

EVALUATING EDUCATION POLICIES WHEN THE RATE OF
RETURN VARIES ACROSS INDIVIDUALS

PEDRO CARNEIRO*

1. INTRODUCTION

In this paper I summarize some recent developments in the literature on the econometrics of program evaluation. In particular, I consider the estimation of mean and distributional treatment effects when the impact of the program varies across individuals. I focus the exposition on the study of the returns to education (therefore, education is the program being considered). I provide examples of two sets of papers that illustrate the theoretical and practical importance of accounting for heterogeneity in program evaluation. The research summarized in this paper has general applicability. It is not limited to the study of education. Instead of education (as the “program” being considered) we could use the framework to analyze job training, unionism, migration, medical care, and other types of programs.

2. EVALUATING PROGRAMS - MEANS AND DISTRIBUTIONS

Take the standard model of potential outcomes:

$$(1) \quad \begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0 \end{aligned}$$

where Y_1 and Y_0 are potential outcomes, say wages, under two different states, which are high school and college in this paper. X is a vector of observable variables (such as experience and test scores) and U_1 and U_0 are mean zero unobservables.¹ The gain of going from high school to college is the return to college:

* The University of Chicago. Email: uctppca@ucl.ac.uk. I learned the material summarized in this paper from close interactions with James Heckman, with whom I worked on many of these topics, and whom I thank for his teachings, encouragement and support. I have also benefited heavily from working with Karsten Hansen and Edward Vytlacil on these same topics. Financial support from Fundação Ciencia e Tecnologia and Fundação Calouste Gulbenkian is gratefully acknowledged. pmcarnei@midway.uchicago.edu

¹ Separability between X and U_s is not essential to all of the methods developed in this paper. Alternatively, we could have:

$$\begin{aligned} Y_1 &= \mu_1(X, U_1) \\ Y_0 &= \mu_0(X, U_0) \end{aligned}$$

$$(2) \quad \begin{aligned} \beta &= Y_1 - Y_0 \\ &= \mu_1(X) - \mu_0(X) + U_1 - U_0 \end{aligned}$$

Returns can vary across individuals because of both X and (U_1, U_0) . The college decision equation (decision equation) is the following:

$$(3) \quad S = 1 \text{ if } \mu_S(X, Z) - U_S > 0$$

where Z are variables that influence the decision to enroll in college but not the potential outcomes:

$$(4) \quad Z \perp\!\!\!\perp U_1, U_0, U_S | X$$

(these are usually called instrumental variables; some examples in the literature on the returns to schooling are distance to college and tuition). In the empirical work reported in the next sections it is also assumed that

$$(5) \quad X \perp\!\!\!\perp U_S$$

(although this assumption is not essential for the analysis; see Heckman and Vytlačil, 2000, 2003). U_S is unobservable and mean zero. For each person, we only observe potential wages in the state that is actually chosen. For example, if a person is a college graduate, we do not observe the wages he would have earned had he not gone to college. Equation (3) is the reduced form for a general economic model of schooling (with income or utility maximizing agents). Observed wages are:

$$(6) \quad \begin{aligned} Y &= SY_1 + (1-S) Y_0 \\ &= \mu_0(X) + S[\mu_1(X) - \mu_0(X) + (U_1 - U_0)] + U_0 \\ &= \mu_0(X) + S\beta(X) + U_0 \end{aligned}$$

where $\beta(X) = \mu_1(X) - \mu_0(X) + (U_1 - U_0)$. For each person we either observe Y_1 or Y_0 , but not both, and therefore we cannot compute $\beta = Y_1 - Y_0$.

The traditional approach to the estimation of (6) assumes that $U_1 = Y_0$ and therefore, conditional on X , everybody has the same return to college (see for example, Griliches, 1977):

$$\beta(X) = \bar{\beta}(X) = \mu_1(X) - \mu_0(X)$$

In this case (6) becomes

$$Y = \mu_0(X) + S\bar{\beta}(X) + U_0$$

However, for simplicity and for easier exposition I will continue with the separable case. I refer the reader to the papers referenced throughout the discussion, in particular Heckman and Vytlačil (2000, 2003), and Carneiro, Heckman and Vytlačil (2001).

The econometric problem in this equation is that S may be correlated with U_0 , which is usually called the ability bias problem. The solution is to find an instrument Z for schooling, a variable correlated with S but not with U_0 . Then we can estimate $\bar{\beta}(X)$, which is the same for every individual (conditional on X). The distribution of returns conditional on X is degenerate at $\bar{\beta}(X)$, and once we remove the conditioning on X , it is just the distribution of a function of X . There is no heterogeneity in returns conditional on X , even though there may still be substantial heterogeneity in returns in the population because of X . Suppose that one of the variables in X is a cognitive test score, and this is an important determinant of both the returns to college (through $\mu_1(X) - \mu_0(X)$) and of the decision to attend in college (through $\mu_S(X, Z)$). Then $f[\beta(X)|S=1] \neq f[\beta(X)|S=0]$, where f is the pdf of $\beta(X)$.

