NOTE

A CAUSAL INTERPRETATION OF EXTENSIVE AND INTENSIVE MARGIN EFFECTS IN GENERALIZED TOBIT MODELS

Kevin E. Staub*

Abstract—This note proposes a new decomposition of average treatment effects on nonnegative outcomes. It represents the total effect as a population-weighted sum of the effects for two groups: those induced to participate by the treatment and those participating regardless of it. The usual decomposition into extensive and intensive margins used in the literature is generally incompatible with such a causal interpretation. The difference between decompositions can be substantial and yield diametrically opposed results.

I. Introduction

MANY outcomes of interest in economics are nonnegative and have a cluster of observations at the value 0.[1] Prominent examples include working hours, health care demand, and expenditure data. Researchers analyzing average treatment effects (ATE) on such outcomes frequently take interest in decomposing the effects into the part attributable to individuals starting to participate (called *extensive margin*) and the part attributable to already participating individuals (called *intensive margin*). The decomposition used is algebraically straightforward, as it is based on factoring the expectation of the outcome variable, say $\mathrm{E}(Y)$, into the participation probability $\Pr(Y > 0)$ and the conditional expectation $\mathrm{E}(Y|Y > 0)$ (McDonald & Moffitt, 1980). The extensive margin is driven by the participation effect (PE), the change in the probability to participate; the intensive margin is driven by the conditional-on-positives (COP) effect, the change in the outcome given participation.

In contrast to the simplicity of the mechanical aspect, endowing the decomposition with a causal interpretation is substantially more problematic. For instance, recent work framing the problem in terms of Rubin's potential outcomes model has pointed out that COP effects do not measure the impact of a treatment on participating individuals; rather, they are hopelessly contaminated by a sort of selection bias, even in experimental settings (Angrist, 2001; Angrist & Pischke, 2009).

In this paper, I propose a conceptually different decomposition of the ATE into extensive and intensive margins. It is based on stratifying the population into groups defined by the joint distribution of potential outcomes (Frangakis & Rubin, 2002), representing the total effect as an average of the treatment effects for interesting groups of the population: those induced to participate by the treatment and those participating regardless of it. Unlike the conventional decomposition, the proposed one is not identified nonparametrically. However, the structure imposed in (nonparametric versions of) the censored regression model and the selection model identifies the decomposition. A numerical example for the tobit model, a special case of both models, shows that the differences between decompositions can be major.

The parameters of interest discussed here are functions of the joint distribution of potential outcomes, which cannot be rewritten as functions of the marginal distributions, a problem considered early on by Heckman, Smith, and Clements (1997) and recently by Fan and Park (2009), among others. While the imposition of model structure to achieve identification places this note closer to literature in the vein of Heckman et al. (1997) than to the partial identification literature (see Fan & Park, 2009, and references therein), I include a brief discussion of simple bounds for the objects of interest. The contribution of this paper lies in a conceptual definition of objects of interest and interpretation, and it is thus close in spirit to Angrist (2001).

II. Nonnegative Variables and Potential Outcomes

Consider the variable Y_i for individuals $i = 1, \ldots, N$, where $Y_i \geq 0$ and $\Pr(Y_i = 0) > 0$. Let Y_{1i} denote the potential outcome for i if i received a binary treatment $T_i (T_i = 1)$ and Y_{0i} if $T_i = 0$, so that as usual, $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})T_i$. A causal treatment effect is a comparison of treatment and control potential outcomes for a common subset of individuals (Frangakis & Rubin, 2002; Rubin, 2006). The focus here will be on the average treatment effect [ATE] $\mathrm{E}(Y_{1i} - Y_{0i})$. Assume that the data come from an ideal randomized controlled trial, so that assignment to treatment is random and compliance is perfect. Then T_i is independent of (Y_{1i}, Y_{0i}), and ATE can be obtained from the prima facie contrast $\mathrm{E}(Y_i|T_i = 1) - E(Y_i|T_i = 0)$. Using

$$\mathrm{E}(Y_i|T_i) = \Pr(Y_i > 0|T_i)\mathrm{E}(Y_i|Y_i > 0, T_i),$$

this contrast can be written as

$$\begin{aligned}&\mathrm{E}(Y_i|T_i = 1) - E(Y_i|T_i = 0)\\ &= \{\Pr(Y_i > 0|T_i = 1) - \Pr(Y_i > 0|T_i = 0)\}\mathrm{E}(Y_i|Y_i > 0, T_i = 1)\\ &+ \{\mathrm{E}(Y_i|Y_i > 0, T_i = 1) - \mathrm{E}(Y_i|Y_i > 0, T_i = 0)\}\Pr(Y_i > 0|T_i = 0).\end{aligned} \tag{1}$$

