NIH Public Access
Author Manuscript
Econometrica. Author manuscript; available in PMC 2010 April 1.

Published in final edited form as:
Econometrica. 2010 January 1; 78(1): 377–394. doi:10.3982/ECTA7089.

Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin[1]

Pedro Carneiro,
Department of Economics University College London IFS CeMMAP Gower Street London WC1E 6BT United Kingdom p.carneiro@ucl.ac.uk phone: +44-20-76795824 fax: +44-20-79162775

James J. Heckman, and
Department of Economics The University of Chicago Geary Institute, University College Dublin Cowles Foundation, Yale University American Bar Foundation 1126 E. 59th Street Chicago, IL 60637 jjh@uchicago.edu phone: 773-702-0634 fax: 773-702-8490

Edward Vytlacil
Department of Economics Yale University P.O. Box 208281 New Haven, CT 06520-8281 edward.vytlacil@yale.edu phone: 203-436-3994 fax: 203-432-6167

Abstract

This paper develops methods for evaluating marginal policy changes. We characterize how the effects of marginal policy changes depend on the direction of the policy change, and show that marginal policy effects are fundamentally easier to identify and to estimate than conventional treatment parameters. We develop the connection between marginal policy effects and the average effect of treatment for persons on the margin of indifference between participation in treatment and nonparticipation, and use this connection to analyze both parameters. We apply our analysis to estimate the effect of marginal changes in tuition on the return to going to college.

Keywords

Marginal Treatment Effect; Effects of Marginal Policy Changes; Marginal Policy Relevant Treatment Effect; Average Marginal Treatment Effect

1 Introduction

The Policy Relevant Treatment Effect (PRTE) is the mean effect of changing from a baseline policy to an alternative policy that provides different incentives to participate in treatment (Heckman and Vytlacil, 2001b, 2005). Identification and estimation of the PRTE is generally difficult. Identification of the PRTE typically requires large support conditions. $\sqrt{N}$-consistent estimation of the PRTE parameter is generally not possible.

In many cases proposed policy reforms are incremental in nature and a marginal version of the PRTE (MPRTE) is all that is required to answer questions of economic interest. This paper develops the MPRTE and establishes how the MPRTE depends on the direction of a proposed

[1]This research was supported by NIH R01-HD32058-3, NSF SES-0832845, NSF SES-024158, NSF SES-05-51089, ESRC RES-000-22-2542, the Geary Institute at University College Dublin, the Leverhulme Trust, and ESRC (NIH R01-HD054702) through the funding of the Centre for Microdata Methods and Practice. The research was conducted in part while Edward Vytlacil was a Visiting Professor at Hitotsubashi University. We thank the editor, two anonymous referees, Hidehiko Ichimura, Richard Robb, Daniel Schmierer, and Azeem Shaikh for very helpful comments. We would like to thank Erica Blom and Sukjin Han for research assistance.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

marginal policy change. We establish that the support conditions required for identifying the MPRTE are very weak. The essential requirement is availability of a continuous instrument. The MPRTE parameter can be represented as a weighted average derivative with weights determined by the marginal policy of interest. The parameter is $\sqrt{N}$-estimable under standard regularity conditions. Thus, the MPRTE parameter is fundamentally easier to identify and estimate than the PRTE parameter.

We connect the MPRTE to the Average Marginal Treatment Effect (AMTE): the mean benefit of treatment for people at the margin of indifference between participation in treatment and nonparticipation. AMTE is compared to marginal cost in an aggregate benefit-cost analysis of a program. We establish a correspondence between MPRTE parameters and AMTE parameters, showing that the effect of a marginal policy change in a particular direction is the same as the average effect of treatment for those at the margin of indifference. We use this correspondence to produce new insights about the AMTE parameter.

The paper proceeds as follows. Section 2 presents the nonparametric selection model that underlies our analysis. Section 3 reviews the analysis of the PRTE by Heckman and Vytlacil (2005) and presents a new interpretation of the PRTE as a function from the space of all possible policies to the space of effects of policies. This interpretation is key to the analysis in Section 4 which introduces and analyzes the marginal policy relevant treatment effect (MPRTE). We discuss identification and estimation of the MPRTE in Sections 5 and 6, respectively. We define and analyze the AMTE in Section 7. Section 8 presents an empirical application of our analysis. Section 9 concludes.

2 Nonparametric Selection Model and Our Assumptions

Assume that there are two potential outcomes (Y_0, Y_1) and a binary treatment choice indicator D. Outcome Y is written in switching regression form $Y = DY_1+(1 - D)\, Y_0$, where Y_1 is the potential outcome that is observed if the agent chooses treatment 1, and Y_0 is the potential outcome that is observed if the agent chooses treatment 0. $Y_1 - Y_0$ is the individual level treatment effect. We keep implicit the conditioning on observed variables X that determine Y_1 or Y_0, and maintain the assumption that D does not determine X (see Heckman and Vytlacil, 2005).

Program participation is voluntary. To link this framework to standard choice models, we characterize the decision rule for program participation by a latent index model:

$$D=1\left[\mu\left(Z\right)-V\geq 0\right] \tag{2.1}$$

where $\mathbf{1}[\cdot]$ is the indicator function taking the value 1 if its argument is true and the value 0 otherwise. From the point of view of the econometrician, Z is observed and V is unobserved.

We maintain the following assumptions:

(A-1) (Y_0, Y_1, V) *is independent of Z;*

(A-2) The distribution of V is absolutely continuous with respect to Lebesgue measure;

(A-3) *The distribution of* $\mu_D(Z)$ is absolutely continuous with respect to Lebesgue measure;

(A-4) $\sup_v E\left(|Y_1| \mid V = v\right) < \infty$, $\sup_v E\left(|Y_0| \mid V = v\right) < \infty$;

(A-5) $0 < \Pr(D = 1) < 1$.

These conditions are discussed in Heckman and Vytlacil (2005, 2007). Under them, Vytlacil (2002) establishes the equivalence between the nonparametric latent index model (2.1) and the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

monotonicity assumption used by Imbens and Angrist (1994). A necessary condition for (A-3) is that Z contains a continuous variable (i.e., that there is a continuous instrument for D). We use this assumption to analyze marginal policy changes, and to identify marginal policy treatment effects. These conditions should be interpreted as conditional on X, which we have kept implicit. For example, (A-1) should be interpreted as saying that (Y_0, Y_1, V) is independent of Z conditional on X, while there can be dependence between (Y_0, Y_1, V) and X.

