Y_j | Z=2100) - E(\ln Y_j | Z=2200)}{E(S | Z=2100) - E(S | Z=2200)} = E[\beta | S(2100) = 1, S(2200) = 0] = E(\beta | 2100 < U_y < 2200). \]

This is the average return for individuals not going to school if $Z=2200$ but going to school if $Z=2100$. They are at the margin at a high level of tuition, which means that they have a high level of ability.

The general formula for any pair of counties is

\[ \hat{\beta}_{IV}^{z,z'} = \frac{E(\ln Y_j | Z=z) - E(\ln Y_j | Z=z')}{E(S | Z=z) - E(S | Z=z')} = E(\beta | S(z) = 1, S(z') = 0). \]

We can make $z$ and $z'$ close and get

\[ E(\beta | U_y = \gamma). \]

Therefore, by varying $Z$ we can trace out how $\beta$ varies with $U_y$. This is the marginal treatment effect. If the MTE is flat, then there is no heterogeneity. Notice that to trace out whole support of $U_y$ we need large support for the instrument.

To put this in practice we first aggregate multiple instruments into a single (cost) index (by modeling the probability of selection):

\[ S = I[Z \gamma = U_y > 0]. \]

By aggregating multiple instruments into a single index and then using it as the instrumental variable, I can get a larger support over which to estimate the MTE (larger in the sense of having extremes that are farther away and having no holes in the middle of the support). Then we construct $P(Z) = P(S | Z)$ and use it as the instrument:

\[ \hat{\beta}_{IV}^{P,p} = \frac{E(\ln Y_j | P=p) - E(\ln Y_j | P=p')}{E(S | P=p) - E(S | P=p')} = \frac{E(\ln Y_j | P=p) - E(\ln Y_j | P=p')}{p-p'}. \]

In the numerator we evaluate the function $E(\ln Y | P)$ in two points ($p$ and $p'$) and then we take the difference. In the denominator we have the difference in the points of evaluation. Notice that this is like a derivative. In summary, the first step is to construct $E(\ln Y | P)$. This is simply a regression of $Y$ on $P$ and can be estimated parametrically or nonparametrically. Then we take the derivative. If the derivative is flat, then selection is not important (heterogeneity plays a small role in the problem). Therefore a simple test of selection on unobservables is the following: "Is $E(\ln Y | P)$ linear in $P$?" Standard instrumental variable assumes no selection on unobservables and therefore imposes linearity in $P$.
In this framework, we can also allow for $X$ (example: test scores), in a parametric way. In this case,

$$
\beta = \alpha_1 - \alpha_9 + \alpha(X) + U_1 - U_9.
$$

A simple formal justification for this estimator of the marginal treatment effect is the following: we can transform the variables in the choice model in (4) such that it becomes

$$
S=1 \text{ if } P(Z) > V_s,
$$

where

$$
P(Z) = 1 - F_{U_s}(-\mu_s(Z)) = \Pr(S = 1| Z),
V_s = 1 - F_{U_s}(U_s).
$$

Notice that $V_s \sim \text{Unif}(0,1)$ (with uniform distribution taking values between 0 and 1). Under these conditions Heckman and Vytlacil (2000, 2003) and Carneiro, Heckman, and Vytlacil (2001) show that we can identify $E(\beta | V_s = v_s)$ by first running a (nonparametric) regression of $\ln Y$ on $P$ and then computing the derivative with respect to $P$. Notice that we can write observed outcomes as

$$
\ln Y = \ln Y_s + (1-S)\ln Y_o - \mu_s(X) + S\beta + U_o + S(U_1 - U_o),
$$

then

$$
E(\ln Y | P = p) = \mu_o + \beta p + E(U_1 - U_o | P > V_s, P = p)p = \mu_o + \beta p + \int E(U_1 - U_o | V_s = v)dv,
$$

and finally,

$$
\frac{\partial E(\ln Y | P = p)}{\partial p} = \beta + E(U_1 - U_o | V_s = p)p = E(\beta | V_s = p).
$$

A continuous instrument $Z$ generates a continuous $P$, which we need to have in order to be able to compute this derivative. $E(\ln Y | P)$ and its derivative can be estimated using standard nonparametric methods.

