Independence, Monotonicity, and Latent Index Models: An Equivalence Result

Edward Vytlacil*

September 10, 2000

Abstract

The selection model and instrumental variable, local average treatment effect (LATE) framework are widely interpreted as alternative, competing frameworks. This note shows that the assumption of an unobserved index crossing a threshold that defines the selection model is equivalent to the independence and monotonicity assumptions at the center of the LATE approach. The underlying assumptions of the two approaches are equivalent.

JEL Numbers: C34, H43
KEYWORDS: selection model, latent index model, instrumental variables, social program evaluation, local average treatment effect.

---

*Stanford University, Department of Economics. I would like to acknowledge James Heckman for his continued support. Many of the ideas in this note were directly motivated from my discussions with him, and this note grew out of our joint research. I would also like to thank him for his many detailed and extremely helpful comments and suggestions on this note. I would like to thank Victor Aguirregabiria, Lars Hansen, Guido Imbens, and especially Jaap Abbring for helpful comments. I would also like to thank seminar participants at the University of California-Irvine, Carnegie-Mellon University, University of Chicago, University College London, Institute for Fiscal Studies, Northwestern University, Olin School of Business at Washington University, Rochester University, University of Southern California, Stanford University, Tel Aviv University, Tilberg University, Univesity of Toulouse, University of Wisconsin, and Yale University. All errors are my own. Financial support from the Alfred P. Sloan Doctoral Dissertation Fellowship, the C. V. Starr Foundation, and the Joint Center for Poverty Research is gratefully acknowledged. Correspondence: Landau Economics Building, 579 Serra Mall, Stanford CA 94305; Email: vytlacil@stanford.edu; Phone: 650-725-7836; Fax: 650-725-5702.

1 Introduction

A common problem in economics is how should we evaluate the effect of a treatment when individuals self-select whether to receive the treatment? This problem arises when trying to evaluate the union/non-union wage differential, the effect of job training on earnings, and the returns to schooling, where being unionized or making a human capital investment is the treatment.

One standard approach to this problem is the use of a selection model as first proposed by Heckman (1976). Under this approach, the researcher models selection into the program by a latent index crossing a threshold, where the latent index is interpreted as the expected net utility of selecting into treatment. However, some statisticians have criticized or even dismissed the use of selection models to estimate treatment effects, arguing that such analysis is inherently driven by distributional and functional form assumptions.[1] This sentiment has been echoed within economics.

The local average treatment effect (LATE) framework is a form of linear instrumental variables (IV) analysis developed by Imbens and Angrist (1994).[2] However, like the selection model approach,[3] and unlike traditional IV analysis,[4] the LATE framework allows for heterogeneous treatment effects, in particular, for the effect of the treatment to vary across individuals with

[1]See, e.g., the comments and discussions on Heckman and Robb (1986) in Wainer (1986), especially those by Tukey and Holland and the Heckman and Robb rejoinder. There is a large literature on the semiparametric estimation of selection models, see, for example, Ahn and Powell (1993). However, as emphasized by Heckman (1990), estimation of treatment effects requires knowledge of the intercept of the outcome equations, and most of the semiparametric literature does not consider estimation of the intercepts of the outcome equations.

[2]See also Angrist, Imbens and Rubin (1996) for a further exposition of the approach for the special case of a binary instrument.

[3]See Heckman and Robb (1986) and Björklund and Moffitt (1987).

[4]Traditional IV approaches are critically dependent on the assumption that either (1) the treatment effect does not vary across individuals, or (2) the treatment effect does vary across individuals, but individuals do not select into treatment based on their idiosyncratic gains from treatment. See, e.g., Heckman and Robb (1986), Heckman (1997), and Heckman and Vytlacil (2000c).

the same observed characteristics and with selection into treatment possibly dependent on the individual-specific treatment effect. In order to allow for heterogeneous treatment effects, the LATE approach requires additional assumptions not imposed in conventional IV analysis, with these additional assumptions stated directly in terms of the underlying counterfactual variables.

On the surface, the LATE assumptions do not seem connected to the assumption of a selection model. They do not directly involve an unobserved index crossing a threshold or the imposition of any structural economic model. The LATE approach has been advanced as being less restrictive than econometric selection models, but has been attacked for not being motivated by or interpretable as an economic model.[5]

This note shows that the assumption of an unobserved index crossing a threshold that defines the selection model is equivalent to the independence and monotonicity assumptions at the center of the LATE approach. In particular, the selection model assumptions imply the LATE assumptions, and given the LATE assumptions, there always exists a selection model which rationalizes the observed and counterfactual data. The LATE assumptions are not weaker than the assumptions of a latent index model, but instead impose the same restrictions on the counterfactual data as the classical selection model if one does not impose parametric functional form or distributional assumptions on the latter. This equivalence result shows that the LATE analysis of Imbens and Angrist can be seen as an application of a latent index model, and thus directly connects their research to the econometric literature on selection models.

