Journal of Econometrics 195 (2016) 1–22

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Identifying the average treatment effect in ordered treatment models without unconfoundedness☆

Arthur Lewbel [a,*], Thomas Tao Yang [b]

[a] Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA, 02467, USA

[b] Research School of Economics, Australian National University, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form
3 May 2016
Accepted 19 May 2016
Available online 28 June 2016

JEL classification:
C14
C21
C26

Keywords:
Average treatment effect
Ordered choice model
Special regressor
Semiparametric
Competition and innovation
Identification

ABSTRACT

We show identification of the Average Treatment Effect (ATE) when treatment is specified by ordered choice in cross section or panel models. Treatment is determined by location of a latent variable (containing a continuous instrument) relative to two or more thresholds. We place no functional form restrictions on latent errors and potential outcomes. Unconfoundedness of treatment does not hold and identification at infinity for the treated is not possible. Yet we still show nonparametric point identification and estimation of the ATE. We apply our model to reinvestigate the inverted-U relationship between competition and innovation, and find no inverted-U in US data.

1. Introduction

We consider a model where possible treatments are specified by an ordered choice model. For example, treatment could be determined by an ordered probit. However, unlike probit, we will not specify the distribution of the latent error term. We also do not specify how outcomes are determined as functions of treatment. We place no functional form restrictions on the joint distribution of latent errors and potential outcomes. Unconfoundedness of treatment (either unconditional or conditional on covariates) does not hold, and identification at infinity for the treated is not possible. Yet we still show nonparametric point identification of the Average Treatment Effect (ATE), and we provide an associated estimator, which converges at parametric rates. We describe application of the estimator to both cross section and panel data, though we focus on panel data. The panel model allows for fixed effects in both the treatment and outcome equations. In general, ordered choice panel models with fixed effects suffer from the incidental parameters problem, leading to slow rates of convergence, but we provide conditions under which our estimator does not suffer from the incidental parameters problem.

Suppose an outcome Y is given by

$$Y = Y_0 + (Y_1 - Y_0)\, D \tag{1.1}$$

where Y_0 and Y_1 are potential outcomes as in Rubin (1974), and D is a binary treatment indicator. Generally, point identification of the average treatment effect (ATE) $E(Y_1 - Y_0)$ requires either (i) conditional or unconditional unconfoundedness of treatment, or (ii) an instrument for D that can drive D to zero and to one with probability one (i.e., identification at infinity), or (iii) functional restrictions on the joint distribution of Y_0, Y_1 and D. In contrast, we provide a novel point identification result, and an associated estimator, for the ATE in a model where none of these conditions hold.

☆ Earlier versions of this paper circulated under the title, "Identifying the Average Treatment Effect in a Two Threshold Model". The authors would like to thank Aamir Hashmi for providing data, and Han Hong, Jim Heckman, Bertan Turhan, Filippo De Marco, Yatfung Wong and four anonymous referees for helpful comments and suggestions. All errors are our own.

* Corresponding author.
E-mail addresses: lewbel@bc.edu (A. Lewbel), tao.yang@anu.edu.au (T.T. Yang).
URLs: https://www2.bc.edu/lewbel/ (A. Lewbel), https://sites.google.com/site/yangtomtao/ (T.T. Yang).

http://dx.doi.org/10.1016/j.jeconom.2016.05.015
0304-4076/

Let V be a continuous instrument that affects the probability of treatment but not the outcome, and let X denote a vector of other covariates. Let D be given by a structure that is identical to one of the choices in an ordered choice model, that is,

$$D = I\left[\alpha_0(X) \leq V + U \leq \alpha_1(X)\right] \tag{1.2}$$

where $I(\cdot)$ is the indicator function that equals one if $\cdot$ is true and zero otherwise, U is a latent error term, $\alpha_0(X)$ and $\alpha_1(X)$ are unknown functions, and the coefficient of V is normalized to equal one. The joint distribution of $(U, Y_0, Y_1 \mid X)$ is assumed to be unknown. Later we provide an extension to a model where V in the treatment Eq. (1.2) is replaced with $\varsigma(V)$ for some unknown function ς.

In the special case of Eq. (1.2) where $\alpha_0(X)$ and $\alpha_1(X)$ are linear with the same slope, this is equivalent to treatment being given by the more standard looking ordered choice specification

$$D = I\left(\delta_0 \leq X'\beta_1 + V + U \leq \delta_1\right)$$

for constants δ_0, δ_1, and β_1. However, we do not impose these linearity restrictions. In addition, unlike standard ordered choice models, we allow the distribution of U to depend on X in completely unknown ways. Equivalently, the covariates X can all be endogenous regressors, with no available associated instruments. The only covariate we require to be exogenous is V.

In the proposed model, treatment and outcomes are confounded (both conditionally and unconditionally), because the unobservable U that affects D can be correlated with Y_0 and Y_1, with or without conditioning on X. No parametric or semiparametric restrictions are placed on the distribution of $(U, Y_0, Y_1 \mid X)$, so treatment effects are not identified by functional form restrictions on the distributions of unobservables. We assume V has large support, but the model is not identified at infinity. This is because both very large and very small values of V drive the probability of treatment close to zero, but no value of V (or of other covariates) drives the probability of treatment close to one. So in this framework none of the conditions that are known to permit point identification of the ATE hold. Even a local ATE (LATE) is not identified in the usual way (e.g., Imbens and Angrist, 1994), because monotonicity of treatment with respect to the instrument cannot hold in the proposed model. Nevertheless, we show that the ATE is identified in our model, using a special regressor argument as in Lewbel (1998, 2000, 2007)). Our results include a test of the large support assumption required for this identification. We construct a very simple estimator of the ATE based on this identification.