Define $P(Z)$ as the probability of receiving treatment given Z, $P(Z) \equiv \Pr(D = 1 \mid Z) = F_V(\mu(Z))$ where $F_V(\cdot)$ is the distribution function of V. We sometimes denote $P(Z)$ by P, suppressing the Z argument. We also use U, defined by $U = F_V(V)$ so that U is distributed unit uniform. In this notation, applying a monotonic transformation to both sides of the argument in equation (2.1) allows us to write that equation as $D = \mathbf{1}[P(Z) \geq U]$.

The marginal treatment effect (MTE) plays a fundamental role in our analysis. MTE is defined as $\text{MTE}(u) \equiv E(Y_1 - Y_0 \mid U = u)$, i.e., the expected treatment effect conditional on the unobservables which determine participation. For values of u close to zero, $\text{MTE}(u)$ is the expected effect of treatment on individuals who have unobservables that make them most likely to participate in treatment and who would participate even if the mean scale utility $\mu(Z)$ is small. See Heckman and Vytlacil (2005, 2007) for more discussion and interpretation of the MTE.

In some of our examples, we consider the following special case of our general model: (B-1) *Suppose* $\mu(Z) = Z\gamma$ and suppose that F_V is strictly increasing. Suppose that the kth component of Z, $Z^{[k]}$, *has a strictly positive coefficient,* $\gamma^{[k]} > 0$. *Let* $\tilde{Z}\tilde{\gamma} = Z\gamma - Z^{[k]}\gamma^{[k]}$. *Suppose that the distribution of* $Z^{[k]}\gamma^{[k]}$ conditional on $\tilde{Z}\tilde{\gamma}$ has a density with respect to Lebesgue measure.

3 The Policy Relevant Treatment Effect

We first present the Policy Relevant Treatment Effect (PRTE) of Heckman and Vytlacil (2001b, 2005) in order to motivate our marginal version of it. We then reformulate the PRTE as a function of the proposed policy change. This enables us to define a sequence of PRTEs corresponding to a sequence of proposed policy changes.

3.1 Review of PRTE

Following Heckman and Vytlacil (2001b, 2005), consider a class of policies that affect P, the probability of participation in a program, but that do not affect potential outcomes or unobservables related to the selection process, (Y_1, Y_0, U).[1] An example from the literature on the economic returns to schooling would be policies that change tuition or distance to school but that do not directly affect potential wages (Card, 2001). We ignore general equilibrium effects.

Let D^* be the treatment choice that would be made after the policy change. Let P^* be the corresponding probability that $D^* = 1$ after the policy change. D^* is defined by $D^* = \mathbf{1}[P^* \geq U]$. Let $Y^* = D^*Y_1 + (1 - D^*)Y_0$ be the outcome under the alternative policy. Following Heckman and Vytlacil (2005), the mean effect of going from a baseline policy to an alternative policy per net person shifted is the Policy Relevant Treatment Effect (PRTE), defined when $E(D) \neq E(D^*)$ as

[1]This restriction can be relaxed to a weaker policy invariance for the distribution of (Y_1, Y_0, U), see Heckman and Vytlacil (2005, 2007).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\frac{E(Y \mid \text{Alternative Policy}) - E(Y \mid \text{Baseline Policy})}{E(D \mid \text{Alternative Policy}) - E(D \mid \text{Baseline Policy})} = \frac{E(Y^*) - E(Y)}{E(D^*) - E(D)} = \int_0^1 \text{MTE}(u)\ \omega_{\text{PRTE}}(u)\, du \tag{3.1}$$

where

$$\omega_{\text{PRTE}}(u) = \frac{F_P(u) - F_{P^*}(u)}{E_{F_{P^*}}(P) - E_{F_P}(P)}. \tag{3.2}$$

The condition $E(D) \neq E(D^*)$ is consistent with the program having a non-monotonic effect on participation as long as the fraction switching into treatment is not exactly offset by the fraction switching out of treatment. The PRTE parameter gives the normalized effect of a change from a baseline policy to an alternative policy and depends on the alternative being considered.[2] Heckman and Vytlacil (2005) show that the PRTE can be identified under strong support conditions. In Section 5, we establish that the marginal version of the PRTE parameter can be identified under much weaker conditions than are required to identify the PRTE. In Section 6, we establish that, unlike the PRTE parameter, the MPRTE parameter is generally estimable at a $\sqrt{N}$ rate.

We define PRTE as the average effect per net person shifted into treatment. With this definition, no normalization is required when taking limits to define the MPRTE, and we obtain an equivalence between the MPRTE and the average effect for individuals at the margin of indifference (AMTE). Under this definition, the MPRTE corresponds to a weighted average derivative and can be estimated by a weighted average derivative estimator.

Alternatively, we could follow Heckman and Vytlacil (2001b) in defining PRTE without normalizing it by the net change in treatment status, i.e., we could define PRTE as $E(Y^*) - E(Y)$. However, some normalization is required when taking limits. If we were to define PRTE in that manner, and then normalize by $E(D^*) - E(D)$ when taking limits, we would obtain the same limits as would be obtained from analyzing the marginal version of PRTE defined by equation (3.1). Using the unnormalized version of PRTE, we can analyze alternative normalizations when taking limits. In footnote 3 below, we discuss one such normalization which results in an alternative version of MPRTE that is a rescaled version of the marginal version of PRTE defined by equation (3.1).

3.2 PRTE as a Function of Proposed Policy Changes

The PRTE depends on a policy change only through the distribution of P^* after the policy change. Given our assumptions, F_{P^*} is sufficient to summarize everything about the proposed policy change that is relevant for calculating the average effect of the policy change. We can thus define the PRTE function as a function mapping the proposed policy change (corresponding to a distribution of P^*) to the resulting per-person change in outcomes. Expressing the PRTE this way is important in the next step of our analysis that uses sequences of PRTEs to define a marginal version of the PRTE.