This note proceeds in the following way. I first introduce the switching regression framework and the necessary notation in Section 2. I define and discuss the assumption of a selection model

[5]See, e.g., the comments by Heckman (1996) and Moffitt (1996) on Angrist, Imbens and Rubin (1996), and the rejoinder by Angrist, Imbens and Rubin (1996).

in Section 3 and the LATE assumptions in Section 4. In Section 5, I show that these two sets of assumptions are equivalent: the selection model implies the independence and monotonicity conditions assumed in LATE analysis, and the independence and monotonicity conditions imply that a selection model may be assumed without imposing any additional restrictions.

2 Framework

Let $(\Omega, \mathcal{F}, \mathbf{P})$ denote the probability space. All random variables will be defined on this common probability space. Let ω denote an element of Ω. Let $(Y_0(\omega), Y_1(\omega))$ denote random variables corresponding, respectively, to the potential outcomes in the untreated and treated states. For those unfamiliar with measure-theoretic notation, it may be most intuitive in the context of this note to think of Ω as denoting the set of all individuals in the universe of interest and of ω as indexing individuals. When translating the LATE assumptions of Imbens and Angrist (1994) into this notation, ω will serve the same role as their individual i subscript.

Let $D(\omega)$ be a random variable for receipt of treatment; $D(\omega) = 1$ denotes the receipt of treatment; $D(\omega) = 0$ denotes nonreceipt. Let $Y(\omega)$ be the measured outcome variable so that

$$Y(\omega) = D(\omega)Y_1(\omega) + (1 - D(\omega))Y_0(\omega).$$

I will assume that $Y_1(\omega)$ and $Y_0(\omega)$ have finite first moments and also that $1 > \mathbf{P}[D(\omega) = 1] > 0$. Let $W(\omega)$ denote a vector of observed covariates. Partition $W(\omega)$ into two subvectors, $W(\omega) = [X(\omega), Z(\omega)]$, where X contains the covariates that directly affect the (Y_1, Y_0) outcomes (as well as possibly affecting the treatment decision), and Z contains the covariates that only affect the treatment decision D. Variables in Z are referred to as instruments or excluded variables. In the rest of this note, I do not explicitly consider the observed random variables X; everything in this note is conditional on X.

Let $\mathcal{Z}$ denote the support of Z, and let z denote a possible realization of Z. For each $z \in \mathcal{Z}$, let $D_z(\omega)$ be the counterfactual variable denoting whether the observation would have received treatment if $Z(\omega)$ had been externally set to z. This defines a collection of random variables, $\{D_z(\omega) : z \in \mathcal{Z}\}$, where the collection of random variables is indexed by the support of Z. Clearly $D(\omega) = D_{Z(\omega)}(\omega)$. To simplify the analysis, I will assume that $\mathcal{Z}$ is finite or countable. This assumption is imposed for expositional purposes only. An appendix containing the derivations for the more general case where $\mathcal{Z}$ is possibly uncountable is available upon request from the author.[6]

Let

$$p(z) = \mathbf{P}[D(\omega) = 1 | Z(\omega) = z]. \tag{1}$$

$p(z)$ is sometimes called the "propensity score" by statisticians and is sometimes called the "choice probability" by economists.[7]

3 Latent Index Selection Models

The latent index assumption is that the treatment choice is determined by an index crossing a threshold. The most conventional form of the classical selection model imposes a linear index assumption, in particular,

$$D = 1[Z\beta \geq U], \tag{2}$$

[6]The analysis of this note trivially generalizes to the case where $\mathcal{Z}$ is uncountable if one imposes that $\{D_z(\omega) : z \in \mathcal{Z}\}$ is separable, i.e., if one imposes that there exists a countable set $\mathcal{Z}^* \subset \mathcal{Z}$, such that for any $z \in \mathcal{Z}$ there exists a sequence $\{z_1, z_2, ...\} \subset \mathcal{Z}^*$ such that $\lim_{k\to\infty} D_{z_k} = D_z$. The appendix shows the equivalence result for the case where $\mathcal{Z}$ is uncountable and a separability assumption has not been imposed.

[7]The term "propensity score" originates in Rosenbaum and Rubin (1983), who analyzed its role in the method of matching. See Heckman and Robb (1986) and Heckman and Vytlacil (2000c) for a discussion of how the propensity score plays a fundamentally different role in matching models versus selection models.

where $1[\cdot]$ is the indicator function and U is assumed to be independent of Z. The form of the selection model that I consider (and show is equivalent to the LATE assumptions) revises equation 2 in two ways. First, I relax the linear index assumption to consider $D = 1[\nu(Z) \geq U]$ for some function ν. Second, I explicitly impose that the latent index equation continues to hold under hypothetical interventions. In other words, I assume not only that the individual's actual choice be described by the latent index model, but also that the same model describes what her choices would have been had her value of Z been externally set to any other value. Formally, I make the following assumptions:[8]

Latent Index Selection Model (S-1): $D_z(\omega) = \mathbf{1}[\nu(z) \geq U(\omega)]$ for some $\nu : \mathcal{Z} \mapsto \Re$, with (i) $\nu(z)$ a measurable, nontrivial function of z,[9] and (ii) $Z(\omega) \perp\!\!\!\perp (U(\omega), Y_0(\omega), Y_1(\omega))$.

