
Four Parameters of Interest in the Evaluation of Social Programs

Author(s): James Heckman, Justin L. Tobias and Edward Vytlačil

Source: *Southern Economic Journal*, Vol. 68, No. 2 (Oct., 2001), pp. 210-223

Published by: Southern Economic Association

Stable URL: <https://www.jstor.org/stable/1061591>

Accessed: 26-11-2019 13:32 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Southern Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *Southern Economic Journal*

Distinguished Guest Lecture

Four Parameters of Interest in the Evaluation of Social Programs

James Heckman,* Justin L. Tobias,† and Edward Vytlačil‡

This paper reviews four treatment parameters that have become commonly used in the program evaluation literature: the average treatment effect, the effect of treatment on the treated, the local average treatment effect, and the marginal treatment effect. We derive simply computed closed-form expressions for these treatment parameters in a latent variable framework with Gaussian error terms. These parameters can be estimated using nothing more than output from a standard two-step procedure. We also briefly describe recent work that seeks to go beyond mean effects and estimate the distributions associated with various outcome gains. The techniques presented in the paper are applied to estimate the return to some form of college education for various populations using data from the National Longitudinal Survey of Youth.

1. Introduction

The problem of evaluating the effectiveness of a social program or a “treatment” is a central problem in social science and medicine. The problem of selection bias could arise in any evaluation. Individuals observed participating in a program or receiving treatment often possess different characteristics than an average person. Evaluating the economic return to a program requires accounting for the nonrandom assignment of individuals into the treated and untreated states.

One popular approach for dealing with selection bias, introduced in Gronau (1974) and Heckman (1974, 1976), is to specify a latent index model that relates the rule for assigning individuals from treatment to the potential treatment outcomes. The latent index has the interpretation of the expected net utility derived from receiving treatment; individuals participate in a program if net utility is positive (or nonnegative) and do not participate if net utility is negative. This approach is based on assumptions about error distributions and allows for de-

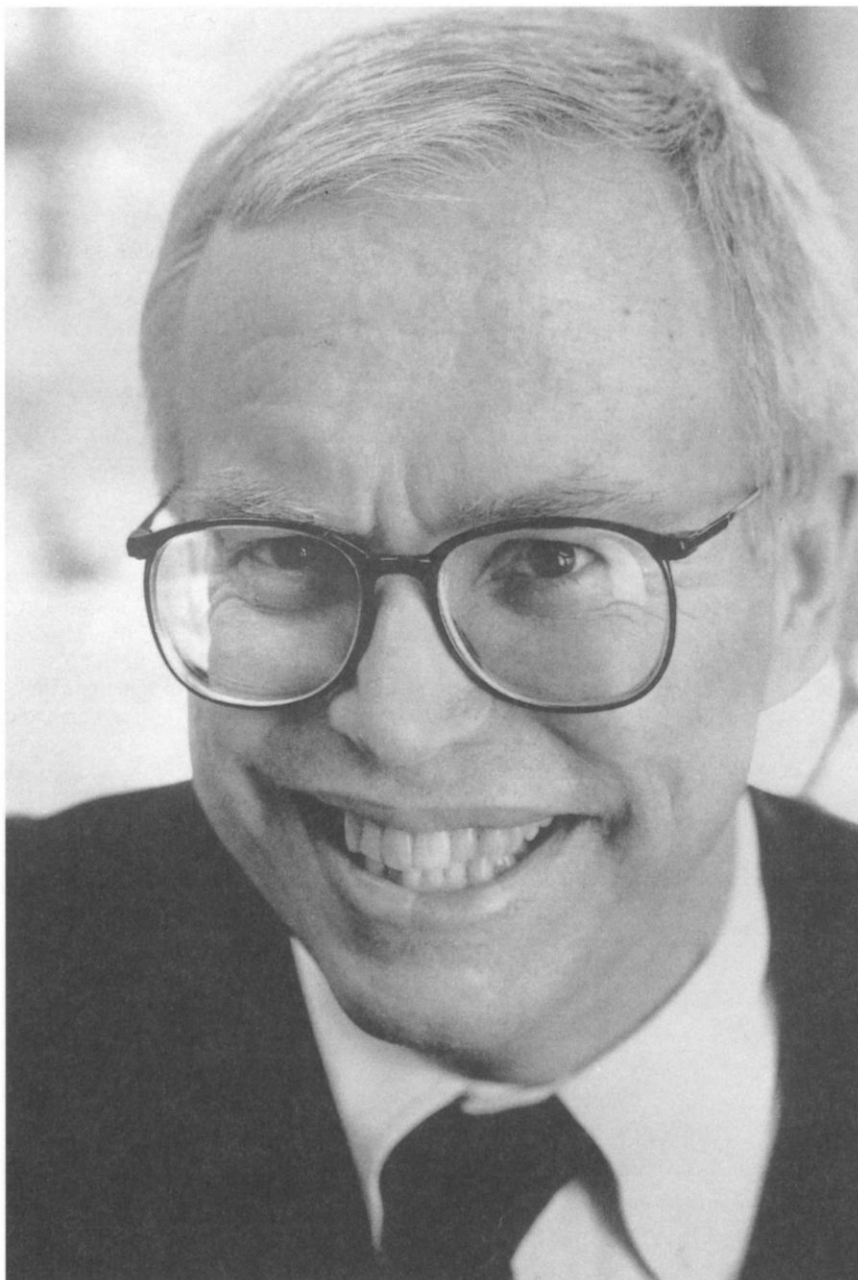
* Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, USA; E-mail jheckman@midway.uchicago.edu; corresponding author.

