Journal of Econometrics 187 (2015) 74–81

Contents lists available at ScienceDirect

Journal of Econometrics

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

Sharp bounds on treatment effects in a binary triangular system

Ismael Mourifié

Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada

ARTICLE INFO

Article history:
Received 24 June 2013
Received in revised form
24 October 2014
Accepted 31 January 2015
Available online 18 February 2015

JEL classification:
C14
C31
C35

Keywords:
Partial identification
Threshold crossing model
Triangular system
Average treatment effect
Endogeneity

ABSTRACT

This paper considers the evaluation of the average treatment effect (*ATE*) in a triangular system with binary dependent variables. I impose a threshold crossing model on both the endogenous regressor and the outcome. The bounds proposed by Shaikh and Vytlacil (2011,SV) on the *ATE* are sharp only under a restrictive condition on the support of the covariates and the instruments, which rules out a wide range of models and many relevant applications. In this setting, I provide a methodology that allows the construction of sharp bounds on the *ATE* by efficiently using the variation of covariates without imposing support restrictions.

0. Introduction

This paper considers the evaluation of the average treatment effect (*ATE*) of a binary endogenous regressor on a binary outcome when a threshold crossing model on both the endogenous regressor and the outcome is imposed. The joint threshold crossing (JTC) model was recently investigated by Shaikh and Vytlacil (2011), but their proposed bounds are sharp only under a critical restriction imposed on the support of the covariates and the instruments. The support condition required is very strong and often fails to hold for a wide range of models. SV takes advantage of the threshold crossing condition imposed on the endogenous regressor to refine the known bounds on the *ATE* in the model with an unrestricted endogenous regressor. However, whenever the support condition fails, their bounds are still valid but no longer sharp, because they do not take full advantage of the threshold crossing condition imposed on the endogenous regressor. I show in this paper how to fully exploit the second threshold crossing restriction imposed on the endogenous regressor without imposing any support restrictions.

Therefore, this paper complements SV's work by providing a methodology that allows the construction of sharp bounds on the *ATE* by efficiently using variation on covariates. The proposed methodology requires only mild regularity conditions on the distribution of unobservable variables and a typical independence assumption between the covariates (except the binary endogenous regressor) and the unobservable variables. Inference of the bounds can easily be carried out using the inferential methods of Chernozhukov et al. (2013) or of Andrews and Shi (2014). The proof of the sharpness of the proposed bounds is based on copula theory and a characterization theorem proposed by Chiburis (2010). Indeed, a similar objective was pursued by Chiburis (2010), however his characterization is not an operational characterization in the sense that it does not allow a direct computation of the identified set based on the knowledge of the observed probabilities in the data because the copula is an infinite-dimensional nuisance parameter. This makes his approach computationally infeasible in most cases of interest. Also, the JTC model is a particular case of the Chesher (2005), and Jun et al. (2010) models. However, their analyses imposed an additional restriction on the joint distribution of the unobservable variables. I do not impose such a restriction.

The rest of the paper is organized as follows. The following section considers joint threshold crossing models, explains why SV's bounds fail to be sharp without their support condition, and proposes a methodology to sharpen their bounds in this case. The second and third sections present a numerical illustration and discuss the inference procedure. The last section concludes, and proofs are collected in the Appendix.

E-mail address: ismael.mourifie@utoronto.ca.

http://dx.doi.org/10.1016/j.jeconom.2015.01.006
0304-4076/

1. Joint threshold crossing model

I adopt the framework of the potential outcome model $Y = Y_1D + Y_0(1 - D)$, where Y is an observed outcome, D denotes the observed binary endogenous regressor, and Y_1, Y_0 are potential outcomes. The potential outcomes and D are as follows:

$$Y_d = 1\{\nu(d, X) > u\}, \quad d = 0, 1$$
$$D = 1\{p(X, Z) > v\}, \tag{1.1}$$

where u and v are normalized to be uniformly distributed $u, v \sim U[0, 1]$, $1\{.\}$ denotes the indicator function, $\nu(0, X)$ and $\nu(1, X)$ are unknown functions of a vector of exogenous regressors X, and $p(X, Z)$ is an unknown function of a vector of exogenous regressors $[X, Z]$. The formal assumption I use in this section may be expressed as follows:

Assumption 1. (X, Z) and (u, v) are statistically independent.