In assumption S-1, without loss of generality, we will impose the normalization that $\nu(z) = p(z)$ and $\mathbf{P}[U(\omega) \leq t] = t$ for any $t \in \mathcal{P}$, where $p(z) = \mathbf{P}[D(\omega) = 1 | Z(\omega) = z]$ and $\mathcal{P}$ is the support of $p(Z(\omega))$.

4 Assumptions of LATE Approach

As an alternative to the assumption of a selection model, the two identifying assumptions of the LATE approach of Imbens and Angrist are:[10]

LATE Independence Assumption (L-1): (i) for all $z \in \mathcal{Z}$, $Z(\omega) \perp\!\!\!\perp (Y_1(\omega), Y_0(\omega), D_z(\omega))$, and (ii) $p(z)$ is a nontrivial function of z.

[8]Recall that the analysis of this note is implicitly conditioning on $X(\omega)$, where $X(\omega)$ are any covariates that directly effect the outcome variables. Thus, assumption (ii) is that $Z(\omega) \perp\!\!\!\perp (U(\omega), Y_0(\omega), Y_1(\omega)) | X(\omega)$.

[9]In particular, $\nu(z)$ is a measurable function such that there exists $(z, z') \in \mathcal{Z} \times \mathcal{Z}$ with $\nu(z) \neq \nu(z')$.

[10]Recall that the analysis of this note is implicitly conditioning on $X(\omega)$, where $X(\omega)$ are any covariates that directly effect the outcome variables. Thus, making the conditioning on $X(\omega)$ explicit, we have, e.g., that $Z(\omega) \perp\!\!\!\perp (Y_0(\omega), Y_1(\omega), D_z(\omega)) | X(\omega)$ for all $z \in \mathcal{Z}$; likewise, the monotonicity condition holds conditional on $X(\omega)$.

LATE Monotonicity Assumption (L-2): For all $(z, z') \in \mathcal{Z} \times \mathcal{Z}$, either $D_z(\omega) \geq D_{z'}(\omega)$ for all $\omega \in \Omega$, or $D_z(\omega) \leq D_{z'}(\omega)$ for all $\omega \in \Omega$.

The independence assumption is *not* that $Z(\omega)$ is independent of $D_z(\omega)$ conditional on $Z(\omega) = z$. ($Z(\omega)$ is independent of $D_z(\omega)$ conditional on $Z(\omega) = z$ from the usual laws of probability). Instead, the assumption is that $Z(\omega)$ is independent of each element of $\{D_z(\omega) : z \in \mathcal{Z}\}$. The monotonicity assumption is *not* that $D_z(\omega)$ is non-increasing or non-decreasing in z. Instead, the assumption is that for any $(z, z') \in \mathcal{Z} \times \mathcal{Z}$, the (weak) ordering between $D_z(\omega)$ and $D_{z'}(\omega)$ is invariant to choice of ω.[11]

To help understand these assumptions, assume the $D_z(\omega)$ are generated by the latent variable model $D_z(\omega) = \mathbf{1}[\zeta(z, U(\omega)) \geq 0]$. If $Z(\omega)$ is independent of $(U(\omega), Y_0(\omega), Y_1(\omega))$, then $Z(\omega)$ is independent of $(D_z(\omega), Y_0(\omega), Y_1(\omega))$ for all $z \in \mathcal{Z}$ and the independence assumption is satisfied. If $Z(\omega)$ is not independent of $U(\omega)$, then in general $Z(\omega)$ will not be independent of $D_z(\omega)$ and the independence assumption will not hold. If the ζ index is separable in z and U, so that $\zeta(z, U(\omega)) = \nu(z) + U(\omega)$, then for any $(z, z') \in \mathcal{Z}$, $\nu(z) \geq \nu(z')$ implies that $D_z(\omega) \geq D_{z'}(\omega)$ for all $\omega \in \Omega$. Thus, the additive separability assumption implies the monotonicity assumption. The monotonicity assumption will not necessarily hold without additive separability. For example, it will not hold in the random coefficient latent index models if a given coefficient is both positive and negative with positive probability.

[11] Manski (1997) also imposes a monotonicity condition in his analysis, although his monotonicity condition is fundamentally different from the one imposed in LATE analysis. The LATE monotonicity condition is a cross-person restriction on the relationship between different hypothetical treatment choices, with the hypothetical treatment choices defined in terms of an instrument. In contrast, the Manski (1997) monotonicity condition is a monotonicity restriction on the relationship between the treatment and the outcome for each given individual.