† Department of Economics, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA; E-mail jtobias@uci.edu.

‡ Department of Economics, Stanford University, Stanford, CA 94305-6072, USA; E-mail vytlacil@stanford.edu.

This research was supported by NSF 97-09-873 and NIH ROI-HD34958-01.

Professor James Heckman presented the Distinguished Guest Lecture at the 2000 Annual Meeting of the Southern Economic Association in Washington, DC. He is the Henry Schultz Distinguished Service Professor of Economics and Director of the Center for Social Program Evaluation at the Harris School of Public Policy at the University of Chicago. He is also a Senior Research Fellow at the American Bar Foundation. He is a fellow of the Econometric Society and the American Statistical Association, and a member of the American Academy of Arts and Sciences and the National Academy of Sciences. He received the John Bates Clark Award in 1983. He shared the 2000 Nobel Memorial Prize in Economic Sciences with Daniel McFadden.



James Heckman, University of Chicago, 2000 Distinguished Guest Lecturer.
Photograph by Matthew Gilson.

pendence between the errors in outcome and choice equations. Although computationally convenient, this approach has been criticized for its reliance on distributional assumptions and lack of robustness to departures from normality (Goldberger 1983; Paarsch 1984; and later work by Glynn, Laird, and Rubin 1986), although the empirical relevance of this criticism is far from clear (Heckman 2001).

In response to these criticisms, recent analysts have adopted a more robust approach and have attempted to identify and estimate various treatment parameters without imposing strong distributional assumptions (see, for example, the local average treatment effect [LATE] analysis of Imbens and Angrist 1994). Although these methods are free of parametric distributional assumptions, they typically estimate only one treatment parameter and are quite limited in the range of policy questions they can answer (Heckman and Vytlačil 2000a). Further, the assumptions imposed in LATE analysis are actually equivalent to those required to specify a nonparametric selection model (Vytlačil 2002).

This paper uses a latent variable framework to unite the recent treatment effect literature with the classical selection bias literature. We obtain simple closed-form expressions for four treatment parameters of interest: the average treatment effect (ATE), the effect of treatment on the treated (TT), LATE (Imbens and Angrist 1994), and the marginal treatment effect (MTE) (Björklund and Moffitt 1987; Heckman 1997; Heckman and Vytlačil 1999, 2000a, b) for the “textbook” Gaussian selection model. Our impression is that despite recent advances in nonparametric and semiparametric estimation of these parameters, many practitioners will continue to use the two-step estimator of Heckman (1976) when confronted with selection bias, and thus it is beneficial to clearly describe simple methods for estimating these parameters in the textbook selection model. For others, these expressions may be used as a starting point to illustrate the empirical importance of selection bias. Throughout this paper, we review other recent work that has relaxed the distributional requirements of this textbook model.

In addition to presenting mean effects, we also discuss how one might approach estimation of the distributions associated with these parameters of interest. The extension to the distributions of outcome gains is not immediate, nor without difficulty, since the distributions of interest depend on the unidentified cross-potential outcome correlation parameter. We briefly mention several approaches for estimating these distributions, and provide the reader with references for further information on this topic.

The plan of this paper is as follows. In the next section, we present a general model of potential outcomes, and define and interpret the various treatment parameters within it. In section 3, expressions for these parameters are derived under the assumption of trivariate normality. In section 4, we briefly discuss how one might approach estimation of the distributions associated with various outcome gains, and thus extend the analysis of mean effects. Section 5 applies the mean effect analysis to estimate various average gains in postschooling earnings resulting from the receipt of some form of college education. Using data from the National Longitudinal Survey of Youth (NLSY) we present point estimates of ATE, TT, LATE and MTE. The paper concludes with a summary in section 6.

2. Treatment Parameters in a Canonical Model

Consider a model of potential outcomes:

$$Y^1 = X\beta^1 + U^1, \quad Y^0 = X\beta^0 + U^0, \quad D^* = Z\theta + U^D. \quad (1)$$

The first two equations denote outcome equations in two possible “states” or “sectors” (college or noncollege in the application of section 5). Without loss of generality, we assume that the first state indexed by the “1” superscript represents the treated state and the “0” superscript denotes the untreated state. Each agent is observed in only one state, so that either Y^1 or Y^0 is observed for each person, but the pair (Y^1, Y^0) is never observed for any given person. What we would like to recover is information about various expected gains from the receipt of treatment, where the gain is denoted by $\Delta \equiv Y^1 - Y^0$.

Let $D(Z)$ denote the observed treatment decision, where $D(Z) = 1$ denotes receipt of treatment and $D(Z) = 0$ denotes nonreceipt. The variable D^* is a latent variable that generates $D(Z)$ according to a threshold crossing rule,

$$D(Z) = 1[D^*(Z) \geq 0] = 1[Z\theta + U^D \geq 0] \quad (2)$$

where $1[A]$ is the indicator function that takes the value 1 if the event A is true and the value 0 otherwise. In an extension of the Roy (1951) model, $D^* = Y^1 - Y^0 - C$, where C represents the cost of participating in the treated state, so that agents choose to receive treatment if the gain from participating in the program minus costs is nonnegative. We also define the following counterfactual choice variables. For any z that is a potential realization of Z , we define the variable $D(z) = 1[z\theta \geq U^D]$. $D(z)$ indicates whether or not the individual would have received treatment had her value of Z been externally set to z , holding her unobserved U^D constant. We require an exclusion restriction and denote by Z_k some element of Z that is not contained in X . By varying Z_k , we can manipulate an individual’s probability of receiving treatment without affecting the potential outcomes. Finally, we assume (U^D, U^1, U^0) is independent of X and Z .