Let $\mathcal{G}$ denote the space of all cumulative distribution functions for random variables that lie in the unit interval such that $\int t dG(t) \neq \int t dF_P(t)$ i.e., all right-continuous, non-decreasing

[2]The PRTE can be interpreted as an economically more explicit version of Stock's (1989) nonparametric policy analysis parameter for a class of policy interventions with explicit agent preferences where the policies evaluated operate solely on agent choice sets.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

functions satisfying $G(t) = 1$ for $t \geq 1$, $G(t) = 0$ for $t < 0$, and such that the first moment of G does not equal the first moment of F_P. Any $G \in \mathcal{G}$ corresponds to a potential distribution of P^* and thus corresponds to a potential alternative policy, with PRTE(G) denoting the corresponding policy effect. We define the PRTE function, PRTE : $\mathcal{G} \mapsto \mathbb{R}$, by

$$\text{PRTE}(G) = \int_0^1 \text{MTE}(u)\, \omega_{\text{PRTE}}(u;G)\, du \tag{3.3}$$

where

$$\omega_{\text{PRTE}}(u, G) = \frac{F_P(u) - G(u)}{E_G(P) - E_{F_P}(P)}. \tag{3.4}$$

In many cases, the class of policy alternatives under consideration can be indexed by a scalar variable. Let $P_0 = P$ denote the baseline probability for $D = 1$. Let $\mathbf{M}$ denote a subset of $\mathbb{R}$ with $0 \in \mathbf{M}$. Let $\{P_\alpha : \alpha \in \mathbf{M}\}$ denote a class of alternative probabilities corresponding to alternative policy regimes with associated cumulative distribution functions F_α. For example, one alternative policy could increase the probability of attending college by an amount α, so that $P_\alpha = P_0 + \alpha$ and $F_\alpha(t) = F_0(t - \alpha)$. An alternative policy could change each person's probability of attending college by the proportion $(1 + \alpha)$, so that $P_\alpha = (1 + \alpha)P_0$ and $F_\alpha(t) = F_0\left(\frac{t}{1+\alpha}\right)$. The policy intervention might have an effect similar to a shift in one of the components of Z, say $Z^{[k]}$. In particular, suppose $Z_\alpha^{[k]} = Z^{[k]} + \alpha$ and $Z_\alpha^{[j]} = Z^{[j]}$ for $j \neq k$. For example, the kth element of Z might be college tuition, and the policy under consideration subsidizes college tuition by the fixed amount α. Suppose that the linear latent index assumption (B-1) holds. Then $P_\alpha(Z) = F_V(Z_\gamma + \alpha_\gamma^{[k]})$ and $F_\alpha(t) = F_{Z_\gamma}\left(F_V^{-1}(t) - \alpha\gamma^{[k]}\right)$. Notice that the first two examples have the form $P_\alpha = q_\alpha(P_0)$ for some function q_α, while the last example has the form $P_\alpha = F_V(Z^{[-k]}\gamma^{-[k]} + q_\alpha(Z^{[k]}\gamma^{[k]})$. We will explore more general examples in the next section.

4 Marginal Policy Changes

The PRTE is defined for a discrete change from a baseline policy to a fixed alternative. We now consider a marginal version of the PRTE parameter that corresponds to a marginal change from a baseline policy. It is expositionally convenient to think of the treatment as college attendance and the policy as a change in tuition. The marginal version of the PRTE depends on the nature of the perturbation that defines the marginal change. For example, a policy change that subsidizes tuition by a fixed amount has different effects than a policy change that subsidizes tuition by a fixed proportion. The limits of these two policies for infinitesimally small subsidies are different.

To define the marginal version of the PRTE, we could consider the limit of $\Delta^{\text{PRTE}}(G)$ as G gets close to F_P in some metric. We could define the marginal PRTE as a directional derivative. For ease of exposition, we do not work with this more general formulation but instead work with a one-dimensional version of it. Thus we do not analyze general perturbations within the function space $\mathcal{G}$ but only one-dimensional curves within $\mathcal{G}$.

Consider a class of alternative policies indexed by α, $\{F_\alpha : \alpha \in \mathbf{M}\}$, where 0 is a limit point of $\mathbf{M}$, and 0 represents the baseline policy so that $F_0 = F_P$. Consider the effect of a marginal change in α in a neighborhood of the current base policy of $\alpha = 0$. When the limit exists, we can define the marginal policy relevant treatment effect as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\text{MPRTE}(\{F_\alpha\}) = \lim_{\tau \to 0} \text{PRTE}(F_\tau),$$

where the MPRTE parameter depends on the class of alternative policies $\{F_\alpha : \alpha \in \mathbf{M}\}$, i.e., on the choice of a particular curve within $\mathscr{G}$. Thus, the MPRTE can be seen as a path-derivative along the path $\{F_\alpha : \alpha \in \mathbf{M}\}$. We impose the following sufficient conditions for the limit to exist: (A-6) *For α in a neighborhood of* 0, F_α has a density f_α with respect to Lebesgue measure, f_α is differentiable in α, $\sup_{t \in (0,1)} \left| \frac{\partial}{\partial \alpha} f_\alpha(t) \right| < \infty$, *and* $\int_0^1 t \frac{\partial}{\partial \alpha} f_\alpha(t)\, dt \neq 0$.

Under assumptions (A-1)-(A-6), MPRTE($\{F_\alpha\}$) = $\lim_{\tau \to 0}$ PRTE(F_τ) exists and is given by

$$\begin{aligned} &\text{MPRTE}(\{F_\alpha\}) \\ &\quad = \int_0^1 \text{MTE}(u)\ \omega_{\text{MPRTE}}\ (u; \{F_\alpha\})\ du, \end{aligned} \tag{4.1}$$

where $\omega_{\text{MPRTE}}(u; \{F_\alpha\})$ is

$$\begin{aligned} &\omega_{\text{MPRTE}}(u; \{F_\alpha\}) \\ &\quad = -\frac{\frac{\partial}{\partial \alpha} F_0(u)}{\frac{\partial}{\partial \alpha} E_{F_0}(P)} \\ &\quad = -\frac{\int_0^u \left(\frac{\partial}{\partial \alpha} f_0(p) \right) dp}{\int_0^1 p \left(\frac{\partial}{\partial \alpha} f_0(p) \right) dp}, \end{aligned} \tag{4.2}$$

where we write $\frac{\partial}{\partial \alpha} F_0(p)$ and $\frac{\partial}{\partial \alpha} f_0(p)$ as shorthand expressions for $\frac{\partial}{\partial \alpha} F_0(p)|_{\alpha=0}$ and $\frac{\partial}{\partial \alpha} f_\alpha(p)|_{\alpha=0}$, respectively. An alternative way to express the form of the weights uses the property that $0 \leq P_\alpha \leq 1$ so that $E(P_\alpha) = \int_0^1 (1 - F_\alpha(t))\, dt$ to obtain

$$\begin{aligned} &\omega_{\text{MPRTE}}(u; \{F_\alpha\}) \\ &\quad = \frac{\frac{\partial}{\partial \alpha} F_0(u)}{\int_0^1 \left(\frac{\partial}{\partial \alpha} F_0(t) \right) dt}, \end{aligned} \tag{4.3}$$

which makes it clear that the weights always integrate to unity.[3]

Just as the PRTE parameter depends on which particular policy counterfactual is being considered, the marginal PRTE parameter depends on which particular class of policy perturbations is being considered. Just as the effect of a fixed dollar amount tuition subsidy