It will be convenient to define the following sets:

$$D_z^{-1}(1) = \{\omega : D_z(\omega) = 1\},$$

$$D_z^{-1}(0) = \{\omega : D_z(\omega) = 0\}.$$

Speaking loosely of ω as an "individual," $D_z^{-1}(1)$ is the set of people who would select into treatment if the instrument were externally set to z, and $D_z^{-1}(0)$ is the set of people who would not select into treatment if the instrument were externally set to z. Using this notation, the monotonicity condition can be equivalently stated as follows:

Monotonicity: For all $(z, z') \in \mathcal{Z} \times \mathcal{Z}$, either $D_z^{-1}(1) \subseteq D_{z'}^{-1}(1)$ or $D_z^{-1}(1) \supseteq D_{z'}^{-1}(1)$.

Given the above assumptions, Imbens and Angrist (1994) show that the linear instrumental variables estimand can be interpreted as a weighted average of treatment effects. In particular, they show that if Z is binary, $\mathcal{Z} = \{0, 1\}$, with $D_1(\omega) \geq D_0(\omega)$, then the linear instrumental variables estimand identifies $E(Y_1 - Y_0 | D_1 = 1, D_0 = 0)$. They show that if Z is not binary, then the linear instrumental variables estimand identifies a particular weighted average of such terms.

5 Equivalence of Identifying Assumptions

The central hypothesis of this note is that the independence and monotonicity assumptions of the LATE approach are equivalent to the assumption of a latent index model as specified in S-1. Since one can trivially show that the latent index model defined by S-1 implies conditions L-1 and L-2, we need only to show that conditions L-1 and L-2 imply a latent index representation of the form given by S-1. The analysis proceeds as follows. Given conditions L-1 and L-2, I show that one can construct a latent index with implied counterfactual treatment variables $\{\tilde{D}_z(\omega)\}_{z \in \mathcal{Z}}$, s.t.

(1) the constructed latent index satisfies all conditions of S-1, and (2) $\tilde{D}_z(\omega) = D_z(\omega)$ for all ω outside of a set of $\mathbf{P}$-measure zero, for all $z \in \mathcal{Z}$.

We now construct the latent index representation implied by the LATE assumptions. The latent index representation will have the form:

$$\tilde{D}_z(\omega) = \mathbf{1}[p(z) \geq U(\omega)].$$

The random variable $U(\omega)$ is by definition a real valued measurable function, and we now construct this function.

We will define the $U(\omega)$ function to take different values depending on whether ω is in the sets N, A, or C, with these sets defined as follows:

$$\begin{aligned} N &\equiv \bigcap_{z \in \mathcal{Z}} D_z^{-1}(0) \\ A &\equiv \bigcap_{z \in \mathcal{Z}} D_z^{-1}(1) \\ C &\equiv \left(N \bigcup A\right)^c = \left(\bigcup_{z \in \mathcal{Z}} D_z^{-1}(0)\right) \bigcap \left(\bigcup_{z \in \mathcal{Z}} D_z^{-1}(1)\right). \end{aligned}$$

By construction, the sets N, A, and C form a partition of Ω. Again loosely speaking of ω as indexing individuals, we have that N is the set of individuals who would not select in for any value of the instrument, A is the set of individuals who would select in for all values of the instrument, and C is the set of individuals who would select in for some values of the instrument but would not for other values. In the terminology of Angrist, Imbens and Rubin (1996), $\omega \in N$ are referred to as *never-takers*, $\omega \in A$ are referred to as *always-takers*, and $\omega \in C$ are referred to as *compliers*.

I will set $U(\omega) = 1$ for $\omega \in N$ and set $U(\omega) = 0$ for $\omega \in A$. For $\omega \in C$, I proceed as follows. Let

$$\mathcal{Z}_0(\omega) = \{z \in \mathcal{Z} : D_z(\omega) = 0\},$$

$$\mathcal{Z}_1(\omega) = \{z \in \mathcal{Z} : D_z(\omega) = 1\}.$$

The sets $\mathcal{Z}_0(\omega)$ and $\mathcal{Z}_1(\omega)$ partition $\mathcal{Z}$, where the partition is a function of ω. $\mathcal{Z}_0(\omega)$ is the set of instrument values such that the individual would not select into treatment if her Z had been externally set to those values; $\mathcal{Z}_1(\omega)$ is the set of instrument values such that the individual would select into treatment if her Z had been externally set to those values. Critical to construction of $U(\omega)$ for $\omega \in C$ is the following result, where $p(z)$ is the propensity score defined in equation (1).

Lemma 1 *Assume L-1 and L-2. Then for all $\omega \in C$,*

$$\sup_{z \in \mathcal{Z}_0(\omega)} p(z) \leq \inf_{z \in \mathcal{Z}_1(\omega)} p(z).$$

Proof: See Appendix A.

Lemma 1 essentially says that, for any fixed $\omega \in C$, the sets $\mathcal{Z}_1(\omega)$ and $\mathcal{Z}_0(\omega)$ can be separated based on the propensity score. In other words, for any fixed individual who would select in for some values of the instrument but would not select in for other values of the instrument, the propensity score corresponding to any instrument value for which the person selects in is always at least as large as the propensity score corresponding to any instrument value for which the person would not select in.