Letting Y denote observed earnings,

$$Y = DY^1 + (1 - D)Y^0. \quad (3)$$

This model has been called the switching regression model of Quandt (1972), Rubin’s model (Rubin 1978), or the Roy model of income distribution (Roy 1951; Heckman and Honoré 1990).¹ To illustrate how a model of this type can be applied to evaluate an interesting policy question, consider the problem of estimating the return to a college education. In this case, Y represents log earnings, Y^1 denotes the log earnings of college graduates, and Y^0 denotes the log earnings of those not selecting into higher education. The latent index maps people into either the “college” (or treated) state and the “no-college” (or untreated) state. To estimate the return to college, we might estimate the expected college log wage premium for given characteristics X , (i.e., $E[Y^1 - Y^0|X]$).² In general, given the model described by Equations 1 and 2, we would like to have methods for estimating various average gains to program participation. In this paper, we examine four such treatment parameters, which measure possibly different average gains to the receipt of treatment. These four parameters are ATE, TT, LATE, and MTE.³

ATE is defined as the expected gain from participating in the program for a randomly chosen individual. As before, we let $\Delta \equiv Y^1 - Y^0$ denote the gain from program participation, and note that the average treatment effect conditional on $X = x$ can be expressed as:

¹ Amemiya (1985) has classified models of this type as generalized tobit models, and refers to the model in Equation 1 as the type 5 tobit model.

² Other applications of this model include Lee (1978) and Willis and Rosen (1979).

³ For a more general discussion of the parameters and the relation among them, see Heckman and Vytlačil (1999, 2000a, b).

$$\text{ATE}(x) = E(\Delta | X = x) = x(\beta^1 - \beta^0). \quad (4)$$

The average treatment effect evaluated at the random variable X is $\text{ATE}(X)$, which defines the treatment parameter as a function of the characteristics X . We can obtain unconditional estimates by integrating Equation 4 over the distribution of X ,

$$\text{ATE} = E(\Delta) = \int \text{ATE}(X) dF(X) \approx \frac{1}{n} \sum_{i=1}^n \text{ATE}(x_i) = \bar{x}(\beta^1 - \beta^0), \quad (5)$$

where n is sample size.

A conceptually different parameter is TT. This is the average gain from treatment for those that actually select into the treatment:

$$\begin{aligned} \text{TT}(x, z, D[z] = 1) &= E(\Delta | X = x, Z = z, D[z] = 1) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D \geq -z\theta, X = x, Z = z) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D \geq -z\theta), \end{aligned} \quad (6)$$

where the third equality follows from the assumption that (U^D, U^1, U^0) is independent of (X, Z) . The value of the TT parameter evaluated at the random variables (X, Z) is $\text{TT}(X, Z, D[Z] = 1)$. As with ATE, we can obtain an unconditional estimate by integrating over the joint distribution of X and Z for those who actually receive treatment. Letting n_t be the number of observations with $D_i = 1$, TT can be approximated as follows:

$$\begin{aligned} \text{TT} &= E(\Delta | D[Z] = 1) = \int \text{TT}(X, Z, D[Z] = 1) dF(X, Z | D[Z] = 1) \\ &\approx \frac{1}{n_t} \sum_{i=1}^n D_i \text{TT}(x_i, z_i, D[z_i] = 1). \end{aligned} \quad (7)$$

The LATE of Imbens and Angrist (1994) estimates an average gain to program participation without explicitly specifying a latent variable framework or imposing a distributional assumption.⁴ LATE is defined as the expected outcome gain for those induced to receive treatment through a change in the instrument from $Z_k = z_k$ to $Z_k = z'_k$. The variable Z_k is assumed to affect the treatment decision (is contained in Z in Equation 1), but not to affect the outcomes Y^1 and Y^0 . Below and throughout this paper, we define the LATE parameter as a change in the index from $Z\theta = z\theta$ to $Z\theta = z'\theta$, where $z'\theta > z\theta$ and z and z' are identical except for their k th coordinate. Because of the latent index structure in Equations 1 and 2, we could equivalently define the treatment parameters in terms of the propensity score, $P(Z) = \Pr(D = 1 | Z) = 1 - F_{U^D}(-Z\theta)$, where F_S denotes the cdf of the random variable S . The LATE parameter is defined as follows:

$$\begin{aligned} \text{LATE}(D[z] = 0, D[z'] = 1, X = x) &= E(\Delta | D[z] = 0, D[z'] = 1, X = x) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | -z'\theta \leq U^D \leq -z\theta, X = x) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | -z'\theta \leq U^D \leq -z\theta), \end{aligned} \quad (8)$$

⁴ The implications of the assumptions imposed in Imbens and Angrist (1994) that permit estimation of the LATE parameter have been examined by Vytlačil (2002). Vytlačil shows that the independence and monotonicity assumptions used by Angrist and Imbens imply a latent variable specification without parametric restrictions.