The assumption that $U_1 = U_0$ is convenient but implausible (see Heckman, 2001). If $U_1 \neq U_0$ then there is another component of heterogeneity in returns, $U_1 - U_0$, which is unobservable. In this case (6) is

$$Y = \mu_0(X) + S\bar{\beta}(X) + U_0 + S(U_1 - U_0)$$

and instrumental variables no longer estimates $\bar{\beta}(X)$, unless $(U_1 - U_0) \perp\!\!\!\perp S|X$ (see Heckman and Vytlačil, 2000, 2003, and Carneiro, Heckman and Vytlačil, 2001).

Furthermore, if $(U_1 - U_0) \perp\!\!\!\perp S|X$ it is not clear that $\bar{\beta}(X)$, is a useful parameter to consider. Suppose, as most of the literature does, that we focus on mean treatment parameters, and that we want to compute the benefits of a particular policy, say, a tuition subsidy. How would we evaluate such a policy? Assume the tuition subsidy induces K individuals to enroll in college. The traditional approach to the problem assumes that the return to college, β , is the same for all individuals (at least conditional on X). Then the benefit of the policy is simply $B = \beta * K$.² However if β varies in the population (even conditional on X) we want to multiply K by the average return to college for those individuals induced to go to college by the tuition subsidy, and therefore we need to know where in the distribution of β are the new entrants coming from. Suppose we have the following version of the model of equation (3):

$$S = 1 \text{ if } \beta - Z > 0$$

where β and Z vary in the population. Let Z be the tuition faced by a given individual under no subsidy, and Z' be the tuition he faces with the subsidy in place.³ Then, for individuals induced to attend college by the subsidy, $\beta - Z \leq 0$

² Or, conditional on X , $B(X) = \beta(X) K(X)$, where $B(X)$ is the benefit of the policy for individuals with a given level of X , $\beta(X)$ is the return to college for those individuals, and $K(X)$ is the number of such individuals induced to attend college by the subsidy. Then $B = \sum_X B(X)$. I will drop X when convenient for simplicity of exposition.

³ There is a distribution of Z and Z' in the population (different individuals face different tuition levels with and without the subsidy).

(they do not attend college without the subsidy, or $S(Z) = 0$, and $\beta - Z' > 0$ (they attend college with the subsidy, or $S(Z') = 1$). Then

$$\begin{aligned} B &= E[\beta | S(Z)=0, S(Z')=1] * K \\ &= E(\beta | Z' < \beta \leq Z) \end{aligned}$$

Many of the standard parameters of the evaluation literature do not in general equal the policy relevant treatment effect (PRTE), or $E(\beta | Z' < \beta \leq Z)$ (see Heckman and Vytlačil, 2001, 2003, and Carneiro, Heckman and Vytlačil, 2001). The average treatment effect (ATE) measures the average return to college for a randomly selected person in the population ($E(\beta)$). Treatment on the treated (TT) is the average return to college for those that attend college ($E(\beta | S=1) = E(\beta / \beta > Z)$). Treatment on the untreated (TUT) measures the average return that those who do not attend college would experience had they gone to college ($E(\beta | S=0) = E(\beta / \beta \leq Z)$). The average marginal treatment effect (AMTE) is the effect of treatment for individuals at the margin between enrolling in college or not ($E(\beta | \beta = Z)$). Individuals who are at the margin will probably be more responsive to different kinds of interventions than inframarginal individuals.

Some important evaluation questions require knowledge of features of the distribution of returns beyond the mean parameters reported above (Heckman, Smith and Clemens, 1997). For example, we may want to know what is the proportion of individuals benefiting from a program (and symmetrically, what is the proportion of individuals who lose from the program), or what is the impact of the program on the distribution of income. Knowledge of distributions of returns allows us to answer a much larger range of questions. We can define parameters analogous to the mean parameters described above. $f(\beta)$ is the density of β in the population, $f(\beta | S=1)$ is the density of β for those who enrolled in college and $f(\beta | S=0)$ is for those who did not enroll in college. $f(\beta | \beta = Z)$ is the density of β for those at the margin and $f(\beta | Z' < \beta \leq Z)$ is the density of β for those induced to enroll in college by the tuition subsidy. Once we have the latter distribution it is possible to compute, for example, the proportion of individuals affected by the policy who benefit from the policy.