We now construct the $U(\omega)$ function:

$$U(\omega) = \begin{cases} 1 & \text{if } \omega \in N, \\ 0 & \text{if } \omega \in A, \\ \inf_{z \in \mathcal{Z}_1(\omega)} p(z) & \text{if } \omega \in C. \end{cases}$$

Given this construction of $U(\omega)$, the following lemma shows that $U(\omega)$ is a random variable.

Lemma 2 *Given L-1 and L-2, we have that $U(\omega)$ is a random variable, i.e., the function U, $U : \Omega \mapsto \Re$, is measurable $\mathcal{F}$.*

Proof: See Appendix A.

The following lemma shows that $Z(\omega)$ is independent of $(U(\omega), Y_0(\omega), Y_1(\omega))$.

Lemma 3 *Given L-1 and L-2, we have,*

$$Z(\omega) \perp\!\!\!\perp (U(\omega), Y_0(\omega), Y_1(\omega)).$$

Proof: See Appendix A.

We now define a selection model using the propensity score, $p(z)$, as the index and the random variable $U(\omega)$ as the threshold:

$$\tilde{D}_z(\omega) = \mathbf{1}[p(z) \geq U(\omega)].$$

From L-1, we have that $p(z)$ is a nontrivial function of z, so that the selection model satisfies condition (i) of S-1. Given L-1 and L-2, Lemma 3 states that the selection model satisfies condition (ii) of S-1. The following theorem shows that the hypothetical choices implied by the selection model agree with the original hypothetical choice variables with probability one.

Theorem 1 *Given L-1 and L-2, we have, for any $z \in \mathcal{Z}$,*

$$D_z(\omega) = \tilde{D}_z(\omega) \quad \textit{w.p.1.}$$

Proof: See Appendix A.

Thus it is possible to construct a latent index selection model that agrees with the original hypothetical choice variables w.p.1 and such that both conditions of S-1 holds. Thus, the independence and monotonicity assumptions imply that there exists a latent index representation for participation in treatment. The LATE conditions and the selection model impose exactly the same restrictions on $(Z(\omega), Y_1(\omega), Y_0(\omega), \{D_z(\omega)\}_{z\in\mathcal{Z}})$. The LATE conditions and the selection model are not only indistinguishable based on observational data, but they cannot be distinguished based on any hypothetical intervention or experiment. The two models are equivalent.[12]

This derivation has used the assumption that $\mathcal{Z}$ is finite or countable. This assumption has been imposed for expositional purposes only, and an appendix containing the derivations for the more general case where $\mathcal{Z}$ is possibly uncountable is available upon request from the author. The analysis of this note has not imposed the assumption that the distribution of the threshold U be absolutely continuous with respect to Lebesgue measure, and the random variable U constructed above from the LATE assumptions will have a distribution that is not in general absolutely continuous with respect to Lebesgue measure. However, this does not imply that selection models with continuous U are more restrictive than the LATE assumptions. An appendix, available upon request from the author, shows that, for a selection model with the distribution of U not absolutely continuous with respect to Lebesgue measure, there will exist on some probability space an alternative selection model with U distributed absolutely continuous with respect to Lebesgue measure and implying the same joint distribution for all variables of interest, i.e., the same distribution of potential outcomes, hypothetical choices and covariates. Finally, the analysis of this note has used $p(z)$ as the latent index of the selection model. This normalization is convenient

[12]This equivalence result is related to a result of Glickman and Normand (2000). They consider the LATE assumptions L-1 and L-2 augmented with the additional assumptions that (1) Z is a scalar random variable and (2) $z \leq z' \Rightarrow D_z(\omega) \leq D_{z'}(\omega)$. Note that assumptions L-1 and L-2 do not imply these conditions, in particular, note that the LATE monotonicity condition does not imply condition (2). They show that LATE, when augmented with these additional assumptions, is equivalent to the latent variable model $D_z = 1[z \geq U]$. This result can be seen as a special case of the equivalence result of this note. In particular, this note shows that LATE (without the additional assumptions) is equivalent to $D_z = 1[\nu(z) \geq U]$. Imposing conditions (1) and (2) implies that ν is invertible so that under their conditions we have $D_z = 1[z \geq V]$ with $V = \nu^{-1}(U)$.

for the derivation. This is only a normalization, and applying any monotonic transformation to $p(z)$ and U will result in an equally valid representation for the latent index. The use of $p(z)$ as a normalization in the above derivation is not dissimilar to proofs of the existence of a utility function. Utility functions are only defined up to monotonic transformations, and proofs by construction for utility functions typically involve representations for the utility function which are convenient for the derivation but not necessarily natural otherwise (see, e.g., Debreu, 1959).