where the third equality follows from the assumption that $(U^D \ U^1 \ U^0)$ is independent of (X, Z) . There are two ways to define the unconditional version of LATE. First, consider

$$\begin{aligned} E(\Delta | D[z] = 0, D[z'] = 1) &= \int \text{LATE}(D[z] = 0, D[z'] = 1, X) \, dF(X) \\ &\approx \frac{1}{n} \sum_{i=1}^n \text{LATE}(D[z] = 0, D[z'] = 1, x_i). \end{aligned} \quad (9)$$

The parameter $E(\Delta | D[z] = 0, D[z'] = 1)$ corresponds to the treatment effect for individuals who would not select into treatment if their vector Z was set to z but would select into treatment if Z was set to z' . An alternative definition of the unconditional version of LATE is as follows. Let $Z^0(Z)$ equal Z but with the k th element replaced by z_k . Let $Z^1(Z)$ equal Z but with the k th element replaced by z'_k . In this notation the second definition of the unconditional version of LATE,

$$\begin{aligned} E(\Delta | D[Z^0\{Z\}] = 0, D[Z^1\{Z\}] = 1) \\ &= \int \text{LATE}(D[Z^0\{Z\}] = 0, D[Z^1\{Z\}] = 1, X) \, dF(X, Z) \\ &\approx \frac{1}{n} \sum_{i=1}^n \text{LATE}(D[Z^0\{z_i\}] = 0, D[Z^1\{z_i\}] = 1, x_i). \end{aligned} \quad (10)$$

This parameter corresponds to the treatment effect for individuals who would not select into treatment if the k th component of the Z vector is set to z_k (all other components of Z unchanged), but would select into treatment if the k th component of the Z vector is set to z'_k (all other components of Z unchanged).

Finally, the MTE parameter (Björklund and Moffitt 1987; Heckman 1997; Heckman and Smith 1998; Heckman and Vytlačil 1999, 2000a, b) is the treatment effect for individuals with a given value of U^D ,

$$\begin{aligned} \text{MTE}(x, u^D) &= E(\Delta | X = x, U^D = u^D) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D = u^D, X = x) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D = u^D), \end{aligned} \quad (11)$$

where the third equality again follows from the assumption that $(U^D \ U^1 \ U^0)$ is independent of X . Evaluation of the MTE parameter at low values of u^D averages the outcome gain for those with unobservables making them least likely to participate, whereas evaluation of the MTE parameter at high values of u^D is the gain for those individuals with unobservables that make them most likely to participate. Because X is independent of U^D , the MTE parameter unconditional on observed covariates can be written as

$$\begin{aligned} \text{MTE}(u^D) &= \int \text{MTE}(X, u^D) \, dF(X) \approx \frac{1}{n} \sum_{i=1}^n \text{MTE}(x_i, u^D) \\ &= \bar{x}(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D = u^D). \end{aligned}$$

The MTE parameter can also be expressed as the limit form of the LATE parameter,

$$\begin{aligned}
& \lim_{z\theta \rightarrow z'\theta} \text{LATE}(x, D[z] = 0, D[z'] = 1) \\
&= x(\beta^1 - \beta^0) + \lim_{z\theta \rightarrow z'\theta} E(U^1 - U^0 | -z'\theta \leq U^D \leq -z\theta, X = x) \\
&= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D = -z'\theta) = \text{MTE}(x, -z'\theta).
\end{aligned}$$

Thus the MTE parameter measures the average gain in outcomes for those individuals who are just indifferent to the receipt of treatment when the $z\theta$ index is fixed at the value $-u^D$.

The four parameters above define different average gains to program participation if U^D is not (mean) independent of $U^1 - U^0$, but the four parameters are identical if U^D is mean independent of $U^1 - U^0$ conditional on $X = x$. In this paper, we derive closed-form solutions and simple estimators for these four parameters in arguably the model most widely used in empirical practice—the Gaussian selection model. Extensions of this standard analysis have been proposed by many, including Lee (1982, 1983), Heckman and Robb (1985, 1986), Heckman (1990), Andrews and Schafgans (1998), Heckman and Vytlačil (2000a), and Heckman, Tobias, and Vytlačil (2000). In the following section, we provide simple expressions for these parameters in the Gaussian selection model, and invite the reader to see the studies above for more general procedures.

3. Simple Expressions for Treatment Parameters in the Gaussian Selection Model

This section derives expressions for ATE, TT, LATE, and MTE as given in Equations 4–11 under the assumption of trivariate normality. Results for this case were first reported in Heckman and Vytlačil (2000b), although they present a more general analysis and do not discuss how estimates of these parameters can be obtained using simple two-step procedures.

Results for the “Textbook” Model

In the “textbook” selection model, we make the assumption of jointly normally distributed errors:

$$\begin{bmatrix} U^D \\ U^1 \\ U^0 \end{bmatrix} \sim N \left(0, \begin{bmatrix} 1 & \sigma_{1D} & \sigma_{0D} \\ \sigma_{1D} & \sigma_1^2 & \sigma_{10} \\ \sigma_{0D} & \sigma_{10} & \sigma_0^2 \end{bmatrix} \right).$$

The variance parameter in the selection equation is normalized to unity without loss of generality. Immediately, we recognize that ATE takes the form in Equation 4, and that the distributional assumption imposed does not change the functional form of this relation. Using the normality assumption, we find that the expression for TT is given as:

$$\text{TT}(x, z, D[z] = 1) = x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0) \frac{\phi(z\theta)}{\Phi(z\theta)}, \quad (12)$$

where $\rho_i \equiv \text{Corr}(U^i, U^D)$, $i = 0, 1$. Under the normalization that the variance of the disturbance term in the selection equation is unity, $\rho_i\sigma_i = \sigma_{iD}$. As previously noted, under independence between U^D and $(U^1 - U^0)$, all treatment parameters are the same. Thus, if $\text{Cov}(U^1 - U^0, U^D)$