[3]An alternative way to define MPRTE that does not require the condition $E(D) \neq E(D^*)$ normalizes by α instead of $E(D) - E(D^*)$: $\lim_{\alpha \to 0} \int_0^1 \text{MTE}(u) \frac{[F_\alpha(u) - F_0(u)]}{\alpha} du$. With this definition the corresponding weights on MTE are given by $\omega_{\text{MPRTE}}\ \omega_{\text{MPRTE}}\ (u; \{F_\alpha\}) = \frac{\partial}{\partial \alpha} F_0(u)$. This change in the normalization only affects the constant of integration for the weights, and results in an alternative MPRTE that is a rescaled version of the MPRTE analyzed in this paper.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

will result in a different PRTE parameter than a proportional tuition subsidy, the limit parameter for a marginal change in an additive tuition subsidy will be different from the limit parameter for a marginal change in a proportional tuition subsidy. Of particular interest to us are the following two cases for which we will define classes of functions. Let $\mathcal{Q}$ denote the set of sequences of functions $\{q_\alpha(\cdot) : \alpha \in \mathbf{M}\}$ such that $q_\alpha(t)$ is strictly increasing in t for any $\alpha \in \mathbf{M}$; such that $q_0(\cdot)$ is the identity function; such that $q_\alpha(t)$ is differentiable in α with $\frac{\partial}{\partial\alpha} q_\alpha(t)$ bounded; such that r_α is a bounded function where r_α denotes the inverse of q_α; and such that r_α is differentiable in α with $\frac{\partial}{\partial\alpha} r_\alpha(t)$ bounded. The following examples can be trivially modified to allow q_α to be monotonically decreasing.

(Ex-1) Suppose that the alternative sequence of policies have the form $P_\alpha = q_\alpha(P_0)$ for some $\{q_\alpha\} \in \mathcal{Q}$. Then $F_\alpha(t) = F_P(r_\alpha(t))$, and

$$\begin{aligned}\omega_{MPRTE}(u;\{F_\alpha\}) &= -\frac{f_P(u)\frac{\partial}{\partial\alpha} r_0(u)}{E_P\left(\frac{\partial}{\partial\alpha} q_0(P)\right)} \\ &= \frac{f_P(u)\frac{\partial}{\partial\alpha} r_0(u)}{\int_0^1 f_P(t)\frac{\partial}{\partial\alpha} r_0(t)\,dt}.\end{aligned} \tag{4.4}$$

(Ex-2) Suppose that the alternative sequence of policies shifts the kth component of Z, $Z^{[k]}$, to $q_\alpha(Z^{[k]})$ *for some* $\{q_\alpha\} \in \mathcal{Q}$. *Suppose that Assumption (B-1) holds. Then*

$$F_\alpha(t) = \int F_{\underset{\sim}{Z}^{[k]}|\underset{\sim}{Z}\gamma}\left[r_\alpha\left(\frac{F_V^{-1}(t)-s}{\gamma^{[k]}}\right)\right] f_{\underset{\sim}{Z}\gamma}(s)\,ds,$$

and

$$\begin{aligned}\omega_{MPRTE}(u;\{F_\alpha\}) &= \frac{E_{\underset{\sim}{Z}\gamma}\left[f_{\underset{\sim}{Z}^{[k]}|\underset{\sim}{Z}\gamma}\left[\frac{F_V^{-1}(u)-\underset{\sim}{Z}\gamma}{\gamma^{[k]}}\right]\frac{\partial}{\partial\alpha} r_0\left(\frac{F_V^{-1}(u)-\underset{\sim}{Z}\gamma}{\gamma^{[k]}}\right)\right]}{\int_0^1 E_{\underset{\sim}{Z}\gamma}\left[f_{\underset{\sim}{Z}^{[k]}|\underset{\sim}{Z}\gamma}\left[\frac{F_V^{-1}(t)-\underset{\sim}{Z}\gamma}{\gamma^{[k]}}\right]\frac{\partial}{\partial\alpha} r_0\left(\frac{F_V^{-1}(t)-\underset{\sim}{Z}\gamma}{\gamma^{[k]}}\right)\right]dt}.\end{aligned} \tag{4.5}$$

The expressions in equations (4.4) and (4.5) look different, but both can be represented as weighted densities

$$\begin{aligned}\omega_{\mathrm{MPRTE}}(u;\{F_\alpha\}) &= f_P(u)\,h(u)/E(h(P))\end{aligned} \tag{4.6}$$

for some function $h(\cdot)$. For (Ex-1), this form is immediate, substituting $h(u) = \frac{\partial}{\partial\alpha} r_0(u)$. For (Ex-2), note that, for $\omega_{\mathrm{MPRTE}}(u;\{F_\alpha\})$ given by equation (4.5),

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\omega_{\text{MPRTE}}(u;\{F_\alpha\}) \neq 0 \Rightarrow \int f_{\underset{Z[k]|Z\gamma}{--}}\left[\frac{F_V^{-1}(u)-s}{\gamma[k]}\right] f_{\underset{Z\gamma}{--}}(s)\,ds \neq 0$$
$$\iff f_{Z\gamma}\left(F_V^{-1}(u)\right) \neq 0$$
$$\iff f_P(u) f_V\left(F_V^{-1}(u)\right) \neq 0$$
$$\Rightarrow f_P(u) \neq 0,$$

$\Pr[P(Z) \leq u]$ where we use the fact that $= \Pr\left[Z\gamma \leq F_V^{-1}(u)\right]$ and the chain rule to obtain $f_P(u) = \frac{f_{Z\gamma}\left(F_V^{-1}(u)\right)}{f_V\left(F_V^{-1}(u)\right)}$. Thus, the weights for (Ex-2) are of the form $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = f_P(u)h(u)/C$ for some function h and some constant C. The weights integrate to one, because $C = \int f_P(u)\,h(u)\,du = E(h(P))$. Thus the weights for both (Ex-1) and (Ex-2) are of the form (4.6). This expression plays an important role when we consider identification and estimation of the MPRTE.

Special cases include $q_\alpha(t) = t + \alpha$ and $q_\alpha(t) = (1 + \alpha)t$, for policy changes that act like constant shifts or proportional shifts either in P or in a component of Z. For example 1, if $q_\alpha(t) = t + \alpha$ then $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = f_P(u)$, so that the MPRTE weights MTE according to the density of $P(Z)$. In contrast, if $q_\alpha(t) = (1 + \alpha)t$, then $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) = uf_P(u)/E(P)$. Thus, for example, the limit form associated with increasing the probability of college attendance by a fixed amount and the limit form associated with increasing the probability of college attendance by a proportional amount produce different weights on MTE. The limit of the latter puts higher weight on MTE for higher u evaluation points, i.e., puts higher weight on MTE for individuals whose unobservables make them less likely to go to college. For example 2, setting $q_\alpha(t) = t$

$$\omega_{\text{MPRTE}}(u;\{F_\alpha\}) = \frac{f_{Z\gamma}\left(F_V^{-1}(u)\right)}{E\left(f_V(Z\gamma)\right)}$$

$+ \alpha$ results in $= \frac{f_P(u) f_V\left(F_V^{-1}(u)\right)}{E\left(f_V(Z\gamma)\right)}$, which again will weight the MTE and thus weight people with different unobserved preferences for treatment differently from the way MTE is weighted in the other examples.