The equivalence result of this paper implies that results developed within the LATE framework apply equally to the selection model framework, and vice versa. The identification analysis for treatment parameters developed under the LATE assumptions (e.g., Imbens and Angrist, 1994) apply to the identification of treatment parameters under the assumption of a selection model. Likewise, identification analysis for treatment parameters under the selection model assumption (e.g., Heckman, 1990, and Heckman and Vytlacil, 1999, 2000b) apply to identification of treatment parameters under the LATE conditions. The sharp bounds on the average treatment effect, $E(Y_1 - Y_0)$, under the LATE assumptions with Y_1, Y_0, and Z binary (Balke and Pearl, 1997), are also the sharp bounds for the average treatment effect under a selection model assumption with Y_1, Y_0, and Z binary. And likewise, the bounds on the average treatment effect shown by Heckman and Vytlacil (2000d) to be sharp under the selection model assumption without requiring Y_1, Y_0, or Z to be binary are also the sharp bounds on the average treatment effect under the LATE assumptions without requiring Y_1, Y_0, or Z to be binary. The relationship between treatment parameters under the selection model assumption (see Heckman and Vytlacil, 2000a) also holds under the LATE assumptions. In addition, there is a large literature within econometrics on the semiparametric and nonparametric estimation of selection models (e.g., Ahn and Powell, 1993, Andrews and Shafgans 1998, and Das and Newey, 2000). The equivalence result of this paper implies that these estimation methods can be applied under the LATE conditions. In either the LATE or selection model approach, auxiliary assumptions are sometimes imposed, and the analysis of this paper may be adapted to show what the auxiliary assumptions stated in terms of one approach translate into in terms of restrictions in the alternative approach.

6 Conclusion

This note shows that the independence and monotonicity assumptions of LATE and the assumption of a latent index selection model are equivalent sets of assumptions if one does not impose parametric restrictions on the selection model. In particular, the two sets of assumptions impose the same restrictions on observed and counterfactual data. This equivalence result shows that LATE analysis can be seen as an application of a latent index model in which the analyst uses the assumption of a selection model to interpret the linear IV estimand. The equivalence result of this paper implies that results developed within the selection model framework apply equally to the LATE framework, and vice versa.

References

[1] Ahn, H. and J. Powell, 1993, "Semiparametric Estimation of Censored Selection Models with Nonparametric Selection Mechanism," *Journal of Econometrics* **58**, 3-29.

[2] Andrews, D. and M. Schafgans, 1998, "Semiparametric Estimation of the Intercept of a Sample Selection Model," *The Review of Economic Studies*, **65**, 497-518.

[3] Angrist, J., G. Imbens, and D. Rubin, 1996, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association,* **91**, 444-455.

[4] Balke, A., and Pearl, J., 1997, "Bounds on Treatment Effects From Studies with Imperfect Compliance," *Journal of the American Statistical Association,* **92**, 1171-1176.

[5] Billingsley, P., 1995, *Probability and Measure*, (New York: John Wiley).

[6] Björklund, A. and R. Moffitt, 1987, "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models," *Review of Economics and Statistics*, **69**, 42-49.

[7] Das, M., and W. Newey, 1998, "Non-parametric Estimation of the Sample Selection Model," unpublished working paper, Massachusetts Institute of Technology.

[8] Debreu, G., 1959, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, (New York: John Wiley).

[9] Glickman, M. E., and S. T. Normand, 2000, "The Derivation of a Latent Threshold Instrumental Variables Model," *Statistica Sinica*, **10**, 517-544.

[10] Heckman, J., 1976, "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Simple Estimator for such Models," *Annals of Economic and Social Measurement,* **5**, 475-492.

[11] ______, 1990, "Varieties of Selection Bias," *American Economic Review*, **80**, 313-318.

[12] ______, 1996, "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association,* **91**, 459-62.

[13] ——, 1997, "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, **32**, 441-462.

[14] Heckman, J. and R. Robb, 1986, "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in H. Wainer, ed., *Drawing Inference from Self-Selected Samples*, (New York: Springer-Verlag).

[15] Heckman, J., and E. Vytlacil, 1999, "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences,* **96**, 4730-4734.

[16] ——, 2000a, "The Relationship Between Treatment Parameters within a Latent Variable Framework," *Economics Letters,* January 2000, 66(1): 33-39.

[17] ——, 2000b, "Local Instrumental Variables," forthcoming in C. Hsiao, K. Morimune, and J. Powell, eds., *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, (Cambridge University Press: Cambridge).

[18] ——, 2000c, "Econometric Evaluations of Social Programs," forthcoming in J. Heckman and E. Leamer, eds., *Handbook of Econometrics, Volume 5*, (North-Holland: Amsterdam).

[19] ——, 2000d, "Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effect," forthcoming in M. Lechner and F. Pfeiffer, eds., *Econometric Evaluations of Active Labor Market Policies in Europe.*

[20] Imbens, G. and J. Angrist, 1994, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, **62**, 467-476.

[21] Manski, C., 1997, "Monotone Treatment Response," *Econometrica,* **65**, 25-37.

[22] Moffitt, R. A., 1996, "Identification of Causal Effects Using Instrumental Variables: Comment," *Journal of the American Statistical Association,* **91**, 462-65.