However, identification of distributions of returns is more difficult than identification of the mean parameters discussed above. Under assumptions (4) and (5) it is possible to identify $f(Y_1)$, $f(Y_0)$ but not $f(Y_1, Y_0)$ (see Heckman and Smith, 1993, 1998).⁴ The reason is that we never observe both Y_1 and Y_0 for the same individual. Therefore it is not possible to identify the different densities listed in the previous paragraph. However it is still possible to identify the different mean parameters described above (under some continuity and support conditions for the instrument Z ; see Heckman and Vytlačil, 2000, 2003). Identification of mean effects is achieved under weaker conditions than identification of distributions of

⁴ It is possible to construct bounds for $f(Y_1, Y_0)$, but in practical applications these bounds turned out to be very wide (see Heckman, Smith and Clemens, 1997).

effects.⁵ Assuming we have the model presented above and there exists a continuous instrument Z , it is possible to identify $E(Y_1|X, U_S)$, $E(Y_0|X, U_S)$ and therefore, $E(Y_1 - Y_0|X, U_S)$ (see, for example, Heckman and Vytlacil, 2000, 2003). Heckman and Vytlacil (2000, 2003) show how it is possible to construct parameters such as ATE, TT, TUT, AMTE and PRTE as weighted averages of $E(Y_1 - Y_0|X, U_S)$, a parameter they call the marginal treatment effect (MTE). For example:

$$E(\beta) = \int E(Y_1 - Y_0 | X=x, U_S=u_S) f(x, u_S) dx du_S$$

$$E(\beta | S=1) = \int E(Y_1 - Y_0 | X=x, U_S=u_S) f(x, u_S | S=1) dx du_S$$

$$E(\beta | S=0) = \int E(Y_1 - Y_0 | X=x, U_S=u_S) f(x, u_S | S=0) dx du_S$$

where $f(X, U_S)$ is the joint density of X and U_S . If it were possible to identify $f(Y_1 - Y_0 | U_S)$ then by the same reasoning it would be possible to construct $f(Y_1 - Y_0)$ and the remaining densities I presented above. However all we can recover from the data under the assumptions presented so far is $f(Y_1 | U_S)$ and $f(Y_0 | U_S)$ (Heckman and Smith, 1993, 1998), which is not enough for recovering $f(Y_1 - Y_0 | U_S)$ since that requires knowledge of the joint density $f(Y_1, Y_0 | U_S)$. Knowledge of this joint density is not necessary for constructing the marginal treatment effect; the marginal densities provide enough information. In section 4 I will present alternative sets of assumptions under which it is possible to recover these densities. Once we recover the joint density of potential outcomes, $f(Y_1, Y_0)$, we can compute distributions of returns to schooling, $f(\beta)$, and different distributional parameters, and we can also analyze the effects of different policies on the distribution of income.

There is a large literature I could draw on and review in this summary, but the size and scope of the paper is much smaller than a complete review of the literature. In the next two sections I choose to summarize two sets of papers where these ideas are applied in the estimation of the returns to schooling. Section 3 focuses on the work of Carneiro, Heckman and Vytlacil (2001) and Carneiro (2002), papers where different mean treatment parameters are estimated. In section 4 I discuss the work of Carneiro, Hansen and Heckman (2001, 2003) which provides estimates of distributions of returns to schooling.

3. ESTIMATING AVERAGE RETURNS TO EDUCATION

Heckman and Vytlacil (2001, 2003) show how the marginal treatment effect ($E(Y_1 - Y_0 | X, U_S)$) can be estimated using the method of local instrumental variables. Carneiro, Heckman and Vytlacil (2001) and Carneiro (2002) apply these ideas

⁵ Obviously, if we can identify distributions of effects, we can identify all the mean effects.

to the estimation of returns to college. The intuition of the method is best explained with an example. Suppose the decision model has the following reduced form:

$$(7) \quad S=1 \text{ if } -Z\gamma + U_S > 0$$

where Z (the instrumental variable) is tuition in county of residence ($\gamma > 0$) and U_S is ‘ability’ and it is unobserved.⁶

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 tuition⁷. We can estimate β by standard IV using tuition as the instrument and data from only counties A and B:

$$\begin{aligned} \hat{\beta}_{IV}^{100,200} &= \frac{E(\ln Y | Z=100) - E(\ln Y | Z=200)}{E(S | Z=100) - E(S | Z=200)} \\ &= E[\beta | S(Z=100) = 1, S(Z=200) = 0] \\ &= E(\beta | 100 * \gamma < U_S \leq 200 * \gamma) \end{aligned}$$

This is the Local Average Treatment Effect for the case where the instrument takes values $Z = \$100$ and $Z = \$200$ (Imbens and Angrist, 1994). 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 or not if they face a Z that 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_S).⁸

Suppose we use two different counties. County C has $Z = \$2100$ and county D has $Z = \$2200$. Using C and D only:

$$\begin{aligned} \hat{\beta}_{IV}^{2100,2200} &= \frac{E(\ln Y | Z=2100) - E(\ln Y | Z=2200)}{E(S | Z=2100) - E(S | Z=2200)} \\ &= E[\beta | S(Z=2100) = 1, S(Z=2200) = 0] \\ &= E(\beta | 2100 * \gamma < U_S \leq 2200 * \gamma) \end{aligned}$$

⁶ The unobservable does not have to be ‘ability’. I call it ability for simplicity but U_S can represent any unobservable (ex: unobservable cost).