$\mathcal{X}$ and $\mathcal{Z}$ denote the respective supports of the variables X and Z. Since $u, v \sim U[0, 1]$ and Y and D are binary, we have the following: $\nu(d, x) = P(Y_d = 1 \mid X = x) = \mathbb{E}[Y_d|X = x]$, and $p(x, z) = P(D = 1 \mid X = x, Z = z)$ for all $(x, z) \in Supp(X, Z)$, where $Supp(X, Z)$ denotes the joint support of (X, Z). The normalization of u is convenient when the potential outcomes are binary since it implies $\mathbb{E}[Y_d \mid X = x] = \nu(d, x)$, and bounds on treatment effect parameters can be derived from bounds on the structural parameters $\nu(1, x)$ and $\nu(0, x)$. Then, we may define the average structural function (*ASF*) and the average treatment effect (*ATE*), respectively, as: $\nu(d, x)$ and $\Delta\nu(x) = \nu(1, x) - \nu(0, x)$. Let $Supp(P \mid X)$ denotes the support of $p(X, Z)$ conditional on X. When no confusion is possible, I shall use the shorthand notation $p = p(x, z)$, $p' = p(x', z')$, where $p(x, z) \in Supp(P|X = x)$ and $p(x', z') \in Supp(P|X = x')$, $P(i, j|x, p) = P(Y = i, D = j|X = x, p(X, Z) = p)$, and $sign(a) = 1\{a > 0\} - 1\{a < 0\}$.

SV used the JTC equations determining Y and D along with additional assumptions to identify the sign of $[\nu(1, x') - \nu(0, x)]$ from the distribution of observed data, and then they took advantage of this information to construct bounds on *ASF* that exploit variation in covariates. However, their strategy provides bounds on the *ASF* that are sharp only whenever $Supp(X, P(X, Z)) = \mathcal{X} \times Supp(P(X, Z))$, namely the "critical support condition". Moreover, whenever $Supp(P \mid X = x) \cap Supp(P \mid X = x')$ is empty or reduced to a singleton, SV's bounds do not take advantage of the threshold restriction imposed on the equation determining D. This "critical support condition" implies that $Supp(P \mid X = x) = Supp(P \mid X = x')$ for all $(x, x') \in \mathcal{X} \times \mathcal{X}$; in other words for all $(x, x') \in \mathcal{X} \times \mathcal{X}$ and $z \in Supp(Z|X = x)$, there exists $z' \in Supp(Z|X = x')$ such that $p(x, z) = p(x', z')$. This type of "perfect matching restriction" is difficult to achieve in many applications. As Chiburis (2010) pointed out, the SV critical support condition tends to hold only when $p(x, z)$ does not depend on x, which is only true in the rare case of a complete dichotomy between variables in the outcome equation and variables in the treatment equation. I will now show how it is possible to sharpen bounds on the *ASF* without imposing the "critical support condition".

1.1. Sharpening the bounds

1.1.1. First main idea

Let us present a simple intuition of the main idea of this paper. We have

$$\nu(0, x) = P(u \le \nu(0, x), v \ge p(x, z)) + P(u \le \nu(0, x), v \le p(x, z)),$$

where $P(u \le \nu(0, x), v \ge p(x, z)) = P(1, 0|x, p)$, but the second term $P(u \le \nu(0, x), v \le p(x, z)) = P(Y_0 = 1, D = 1|X = x, Z = z)$ is the unobserved counterfactual. SV proposed bounding this counterfactual by exploiting variation in covariates. Indeed, SV's idea suggests that we may bound the unobserved counterfactual for untreated individuals ($D = 0$) with characteristic x by using information on treated individuals ($D = 1$) with different characteristics x' whenever they have exactly the same probability of being treated. In fact, if we have a treated individual with characteristic x' belonging to the set $\Delta_p(x) = \{x' : \nu(0, x) \le \nu(1, x')\} \cap \{x' : \exists p' \in Supp(P \mid X = x'),\ p(x, z) = p(x', z')\}$, the proposed bounds of SV for the unobserved counterfactual can be summarized as follows:

$$P(u \le \nu(0, x), v \le p(x, z)) \le \begin{cases} P(u \le \nu(1, x'), v \le p(x', z')) & \text{if } x' \in \Delta_p(x) \\ p(x, z) & \text{if } \Delta_p(x) = \varnothing, \end{cases}$$