[23] Rosenbaum, P. R., and Rubin, D., 1983, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, **70**, 41-55.

[24] Wainer, H., 1986, ed., *Drawing Inference from Self-Selected Samples*, (New York: Springer-Verlag).

A Proofs

Proof. (Lemma 1)

Proof by contradiction. Assume there exists $\omega \in C$ s.t. $\sup_{z\in\mathcal{Z}_0(\omega)} p(z) > \inf_{z\in\mathcal{Z}_1(\omega)} p(z)$. Then there exists (z, z') s.t. $D_z(\omega) = 0$, $D_{z'}(\omega) = 1$, and $p(z) > p(z')$. By the monotonicity and independence conditions (L-1 and L-2), $p(z) > p(z')$ implies $D_z^{-1}(1) \supseteq D_{z'}^{-1}(1)$, which contradicts the assumption that $D_z(\omega) = 0$, $D_{z'}(\omega) = 1$.

Proof: (Lemma 2)

First consider the sets A, N, and C. $D_z^{-1}(1)$ is a measurable set for any $z \in \mathcal{Z}$, $\mathcal{Z}$ is finite or countable by assumption, and thus the sets A, N, and C are measurable.

We now show that U is a measurable function. For any $t \in [0,1]$, $\{\omega \in C : \inf_{z\in\mathcal{Z}_1(\omega)} p(z) < t\} = \bigcup_{z\in\mathcal{Z}(t)} \left[C \bigcap D_z^{-1}(1)\right]$, where $\mathcal{Z}(t) = \{z \in \mathcal{Z} : p(z) < t\}$. Since $D_z^{-1}(1)$ is a measurable set for any $z \in \mathcal{Z}$ and $\mathcal{Z}$ is finite or countable by assumption, we have that $\bigcup_{z\in\mathcal{Z}(t)} \left[C \bigcap D_z^{-1}(1)\right]$

is measurable $C \bigcap \mathcal{F}$ for any $t \in [0,1]$. Thus, the restriction of U to C is measurable. The restrictions of U to N and to A are trivially measurable. Finally the sets N, A, and C are measurable and form a partition of Ω, and we thus have that $U(\omega)$ is a random variable, i.e., the function U, $U : \Omega \mapsto \Re$, is measurable $\mathcal{F}$.

Proof: (Lemma 3)

We first show that $U(\omega)$ is independent of $Z(\omega)$ conditional on $(Y_0(\omega), Y_1(\omega))$. For any fixed t, consider

$$\mathbf{P}[U(\omega) < t | Z(\omega), Y_0(\omega), Y_1(\omega)].$$

Define $\mathcal{Z}(t) = \{z \in \mathcal{Z} : p(z) < t\}$. If $\mathcal{Z}(t)$ is nonempty, then using the construction of $U(\omega)$ and Lemma 1, we have that

$$\mathbf{P}[U(\omega) < t | Z(\omega), Y_0(\omega), Y_1(\omega)] = \mathbf{P}[\bigcup_{z \in \mathcal{Z}(t)} D_z^{-1}(1) | Z(\omega), Y_0(\omega), Y_1(\omega)].$$

Let $\mathbb{N}$ denote the set of natural numbers. By L-2, we can construct a sequence $\{z_j \in \mathcal{Z} : j \in \mathbb{N}\}$, such that $D_{z_j}^{-1}(1) \subseteq D_{z_{j+1}}^{-1}(1)$ for all $j \in \mathbb{N}$ and $\bigcup_{j \in \mathbb{N}} D_{z_j}^{-1}(1) = \bigcup_{z \in \mathcal{Z}(t)} D_z^{-1}(1)$. Thus

$$\begin{aligned}
\mathbf{P}[U(\omega) < t | Z(\omega), Y_0(\omega), Y_1(\omega)] &= \mathbf{P}[\bigcup_{z \in \mathcal{Z}(t)} D_z^{-1}(1) | Z(\omega), Y_0(\omega), Y_1(\omega)] \\
&= \lim_{j \to \infty} \mathbf{P}[D_{z_j}^{-1}(1) | Z(\omega), Y_0(\omega), Y_1(\omega)] \\
&= \lim_{j \to \infty} \mathbf{P}[D_{z_j}^{-1}(1) | Y_0(\omega), Y_1(\omega)],
\end{aligned}$$

where the first equality follows from the definition of $U(\omega)$, the second equality follows from continuity from below, and the third equality follows from L-1. Following the same reasoning,

$$\mathbf{P}[U(\omega) < t | Y_0(\omega), Y_1(\omega)] = \lim_{j \to \infty} \mathbf{P}[D_{z_j}^{-1}(1) | Y_0(\omega), Y_1(\omega)],$$

and thus

$$\mathbf{P}[U(\omega) < t | Z(\omega), Y_0(\omega), Y_1(\omega)] = \mathbf{P}[U(\omega) < t | Y_0(\omega), Y_1(\omega)]$$