⁷ Tuition can vary across counties if there are barriers to student migration. Tuition variation could also reflect quality variation which would invalidate the use of tuition as an instrumental variable but we abstract from quality considerations in this example and in the rest of the paper. For an example see Carneiro and Heckman (2002).

⁸ If individuals switch when tuition varies very little at such a low range then that means that even though they are facing tuition at a very low level these individuals still decide not to enroll in college unless tuition decreases to an even lower level. Therefore they are likely to have low levels of ability (or high levels of unobserved cost).

This is the average return for individuals who do not go to school if $Z = \$2200$ but who decide to enroll 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 expression for any pair of counties with tuition (z, z') is:

$$\begin{aligned} \hat{\beta}_{IV}^{z,z'} &= \frac{E(\ln Y|Z=z) - E(\ln Y|Z=z')}{E(S|Z=z) - E(S|Z=z')} \\ (8) \quad &= E[\beta | S(Z=z) = 1, S(Z=z') = 0] \\ &= E(\beta | z^* \gamma < U_S \leq z'^* \gamma) \end{aligned}$$

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

$$E(\beta / U_S = z\gamma)$$

By varying Z we can trace out how β varies with U_S . This is the MTE. Notice that to trace out the whole support of U_S we need large support for the instrument. In practical applications of the method we can have many instruments. We aggregate them into a single index through a regular first stage regression such as (7), where Z can be a vector of instrumental variables. It is convenient to pick the index to be the predicted probability of attending college conditional on Z . Let $U_S^* = -U_S$, $P(Z) = \Pr(S=1|Z) = F_{U_S^*}(-Z\gamma)$ and $V_S = F_{U_S^*}(U_S^*)$ where $F_{U_S^*}(\cdot)$ is the cdf of (U_S^*) .⁹ These are just monotonic transformation of variables (see Heckman and Vytlačil, 2001, 2003). The choice model becomes:

$$(9) \quad S = 1 \text{ if } P(Z) > V_S$$

and (8) becomes:

$$\begin{aligned} \hat{\beta}_{IV}^{p,p'} &= \frac{E(\ln Y|P=p) - E(\ln Y|P=p')}{E(S|P=p) - E(S|P=p')} \\ &= \frac{E(\ln Y|P=p) - E(\ln Y|P=p')}{p - p'} \end{aligned}$$

Notice that this is similar to the expression for a derivative (once you take the limit). On the numerator we have the difference of a function of P ($E(\ln Y|P)$) evaluated at two different values, p and p' . In the numerator we have the difference between p and p' . This suggests that an estimator of MTE is simply:

⁹ These are just monotonic transformation of variables (see Heckman and Vytlačil, 2001, 2003). The choice model becomes:

$$S = 1 \text{ if } P(Z) > V_S.$$

Notice that V_S needs to be read in the opposite way as U_S : the higher the U_S (the higher the unobserved ability or benefit, or the lower the unobserved cost), the lower the V_S . Individuals with low V_S are more likely to go to college than individuals with high U_S .

$$\left. \frac{\partial E(\ln Y/P)}{\partial P} \right|_{P=p} = E(Y_1 - Y_0 | V_S = p)$$

(for a formal proof see Heckman and Vytlačil, 2000, 2003). Finally, we can construct the different mean treatment parameters as weighted averages of the MTE. For example:

$$E(Y_1 - Y_0 | P' \leq V_S < P) = \int E(Y_1 - Y_0 | V_S = v) f(v/P' \leq V_S < P) dv$$

$f(v/P' \leq V_S < P)$ can be easily constructed from equation (9) since we observe $f(P, P')$ in the data and by construction V_S is a uniform random variable independent of P (see Heckman and Vytlačil, 2000, 2003; Carneiro, Heckman and Vytlačil, 2001; Carneiro, 2002).¹⁰