This is the usual decomposition applied to limited dependent variables like Y_i in tobit (Tobin, 1958) or Cragg (1971) models (McDonald & Moffitt, 1980; see also Cameron & Trivedi, 2005, Greene, 2008, and Wooldridge, 2002). The first term on the right-hand side of equation (1) is the extensive margin effect, which weights the PE—the term in braces—by the expected Y_i conditional on participation; the second term is the intensive margin effect, which weights the COP (in braces) by the probability of participation given $T_i = 0$.

Using the independence of T_i from Y_{1i} and Y_{0i}, the decomposition can be rewritten in terms of potential outcomes as

$$\begin{aligned}\mathrm{E}(Y_{1i}) - E(Y_{0i}) = &\{\Pr(Y_{1i} > 0) - \Pr(Y_{0i} > 0)\}\mathrm{E}(Y_{1i}|Y_{1i} > 0)\\ &+ \{\mathrm{E}(Y_{1i}|Y_{1i} > 0) - \mathrm{E}(Y_{0i}|Y_{0i} > 0)\}\Pr(Y_{0i} > 0).\end{aligned} \tag{2}$$

While the left-hand side, ATE, is a causal effect, the second component on the right-hand side is not. COP compares the average outcome

Received for publication November 10, 2010. Revision accepted for publication September 21, 2012.

* University of Melbourne.

I thank three referees and the editor, Alberto Abadie, for very helpful comments. I am indebted to Gregori Baetschmann, Leon Bettendorf, Stefan Boes, Colin Cameron, Bo Honoré, and Rainer Winkelmann for useful suggestions at different stages of the paper. The paper benefited from comments of participants in the annual meetings of the Midwest Econometrics Group in St. Louis, Missouri, the Swiss Society of Economics and Statistics in Lucerne, the Royal Economic Society in London, and the European Meeting of the Econometric Society in Oslo.

[1] Sometimes variables with these features are referred to as *corner solution outcomes* (Wooldridge, 2002).

under treatment of one subpopulation (that with positive outcomes under treatment) to the average control outcome of a potentially different subpopulation (that with positive outcomes without treatment). Unless treatment does not induce any individuals to switch from positive to 0 or from 0 to positive outcomes, the subpopulations involved in the COP contrast will be different.

One approach to addressing this problem would be to distill a causal effect for the subpopulation with $Y_{1i} > 0$.[2] However, a treatment effect for that population is not an object of interest in a decomposition into extensive and intensive margins: The ATE for individuals with $Y_{1i} > 0$ mixes the ATE for the very two subpopulations the decomposition sets out to discriminate—the one participating even without treatment and the one participating because of the treatment.

A. *Decomposition Based on Joint Outcomes*

Consider the following stratification of individuals into nonoverlapping and exhaustive subpopulations based on their *joint* distribution of potential outcomes, (Y_{0i}, Y_{1i}):

Group	Name	Potential Outcomes
NP	Nonparticipants	$(Y_{0i} = 0, Y_{1i} = 0)$
S_1	Switchers	$(Y_{0i} = 0, Y_{1i} > 0)$
S_2	Switchers	$(Y_{0i} > 0, Y_{1i} = 0)$
P	Participants	$(Y_{0i} > 0, Y_{1i} > 0)$

Basing the definition of intensive and extensive margin effects on these groups clarifies their meaning substantially. The intensive margin effect is the contribution to the ATE of group P. Similarly, the extensive margin is the ATE contribution of switchers—those changing their participation status (groups S_1 and S_2). These are the objects of interest when decomposing causal effects into extensive and intensive margins. When researchers write about them, it is this that they mean (although they rarely state it so explicitly). For instance, take the labor economics example of working hours. The effect of a policy intervention increasing average working hours in the economy can be decomposed into (a) the average change in hours worked of those working regardless of the intervention, plus (b) the average hours worked by those joining the workforce because of the intervention, minus (c) the average hours worked by those leaving the workforce because of the intervention, with the groups in these three categories being weighted by their population fraction.