5 Identification

Equations (3.3)-(3.4) and (4.1)-(4.2) show that both the PRTE and MPRTE treatment parameters can be expressed in the form

$$\text{Treatment Parameter}(j) = \int_0^1 \text{MTE}(u)\,\omega_j(u)\,du, \tag{5.1}$$

where $\omega_j(u)$ is the weighting function corresponding to treatment parameter j, and where ω_j will depend on which policy change/marginal policy change is being considered. Heckman and Vytlacil (1999,2005) show that, under our assumptions, the standard treatment parameters can also be expressed in this form and that the form can be used for identification. In particular, the weights are often easy to identify, in which case equation (5.1) implies identification of the parameter if we can identify the MTE at any evaluation point for which the weights are non-zero. Under our assumptions, Heckman and Vytlacil (1999,2005) show that the MTE can be identified by the method of local instrumental variables (LIV) at any p in the support of $P(Z)$:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\frac{\partial}{\partial p}E(Y|P(Z)=p)=\text{MTE}(p). \tag{5.2}$$

Any parameter that can be represented as a weighted average of MTE can be identified by first estimating MTE over the appropriate support and then integrating the identified MTE function using the appropriate weights. In order to identify parameter j, not only does $P(Z)$ have to be a continuous random variable, but in addition the support of $P(Z)$ must contain all values of u such that $\omega_j(u) \neq 0$. For the standard treatment parameters, identification requires strong conditions on the support of the distribution of $P(Z)$. For example, for the average treatment effect, $E(Y_1 - Y_0)$, the weights are given by $\omega_{ATE}(u) = 1$ for $u \in [0, 1]$ so that the required support condition is that the support of $P(Z)$ is the full unit interval. It is possible to point-identify the parameters under only slightly weaker conditions than those required by this strategy, as shown in Heckman and Vytlacil (2001a,2007).

Consider identification of PRTE. Suppose that we can identify F_{P*} and thus the weights. In order to identify the parameter using the strategy just discussed, the support of $P(Z)$ must contain all values of u such that $F_P(u) - F_{P*}(u) \neq 0$. Thus, if the support of P^* is not contained in the support of P, it is not possible to identify the PRTE.

For example, suppose that the largest estimated probability of attending college is strictly less than 1. For analyzing a tuition subsidy policy, it is possible that the largest probability of attending college under a tuition subsidy will be greater than the largest probability of attending college without a tuition subsidy so that the support condition for identifying the corresponding PRTE parameter is violated. More generally, in examples (Ex-1) and (Ex-2) in Section 4, the PRTE parameters will not be identified unless the support of P is the full unit interval. Thus, the very strong support conditions required for identification of the standard treatment parameters are also required for identification of the PRTE parameter.

In contrast, the MPRTE parameter can generally be identified under weaker assumptions. The MPRTE weights are non-zero only if the density of $P(Z)$ is nonzero. Consider the classes of MPRTE parameters produced from examples (Ex-1) and (Ex-2). For policy counterfactuals that act like transformations of $P(Z)$ as in (Ex-1), or act like transformations of Z as in (Ex-2), we have from equation (4.6) that $\omega_{\text{MPRTE}}(u; \{F_\alpha\}) \neq 0 \Rightarrow f_P(u) \neq 0$, so that the MPRTE parameters of (Ex-1) and (Ex-2) can be identified without any additional support conditions even though the corresponding PRTE parameters cannot be identified for any α. In these examples, the marginal PRTE parameters place positive weight on MTE(u) for values of u where the density of P is positive, i.e., they only place positive weight on MTE(u) for values of u where we can use local instrumental variables to point identify MTE(u). Thus, even though in each example the PRTE is not identified for any value of $\alpha \neq 0$ without large support assumptions, the marginal PRTE is identified using the assumption that P is a continuous random variable. The MPRTE parameter is thus fundamentally easier to identify than either the conventional treatment parameters or the PRTE parameters.[5]

[4]LIV can be interpreted as the limit form of the Imbens and Angrist (1994) LATE parameter (see Heckman and Vytlacil, 1999). The ideas of the marginal treatment effect and the limit form of LATE were first introduced in the context of a parametric normal generalized Roy model by Björklund and Moffitt (1987), and were analyzed more generally in Heckman (1997). Angrist, Graddy, and Imbens (2000) also define and develop a limit form of LATE.

[5]As is clear from our analysis, Assumption (A-4) can be relaxed to only require that $E(|Y_1| \mid V = v)$ and $E(|Y_0| \mid V = v)$ are bounded for v contained in any arbitrarily small enlargement of the support of $F_V^{-1}(P(Z))$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

6 Issues in Estimation

In addition to the support requirement, an additional difficulty in estimating the standard treatment parameters is that under our assumptions they are not $\sqrt{N}$-estimable. Suppose that P is known. Using equations (5.1) and (5.2), we obtain

$$\text{Treatment Parameter } (j) = \int \frac{\partial}{\partial p} E(Y|P(Z)=p)\ \omega_j(p)\ dp = E\left(g'(P)\, q_j(P)\right), \tag{6.1}$$

where $g'(p) = \frac{\partial}{\partial p} E(Y \mid P(Z)=p)$ and $q_j(p) = \omega_j(p)/f_P(p)$, and we assume that $w_j(p) \neq 0 \Rightarrow f_P(p) \neq 0$ as is required for identification. A critical requirement for weighted average derivative estimators to be $\sqrt{N}$-consistent is that $q_j(p)f_P(p)$ vanish on the boundary of the support of P (see Newey and Stoker, 1993). Requiring that $q_j(p)f_P(p) = 0$ on the boundaries of the support of $P(Z)$ is equivalent to requiring that $\omega_j(p) = 0$ at the boundaries of the support of P. This requirement is not satisfied by conventional treatment parameters. For example, the average treatment effect imposes the requirement that $\omega_{\text{ATE}}(p) = 1\ \forall\ p \in [0, 1]$, so that $\omega_{ATE}(p) \neq 0$ at the boundaries of the support and $E(g'(P)q_{\text{ATE}}(P))$ is not $\sqrt{N}$-estimable under our assumptions. A parallel analysis applies to the PRTE parameters. From equation (3.4), $\omega_{PRTE}(p;G) = 0$ at the boundaries of the support of P only if $F_P(t) - G(t) = 0$ at the boundaries of the support of P. This requires equality of the supremum of the supports of the two distributions, so that a necessary condition for $\sqrt{N}$-consistent estimation of the PRTE is that the policy counterfactual does not increase or decrease the largest value of the probability of participation.