Adding X to the model is a simple (but important) extension. Carneiro, Heckman and Vytlačil (2001) estimate $E(Y_1 - Y_0 | X, V_S)$ using white males from NLSY79¹¹ with a high school degree ($S = 0$) or above ($S = 1$). Wages are measured in 1992. The relevant variable in X is AFQT scores, which is an important determinant of the return to college. The instrumental variables Z are number of siblings, father's education, distance to college, tuition and local labor market variables. These estimates correspond to returns to one year of college, and are obtained by dividing gross returns by 3.5, the difference in the average years of schooling for individuals that attend college and those that do not. Figure 1 shows the estimate of $E(Y_1 - Y_0 | X, V_S)$. Returns are highest for individuals with the highest level of AFQT scores and the lowest level of V_S (notice that the lowest V_S the more likely is an individual to attend college).¹² The relationship between $E(Y_1 - Y_0 | X, V_S)$ and V_S is not monotonic. Returns are decreasing with V_S for low values of this variable indicating that individuals who are more likely to attend college have higher returns. However, returns are increasing with V_S for high values of V_S . Some individuals have high returns and still do not enroll in college. This may happen if they face high psychic costs of schooling or if they are credit constrained. A simple formal test for selection rejects the null hypothesis that $E(Y_1 - Y_0 | X, V_S)$ does not vary with V_S (i.e., that selection on unobservables is not important). Table 1 shows estimates of ATE, TT, TUT, AMTE and PRTE.¹³ The simulated policy is a \$1000 tuition subsidy. The table also present OLS and linear IV estimates of the returns to college, where the instrument is P . Neither ATE, TT or TUT correspond to PRTE. AMTE is very similar to PRTE. OLS and linear IV do not estimate any of the above

¹⁰ By assumption, P, P' are independent of V_S .

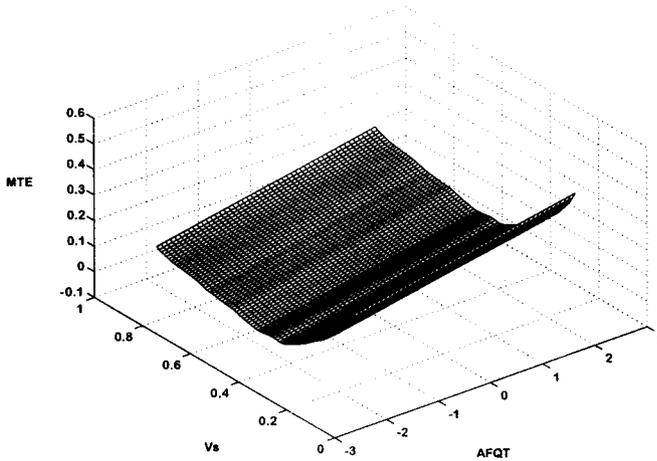
¹¹ NLSY79 is the 1979 National Longitudinal Survey of Youth (see BLS, 2001).

¹² Recall that $V_S = F_{U'_S}(U'_S)$, where $U'_S = -U_S$. Therefore, V_S is negatively correlated with U_S (if U_S is ability, then a if an individual has a high level of V_S then that means that he has low ability). This transformation is only done for notational simplicity.

¹³ Ichimura and Taber (2001) suggest an alternative estimator of policy effects, which they apply in the estimation of the effects of tuition subsidies.

parameters, and in particular, they do not estimate PRTE. AMTE (the marginal person) is below TT (the average person in college), which means that as we expand education the quality of the marginal entrant declines.

FIGURE 1
ESTIMATE OF THE MARGINAL TREATMENT EFFECT



The estimated function in this figure comes from a regression of log wages on experience, experience squared, corrected AFQT, $P \cdot AFQT$ (where P , or P^* in the table, is the predicted probability of attending college), and $K(P)$, a nonparametric function of P . $K(P)$ is estimated by local linear regression. Let γ be the coefficient on $P \cdot AFQT$. Then the function graphed above is $E(Y_1 - Y_0 | X, V_S) = \gamma \cdot AFQT + E(U_1 - U_0 | V_S)$ (in this case the only relevant X variable is AFQT), where $E(U_1 - U_0 | V_S)$ is the first derivative of $K(P)$ with respect to P . $E(Y_1 - Y_0 | X, V_S)$ is divided by 3.5 to account for the fact that individuals that attend college have on average 3.5 more years of schooling than those who do not (therefore these correspond to estimates of returns to one year of college).
Source: Carneiro, Heckman and Vytlačil (2001)

TABLE 1
RETURNS TO COLLEGE

$E(\beta)$	0.1625 (0.0319)
$E(\beta S = 1)$	0.1807 (0.0572)
$E(\beta S = 0)$	0.1456 (0.0415)
$E(\beta V_S = P)$	0.1502 0.0345
$E(\beta P' \leq V_S < P)$	0.1505 (0.0289)
OLS	0.0760 (0.0068)
IV	0.1256 (0.0227)

The OLS and IV estimate are evaluated at the average level of AFQT for individuals induced to enroll in college by a \$1000 tuition subsidy. Therefore these figures are directly comparable with the policy relevant treatment effect.

Source: Carneiro, Heckman and Vytlačil (2001).