Thus, the decomposition of the ATE based on the joint distribution of potential outcomes is

$$\begin{aligned} &\mathrm{E}(Y_{1i}) - \mathrm{E}(Y_{0i}) = \mathrm{E}_{Y_{1i},Y_{0i}}\left[\mathrm{E}(Y_{1i} - Y_{0i})|Y_{1i}, Y_{0i}\right] \\ &= \mathrm{E}(Y_{1i}|Y_{0i} = 0, Y_{1i} > 0)\Pr(Y_{0i} = 0, Y_{1i} > 0) \\ &\quad + \mathrm{E}(-Y_{0i}|Y_{0i} > 0, Y_{1i} = 0)\Pr(Y_{0i} > 0, Y_{1i} = 0) \\ &\quad + \mathrm{E}(Y_{1i} - Y_{0i}|Y_{0i} > 0, Y_{1i} > 0)\Pr(Y_{0i} > 0, Y_{1i} > 0). \end{aligned} \tag{3}$$

The expectation over (Y_{1i}, Y_{0i}) is with respect to the four events NP, S_1, S_2, and P. The last term is the intensive margin effect (IME), and the first two are the extensive margin effect (EME). The effect for individuals in group NP is always 0.

[2] Angrist (2001) and Angrist and Pischke (2009) write COP as this causal effect, $\mathrm{E}(Y_{1i} - Y_{0i}|Y_{1i} > 0)$, plus a selection bias term.

B. *Nonparametric Identification*

Group membership being unobserved, the population fractions are not identified from the data on treatment and outcome. This is a common occurrence in principal stratification contexts such as this one (Frangakis & Rubin, 2002; Rubin, 2006). To see this, denote the population fraction for P as π^P, the one for S_1 as π^{S_1}, and so on. Then,

$$\begin{aligned} &\Pr(Y_i = 0|T_i = 0) = \pi^{NP} + \pi^{S_1},\ \Pr(Y_i = 0|T_i = 1) = \pi^{NP} + \pi^{S_2}, \\ &\Pr(Y_i > 0|T_i = 0) = \pi^{P} + \pi^{S_2},\ \Pr(Y_i > 0|T_i = 1) = \pi^{P} + \pi^{S_1}. \end{aligned}$$

Restricting one of the population fractions to 0 identifies the remaining three fractions from these conditional probabilities. Often researchers choose models possessing some monotonicity assumption on the way treatment affects outcomes (Manski, 1997), which can typically lead to the elimination of one group out of S_1 and S_2. Thus, for the rest of the discussion, suppose $\pi^{S_2} = 0$ as a result of some particular monotone treatment response assumption (MTR).[3]

An example of MTR is when the causal effect is nonnegative for all: $Y_{1i} - Y_{0i} \geq 0, \forall i$. Call this assumption MTR1. MTR1 is embedded in the tobit model. The elimination of S_2 is compatible with other MTR assumptions, such as the one implicit in the selection model of Heckman (1979) and two-part models of Cragg (1971) or Duan et al. (1983): the causal effect is nonnegative for switchers (if there exist switchers, they are all members of S_1), and is either positive, 0, or negative for all members of group P. Call this assumption MTR2. Often assumptions such as these are motivated by economic theory, and for many applications it might be plausible to impose them.

Defining mean potential outcomes of switchers as $(\bar{Y}_0^{S_1}, \bar{Y}_1^{S_1})$ and those of participants as $(\bar{Y}_0^{P}, \bar{Y}_1^{P})$—for instance, $\bar{Y}_1^{S_1} = \mathrm{E}(Y_{1i}|Y_{0i} = 0, Y_{1i} > 0)$—the decomposition of ATE in equation (3) can be written as

$$\mathrm{ATE} = \pi^{S_1}\bar{Y}_1^{S_1} + \pi^{P}(\bar{Y}_1^{P} - \bar{Y}_0^{P}) = \pi^{S_1}\mathrm{ATE}^{S_1} + \pi^{P}\mathrm{ATE}^{P}. \tag{4}$$