In contrast, the MPRTE parameters are $\sqrt{N}$-estimable under standard regularity conditions. Consider examples of the form of (Ex-1), policy changes that act like transformations of the P, or examples of the form of (Ex-2), policy changes that act like transformations of Z. From equation (4.6), $\omega_{MPRTE}(\cdot; \{F_\alpha\}) \neq 0$ only if $f_P(\cdot) \neq 0$, and so we again have that the weights will equal zero on the boundaries of the support of $P(Z)$ if f_P goes to zero on the boundary of the support. In each of these cases, the marginal PRTE parameter is given by a weighted average derivative, $E(g'(P)q(P))$ with weights such that $q_j(p)f_P(p)$ equals zero on the boundary of the support of P if $f_P(p) = 0$ on the boundary of the support of P. Thus, each of these parameters is $\sqrt{N}$-estimable under appropriate regularity conditions (see Newey and Stoker, 1993).

Our discussion thus far ignores three issues. First, in general $P(Z)$ is not known but must be estimated. Second, the weights sometimes need to be estimated. Third, we have not analyzed how to deal with observed regressors X that enter the outcome equations for Y_0 and Y_1. If X contains only discrete elements, then the estimation theory just presented is still valid conditional on X. If X contains continuous elements, then we may instead work with marginal PRTE parameters averaged over the distribution of X to obtain $\sqrt{N}$-estimates. One example of an estimator of the weighted average derivative that applies to our problem is the sieve minimum distance estimator of Ai and Chen (2007). We next consider identification and estimation of average marginal treatment effects.

7 The Average Marginal Treatment Effect

We now relate the MPRTE parameter to the Average Marginal Treatment Effect (AMTE), the average effect of treatment for the marginal person who is indifferent between participation and nonparticipation. More precisely, for a given choice of how to measure the distance between P and U, and thus for a given choice of how to measure how close an individual is to being indifferent between treatment or not, we define the AMTE as the average effect of treatment for those who are arbitrarily close to being indifferent between treatment or not.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

While the MPRTE depends on the direction of the marginal policy change, the AMTE parameter depends on the metric by which one measures how close individuals are to being indifferent. We show an equivalence between the MPRTE and AMTE parameters. Choosing a particular distance measure for the AMTE is equivalent to examining a particular policy direction for the MPRTE. The effect of a marginal policy change in a particular direction is equal to the average effect of treatment for those at the margin of indifference in the precise sense that a marginal policy change in that direction would change their treatment participation decision.

Consider the average effect of treatment for those who are close to being indifferent between treatment or not. For any metric $m(\cdot, \cdot)$, we have

$$E(Y_1 - Y_0 | m(P, U) \leq \epsilon) = \int_0^1 \text{MTE}(u) \frac{\Pr[m(P, u) \leq \epsilon]}{\Pr[m(P, U) \leq \epsilon]} du.$$

Suppose the metric $m(\cdot, \cdot)$ is such that, for some strictly monotonic and differentiable function q, $m(P, U) = |q(P) - q(U)|$. Let r denote the inverse of q. Then, under the regularity conditions that allow us to interchange limits and integration, the average effect for those arbitrarily close to indifference between treatment or not is

$$\lim_{\epsilon \to 0} E(Y_1 - Y_0 | m(P, U) \leq \epsilon) = \int_0^1 \text{MTE}(u) \omega_{AMTE}(u) du,$$

where

$$\omega_{AMTE}(u) = \frac{f_P(u) r'(u)}{\int f_P(u) r'(u) du} = \frac{f_P(u) r'(u)}{E(r'(P))}. \quad (7.1)$$

The form of the weights depends on the choice of a metric. Choosing different ways to measure the distance between P and U, and thus different ways of measuring how close individuals are to being indifferent between treatment or not, produces different weighting functions. For any choice of q, the weights will be positive, will integrate to unity, and will only be non-zero where the density of P is non-zero, but otherwise the weights can be shifted around arbitrarily by picking alternative metrics. One might think that there is a natural way to pick a particular metric. Taking $m(P, U) = |P - U|$ seems more natural than taking $m(P, U) = |q^P - q^U|$ for general $q(\cdot)$ including $F_V^{-1}(\cdot)$. However, under assumption (B-1), it is not any more natural to pick $m(P, U) = |F_V^{-1}(P) - F_V^{-1}(U)| = |Z\gamma - V|$. Yet, the two choices of metrics $(P, U) = |P - U|$ than it is to pick give different limit results.

Ambiguity over which metric to use for the AMTE can be resolved by connecting the AMTE to the marginal PRTE. The connection is natural. Marginal policy changes only affect people who are indifferent between treatment or not. Thus there is a close connection between the two types of parameters. An arbitrary choice of a metric to define AMTE lacks an economic motivation. The choice of which marginal policy to study to define the AMTE is well

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

motivated. Comparing equations (4.6) and (7.1), we see a duality between MPRTE parameters and alternative definitions of the AMTE. For example, note that AMTE taking q to be the identity function is exactly the MPRTE expression for policy alternatives of the form $P_\alpha = P + \alpha$. If (B-1) holds, then AMTE taking $q(\cdot) = F_V^{-1}(\cdot)$ is exactly the MPRTE expression for policy alternatives of the form of changing one component of Z by α. Different policy alternatives define different AMTE parameters.

Consider an analysis of college attendance. Each marginal PRTE parameter defines a different set of "marginal people" who are indifferent between college participation or not and who would change their college participation in response to a marginal change in the policy. Notice that any inframarginal individual with $\mu(Z) > V$ will not be affected by any marginal policy change. Only individuals with $\mu(Z) = V$ will be affected by a marginal change in the policy. Each marginal PRTE parameter will correspond to a different valid definition of the average marginal treatment effect. The average effect of college attendance on those who are on the margin defined by an infinitesimal level shift in college tuition is different from the average effect of college attendance on those who are on the margin defined by an infinitesimal proportional shift in college tuition.

Since each average marginal treatment effect can be equated with a marginal PRTE parameter, we can use the analysis of Section 4 to define the AMTE so that it can be identified without large support assumptions. In addition, the parameter will be $\sqrt{N}$-estimable under appropriate regularity conditions. Thus, under our assumptions, the AMTE parameter is fundamentally easier to identify and to estimate than are the conventional treatment parameters.