4. ESTIMATING DISTRIBUTIONS OF RETURNS TO EDUCATION

The estimation of distributions of returns is more complex. Heckman and Smith (1993, 1998) and Heckman, Smith and Clemens (1997) show how under different assumptions it is possible to estimate $f(Y_1, Y_0)$ from knowledge of $f(Y_1)$ and $f(Y_0)$. Assuming absolutely continuous and strictly increasing marginal distributions, they postulate that quantiles are perfectly ranked so $Y_1 = F_{1,x}^{-1}(F_{0,x}(Y_0))$ where $F_{1,x} = F_1(y_1|X)$ and $F_{0,x} = F_0(y_0|X)$. An alternative assumption is that people are perfectly inversely ranked so the best in one distribution is the worst in the other: $Y_1 = F_{1,x}^{-1}(1 - F_{0,x}(Y_0))$. More generally, one could associate quantiles across distributions more freely. Heckman, Smith and Clements (1997) use Markov transition kernels which stochastically map quantiles of one distribution into quantiles of another. They define a pair of Markov kernels $M(y_1, y_0|X)$ and $\tilde{M}(y_0, y_1|X)$ such that

$$F_1(y_1|X) = \int M(y_1, y_0|X) dF_0(y_0|X)$$

$$F_0(y_0|X) = \int \tilde{M}(y_0, y_1|X) dF_1(y_1|X)$$

Allowing these operators to be degenerate produces a variety of deterministic transformations, including the two previously presented, as special cases of a general mapping. Different (M, \tilde{M}) pairs produce different joint distributions. These stochastic or deterministic transformations supply the missing information needed to construct the joint distributions.

A perfect ranking (or perfect inverse ranking) assumption is convenient, but it imposes a strong and arbitrary dependence across distributions. The empirical analysis in Carneiro, Hansen and Heckman (2003), summarized in this paper, shows that this assumption is at odds with data on the returns to education. They implement an alternative approach. Aakvik, Heckman and Vytlačil (1999, 2003) build on Heckman (1990) and Heckman and Smith (1998) by postulating a factor structure connecting (U_0, U_1, U_s) . The work of Carneiro, Hansen and Heckman (2003) builds on the analysis of Aakvik, Heckman and Vytlačil (1999, 2003) so I describe its essential idea. Suppose that the unobservables follow a factor structure:

$$U_0 = \alpha_0 \theta + \epsilon_0, U_1 = \alpha_1 \theta + \epsilon_1, U_s = \alpha_s \theta + \epsilon_s$$

where θ is independent of $(\epsilon_0, \epsilon_1, \epsilon_s)$ and the ϵ 's are mutually independent. In their setup, θ is a scalar θ can be an unobservable trait like ability or motivation that affects all outcomes. Because the factor loadings, $\alpha_0, \alpha_1, \alpha_s$, may be different, the factors may affect outcomes and choices differently. Recall that one can identify $F(U_0, U_s)$ and $F(U_1, U_s)$ under the conditions specified in Heckman and Smith (1998). Thus, one can identify $COV(U_0, U_s) = \alpha_0 \alpha_s \sigma_\theta^2$ and $COV(U_1, U_s) = \alpha_1 \alpha_s \sigma_\theta^2$ assuming finite variances and assuming $E(\theta) = 0, E(\theta^2) = \sigma_\theta^2$. With some normalizations (e.g., $\sigma_\theta^2 = 1, \alpha_s = 1$) and conditions

specified in Carneiro, Hansen and Heckman (2003) we can nonparametrically identify the distribution of θ and the distributions of $\varepsilon_0, \varepsilon_1, \varepsilon_S$ (the last up to scale). With the $\alpha_0, \alpha_1, \alpha_S$, and the distributions of $\theta, \varepsilon_0, \varepsilon_1, \varepsilon_S$ in hand, we can construct the joint distribution $f(Y_0, Y_1 | X)$.¹⁴ Carneiro, Hansen and Heckman (2003) build on this basic idea and extend it to a more general setting. They consider a model with multiple factors, multiple treatments and multiple time periods. Outcome measures may be discrete or continuous. They follow the psychometric literature by adjoining measurement equations to outcome equations to pin down the distribution of θ . In particular, they use cognitive test scores as additional measures in the system. With this framework they can estimate all pairwise treatment effects in a multiple outcome setting. They also consider the benefits for identification of having access to imperfect measurements on vector θ which are observed for all persons independent of their treatment status. Their model integrates the LISREL framework of Jöreskog (1977) into a model of discrete choice and a model of multiple treatment effects.