To illustrate the model and foreshadow our later empirical application, suppose the outcome Y is a measure of innovation in an industry and $D = 1$ when a latent measure of competitiveness in the industry lies between two estimated thresholds, otherwise $D = 0$. According to the "Inverted-U" theory in Aghion et al. (2005) (hereafter ABBGH), industries with intermediate levels of competitiveness have more innovation than those with low levels or high levels of competition. As in Revenga (1990, 1992), Bertrand (2004), and Hashmi (2013), we use a source-weighted average of industry exchange rates as an instrumental variable for competition, which we take to be our special regressor V. This instrument is computed from the weighted average of the US dollar exchange rate with the currencies of its trading partners. When V is low, products from the US become relatively cheaper, thereby reducing competition by driving out competitors. The treatment effect we estimate is therefore the gains in innovation that result from facing moderate (rather than low or high) levels of competition.

With Eqs. (1.1) and (1.2), one has $D = 0$ if the latent variable is either above the upper threshold or below the lower threshold. In many applications we would want to distinguish between those two cases. We would then have a standard ordered choice model for treatment, that is,

$$\begin{aligned} D_0 &= I\left[V + U < \alpha_0(X)\right], \\ D_1 &= I\left[\alpha_0(X) \leq V + U < \alpha_1(X)\right], \\ D_2 &= I\left[\alpha_1(X) \leq V + U\right] \end{aligned} \tag{1.3}$$

so an individual receives treatment j for $j = 0, 1, 2$ if $D_j = 1$. Letting W_j denote the potential outcome of an individual who receives treatment j, we would now have

$$Y = D_0 W_0 + D_1 W_1 + D_2 W_2. \tag{1.4}$$

In particular, W_0 is the potential outcome when lying below the lower threshold and W_2 is the potential outcome when lying above the upper threshold. The earlier model of Eqs. (1.1) and (1.2) are the special case of this model where $D = D_1$, $Y_1 = W_1$, and

$$Y_0 = D_0 W_0 + D_2 W_2.$$

In this standard ordered choice model for treatment, the goal would be identification of the means of three potential outcomes, $E\left(W_j\right)$ for $j = 0, 1, 2$, corresponding to low, medium, and high values of the unobserved latent variable that determines selection.

In an extension section, we show identification of this ordered choice model, and identification of a more general model having any number of choices J instead of just the above case of $J = 3$. For $J = 3$, identification and estimation of $E(W_1)$ is identical to identification and estimation of $E(Y_1)$. But, unlike identification of $E(Y_0)$, identification of the separate extreme potential outcomes $E(W_0)$ and $E(W_2)$ requires identification at infinity arguments as in Heckman et al. (2006). This means that, unlike estimation of our original treatment effect $E(Y_1 - Y_0)$, estimation of treatment effects like $E(W_1 - W_0)$ or $E(W_1 - W_2)$ will converge at slower than parametric rates. In contrast, identification at infinity is not needed for identification of $E(Y_0)$ and $E(Y_1)$, and indeed is not even possible for identification of $E(Y_1)$.

With this extension, our method can be applied to most situations where treatment is defined by ordered response. For example, one might consider returns to education where the three possible treatments correspond to dropping out of school (the low group), completing high school (the middle group), and completing college (the high group). In our later empirical application, this extension will help us to distinguish between competing alternatives to the inverted-U hypothesis.

Even without this extension, our estimator is potentially useful in applications where one wants to assess the impact of a treatment defined as a moderate level of some activity, versus low or high levels. Many such treatments exist. For example, one might want to assess the effects of moderate levels of BMI or of alcohol consumption on a variety of health outcomes (see, e.g., Cao et al., 2014; Koppes et al., 2005; Solomon et al., 2000). Other examples are the effect of moderate levels of financial development on the growth rates of countries (see Cecchetti and Kharroubi, 2012) or the effects of moderate levels of financial regulation on measures of financial instability (see Huang, 2015).

Our empirical application uses panel data. We extend our method to show identification of $E\left(Y_{jit}\right)$, and hence of $E\left(Y_{1it} - Y_{0it}\right)$, in the panel data model

$$Y_{it} = \widetilde{a}_i + \widetilde{b}_t + \left(1 - D_{it}\right) Y_{0it} + Y_{1it} D_{it}, \tag{1.5}$$

$$D_{it} = I(\alpha_0(x_{it}) \leq a_i + b_t + V_{it} + U_{it} \leq \alpha_1(x_{it})), \tag{1.6}$$

where $a_i, \widetilde{a}_i, b_t, \widetilde{b}_t$ are individual and time dummies in the selection and outcome equations. To interpret Eq. (1.5), define $Y^*_{it} = Y_{it} - \widetilde{a}_i - \widetilde{b}_t$, so Y^*_{it} is the outcome Y_{it} after time and individual specific fixed effects have been removed. Eq. (1.5) then becomes $Y^*_{it} = \left(1 - D_{it}\right) Y_{0it} + Y_{1it} D_{it}$, so Y_{0it} and Y_{1it} are the potential outcomes, not of Y_{it}, but of Y^*_{it}. With this construction, then the