One reason why it is useful to study the MTE is that it is a natural way to look at heterogeneity. But there is one other important reason. Heckman and Vytlacil (1999, 2000, 2001b, 2003) and Carneiro, Heckman, and Vytlacil (2002) show how we can compute different evaluation parameters by constructing different weighted averages of the MTE. For example, to compute the average return to schooling in the population one can weight the MTE by the distribution of $(X, U_s)$ in the population. For treatment on the treated, one weights MTE by the distribution of $(X, U_s)$ for the individuals that go to college. For the policy-relevant parameter we weight MTE by the distribution of $(X, U_s)$ for those individuals induced to go to college by the policy. These weights are interesting objects in themselves because they tell us the distribution of observable and unobservable determinants of returns for different groups in the population. For example, by considering them together with the MTE we can have an idea of whether a given policy benefits mostly individuals with positive returns or mostly individuals with negative returns to college (or high vs. low returns).

To calculate these weights we need to estimate the joint distribution of $(X, Z, U_s)$. However, under assumptions (6) and (7),

$$
f_{Z,X,U_s}(x,x,u_s) = f_x(x)f_{U_s}(u_s)f_z(x \mid X = x).
$$

Both $f_x(x)$ and $f_z(x \mid X = x)$ can be directly estimated from the data. $f_{U_s}(u_s)$ can in principle be estimated nonparametrically from (4) (Matzkin 1992) although discrete choice models are commonly estimated by assuming a parametric functional form for $f_{U_s}(u_s)$ (such as a normal, a logistic, or a uniform). In this paper the empirical results I present are for a normal. Notice that even though one does not know a priori which individuals will be affected by a given policy, it is possible to estimate the distribution of $(X, U_s)$ for these individuals by making use of the schooling model. Suppose that one of the variables in $Z$ is tuition and the policy we are interested in is a tuition subsidy of $\lambda$ given for each person that decides to attend college. Furthermore, for simplicity of exposition, assume that we can write (4) as

$$
S = 1[X_{Y_x} + Z + U_s > 0]
$$

(assuming $Y_x$ for simplicity). Therefore an individual that chooses not to go to college without the subsidy may be induced to enroll in college once the subsidy is in place if

$$
X_{Y_x} + Z + U_s < 0, X_{Y_x} + (Z - \lambda) + U_s > 0
$$

or if

$$
-Z < X_{Y_x} + U_s < -Z.
$$

This condition defines the set of $(X, U_s)$ values such that individuals who have $(X, U_s)$ in this set switch from one schooling level to another in response to the policy, for a given value of $Z$. Once we estimate $f(Z, X, U_s)$, we can compute the policy weight. The class of policies that can be evaluated with this method have the following characteristics: 1) the policy cannot affect the outcome equation, either through $X$, $U_s$, and $U_o$ or through the functions $\mu_s(X)$ and $\mu_u(X)$; it can only operate through the choice equation (this rules out general equilibrium effects); 2) the policy has to operate through one of the $Z$ variables that are observed in the data; and 3) the policy cannot change $Z$ to values outside the support of the observed data, unless we can extrapolate by parameterizing in advance the relationship between $Z$, $S$, $X$, and $U_s$ outside the support of observed data (Heckman 2001). The idea behind this exercise is that variation in $Z$ in the cross section mimics variation in the policy. For example, by using LATE we can estimate the average
return to college for individuals that would go to college if they faced a net tuition of $1000 but not if they faced a net tuition of $2000:

$$
\beta_p = \frac{E(lnY | Z = 1000) - E(lnY | Z = 2000)}{E(S | Z = 1000) - E(S | Z = 2000)} \\
= E[\beta | S(1000) = 1, S(2000) = 0] \\
= E(\beta | 1000 < U < 2000).
$$

This parameter measures the effect of giving a tuition subsidy of $1000 to individuals that currently face a net tuition of $2000. Using this method we can estimate the effect of giving a subsidy of $1000 to individuals facing a tuition of $2100, $2200, $2300, etc., and therefore we can estimate the overall effect of a tuition subsidy of $1000. Using cross sectional variation in the data to mimic policy variation is the way structural models usually work. This is also the basis for the estimator proposed by Ichimura and Taber (2000, 2002), which is related to the estimator of policy effects used in this paper.