8 Applying the Analysis to Data

Following Carneiro, Heckman, and Vytlacil (2006), we estimate the MTE for a sample of white males from the NLSY.[6] Using this sample, we also estimate the weights needed to construct the MPRTE parameters corresponding to policy perturbations in alternative directions. We group individuals in two groups: persons with a high school education or below who do not go to college ($D = 0$) and persons with some college or above ($D = 1$). The outcome variable is the log of the average of non-missing values of the hourly wage between 1989 and 1993, which we interpret as an estimate of the log hourly wage in 1991. We estimate the selection probability $P(Z)$ using a logit. We assume that $Y_1 = X\beta_1 + U_1$, $Y_0 = X\beta_0 + U_0$, and that U_1 and U_0 are independent of X, so that the marginal treatment effect is given by $MTE(x, u) = x(\beta_1 - \beta_0) + E(U_1 - U_0 | U = u)$.[7]

In order to compute the MTE, we estimate $E(Y|X, P) = X\beta_0 + PX(\beta_1 - \beta_0) + K(P)$, and we take its derivative with respect to P: $\text{MTE}(x, p) = \frac{\partial E(Y|X=x, P=p)}{\partial p}$. We estimate β_1, β_0 applying the partially linear regression method of Robinson (1988). $K(\cdot)$ is estimated using locally quadratic regression.[8] We trim 2% of the observations from the extremes of the distribution of P, which means that we are only able to identify the MTE at evaluation points between 0.1 and 0.93. Thus, we can only identify parameters over this support. We present annualized

[6] See the Web Appendix A at http://jenni.uchicago.edu/evaluating_marginal_policy/.

[7] The components of Z are: AFQT and its square, mother's years of schooling and its square, permanent local earnings (average earnings between 1973 and 2000) in the county of residence at age 17 and its square, permanent local unemployment (average unemployment between 1973 and 2000) in state of residence at age 17 and its square, the presence of a four year college in the county of residence at age 17, average tuition in public four year colleges in the county of residence at age 17, log average wage in the county of residence at age 17, unemployment rate in the state of residence at age 17. The latter four variables are excluded from the set of variables in the wage equations (X) for $D = 1$ and $D = 0$, which are: years of experience and its square, log average wage in the county of residence in 1991, unemployment rate in the state of residence in 1991, AFQT and its square, mother's years of schooling and its square, permanent local earnings (average earnings between 1973 and 2000) in the county of residence at age 17 and its square, and permanent local unemployment (average unemployment between 1973 and 2000) in state of residence at age 17 and its square.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

returns, obtained by dividing the MTE (and the parameters it generates) by 4 (corresponding to 4 years of college).

Figure 1 plots three alternative MPRTE weights. We take two cases based on example 1, where $q_\alpha^A(P) = P + \alpha$ and $q_\alpha^C(P) = (1+\alpha)P$. A third case is taken from example 2 (*B*), where $q_\alpha^B\left(Z^{[k]}\right) = Z^{[k]} + \alpha$.[9] ω^C_{MPRTE} puts more weight than ω^A_{MPRTE} on higher values of *U*, and both sets of weights are zero in the region of *U* for which we cannot identify the MTE. The intuitive reason for this pattern is that a proportional change in *P* results in larger absolute changes for high levels of *P* than for low levels of *P*, while an additive change in *P* results in uniform changes in *P* across the whole distribution. The MPRTE for a marginal proportional change in *P* is estimated to be 0.1296 while the MPRTE for a marginal additive change in *P* is estimated to be 0.0880, since the former puts more weight on high values of *U* where the MTE is estimated to be smaller.ω^B_{MPRTE} puts more weight than either ω^A_{MPRTE} or ω^C_{MPRTE} in the center of the support of *U*, and the MPRTE is estimated to be 0.1274. Figure 1 also shows that corresponding to each MPRTE weight there is an AMTE weight, defined by different choices of metric for measuring the distance between *P* and *U*. For case *A*, $m(P, U) = |P - U|$. For case *B*, $m(P, U) = \left|F_V^{-1}(P) - F_V^{-1}(U)\right| = |Z\gamma - V|$. For case *C*, $m(P, U) = \left|\frac{P}{U} - 1\right|$.

9 Summary and Conclusions

This paper extends the analysis of Heckman and Vytlacil (1999, 2005, 2007) by using the marginal treatment effect (MTE) to identify the effect of a marginal policy change, and to identify the average effect of treatment on individuals who are indifferent between treatment or not. Conventional treatment parameters and the Policy Relevant Treatment Effect require large support conditions for identification, and often are not $\sqrt{N}$-estimable. Under our assumptions the marginal policy effect parameters and the average marginal treatment effects are generally identified without large support conditions and are $\sqrt{N}$-estimable. An analysis of the effect of marginal changes in tuition policies on college attendance illustrates the empirical relevance of this analysis.

References

Ai C, Chen X. Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. Journal of Econometrics 2007;141:5–43.

Angrist J, Graddy K, Imbens G. The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish. Review of Economic Studies 2000;67:499–527.

Björklund A, Moffitt R. The Estimation of Wage Gains and Welfare Gains in Self-Selection. Review of Economics and Statistics 1987;69:42–49.

[8] We first run kernel regressions of each *X* on $\hat{P}$ using a bandwidth of 0.05 and trimming 2% of the observations, and we compute the resulting residuals. We then estimate a linear regression of the outcome variable on these residuals and obtain estimates $\hat{\beta}_1$ and $\hat{\beta}_0$. In order to estimate $K(\cdot)$ and its first derivative we run a locally quadratic regression of $Y - X\hat{\beta}_0 + \hat{P}X\hat{\beta}_1 - \beta_0)$ on $\hat{P}$, using a bandwidth of 0.322 (determined by cross validation), and trimming 2% of the observations. $\hat{K'}(\cdot)$ is constructed by taking the coefficient on the linear term of the locally quadratic regression.

[9] In the empirical work we impose an index sufficiency restriction when estimating $f_{P|X}$: $f_{P|X\delta}$. To estimate the scalar index $X\delta$, we use the fact that $f_{P|X} = f_{P|X\delta}$ implies that $E(P|X) = E(P|X\delta$ to estimate δ using semi-parametric least squares (Ichimura, 1993). We estimate $f_{P|X\delta}$ by running a locally linear regression of $\left(\frac{1}{h}\right) * K\left(\frac{\widehat{P} - p}{h}\right)$ on $X\hat{\delta}$, where $K(\cdot)$ is a standard normal density and $h = 1.06 * \left[\widehat{Var}(P)\right]^{\frac{1}{2}} * n^{(-\frac{1}{5})}$. We use a bandwidth equal to $1.06 * \left[\widehat{Var}(I(X\delta))\right]^{\frac{1}{2}} * n^{(-\frac{1}{5})}$. The figure fixes $X\delta$ at $\bar{X}\delta$ The weights are rescaled so that both the MTE and the weights fit in the figure.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Card D. Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. Econometrica 2001;69:1127–1160.

