

## Commentary

# Models of treatment effects when responses are heterogeneous

Robert A. Moffitt<sup>†</sup>

Department of Economics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218

Heckman and Vytlacil (1) synthesize and extend a recent body of research in economics and statistics on the identification and estimation of treatment effects when the subjects have heterogeneous responses. This research has demonstrated the importance of distinguishing between several different types of treatment effects and the need to establish firmly the relationship between the different types. Heckman and Vytlacil (H&V) distinguish between four different types: average treatment effects (ATE), effects of the treatment on the treated (TT), local average treatment effects (LATE), and local instrumental variable (LIV) treatment effects. Each is conceptually different, and each has a different set of conditions for identification. The research discussed by H&V represents a genuine advance and clarification of concepts in models of treatment effects.

The framework set up by H&V is heavily influenced by economic and econometric terminology but has major elements that are drawn from statistics as well. The reader may be led to suppose that the concepts and methods are applicable only to observational data and not to randomized clinical trials (RCTs), but this is not the case. All parts of their analysis are equally applicable to both. The application to RCTs is most easily seen by drawing an analogy with intent-to-treat models, as discussed by Angrist *et al.* (2). Although, in an RCT, the randomization creates a treatment-assignment dummy variable that economists call an “instrument,” similar variables often exist in observational data—natural experiments, quasi-experiments, or, more generally, what economists merely call exogenous identifying variables—and these can satisfy the same conditions as the treatment assignment variable in an RCT, and, hence, the same methods apply.

The major restriction in the models of H&V is the index function restriction, which is the assumption that the propensity to “participate” or “take up” the treatment, when it is offered, can be described by a single function with a single unobservable. H&V note correctly in their conclusions that most of their results can be obtained, in modified form, without an index function restriction, but their analysis in the main makes that assumption. While imposing restrictions, the index function model has great expository and intuitive value in the analysis of treatment effects models.

In this commentary, I demonstrate that, with a few additional restrictions, most of the major points discussed by H&V can be given a simple graphical interpretation that has the same virtue of expository usefulness. Denote by  $\alpha_i$  the treatment effect for subject  $i$ , often called a “random coefficient” in econometric models but equivalent to the heterogeneous treatment response in treatment effect models in statistics. Fig. 1 shows a hypothetical density of  $\alpha_i$  in the population, a density assumed to have a mean of  $\bar{\alpha}$ . This figure appears in Björklund and Moffitt (3), where  $\alpha_i$  is given the random coefficient interpretation just referred to. Björklund and Moffitt call  $\alpha_i$  the “gain” to the treatment for individual  $i$ . Now make the additional restriction that selection takes place strictly on  $\alpha_i$  and that, for illustration, selection is positive: subjects with higher values of  $\alpha_i$  are more likely to take up the treatment if

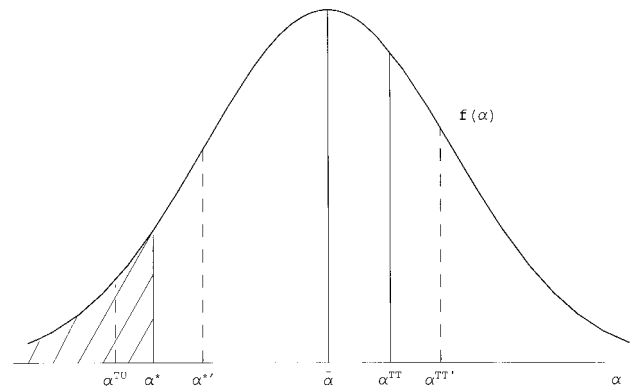


FIG. 1. Density of treatment gains in a population.

offered, i.e., more likely to participate.<sup>‡</sup> Denote the cutoff value of  $\alpha_i$  as  $\alpha^*$ , above which subjects participate and below which subjects do not.

Fig. 1 can be used to illustrate all four estimators. The parameter  $\bar{\alpha}$  is equivalent to the ATE. The treatment effect on the treated is shown in Fig. 1 as  $\alpha^{TT}$  and is equal to  $E(\alpha_i | \alpha_i > \alpha^*)$ , the mean gain of those who are participants. For future use, Fig. 1 also shows the treatment effect on the untreated,  $\alpha^{TU} = E(\alpha_i | \alpha_i < \alpha^*)$ , the mean gain that those who are nonparticipants would have if they received the treatment. This quantity is an unobservable. Note that  $\bar{\alpha} = P\alpha^{TT} + (1 - P)\alpha^{TU}$ , where  $P = \text{Prob}(\alpha_i > \alpha^*)$ , by construction.

The LATE estimator is representable in Fig. 1 by a discrete change in  $\alpha^*$  moving it to  $\alpha^{*'}$ , which necessarily shifts  $\alpha^{TT}$  as well, moving it to  $\alpha^{TT'}$ . The LATE estimator equals  $\alpha^{TT} - \alpha^{TT'}$  divided by the area under the curve between  $\alpha^*$  and  $\alpha^{*'}$ , which equals the change in the probability of participating. The estimator termed by H&V the LIV is the limit of this change as  $\alpha^* \rightarrow \alpha^{*'}$ . This estimator is termed the “marginal” gain by Björklund and Moffitt (3) to emphasize that it is the treatment gain of the marginal subject just on the edge of participating and not participating. A small expansion or contraction of the program will bring these subjects into the program or push them out of it and will lead to a change in the mean gain of participants accordingly. In their empirical example, Björklund and Moffitt analyzed a program that had a positive mean gain for those in the program (positive  $\alpha^{TT}$ ) but a negative marginal gain, implying that the program was “too large”—it had been overextended to include subjects who were made worse off by it. Thus the difference between the two can be important.