The policy just presented is very simple and consists of a change in Z that is uniform across all levels of Z (subtracting \( \lambda \) to each value of Z). Although this is the set of policies considered in this paper, it is possible to study much more general policies, provided that they are subject to the conditions specified in the previous paragraph. For example, the subsidy could be proportional (in which case the tuition faced by each individual after the policy is implemented is \( \lambda Z \), instead of \( Z - \lambda \), where \( \lambda \) is the subsidy rate). Or the subsidy could be targeted to individuals with a given level of \( X \) and Z. More general policies are considered in Iheckman and Vytlacil (2001b, 2003) and in Carneiro, Heckman, and Vytlacil (2001).

I now present estimates of the MTE and of different policy evaluation parameters across different samples, using the framework just described. The figure on this page shows an estimate of the MTE for white males in the NLSY79. Returns are high for individuals who have unobservables that make them either very likely to go to school (high \( U_s \)) or very unlikely to go to school (low \( U_s \)), leading to a U-shaped function. In a Roy model, where people select into the sector where their gain is highest, we expect returns to be higher for individuals who are more likely to go to school. This would lead to a rising MTE as a function of \( U_s \). However, there are costs of going to college, and if costs and returns of going to college are positively correlated then it is possible to have segments of the MTE where average returns increase and at the same time the likelihood of going to college decreases (leading to a declining MTE). Therefore, the U-shaped pattern for the MTE may result from the intersection of two different populations, both of them with monotone MTEs in terms of \( X \) and \( U_p \) but one in which it is increasing in these variables and the other in which it is decreasing (I illustrate this in the appendix).

Individuals in the latter population have high returns to college but still choose not to enroll. This may be because they face high psychic costs of schooling or because of credit constraints. Those in the middle of the distribution of \( U_s \) have the lowest returns.\(^\text{10}\) Returns are also the highest for those individuals with high AFQT ("A" in the figure) and high \( U_s \).\(^{\text{11}}\) In summary, heterogeneity in returns is an important feature of the data. Returns vary across individuals with different levels of observables (AFQT in this case) and unobservables (\( U_s \)). The magnitude of this heterogeneity is very large. Returns to one year of college can be as high as 40 percent and as low as 0 percent.\(^{\text{12}}\)

The table on the following page summarizes the results across datasets. It presents estimates of four parameters across samples. The first row shows the average return in the population, the second row shows the average return for those enrolled in college, the third shows the average return for those not enrolled in college, and the fourth shows the average return for the marginal individual. In four of the five datasets analyzed,\(^{\text{13}}\) the marginal person has a return below the return for the average person that went to college. The magnitude of this difference is substantial. Most policies will affect individuals at the margin, so the return for the marginal person is a more relevant parameter to evaluate policies that expand college enrollment than the return for the average person. In the dissertation I present estimates of the returns for individuals induced to enroll in college by different amounts of a tuition subsidy, using the methods of Heckman and Vytlacil (2001b). They are substantially different from the return to college for the average student.

In summary, this paper makes four main points:

1) Heterogeneity and selection are important in the returns to college. The average individual in the...
<table>
<thead>
<tr>
<th></th>
<th>NLSY79 (whites)</th>
<th>HSB (whites)</th>
<th>PSID (whites)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>( E(\beta) )</td>
<td>0.1625</td>
<td>0.2086</td>
<td>0.1702</td>
</tr>
<tr>
<td></td>
<td>(0.0319)</td>
<td>(0.0325)</td>
<td>(0.0393)</td>
</tr>
<tr>
<td>( E(\beta</td>
<td>S = 1) )</td>
<td>0.1807</td>
<td>0.2646</td>
</tr>
<tr>
<td></td>
<td>(0.0572)</td>
<td>(0.0566)</td>
<td>(0.0701)</td>
</tr>
<tr>
<td>( E(\beta</td>
<td>S = 0) )</td>
<td>0.01456</td>
<td>0.1597</td>
</tr>
<tr>
<td></td>
<td>(0.0415)</td>
<td>(0.0448)</td>
<td>(0.0908)</td>
</tr>
<tr>
<td>( E[\beta</td>
<td>\mu_s(X, Z) + U_s = 0] )</td>
<td>0.1501</td>
<td>0.2108</td>
</tr>
<tr>
<td></td>
<td>(0.0296)</td>
<td>(0.0324)</td>
<td>(0.0984)</td>
</tr>
</tbody>
</table>

NOTE: Standard errors are bootstrapped. NLSY79 = National Longitudinal Survey of Youth of 1979 dataset; HSB = High School and Beyond dataset; and PSID = Panel Study of Income Dynamics dataset.