where $P(u \le \nu(1, x'), v \le p(x', z')) = P(1, 1|x', p')$. However, this idea is not sufficient to provide sharp bounds. My argument relies on the fact that under the threshold crossing model assumption imposed on the treatment (D), we may bound the unobserved counterfactual $\mathbb{P}(Y_0 = 1, D = 1|x, z)$ by using information on treated individuals with different characteristics x' even if they have different probabilities of being treated. In fact, if we have a treated individual with characteristic x' belonging to the subset $\tilde{\Delta}_p(x) = \{x' : \nu(0, x) \le \nu(1, x')\} \cap \{x' : \exists p' \in Supp(P \mid X = x'),\ p(x, z) \le p(x', z')\}$, the unobserved counterfactual may be bounded as follows:

$$P(u \le \nu(0, x), v \le p(x, z)) \le \begin{cases} P(1, 1|x', p') & \text{if } x' \in \tilde{\Delta}_p(x) \\ p(x, z) & \text{if } \tilde{\Delta}_p(x) = \varnothing. \end{cases}$$

When $p(x, z) \notin Supp(P \mid X = x) \cap Supp(P \mid X = x')$, we cannot identify $P(u \le \nu(1, x'), v \le p(x, z))$ from the data. In this case, SV proposed bounding $P(u \le \nu(1, x'), v \le p(x, z))$ from above by $P(v \le p(x, z)) = p(x, z)$. However, whenever it is possible to find $x' \in \tilde{\Delta}_p(x)$, I propose bounding $P(u \le \nu(1, x'), v \le p(x, z))$ from above by $P(u \le \nu(1, x'), v \le p(x', z')) = P(1, 1|x', p')$, which may be lower than $P(v \le p(x, z)) = p(x, z)$ in many cases. Since $\Delta_p(x) \subseteq \tilde{\Delta}_p(x)$, it is easy to see that we may obtain an improvement over SV's bounds by using $\tilde{\Delta}_p(x)$ instead of $\Delta_p(x)$, especially when $\Delta_p(x)$ is empty or $\Delta_p(x)=\{x\}$. When $Supp(P|X = x) = Supp(P|X = x')$, we have $\tilde{\Delta}_p(x) = \Delta_p(x)$; this fact explains why the SV bounds would be sharp when $Supp(P|X = x) = Supp(P|X = x')$. Hereafter, I adopt the convention that the supremum over the empty set is zero and the infimum over the empty set is one. Before formalizing this idea, I will define some subsets summarized in Table 1.

Table 1
Collection of sets.

$\mathbf{P}^+(x', p) = \{p(x', z') = p' \in Supp(P|X = x') : p \le p'\}$
$\mathbf{P}^-(x', p) = \{p(x', z') = p' \in Supp(P|X = x') : p \ge p'\}$
$\boldsymbol{\Omega}^+_{\mathbf{d}_1\mathbf{d}_2}(x) = \{x' : \nu(d_1, x) \le \nu(d_2, x')\}$
$\boldsymbol{\Omega}^-_{\mathbf{d}_1\mathbf{d}_2}(x) = \{x' : \nu(d_1, x) \ge \nu(d_2, x')\}$

Indeed, for all $x' \in \boldsymbol{\Omega}^+_{\mathbf{01}}(x)$ and $p(x', z') \in \mathbf{P}^+(x', p)$, we have:

$$\begin{aligned}\nu(0, x) &= P(u \le \nu(0, x), v \ge p(x, z)) \\ &\quad + P(u \le \nu(0, x), v \le p(x, z)) \\ &\le P(u \le \nu(0, x), v \ge p(x, z)) + P(u \le \nu(1, x'), v \le p(x, z)) \\ &\le P(u \le \nu(0, x), v \ge p(x, z)) \\ &\quad + \min[P(u \le \nu(1, x'), v \le p(x', z')), p(x, z)].\end{aligned}$$

Therefore,

$$\nu(0, x) \le P(1, 0|x, p) + \min[\inf_{x' \in \boldsymbol{\Omega}^+_{\mathbf{01}}(x)} \inf_{p' \in \mathbf{P}^+(x', p)} P(1, 1|x', p'), p]. \tag{1.2}$$