Carneiro, Hansen and Heckman (2003) estimate the distribution of returns to college using white males in NLSY79. They compute present values of earnings for high school and college graduates, which they use as outcome measures. The return to college for an individual is the gain in the present value of earnings from moving him from high school to college. Figure 2 presents the estimated density of the returns to college for high school and college graduates. College graduates have higher returns to college than high school graduates (their density of returns is to the right of the density of returns for high school graduates). However, notice that a substantial portion of college graduates (7%) have negative returns to college (14% of high school graduates have negative returns to college). Conversely, there is a large overlap in the distributions of returns for high school and college graduates so that a large number of high school graduates would have benefited substantially from going to college if they had gone to college (instead of just completing high school). In the paper there is a discussion of the reasons for so many apparent “mistakes”, and two important lessons emerge. The first one is that there is considerable uncertainty in the returns to college at the time the college decision is made (*ex-ante* uncertainty). Therefore, the apparent “mistakes” just reflect *ex-post* realizations of this uncertainty.¹⁵ In other words, people make mistakes because they do not know the future with certainty and some of them get a wage draw *ex-post* that is worse than the wage draw they would have had if they had chosen another schooling level. However, the second lesson is that the most important factor driving the college decision is utility or psychic cost of schooling.¹⁶ Enrollment in college has relatively little to do with *ex-ante*

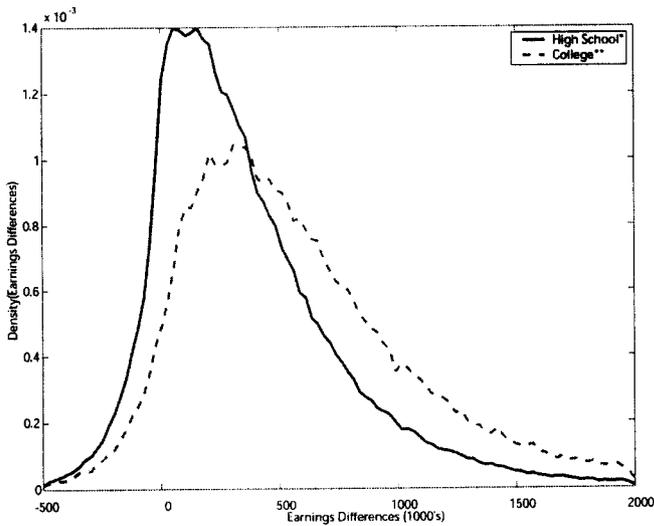
¹⁴ Aakvik, Heckman and Vytlacil (1999) present other sets of identifying assumptions.

¹⁵ Carneiro, Hansen and Heckman (2003) estimate the amount of uncertainty in the returns to college an individual faces at the time the college decision is made.

¹⁶ It can also be some other unobserved cost not captured by the observable variables in our dataset.

computations of the returns to college and of the uncertainty in the returns to college. Table 2 shows the probability of being in decile *i* of the college potential discounted earnings distribution conditional on being in decile *j* of the high school potential earnings distribution. It shows that neither an independence assumption across counterfactual outcomes, which is the Veil of Ignorance assumption used in applied welfare theory, (see, e.g., Sen, 1973) or in aggregate income inequality decompositions (DiNardo, Fortin, and Lemieux, 1996) nor a perfect ranking assumption, which are sometimes used to construct counterfactual joint distributions of outcomes (see e.g. Heckman, Smith, and Clements, 1997 or Athey and Imbens, 2002), are satisfied in the data. There is a strong positive dependence between potential outcomes in each counterfactual state, but it is not perfect dependence. The number of nonzero elements outside the diagonal of Table 2 is substantial. Any analysis of distributional effects of social programs needs to take this dependence into account.

FIGURE 2
 DENSITY OF GROSS LIFETIME EARNINGS DIFFERENCES
 (College-High School)



* $E(PV_c - PV_h | \text{Choice} = \text{High School})$

** $E(PV_c - PV_h | \text{Choice} = \text{College})$

$$PV_h = \sum_a \frac{1}{(1+0.03)^a} Y_{h,a}$$

PV denotes present value, c is college and h is high school, a is age and Y is earnings.

Source: Cancian, Hansen and Heckman (2003)

TABLE 2

$\Pr(d_i < PV_c \leq d_{i+1} | d_j < PV_h \leq d_{j+1})$ where d_i is the i^{th} decile of the college earnings distribution and d_j is the j^{th} decile of the high school earnings distribution*

High school deciles	College deciles									
	1	2	3	4	5	6	7	8	9	10
1	0.7436	0.1936	0.0459	0.0121	0.0035	0.0009	0.0003	0.0001	0.0000	0.0000
2	0.1846	0.3799	0.2372	0.1173	0.0503	0.0206	0.0072	0.0022	0.0006	0.0001
3	0.0482	0.2219	0.2640	0.2078	0.1337	0.0727	0.0344	0.0131	0.0036	0.0005
4	0.0154	0.1108	0.1944	0.2172	0.1902	0.1371	0.0806	0.0389	0.0134	0.0021
5	0.0055	0.0535	0.1240	0.1781	0.1986	0.1807	0.1372	0.0819	0.0341	0.0065
6	0.0019	0.0253	0.0732	0.1274	0.1706	0.1917	0.1826	0.1359	0.0740	0.0175
7	0.0006	0.0103	0.0382	0.0788	0.1271	0.1728	0.2011	0.1926	0.1357	0.0427
8	0.0001	0.0038	0.0171	0.0422	0.0802	0.1300	0.1816	0.2257	0.2173	0.1020
9	0.0000	0.0008	0.0053	0.0165	0.0379	0.0740	0.1288	0.2082	0.2919	0.2365
10	0.0000	0.0000	0.0006	0.0026	0.0079	0.0194	0.0465	0.1015	0.2294	0.5921