The results of this paper rely on the validity of the instrumental variables used, after controlling for measured cognitive ability.

As in other studies of the returns to schooling that rely on the method of instrumental variables, the estimates I obtain are not very precise. It is usually found that IV estimates of the returns to education are higher than OLS estimates of the same returns but that these two parameters are not statistically significantly different from each other. Likewise, I find that matching and local instrumental variables of estimates of the return to schooling are not estimated with as much precision as in the case of the returns to schooling.

Estimates of the Return to Schooling—Local Instrumental Variables Using Local Linear Regression

1) The results of this paper rely on the validity of the instrumental variables used, after controlling for measured cognitive ability.

2) As in other studies of the returns to schooling that rely on the method of instrumental variables, the estimates I obtain are not very precise. It is usually found that IV estimates of the returns to education are higher than OLS estimates of the same returns but that these two parameters are not statistically significantly different from each other. Likewise, I find that matching and local instrumental variables of estimates of the return to schooling are not estimated with as much precision as in the case of the returns to schooling.

3) Since heterogeneity is important we need to clearly define the particular return we are interested in. The average return to schooling in the population is not the right parameter to evaluate a tuition policy what we call the policy-relevant parameter. Even though many economists focus on the former, the latter is what we want. We would also think that to evaluate the benefits of different tuition policies we would need to estimate different parameters, one for each policy. However, I show that, empirically, the relevant return to schooling required to evaluate a broad range of different policies is not very different across them, although this is not true for all policies.

For a broad range of policies, we need only one number, the return to schooling for the marginal person, even if these policies have very different effects on college enrollment. This is a surprising result. It is an empirical result about the quantitative magnitude of the effects of different policies, not a theoretical statement about evaluating policies in other fields of social policy. Theoretically, it can only happen in special cases, which seems to be empirically important. It cannot be generalized to other areas of economics and policy evaluation where heterogeneity and self-selection are thought to be important. It breaks down when we consider large scale policies. Neither OLS nor linear IV (using as the instrumental variable an index of commonly used instruments variables in the literature) estimate policy-relevant parameters. However, IV performs much better than OLS.

4) Most of the instrumental variables used in this literature are either correlated with measured cognitive ability—and therefore are not valid instruments if ability is not included in the model—or weakly correlated with schooling. When family background variables are included in the model as a proxy for ability, the correlation between the instrument and ability becomes, in some cases, smaller, but so does the correlation between the instrument and schooling. I do not elaborate on this point in this summary.

However, there are some important caveats to the results to consider:

1) The results of this paper rely on the validity of the instrumental variables used, after controlling for measured cognitive ability.

2) As in other studies of the returns to schooling that rely on the method of instrumental variables, the estimates I obtain are not very precise. It is usually found that IV estimates of the returns to education are higher than OLS estimates of the same returns but that these two parameters are not statistically significantly different from each other. Likewise, I find that matching and local instrumental variables of estimates of the return to schooling are not estimated with as much precision as in the case of the returns to schooling.

3) Individual returns are determined by both observable and unobservable components of ability (or abilities), which differ in the population. It is important to account for both. Ignoring selection on unobservables may lead to large biases. The distributions of these abilities are very different across the populations of high school graduates, college graduates, and individuals at the margin. Therefore the distributions of returns to schooling are also very different across these different groups.

4) Most of the instrumental variables used in this literature are either correlated with measured cognitive ability—and therefore are not valid instruments if ability is not included in the model—or weakly correlated with schooling. When family background variables are included in the model as a proxy for ability, the correlation between the instrument and ability becomes, in some cases, smaller, but so does the correlation between the instrument and schooling. I do not elaborate on this point in this summary.

However, there are some important caveats to the results to consider:

1) The results of this paper rely on the validity of the instrumental variables used, after controlling for measured cognitive ability.