While ATE, π^{S_1}, π^P, and $\bar{Y}_0^P$ are all identified from the data,[4] the problem is identification of $\bar{Y}_1^{S_1}$ and $\bar{Y}_1^P$ for which the data provide only one quantity, $\mathrm{E}(Y_i|Y_i > 0, T_i = 1)$:

$$\begin{aligned} \mathrm{E}(Y_i|Y_i > 0, T_i = 1) &= \omega^{S_1}\bar{Y}_1^{S_1} + (1 - \omega^{S_1})\bar{Y}_1^{P}, \\ \omega^{S_1} &= \frac{\pi^{S_1}}{\pi^{S_1} + \pi^{P}}. \end{aligned}$$

Thus, it is not possible to attribute a fraction of ATE to the extensive or intensive margin without making more assumptions.[5] However, under MTR1 or MTR2, simple informative bounds for ATE^{S_1} and ATE^P are available.

For instance, assume MTR1 holds ($Y_{1i} - Y_{0i} \geq 0\ \forall i$), implying ATE is positive. The domain of ATE^{S_1} and ATE^P is the positive real line

[3] The case $\pi^{S_1} = 0$ is symmetric and will be omitted for brevity. Note that assuming $\pi^{S_2} = 0$ is equivalent to (a) assuming that either π^{S_1} or π^{S_2} is 0, and (b) $\Pr(Y_i > 0|T_i = 1) - \Pr(Y_i > 0|T_i = 0) > 0$. The latter can be verified in the data.

[4] $\pi^{S_1} = \Pr(Y_i > 0|T_i = 1) - \Pr(Y_i > 0|T_i = 0)$, $\pi^P = \Pr(Y_i > 0|T_i = 0)$, and $\bar{Y}_0^P = \mathrm{E}(Y_i|Y_i > 0, T_i = 0)$.

[5] To the extent that group membership is time invariant, the availability of panel data would identify the decomposition because individuals with transitions between 0 and positive outcomes could be identified as switchers. However, if group membership is partly determined by idiosyncratic shocks, panel data do not solve the identification problem.

$(0, \infty)$. Substituting the limits of these intervals into equation (4), the identification regions for the objects of interest reduce to

$$\text{ATE}^{S_1} \in \left(0, \text{ATE}/\pi^{S_1}\right), \quad \text{ATE}^{P} \in \left(0, \text{ATE}/\pi^{P}\right), \tag{5}$$

which are strictly smaller than the domains.

Now consider bounds under MTR2. MTR2 stated that $\pi^{S_2} = 0$ and that treatment effects for all i in group P had the same sign. Thus, this assumption allows both negative or positive ATE, and the bounds depend on ATE's sign.

If ATE $>$ 0, this is possible even with $\text{ATE}^P < 0$ (since ATE^{S_1} is positive by definition). In this case, the domain of ATE^P widens to $(-\bar{Y}_0^P, \infty)$, leading to a larger interval for ATE^P:

$$\text{ATE}^{S_1} \in \left(0, \text{ATE}/\pi^{S_1}\right), \quad \text{ATE}^{P} \in \left(-\bar{Y}_0^P, \text{ATE}/\pi^{P}\right). \tag{6}$$

In contrast, if ATE < 0, this means ATE^P must be negative with domain $(-\bar{Y}_0^P, 0)$. For ATE^{S_1} the domain is still the positive real line. The identification regions are

$$\text{ATE}^{S_1} \in \left(0, (\text{ATE} + \pi^P \bar{Y}_0^P)/\pi^{S_1}\right), \quad \text{ATE}^{P} \in \left(\text{ATE}/\pi^{P}, 0\right). \tag{7}$$

Thus, for a given absolute value of ATE, the bound for ATE^{S_1} is narrower in this case than in equations (5) and (6). Note that $-\bar{Y}_0^P = \text{ATE}/\pi^P - \pi^{S_1}\text{ATE}^{S_1}/\pi^P - \bar{Y}_1^P < \text{ATE}/\pi^P$, and therefore, as in equations (5) and (6), the bound on ATE^P in equation (7) is informative.[6]

The bounds in equations (5), (6), and (7) are sharp.[7] The line $\text{ATE}^{S_1} = \text{ATE}/\pi^{S_1} - \pi^P/\pi^{S_1}\text{ATE}^P$ in the $(\text{ATE}^{S_1}, \text{ATE}^P)$-plane provides sharp bounds on the components jointly.