In general, only  $\alpha^{TT}$ , and changes in that quantity, are identified. This is clear from Fig. 1. Denote the difference between the mean outcome of the subjects of a treatment group that is offered the program and that divides up into

The companion to this Commentary begins on page 4730 in issue 8 of volume 96.

<sup>†</sup>To whom reprint requests should be addressed. e-mail: moffitt@jhu.edu.

<sup>‡</sup>This selection can either arise from voluntary actions on the part of the subjects or from decisions by the program operators to admit and deny applicants on the basis of  $\alpha_i$ . This model is neutral on this issue.

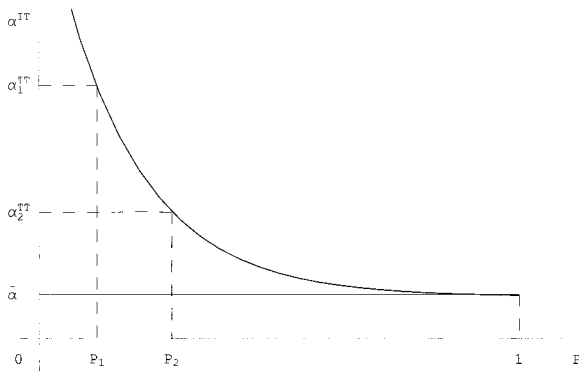


FIG. 2. Relationship between participation rate and mean gain of participants.

participants and nonparticipants in the manner described by Fig. 1, and the mean outcome of a comparison group that is not offered the program at all, as  $\alpha^{UNADJ}$ . Then, it can be shown that  $\alpha^{TT} = \alpha^{UNADJ}/P$ , where  $P$  is the participation rate in the treatment group. Multiple treatment groups with different participation rates can be used with this formula to calculate the LATE estimator and, in the limit, the LIV estimator. Only if  $P = 1$  in one of the treatment groups can  $\bar{\alpha}$  be identified, as is clear from Fig. 1 (“identification at infinity”).

Fig. 2 graphically illustrates this result by showing the relationship between  $P$  and  $\alpha^{TT}$ . By the positive and monotonic selection assumed here, the value of  $\alpha^{TT}$  falls and approaches  $\bar{\alpha}$  as  $P \rightarrow 1$ . Treatment groups with different values of  $P$  form data points along the curve. Two data points are shown in Fig. 2. With two data points, the LATE estimator can be calculated. If the full curve were estimable, what Björklund and Moffitt term the marginal gain and what H&V term the LIV can be calculated as the slope of the curve in Fig. 2. §

Fig. 2 suggests what a research agenda on a program should aim to achieve. Estimates from multiple RCTs, observational studies, and other analyses that contain treatment groups with different values of  $P$  yield data points on the curve. Nonpara-

§ In a personal communication from H&V, the authors show that the ATE and LATE can be graphically obtained from the LIV by integration.

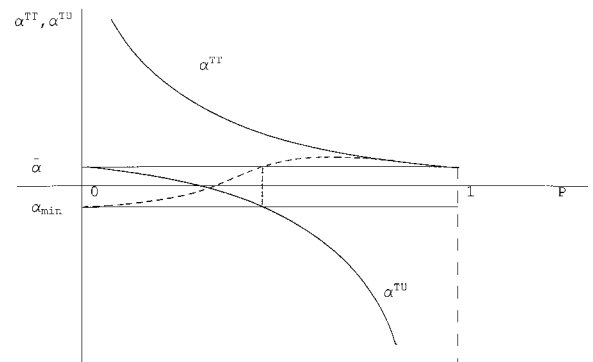


FIG. 3. Mean treatment gains with an assumed lower bound.

metric estimation and related smoothing techniques permit the estimation of the function, at least between the maximum and minimum values of  $P$  in the different studies. Extrapolation is required beyond those points.

The bounding concepts described in H&V also can be illustrated graphically. Fig. 3 shows both  $\alpha^{TT}$  and  $\alpha^{TU}$ . The latter is an unobservable, as previously noted, and hence, the  $\alpha^{TU}$  line cannot be estimated. This can be viewed as the reason that  $\bar{\alpha}$  cannot be estimated as well, for  $\bar{\alpha} = P\alpha^{TT} + (1 - P)\alpha^{TU}$  is an equation with two estimable quantities,  $P$  and  $\alpha^{TT}$ , but two unknowns,  $\bar{\alpha}$  and  $\alpha^{TU}$ . But if a (say) lower bound  $\alpha_{MIN}$  is established for  $\alpha^{TU}$ , a lower bound on  $\bar{\alpha}$  can be established as well (note that it is not the bound on the individual values that matter but only the bound on their difference for the untreated; this is implicit in the formula for the width of the bound). Inserting  $\alpha_{MIN}$  in for  $\alpha^{TU}$  in the weighted average formula for  $\bar{\alpha}$  yields the desired bound for  $\bar{\alpha}$ . It is illustrated by the dotted line in Fig. 3. Upper bounds could be similarly illustrated.

The results in the work by H&V provide a fruitful set of concepts and methods for applications in future research in the social sciences. The different concepts have seen relatively little use to date in applications but should see more in the future.

1. Heckman, J. J. & Vytlacil, E. J. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4730–4734.
2. Angrist, J., Imbens, G. & Rubin, D. (1996) *J. Am. Stat. Assoc.* **91**, 444–472.
3. Björklund, A. & Moffitt, R. (1987) *Rev. Econ. Stat.* **69**, 42–49.