Carneiro, P.; Heckman, JJ.; Vytlacil, EJ. Estimating Marginal and Average Returns to Education. University of Chicago, Department of Economics; 2006. Unpublished manuscript

Heckman JJ. Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. Journal of Human Resources 1997;32:441–462. addendum published vol. 33 no. 1 (Winter 1998).

Heckman, JJ.; Vytlacil, EJ. Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects; Proceedings of the National Academy of Sciences; 1999. p. 4730-4734.

Heckman, JJ.; Vytlacil, EJ. Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect. In: Lechner, M.; Pfeiffer, F., editors. Econometric Evaluation of Labour Market Policies. Center for European Economic Research; New York: 2001a. p. 1-15.

Heckman JJ, Vytlacil EJ. Policy-Relevant Treatment Effects. American Economic Review 2001b; 91:107–111.

Heckman JJ, Vytlacil EJ. Structural Equations, Treatment Effects and Econometric Policy Evaluation. Econometrica 2005;73:669–738.

Heckman, JJ.; Vytlacil, EJ. Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments. In: Heckman, J.; Leamer, E., editors. Handbook of Econometrics. Vol. 6B. Elsevier; Amsterdam: 2007. p. 4875-5144.

Ichimura H. Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. Journal of Econometrics 1993;58:71–120.

Imbens GW, Angrist JD. Identification and Estimation of Local Average Treatment Effects. Econometrica 1994;62:467–475.

Newey W, Stoker TM. Efficiency of Weighted Average Derivative Estimators and Index Models. Econometrica 1993;61:1199–1223.

Robinson PM. Root-N-Consistent Semiparametric Regression. Econometrica 1988;56:931–954.

Stock JH. Nonparametric Policy Analysis. Journal of the American Statistical Association 1989;84:565–575.

Vytlacil EJ. Independence, Monotonicity, and Latent Index Models: An Equivalence Result. Econometrica 2002;70:331–341.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

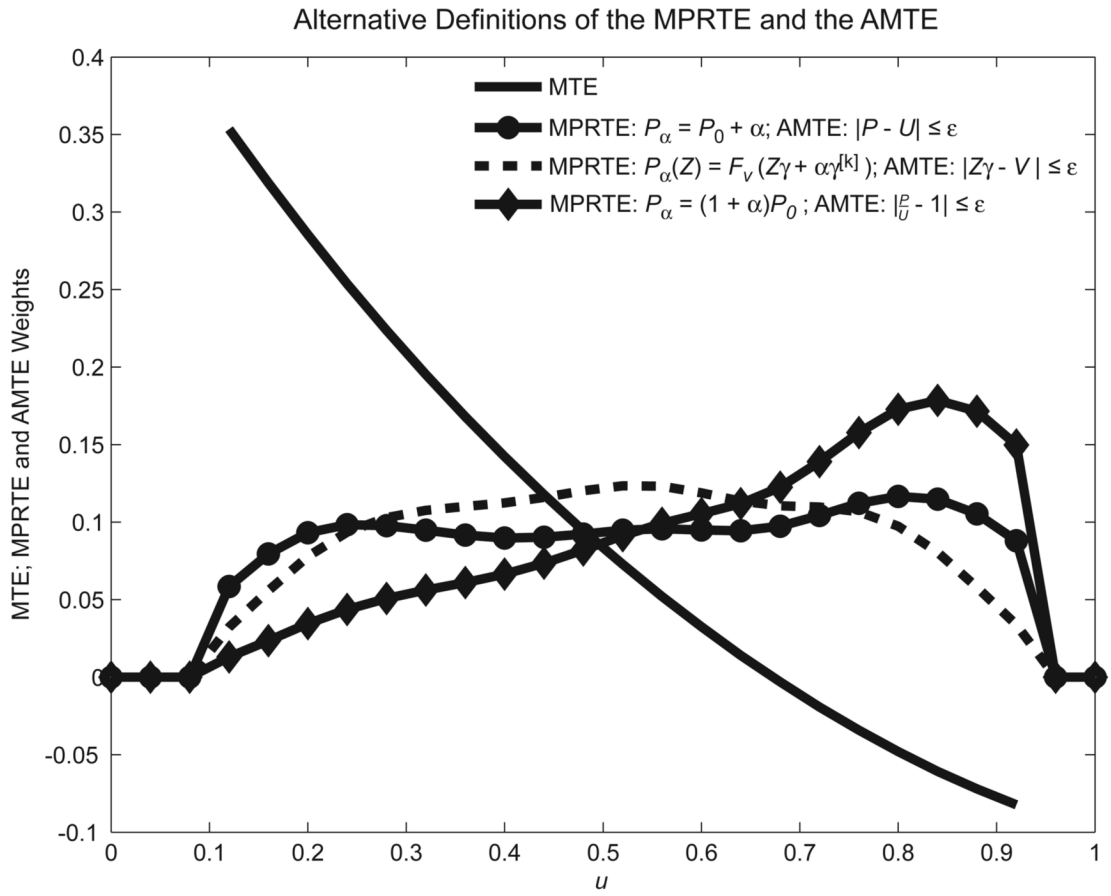


Figure 1.

This figure plots the marginal treatment effect ($MTE = E(Y_1 - Y_0|X = x, U = u)$) and the marginal policy relevant treatment effect (MPRTE) weights for three types of policy shifts: $P_\alpha^A = P + \alpha$, $P_a^B(Z) = F_V\left(Z\gamma + \alpha\gamma^{[k]}\right)$, and $P_\alpha^C = (1+\alpha)P$, where P is the probability of receiving treatment conditional on the observed covariates (Z). The equivalent definitions of the AMTE correspond, respectively, to the following three metrics for measuring the distance between P and U ($m(P, U)$): $|P - U|$, $|Z\gamma - V|$, and $\left|\frac{P}{U} - 1\right|$. The MTE, the MPRTE, and the AMTE weights are evaluated at values of X such that $E(Y_1 - Y_0|X) = 0.13$. The average marginal policy effects are 0.1296 for the additive shift in P (A), 0.1274 for the additive shift in Z (B), and 0.0880 for the proportional shift in P (C). Estimated from NLSY data (see Carneiro, Heckman, and Vytlacil, 2006, for details on estimating MTE).