= 0 or $\rho_1\sigma_1 = \rho_0\sigma_0$, TT reduces to ATE in Equation 4. In this case, people are not selecting into program on the basis of their unobserved (by the econometrician) gain, and all the treatment parameters reduce to ATE. If $\text{Cov}(U^1 - U^0, U^D) > 0$, then $\text{TT} > \text{ATE}$. If this condition is true, people are selecting into treatment on the basis of their idiosyncratic gain to treatment, and thus the gain from program participation for those observed in the treated state will exceed the gain for the average person. Also note that as $z\theta \rightarrow \infty$, $\text{TT} \rightarrow \text{ATE}$. In this case, the probability of receiving treatment is one given the observable characteristics $Z = z$ and thus there is no selection problem. In this case, the conditioning information $D = 1$ is redundant given the characteristics $Z = z$ and thus the two parameters in Equations 4 and 6 are equal.

Using standard results (e.g., Cramer 1946 or Johnson, Kotz, and Balakrishnan 1992), the LATE parameter can easily be derived using the fact that if $(y, z) \sim N(\mu_y, \mu_z, \sigma_y, \sigma_z, \rho)$ and $b > a$, then

$$E(y|a \leq z \leq b) = \mu_y + \rho\sigma_y \frac{(\Phi[\alpha] - \Phi[\beta])}{(\Phi[\beta] - \Phi[\alpha])},$$

where $\alpha = (a - \mu_z)/\sigma_z$, $\beta = (b - \mu_z)/\sigma_z$. Thus,

$$\begin{aligned} \text{LATE}(x, D[z] = 0, D[z'] = 1) &= E(Y^1 - Y^0 | x, -z'\theta \leq U^D \leq [-z\theta]) \\ &= x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0) \frac{\Phi(z'\theta) - \Phi(z\theta)}{\Phi(z'\theta) - \Phi(z\theta)}. \end{aligned} \quad (13)$$

The MTE corresponds to the expected outcome gain for those individuals who are just indifferent to the receipt of treatment at the given value of the unobservable u^D . Formally,

$$\begin{aligned} \text{MTE}(x, u^D) &= x(\beta^1 - \beta^0) + E(U^1 - U^0 | U^D = u^D) \\ &= x(\beta^1 - \beta^0) + E(U^1 | U^D = u^D) - E(U^0 | U^D = u^D) \\ &= x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0)u^D. \end{aligned}$$

We also note that MTE can be regarded as the limit form of LATE,⁵

$$\begin{aligned} \text{MTE}(x, u^D) &= x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0) \lim_{t \rightarrow -u^D} \left[\frac{\Phi(-u^D) - \Phi(t)}{\Phi(-u^D) - \Phi(t)} \right] \\ &= x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0) \lim_{t \rightarrow -u^D} \left[\frac{(\Phi[-u^D] - \Phi[t])/(-u^D - t)}{(\Phi[-u^D] - \Phi[t])/(-u^D - t)} \right] \\ &= x(\beta^1 - \beta^0) + (\rho_1\sigma_1 - \rho_0\sigma_0)u^D. \end{aligned} \quad (14)$$

Evaluating MTE when u^D is large corresponds to the case where the average outcome gain is evaluated for those individuals with unobservables making them most likely to participate, (and conversely when u^D is small). When $u^D = 0$, $\text{MTE} = \text{ATE}$ as a consequence of the symmetry of the normal distribution.

Estimation

It is important to recognize that the expressions above can be consistently estimated using nothing more than the output from a two-step procedure. Specifically, one can consistently estimate these parameters as follows:

⁵ The last line in this derivation follows from L'Hopital's rule.

- (i) Obtain $\hat{\theta}$ from a probit model on the decision to take the treatment.
- (ii) Compute the appropriate selection correction terms evaluated at $\hat{\theta}$ (i.e., $\phi[Z_i\hat{\theta}]/\Phi[Z_i\hat{\theta}]$ when $D_i = 1$, and $\phi[Z_i\hat{\theta}]/[1 - \Phi\{Z_i\hat{\theta}\}]$ when $D_i = 0$.)
- (iii) Run treatment-outcome-specific regressions (for the groups $[i: D_i = 1]$ and $[i: D_i = 0]$) with the inclusion of the appropriate selection-correction terms obtained from the previous step.
- (iv) Given $\hat{\beta}^0$, $\hat{\beta}^1$, $\widehat{\rho_1\sigma_1}$, and $\widehat{\rho_0\sigma_0}$ obtained from step (iii), and $\hat{\theta}$ from step (i), use these parameter estimates to obtain point estimates of the treatment parameters for given X , Z , and Z' . Alternatively, one could integrate over the distribution of the characteristics to obtain unconditional estimates, as suggested in section 2.

4. Going Beyond Mean Effects

The previous section, and most of the evaluation literature, has focused on the estimation of mean or average returns to the receipt of treatment. In this section we briefly describe some recent methods that can be used to determine the distributions associated with various outcome gains resulting from the receipt of treatment.