2) As in other studies of the returns to schooling that rely on the method of instrumental variables, the estimates I obtain are not very precise. It is usually found that IV estimates of the returns to education are higher than OLS estimates of the same returns but that these two parameters are not statistically significantly different from each other. Likewise, I find that matching and local instrumental variables of estimates of the return to schooling are not estimated with as much precision as in the case of the returns to schooling.

3) Since heterogeneity is important we need to clearly define the particular return we are interested in. The average return to schooling in the population is not the right parameter to evaluate a tuition policy what we call the policy-relevant parameter. Even though many economists focus on the former, the latter is what we want. We would also think that to evaluate the benefits of different tuition policies we would need to estimate different parameters, one for each policy. However, I show that, empirically, the relevant return to schooling required to evaluate a broad range of different policies is not very different across them, although this is not true for all policies.

For a broad range of policies, we need only one number, the return to schooling for the marginal person, even if these policies have very different effects on college enrollment. This is a surprising result. It is an empirical result about the quantitative magnitude of the effects of different policies, not a theoretical statement about evaluating policies in other fields of social policy. Theoretically, it can only happen in special cases, which seems to be empirically important. It cannot be generalized to other areas of economics and policy evaluation where heterogeneity and self-selection are thought to be important. It breaks down when we consider large scale policies. Neither OLS nor linear IV (using as the instrumental variable an index of commonly used instruments variables in the literature) estimate policy-relevant parameters. However, IV performs much better than OLS.

4) Most of the instrumental variables used in this literature are either correlated with measured cognitive ability—and therefore are not valid instruments if ability is not included in the model—or weakly correlated with schooling. When family background variables are included in the model as a proxy for ability, the correlation between the instrument and ability becomes, in some cases, smaller, but so does the correlation between the instrument and schooling. I do not elaborate on this point in this summary.

However, there are some important caveats to the results to consider:

1) The results of this paper rely on the validity of the instrumental variables used, after controlling for measured cognitive ability.

2) As in other studies of the returns to schooling that rely on the method of instrumental variables, the estimates I obtain are not very precise. It is usually found that IV estimates of the returns to education are higher than OLS estimates of the same returns but that these two parameters are not statistically significantly different from each other. Likewise, I find that matching and local instrumental variables of estimates of the return to schooling are not estimated with as much precision as in the case of the returns to schooling.
statistically different. However, the magnitude of the numbers is consistently very different.

3) General equilibrium effects of policies are not considered, and for the particular case of tuition policies they are potentially very important (Heckman, Lochner, and Taber 1998).

4) There is no consideration of school quality nor of how students are allocated across schools. Tuition policies are assumed to operate on a single dimension: college enrollment. One obvious problem is that if tuition is correlated with school quality and if school quality is a determinant of the returns to schooling, then tuition may no longer be a valid instrument (Carneiro and Heckman 2002). Furthermore there is a lot of heterogeneity in types of college, degrees offered, and types of students, which may cause my estimates of the returns to college to be difficult to interpret. Related to this last point is that I aggregate individuals with different amounts of college education into a single category. However, this latter problem can be handled by an extension of the model to a multi-outcome setting, which is the subject of another paper (Carneiro and Heckman 2001).

5) Finally, preliminary analysis of blacks and Hispanics shows different results. The results of this paper are not generalizable to other racial groups. These need to be studied separately.

I conclude with two additional notes, both related to important extensions of this work that are currently in progress. The first one concerns the simulation of policies. The choice model estimated in this paper is very simple, and therefore the way the policies can operate is very limited. That may be driving part of the result that one number is all we need to evaluate a broad range of policies. More analysis is needed, and better models of schooling attainment need to be incorporated. Furthermore, local average tuition in the county of residence at 17 (the tuition variable used in this paper) may not be a good approximation to the tuition faced by each individual. The second point concerns important recent developments in the study of heterogeneity in returns. The evidence in this paper (and in many others) suggests that heterogeneity is important. Although the focus of the literature (and also of this paper) has been on the estimation of different mean returns, much more can be learned from the estimation of distributions of returns. Carneiro, Hansen, and Heckman (2001, 2003) estimate distributions of returns to schooling using a very different methodology from the one in the paper (which was developed to deal with means). They show that there is substantial dispersion in the returns to schooling and that many individuals make a mistake by going to college (they have negative returns). Moreover, there is substantial uncertainty about wages and returns at the time agents make their schooling decisions, which helps explain why there are so many mistakes ex post.