III. Decomposing ATE in Some Structural Models

The decomposition based on joint potential outcomes is point-identified for two general classes of models often used in econometric studies, which impose some structure on the way treatment affects outcomes. These models do not require parametric restrictions on the distribution of the errors or on the way treatment (and other exogenous regressors) enters the model, although in practice, researchers sometimes prefer using parametric versions of these models. After briefly discussing the two nonparametric models, this section provides an illustrative example of the decomposition for the tobit model, an important parametric special case of both general models.

A. *Censored Regression Model*

Consider the model,

$$Y_i = \max[0, m(X_i) + U_i], \tag{8}$$

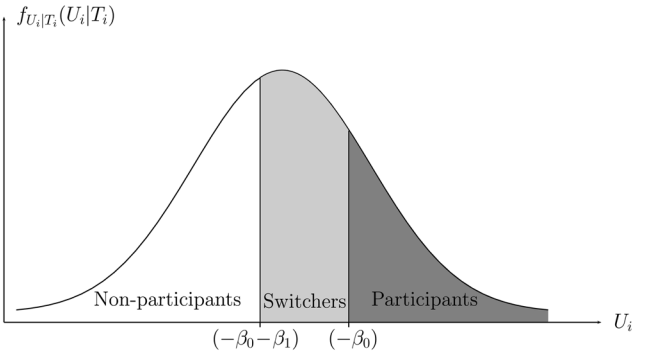
where X_i is a vector of regressors and U_i is an unobserved error independent of X_i with $\text{E}(U_i) = 0$; both the function $m(\cdot)$ and the distribution of the errors $F_U(\cdot)$ are unknown. Lewbel and Linton (2002) show that the unknown functions are identified and provide estimators for them.

In the simplest case considered here, X_i contains only the binary treatment indicator, so without loss of generality, $m(T_i) = \beta_0 + \beta_1 T_i$.

[6] Note that under MTR2, the bounds depend discontinuously on whether ATE is positive or negative. This might make inference difficult when ATE is close to 0.

[7] It is easy to verify that the following DGP, for instance, attains all values in the bounds: for a fraction π^{NP} of the population $Y_{1i} = Y_{0i} = 0$; for a fraction $\pi^{S_1}, Y_{1i} = \alpha_1 > 0, Y_{0i} = 0$; for a fraction $\pi^P, Y_{1i} = \beta_0 + \beta_1 > 0, Y_{0i} = \beta_0 > 0$; and $\pi^{NP} + \pi^{S_1} + \pi^P = 1$.

FIGURE 1.—POPULATION GROUPS BY U_i IN THE CENSORED REGRESSION MODEL



The censored regression model in this figure is $Y_i = \max(0, \beta_0 + \beta_1 T_i + U_i)$, with $\beta_1 > 0$.

Consider the case $\beta_1 > 0$, which implies that Y_i is nondecreasing in T_i, conforming to MTR1. Population group membership is completely determined by the realization of U_i, as illustrated in figure 1. The first column in table 1 contains some features of the decompositions in this model.[8] The population fractions of switchers and participants in the causal decomposition, equation (4), coincide with the terms used in the conventional decomposition, equation (2). However, decomposition (2) fails to attribute them the correct ATE. For instance, for switchers, decomposition (2) assigns as ATE the average Y_i in the population of switchers and participants (see the $\widetilde{\text{ATE}}^{S_1}$ row in table 1). Thus, this invariably overestimates their contribution, as switchers' U_i are in the bottom tail of the error distribution among those with $Y_{1i} > 0$.

B. *Selection Model and the Two-Part Model*

The selection model can be understood as a generalization of the censored regression model. One general version of it is

$$Y_i = \mathbb{1}[g(X_i, Z_i) + V_i \geq 0] \times (m(X_i) + U_i), \tag{9}$$

where $\mathbb{1}(\cdot)$ is the indicator function; U_i, V_i are independent of X_i, Z_i; and $m(X_i) + U_i$ is positive whenever the indicator function equals 1. Compared to equation (8), this model allows the probability of participation to be driven by a different function $g(\cdot)$ and by different unobservables. Thus, effects for switchers and participants might be differently signed, as in MTR2. In turn, identification requires an additional continuous variable, Z_i, which affects the participation probability only. Identification, regularity conditions, and estimation of equation (9) are discussed in Das, Newey, and Vella (2003). Tight parameterizations of equation (9), such as in Heckman (1979), can eliminate the need for Z_i for identification. The second column in table 1 displays the components of the decomposition corresponding to the selection model. The expressions closely resemble those corresponding to the censored regression model, but the increased flexibility stemming from having two errors can lead to the conventional decomposition over- or underestimating either causal margin.