* Thus the number in row j column i is the probability that a person with a potential high school earnings in the j^{th} decile of the high school earnings distribution has potential college earnings in the i^{th} decile of the college earnings distribution.

Source: Carneiro, Hansen and Heckman (2003).

5. CONCLUSION

In this paper I summarize two sets of papers that present methods for estimating the effects of different economic activities when these effects vary across individuals. Even though this is far from a complete review of the literature on evaluation, it illustrates the theoretical and empirical importance of accounting for heterogeneity in the evaluation of different policies. The papers summarized here provide simple examples where heterogeneity is important and of how we can account for heterogeneity in the evaluation of different policies.

REFERENCES

- Aakvik, A., J. Heckman and E. Vytlačil (1999), "Training Effects on Employment when the Training Effects are Heterogeneous: An Application to Norwegian Vocational Rehabilitation Programs", manuscript, University of Chicago.
- Aakvik, A., J. Heckman and E. Vytlačil (2003), "Treatment Effects For Discrete Outcomes when Responses To Treatment Vary Among Observationally Identical Persons: An Application to Norwegian Vocational Rehabilitation Programs", NBER Working Paper N° T0262, forthcoming in *Journal of Econometrics*.
- Athey, S. and G. Imbens (2002), "Identification and Inference in Nonlinear Difference-In-Differences Models", NBER Technical Working Paper T0280.
- Bureau of Labor Statistics (2001), *NLS Handbook 2001*. Washington, D.C.: U.S. Department of Labor.
- Carneiro, P. (2002), "Heterogeneity in the Returns to Schooling: Implications for Policy Evaluation", Ph.D. dissertation, University of Chicago.

- Carneiro, P., K. Hansen, and J. Heckman (2001), "Removing the Veil of Ignorance in Assessing The Distributional Impacts of Social Policies", *Swedish Economic Policy Review*, 8, 273-201.
- Carneiro, P., K. Hansen, and J. Heckman (2003), "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice", *International Economic Review*, 44(2), 361-422.
- Carneiro, P. and J. Heckman (2002), "The Evidence on Credit Constraints in Post Secondary Schooling", *Economic Journal*, 112, 705-734.
- Carneiro, P., J. Heckman and E. Vytlačil (2001), "Understanding What Instrumental Variables Estimate: Estimating the Marginal and the Average Returns to Education", working paper, University of Chicago.
- Di Nardo, J., N. M. Fortin and T. Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach", *Econometrica*, 64, 1001-1044.
- Griliches, Z. (1977), "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica*, 45(1), 1-22.
- Heckman, J. (1990), "Varieties of Selection Bias". *American Economic Review* 80(2), 313-18.
- Heckman, J. (2001), "Micro Data, Heterogeneity and the Evaluation of Public Policy: Nobel Lecture". *Journal of Political Economy* 109(4): 673-748.
- Heckman, J. and J. Smith (1993), "Assessing the Case for Randomized Evaluation of Social Programs", in *Measuring Labour Market Measures: Evaluating The Effects of Active Labour Market Policy Initiatives*, K. Jensen and P.K. Madsen, eds., (Copenhagen: Ministry Labour).
- Heckman, J. and J. Smith (1998), "Evaluating the Welfare State", in S. Strom, ed., *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, Econometric Society Monograph Series, (Cambridge: Cambridge University Press).
- Heckman, J., J. Smith, and N. Clements (1997), "Making the Most out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts", *Review of Economic Studies* 64, 487-535.
- Heckman, J. and E. Vytlačil (2000), "Local Instrumental Variables," in C. Hsiao, K. Morimune, and J. Powells, (eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, (Cambridge: Cambridge University Press).
- Heckman, J. and E. Vytlačil (2003), "Structural Equations, Treatment Effects and Econometric Policy Evaluation", forthcoming in *Econometrica*.
- Ichimura, H. and C. Taber (2001), "Direct Estimation of Policy Effects".
- Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467-475.
- Jöreskog, K. (1977), "Structural Equations Models In The Social Sciences: Specification, Estimation and Testing." In *Applications of Statistics*, edited by P.R. Krishnaiah. (Amsterdam: North Holland).
- Sen, Amartya Kumar (1973), *On Economic Inequality*, Oxford, Clarendon Press.