Notes

I would like to thank James Heckman for his continued support and encouragement. I also thank him for helpful discussions and detailed comments on this paper. I am grateful for comments from Alberto Abadie, Rita Almeida, Mark Duggan, Michael Greenstone, Karsten Hansen, Hidehiko Ichimura, Joao Hrtek, Tom Kane, Steve Levitt, Dimitriy Masterov, Costas Meghir, Kathleen Mullen, Salvador Navarro-Lozano, Derek Neal, Rodrigo Soares, Robert Townsend, Pedro Vicente, Edward Vytlacil and workshop participants at the University of Chicago, the 2002 Latin American and Caribbean Economic Association (LACEA) meetings, Carnegie Mellon, Ohio State, Brown and Georgetown universities, UCLA, the World Bank, Rand Corporation, and University College London. Jingjing Hsee, Dimitry Masterov, Sergio Urzua, and especially Maria Isabel Larenas and Maria Victoria Rodriguez provided excellent research assistance. Financial support from Fundacao Ciencia e Tecnologia and Fundacao Calouste Gulbenkian is gratefully acknowledged. All the references can be found in the original dissertation. I omit a reference list from this summary. E-mail: p.carneiro@ucl.ac.uk.

1. $\beta$ is the percentage increase in earnings due to an additional year of schooling.

2. In particular, suppose we have a policy that consists of a tuition subsidy we want to evaluate. The relevant return in this case is the return to schooling for those induced to go to school by the subsidy, not the average return in the population.

3. However, Vytlacil (2002) shows that the assumptions underlying the LATE parameter of Imbens and Angrist (1994) are the same assumptions of a nonparametric selection model, and therefore the two approaches are equivalent.

4. Carneiro, Heckman, and Vytlacil (2001) show how (4) can be justified by an economic model where the agent chooses the level of schooling that maximizes his present value of earnings.

5. I assumed that $Z$ was independent of $U_i$ and $U_o$ conditional on $X$ (see condition (6)).

6. Typical applications of selection models assume a parametric form for the unobservables in the choice and outcome equations (for example, multivariate normal), which in turn implies a parametric form for the MTE. In contrast I estimate the MTE nonparametrically.

7. Typical applications of instrumental variables (for example, linear instrumental variables) use global variation instead of
local variation in the instrument (such as linear regression of $Y$ on $X$ using global variation in $X$ to fit a line while a nonparametric regression of $Y$ on $X$ uses local variation in $X$ to fit a curve).

8. The unobservable does not have to be "ability." I call it ability for simplicity's sake, but $U_i$ can represent any unobservable (example: unobservable cost).

9. If individuals switch when tuition varies little at such a low range, then that means that even though they are facing tuition at a very low level before the change, these individuals still decide not to enroll in college. Therefore they are likely to have low levels of ability.

10. The same patterns are found for white females in the NLSY79 and for white males in the HSB. Even when test scores are excluded, the MTE has this U-shaped pattern. However, for the PSID the MTE is rising.

11. One possible story for a U-shaped MTE is the following: The rising section at the end is the usual story: individuals with higher ability are more likely to go to school and they have higher returns. Individuals with very low ability are very unlikely to go to school. However, because they have such low ability and possibly have learned very little during high school, teaching them some basic skills may generate very large returns, generating the declining segment in the MTE. While the latter is suggestive of a decreasing returns story (those with high levels of human capital benefit less from learning than those with low levels of human capital), the former is suggestive of a complementarity story (those with higher ability benefit more from further learning).

12. The numbers presented in the figure are gross returns. To obtain annual returns we divide them by 3.5, the difference between the average years of schooling of individuals who attended college and high school graduates.

13. In the NLSY79 we include AFQT scores in the outcome equations and we use as instruments number of siblings, parental education, tuition at local college at age 17, distance to nearest college at age 14, local unemployment rate in county of residence at age 17, and average blue collar wage in state of residence at age 17. Including family background variables in the wage equation (so that they are not instruments anymore) did not change the results significantly but decreased the precision of the estimates. For HSB we include math test scores in the wage equation and we use as instruments family background variables. For the PSID we do not use any test score and we use as instruments family background variables and average tuition in the state of residence at 17.