It is clear that knowledge of the joint distribution of outcomes affords the estimation of a wealth of interesting and policy-relevant parameters that cannot be determined from only mean effects. For example, even if the average treatment effect is large and positive, it is still possible that a substantial fraction of individuals receive negative effects from the treatment. To characterize the effectiveness of the program or treatment, we would like to have some knowledge of the dispersion of the various outcome gains. In addition, the following questions seem to be of primary interest in the evaluation of a given program: (i) What is the probability that a randomly chosen person will benefit positively from the receipt of treatment? (ii) What is the probability of a positive treatment effect for those actually selecting into the treatment? (iii) What is the probability that the treated receive a higher return to treatment than an average person? As emphasized by Heckman, Smith, and Clements (1997) and Heckman and Smith (1998), all of these parameters are interesting, and their investigation requires more than just knowledge of mean effects.⁶

Without further restrictions, the distribution of treatment effects is not identified. One restriction is to impose the Roy model, in particular, that $D = 1[Y^1 - Y^0 \geq 0]$, so that $Z\theta = X(\beta^1 - \beta^0)$ and $U^D = U^1 - U^0$. Under these restrictions, and additional support conditions, Heckman and Honoré (1990) show that it is possible to identify the joint distribution of (U^1, U^0) and thus to identify the distribution of treatment effects. This analysis was further extended by Heckman, Smith, and Clements (1997), Heckman and Smith (1998), and Hansen (1999). Another restriction that will allow identification of the distribution of treatment effects is to assume that (U^1, U^0, U^D) follow a factor structure (see Aakvik, Heckman, and Vytlačil 2000; Hansen and Heckman 2001).⁷ All of these methods provide ways for “solving” the unidentified parameter problem, and enable estimation of the distributions of interest.

⁶ See Heckman, Smith, and Clements (1997) and Heckman and Smith (1998) for a detailed discussion of different questions of interest related to the distribution of treatment effects, and for the connection between the distribution of treatment effects and various criteria for evaluating social programs.

⁷ Carneiro, Heckman, and Vytlačil (2001) propose a test for the factor-structure assumption.

Another approach is to not impose additional structure but instead place bounds on the unidentified parameters of interest. Particularly relevant is work in the Bayesian literature, where Vijverberg (1993), Koop and Poirier (1997), and Poirier and Tobias (2001a, b) note that although the cross-regime correlation parameter is unidentified, one can learn about this parameter through information contained in the identified parameters.⁸ Specifically, the positive definiteness of the 3×3 covariance matrix of the disturbance terms adds an additional source of information, and this information may help us to learn (or at least bound) the values of this parameter. That is, knowledge of the identified correlations can “spill over” and thereby update our beliefs about the unidentified cross-regime correlation. Koop and Poirier (1997) and Poirier and Tobias (2001b) argue that the marginal posteriors and priors for this unidentified correlation can differ substantially, and thus in many situations, learning takes place about this parameter. Despite this, the prior will still affect the behavior of the unidentified parameter even in large samples,⁹ and thus it is necessary to describe sensitivity of results to the choice of a prior.

5. The Returns to College

To illustrate how point estimates of the mean effects are obtained and interpreted, we apply the techniques discussed in section 3 to estimate the return to some form of college education, noting that the problem of selection bias has long been recognized as important in assessing the returns to schooling (see, e.g., Willis and Rosen 1979). Our data are taken from the NLSY. In our analysis, Y^1 denotes the log of 1991 hourly earnings for those individuals completing at least 13 years of schooling by 1991, and Y^0 is the log of hourly wages for those with 12 or fewer years of schooling. The sample is restricted to white males who are not enrolled in school in the current year and report hourly earnings between \$1 and \$100. Observations are also deleted when other explanatory variables used in the analysis are missing, resulting in a final sample of 1230 observations.

The variables in X include an intercept, two indicators for residence in the Northeast and South,¹⁰ potential labor market experience and its square,¹¹ an indicator for residence in an urban area, the local unemployment rate, and a measure of “ability” denoted as g . This ability measure is constructed from the 10 component tests of the Armed Services Vocational Aptitude Battery provided in the NLSY. Because people vary in age at the time of the test, each component test is first regressed on age. The residuals from this regression are then standardized, and g is defined as the first principal component of the standardized residuals.¹² We choose a parsimonious specification for the variables in the selection equation (Z), which includes an intercept, g , indicator variables denoting if the respondent’s mother and father attended college, an indicator for residence in an urban area at age 18, and number of siblings. The last variable serves

⁸ Other work that develops bounds on the distribution of treatment effects includes Heckman, Smith, and Clements (1997) and Heckman and Smith (1998).

⁹ See Poirier and Tobias (2001b) for a more complete description of the role of the prior.

¹⁰ The NLSY provides four regional variables—Northcentral, Northeast, South, and West.

¹¹ Potential experience is defined as $\text{Age} - \text{Years of Schooling} - 6$.

¹² For more on the construction and use of this ability measure, see Cawley et al. (1997).