An interesting restriction on the distribution of the errors is assuming that U_i and V_i are independent, in which case equation (9) is known as the two-part model (Cragg, 1971; Duan et al., 1983).[9] Then the conditional error expectations in table 1 are the same for switchers

[8] For compactness, table 1 uses the notation m_0 and m_1 for $m(0) = \beta_0$ and $m(1) = \beta_0 + \beta_1$.

[9] Z_i is not needed for identification under independence of U_i and V_i.

TABLE 1.—FEATURES OF DECOMPOSITIONS IN STRUCTURAL MODELS

	Censored Regression Model (Equation [8])	Selection Model (Equation [9])	Tobit Model (Equation [10])
Population fractions of groups (in both decompositions)			
π^{S_1}	$F_U(-m_0) - F_U(-m_1)$	$F_V(-g_0) - F_V(-g_1)$	$\Phi_1 - \Phi_0$
π^{P}	$1 - F_U(-m_0)$	$1 - F_V(-g_0)$	Φ_0
Group-specific ATE in causal decomposition (equation [4])			
ATE^{S_1}	$m_1 + \text{E}[U \mid -m_1 < U < -m_0]$	$m_1 + \text{E}[U \mid -g_1 < V < -g_0]$	$\beta_0 + \beta_1 + \sigma \frac{\phi_1 - \phi_0}{\Phi_1 - \Phi_0}$
ATE^{P}	$m_1 - m_0$	$m_1 - m_0$	β_1
Corresponding quantities in conventional decomposition (equation [2])			
$\widetilde{\text{ATE}}^{S_1}$	$m_1 + \text{E}[U \mid U > -m_1]$	$m_1 + \text{E}[U \mid V > -g_1]$	$\beta_0 + \beta_1 + \sigma\phi_1/\Phi_1$
$\widetilde{\text{ATE}}^{P}$	$m_1 - m_0 + \text{E}[U \mid U > -m_1] - \text{E}[U \mid U > -m_0]$	$m_1 - m_0 + \text{E}[U \mid V > -g_1] - \text{E}[U \mid V > -g_0]$	$\beta_1 + \sigma\phi_1/\Phi_1 - \sigma\phi_0/\Phi_0$

$m_T = m(T)$, $g_T = g(T, Z)$, $\Phi_T = \Phi((\beta_0 + \beta_1 T)/\sigma)$, and $\phi_T = \phi((\beta_0 + \beta_1 T)/\sigma)$ for $T = 0, 1$. $\Phi(\cdot)$ is the standard normal cdf, $\phi(\cdot)$ the standard normal pdf.

and participants, and the two decompositions coincide. However, the independence assumption might be unwarranted in most applications. While randomization prevents dependence between treatment and the errors, there is no experiment that could possibly break the potential dependence between U_i and V_i—and applications where the researcher can be certain that this dependence is absent seem difficult to envision.

C. An Example: Tobit Model

Consider a numerical example to illustrate the difference that using the decomposition based on joint potential outcomes can make. Suppose the DGP is the tobit model:

$$Y_i = \max[0, \beta_0 + \beta_1 T_i + U_i], \quad U_i \sim N(0, \sigma^2). \tag{10}$$

Features of the decompositions for equation (10) are depicted in the third column of table 1.