Table 1. Coefficients and Standard Errors for Application of Section 5

Variable	Coefficient	Standard Error
College state		
Constant	1.85	0.225
g (Ability)	0.092	0.053
Northeast	0.124	0.055
South	0.059	0.057
Experience	0.098	0.044
Experience ²	−0.004	0.003
Urban	0.326	0.072
Unemp. Rate	−0.002	0.002
λ ($Z\hat{\theta}$)	−0.165	0.081
No-college state		
Constant	1.89	0.424
g (Ability)	0.191	0.036
Northeast	0.126	0.057
South	−0.046	0.053
Experience	0.043	0.067
Experience ²	−0.001	0.003
Urban	0.136	0.051
Unemp. Rate	0.001	0.002
$-\lambda$ ($-Z\hat{\theta}$)	0.097	0.094
Selection equation		
Constant	−0.478	0.149
MomCollege	0.541	0.112
DadCollege	0.603	0.097
Numsibs	−0.069	0.024
g (Ability)	0.754	0.048
Urban18	0.096	0.131

as our primary exclusion restriction and is assumed to affect the college entry decision without affecting postschooling earnings.¹³

We begin by describing the computation of the mean treatment parameters, as presented in section 2. Point estimates of the ATE are obtained by averaging the conditional treatment effects (given X) over the sample distribution of characteristics, as in Equation 5. For TT, point estimates are obtained as in Equation 7 by averaging over the joint distribution of characteristics (given X and Z) for the subsample that actually selects into college. To estimate LATE, we average over the joint distribution of characteristics after setting the number of siblings variable in $Z = z$ equal to four, and equal to 0 in $Z = z'$ (this is the second form of the unconditional LATE parameter previously discussed). This estimates the average college log wage premium for persons induced to attend college when the number of siblings has been lowered from four to zero. Finally, for each value of U^D , we construct the MTE parameter not conditioning on observable characteristics by averaging $MTE(X, u^D)$ over the sample distribution of X characteristics. In section 2, we regarded MTE as a function of u^D , and suggested plotting the effect

¹³ The number-of-siblings variable was found to be a significant determinant of the college entry decision, but was not significant at the 5% level when included as a regressor in the outcome equations for the college and no-college states. Other variables, such as distance to college, the local unemployment rate at age 18, and a state-level tuition variable were also constructed and investigated as potential instruments. These variables were found to have surprisingly little power in explaining the college entry decision for these data and thus we selected number of siblings as our instrument.

over the support of U^D . Because the result is linear in u^D , we simply report the slope of that effect. Point estimates of the treatment parameters are scaled by the difference in average years of schooling across the college and no-college groups (≈ 3.8) to estimate the return to schooling.

Results of the mean effect analysis indicate that a randomly chosen person might expect to receive a 9% increase in hourly wages resulting from the receipt of some form of higher education. Those actually selecting into college receive about a 4% increase in hourly wages, whereas those induced to attend college as a result of having no siblings (relative to four siblings) approximately receive an 8% increase in hourly wages.¹⁴ The estimated MTE parameter is linear in U^D with a slope equal to -0.07 . This negative slope indicates that individuals with unobservables making them most likely to enroll in college receive the smallest return to a college education. We also test and reject the hypothesis of a constant MTE (i.e., reject $H_0: \text{Cov}[U^D, U^1 - U^0] = 0$), and conclude that selection is an important feature of this data. Similar results are reported in Carneiro, Heckman, and Vytlačil (2001) and Heckman (2001).

The methods used here are easily implemented and can be applied to estimate a variety of average gains to program participation in the presence of selection bias. Although these results depend on the normality assumption, such an assumption can be relaxed using more general parametric models (e.g., Heckman, Tobias, and Vytlačil 2000), or using the more general semiparametric and nonparametric techniques described in the references above. Nonetheless, the simple expressions obtained for the standard Gaussian selection model offer a useful starting point for research in program evaluation.

6. Conclusion

This paper reviewed and provided simple expressions for four parameters commonly used to evaluate the effectiveness of a given program or treatment: ATE, TT, LATE, and the MTE. These expressions were obtained for the “textbook” selection model. The appeal of the approach described in this paper is that practitioners can obtain consistent estimates of these parameters using nothing more than a standard two-step estimator.

The modern approach to program evaluation focuses on the estimation of narrowly defined parameters without having to impose strong distributional assumptions. The approach adopted in this paper permits estimation of a variety of policy-relevant parameters as well as estimation of the four treatment effects listed above, rather than one or the other parameters featured in the recent treatment effect literature. In addition, we briefly described several methods for going beyond estimation of mean parameters to enable researchers to estimate the distributions associated with various outcome gains.

The methods presented in this paper were applied to estimate the return to a college education. Using data from the NLSY, we obtained point estimates of ATE, TT, LATE, and MTE for the textbook Gaussian selection model. Our mean parameter results indicated that the receipt of some form of college education would lead to an expected increase in hourly wages equal to 9% for a randomly selected person, and equal to 4% for those actually selecting into higher education.

¹⁴ The estimated standard errors associated with the point estimates of the ATE, TT, and LATE parameters above were 0.03, 0.04, and 0.03, respectively, and were computed using the nonparametric bootstrap.

References

- Aakvik, A., J. Heckman, and E. Vytlačil. 2000. Treatment effects for discrete outcomes when responses to treatment vary among observationally identical persons: An application to Norwegian vocational rehabilitation programs. Unpublished manuscript, University of Chicago.
- Amemiya, Takeshi. 1985. *Advanced econometrics*. Cambridge, MA: Harvard University Press.
- Andrews, D., and M. Schafgans. 1998. Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65:497–518.
- Björklund, A., and R. Moffitt. 1987. The estimation of wage gains and welfare gains in self-selection models. *Review of Economics and Statistics* 69:42–9.
- Carneiro, P., J. Heckman, and E. Vytlačil. 2001. Estimating the return to education when it varies among individuals. Unpublished working paper, University of Chicago.
- Cawley, John, Karen Conneely, James Heckman, and Edward Vytlačil. 1997. Cognitive ability, wages, and meritocracy. In *Intelligence, genes and success: Scientists respond to the bell curve*, edited by Bernie Devlin, Stephen E. Fienberg, Daniel P. Resnick, and Kathryn Roeder. New York: Springer, pp. 179–92.
- Cramer, Harold. 1946. *Mathematical methods of statistics*. Princeton: Princeton University Press.
- Glynn, Robert, Nan Laird, and Donald Rubin. 1986. Selection models versus mixture modeling with nonignorable non-response. In *Drawing inference from self-selected samples*, edited by Howard Wainer. New York: Springer, pp. 115–42.
- Goldberger, Arthur. 1983. Abnormal selection bias. In *Studies in econometrics, time series, and multivariate statistics*, edited by Samuel Karlin, Takeshi Amemiya, and Leo Goodman. New York: Academic Press.
- Gronau, Reuben. 1974. Wage comparisons—A selectivity bias. *Journal of Political Economy* 82:1119–43.
- Hansen, Karsten. 1999. A semiparametric Bayesian analysis of the Roy model. Unpublished working paper, University of Chicago.
- Hansen, Karsten, and James Heckman. 2001. The formulation and estimation of panel data treatment effects. Unpublished working paper, University of Chicago.
- Heckman, James. 1974. Shadow prices, market wages and labor supply. *Econometrica* 42:679–94.
- Heckman, James. 1976. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5:475–92.
- Heckman, James. 1990. Varieties of selection bias. *American Economic Review* 80:313–8.
- Heckman, James. 1997. Instrumental variables: A study of implicit behavioral assumptions in one widely used estimator. *Journal of Human Resources* 32:441–62.
- Heckman, James. 2001. Microdata, heterogeneity and the evaluation of public policy. *Journal of Political Economy* 109: 673–748.
- Heckman, James, and Bo Honoré. 1990. The empirical content of the Roy model. *Econometrica* 58:1121–49.
- Heckman, J., and R. Robb. 1985. Alternative methods for evaluating the impact of interventions. In *Longitudinal analysis of labor market data*, edited by J. Heckman and B. Singer. New York: Cambridge University Press, pp. 156–245.
- Heckman, J., and R. Robb. 1986. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In *Drawing inference from self-selected samples*, edited by H. Wainer. Berlin: Springer-Verlag, pp. 63–107.
- Heckman, James, Jeffrey Smith, and Nancy Clements. 1997. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *Review of Economic Studies* 64:487–535.
- Heckman, James, and Jeffrey Smith. 1998. Evaluating the welfare state. In *Econometrics and economic theory in the 20th century: The Ragnar Frisch centennial*, edited by S. Strom. Cambridge, UK: Cambridge University Press, pp. 241–318.
- Heckman, James, Justin Tobias, and Edward Vytlačil. 2000. Simple estimators for treatment parameters in a latent variable framework with an application to estimating the returns to schooling. NBER Working Paper No. 7950.
- Heckman, James, and Edward Vytlačil. 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences U.S.A.* 96:4730–4.
- Heckman, James, and Edward Vytlačil. 2000a. Local instrumental variables. In *Nonlinear statistical inference: Essays in honor of Takeshi Amemiya*, edited by C. Hsiao, K. Morimune, and J. Powell. Cambridge: Cambridge University Press.
- Heckman, James, and Edward Vytlačil. 2000b. The relationship between treatment parameters within a latent variable framework. *Economics Letters* 66:33–9.
- Imbens, Guido, and Joshua Angrist. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75.
- Johnson, Norman, Samuel Kotz, and N. Balakrishnan. 1992. *Continuous univariate distributions*. New York: John Wiley and Sons.

- Koop, Gary, and Dale Poirier. 1997. Learning about the across-regime correlation in switching regression models. *Journal of Econometrics* 78:217–27.
- Lee, Lung-Fei. 1978. Unionism and wage rates: A simultaneous model with qualitative and limited dependent variables. *International Economic Review* 19:415–34.
- Lee, Lung-Fei. 1982. Some approaches to the correction of selectivity bias. *Review of Economic Studies* 49:355–72.
- Lee, Lung-Fei. 1983. Generalized econometric models with selectivity. *Econometrica* 51:507–12.
- Lydall, Harold. 1968. *The structure of earnings*. Oxford: Clarendon Press.
- Paarsch, Harry J. 1984. A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics* 24:197–213.
- Poirier, Dale, and Justin Tobias. 2001a. Across-regime covariance restrictions in treatment response models. Unpublished paper, University of California-Irvine.
- Poirier, Dale, and Justin Tobias. 2001b. On the predictive distributions of outcome gains in the presence of an unidentified parameter. Unpublished paper, University of California-Irvine.
- Quandt, Richard. 1972. Methods for estimating switching regressions. *Journal of the American Statistical Association* 67:338, 306–10.
- Roy, A. D. 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3:135–46.
- Rubin, Donald. 1978. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics* 6:34–58.
- Vijverberg, W. P. M. 1993. Measuring the unidentified parameter of the extended Roy model of selectivity. *Journal of Econometrics* 57:69–89.
- Vytlacil, Edward. 2002. Independence, monotonicity, and latent variable models: An equivalence result. *Econometrica*. In press.
- Willis, Robert, and Sherwin Rosen. 1979. Education and self-selection. *Journal of Political Economy* 87:S7–36.