Let $\beta_0 = 0, \beta_1 = 1, \sigma^2 = 1$. Then the ATE is about 0.68. The conventional decomposition assigns about 0.24 to the intensive and 0.44 to the extensive margin effect. In contrast, the decomposition into causally meaningful margins reveals that 0.5 is due to the intensive and only 0.18 due to the extensive margin effect. The intensive margin contribution, which was only 36% using the old decomposition, is thus really 73%.[10]

Similarly stark discrepancies are possible in practice. For instance, McDonald and Moffitt's (1980) application examined the effect of a negative income tax on working hours reductions by estimating a tobit model. Using their decomposition, it assigned 22% of the estimated reduction in working hours to the extensive margin. A follow-up article by Moffitt (1982) reevaluated the same data. In this article, he modified the tobit model to account for a model of labor market frictions. Incidentally, this leads to the same formulas for the decomposition as the ones using the decomposition based on joint outcomes presented in table 1. Applying this decomposition, he now concluded that the extensive margin was responsible not for 22% but for 84% of the reduction. This paper shows that even in the absence or misspecification of the specific labor market frictions model postulated in Moffitt (1982), the causal extensive margin contribution is 84%.

IV. Conclusion

The analysis can easily be extended in several dimensions. First, the arguments presented here carry over to the average treatment effect on the treated if treatment is assumed independent from Y_0 only. Second, the models of section III readily allow for a continuous treatment and the inclusion of covariates. Third, the causal decomposition can be applied to other mixture models estimated from cross-section data where changes between outcomes of 0 and participation are of special interest; examples include hurdle models for count data or mixtures of a binary (participation) variable with an ordered response.

Finally, note that distributions of covariates X_i conditional on group membership are identified under MTR, since then random samples of nonparticipants and participants are available (individuals with $Y_i = 0, T_i = 1$ and with $Y_i > 0, T_i = 0$, respectively).[11] Such knowledge about group characteristics might be especially relevant for policies targeted at specific population groups.

[10] The bounds on $\text{ATE}^{S_1}(=1)$ and $\text{ATE}^{P}(=0.54)$ under MTR1 discussed in the previous section are $\text{ATE}^{S_1} \in (0; 1.99)$ and $\text{ATE}^{P} \in (0; 1.36)$.

[11] The distribution of X_i for switchers can be recovered indirectly from the unconditional distribution, the distributions for nonparticipants and for participants, and the population fractions.

REFERENCES

Angrist, Joshua D., "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics* 19:1 (2001), 2–16.

Angrist, Joshua D., and Jörn-Steffen Pischke, *Mostly Harmless Econometrics* (Princeton, NJ: Princeton University Press, 2009).

Cameron, A. Colin, and Pravin K. Trivedi, *Microeconometrics* (Cambridge: Cambridge University Press, 2005).

Cragg, John G., "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica* 39:5 (1971), 829–844.

Das, Mitali, Whitney K. Newey, and Francis Vella, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies* 70:1 (2003), 33–58.

Duan, Naihua, Willard G. Manning Jr., Carl N. Morris, and Joseph P. Newhouse, "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics* 1:2 (1983), 115–126.

Fan, Yanqin, and Sang So Park, "Partial Identification of the Distribution of Treatment Effects and Its Confidence Sets" (pp. 3–70), in Qi Li and Jeffrey S. Racine (eds.), *Nonparametric Econometric Methods* (Bingley, UK: Emerald Group Publishing, 2009).

Frangakis, Constantine E., and Donald B. Rubin, "Principal Stratification and Causal Inference," *Biometrics* 58 (2002), pp. 21–29.

Greene, William H., *Econometric Analysis*, 6th ed. (Upper Saddle River, NJ: Prentice Hall, 2008).

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica* 47:1 (1979), 153–161.

Heckman, James J., Jeffrey Smith, & Nancy Clements, "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies* 64 (1997), 487–535.

Lewbel, Arthur, and Oliver Linton, "Nonparametric Censored and Truncated Regression," *Econometrica* 70 (2002), 765–779.
Manski, Charles F., "Monotone Treatment Response," *Econometrica* 65 (1997), 1311–1334.
McDonald, John F., and Robert A. Moffitt, "The Uses of Tobit Analysis," this REVIEW 62 (1980), 318–321.
Moffitt, Robert A., "The Tobit Model, Hours of Work and Institutional Constraints," this REVIEW 64 (1982), 510–515.
Rubin, Donald B., "Causal Inference through Potential Outcomes and Principal Stratification: Application to Studies with 'Censoring' due to Death," *Statistical Science* 21 (2006), 299–309.
Tobin, James, "Estimation of Relationships for Limited Dependent Variables," *Econometrica* 26:1 (1958), 24–36.
Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data* (Cambridge, MA: MIT Press, 2002).