for any t s.t. $\mathcal{Z}(t)$ is nonempty. If $\mathcal{Z}(t)$ is empty, then $\{\omega : U(\omega) < t\} = \bigcap_{z \in \mathcal{Z}} D_z^{-1}(1)$, and a parallel argument shows that $\mathbf{P}[U(\omega) < t | Z(\omega), Y_0(\omega), Y_1(\omega)] = \mathbf{P}[U(\omega) < t | Y_0(\omega), Y_1(\omega)]$ for t s.t. $\mathcal{Z}(t)$ is empty. Thus, $U(\omega) \perp\!\!\!\perp Z(\omega) | (Y_1(\omega), Y_0(\omega))$. Using that $Z(\omega) \perp\!\!\!\perp (Y_0(\omega), Y_1(\omega))$ by assumption L-1, we now have that $Z(\omega) \perp\!\!\!\perp (U(\omega), Y_1(\omega), Y_0(\omega))$.

Proof: (Theorem 1)

Define $\tilde{D}_z(\omega) = \mathbf{1}(p(z) \geq U(\omega))$. Recall that $\mathcal{P}$ denotes the support of $p(Z(\omega))$. Pick any $z \in \mathcal{Z}$. Consider

$$\begin{aligned}
&\mathbf{P}[\omega : D_z(\omega) \neq \tilde{D}_z(\omega)] \\
&\quad = \mathbf{P}[\omega \in N : D_z(\omega) \neq \tilde{D}_z(\omega)] + \mathbf{P}[\omega \in A : D_z(\omega) \neq \tilde{D}_z(\omega)] + \mathbf{P}[\omega \in C : D_z(\omega) \neq \tilde{D}_z(\omega)].
\end{aligned}$$

We will prove that each of the three terms on the right hand side of this expression is zero.

Consider the first term of the expression. If $\{1\} \in \mathcal{P}$, then $\Pr[N] = 0$ and trivially $\mathbf{P}[\omega \in N : D_z(\omega) \neq \tilde{D}_z(\omega)] = 0$ for all $z \in \mathcal{Z}$. Now assume $\{1\} \notin \mathcal{P}$ so that $p(z) < 1$ for any $z \in \mathcal{Z}$. For $\omega \in N$, we have $D_z(\omega) = 0$, and $\tilde{D}_z(\omega) = \mathbf{1}(p(z) \geq 1) = 0$ for any $z \in \mathcal{Z}$, so $D_z(\omega) = \tilde{D}_z(\omega)$ for all $\omega \in N$, and thus $\mathbf{P}[\omega \in N : D_z(\omega) \neq \tilde{D}_z(\omega)] = 0$ for all $z \in \mathcal{Z}$.

Consider the second term of the expression. For any $\omega \in A$, $D_z(\omega) = 1$, $\tilde{D}_z(\omega) = \mathbf{1}(p(z) \geq 0) = 1$, so that $D_z(\omega) = \tilde{D}_z(\omega)$ for all $\omega \in A$. Thus, $\mathbf{P}[\omega \in A : D_z(\omega) \neq \tilde{D}_z(\omega)] = 0$ for all $z \in \mathcal{Z}$.

Now consider the third term of the expression. If $\mathbf{P}[\omega \in C] = 0$, then trivially $\Pr[\omega \in C : D_z(\omega) \neq \tilde{D}_z(\omega)] = 0$. Now assume $\mathbf{P}[\omega \in C] > 0$. We have $D_z(\omega) = 0$ and $\tilde{D}_z(\omega) = 0$ if z is s.t. $p(z) < U(\omega)$, and $D_z(\omega) = 1$ and $\tilde{D}_z(\omega) = 1$ if z is s.t. $p(z) > U(\omega)$. But, if z is s.t. $p(z) = U(\omega)$, then $\tilde{D}_z(\omega) = 1$ and $D_z(\omega)$ may equal 0. The event $\tilde{D}_z(\omega) = 1$ and $D_z(\omega) = 0$ will occur for some z values if $\sup_{z \in \mathcal{Z}_0(\omega)} p(z) = \inf_{z \in \mathcal{Z}_1(\omega)} p(z)$. Thus, $\{\omega \in C : D_z(\omega) \neq \tilde{D}_z(\omega)\} = \{\omega \in C : U(\omega) = p(z), D_z(\omega) = 0\}$. So we need to show that $\{\omega \in C : U(\omega) = p(z), D_z(\omega) = 0\}$ is a set of zero measure in the case where $\mathbf{P}[\omega \in C] > 0$. We will now show $\mathbf{P}[\omega : U(\omega) = p(z), D_z(\omega) = 0 | \omega \in C] = 0$ by a proof by contradiction. Let $\mathbf{P}[\omega : U(\omega) = p(z), D_z(\omega) = 0 | \omega \in C] = r$, and assume $r > 0$. There are two cases to consider, first where there does not exist $z' \in \mathcal{Z}_1(\omega)$, s.t. $p(z') = \inf\limits_{z \in \mathcal{Z}_1(\omega)} p(z)$ (the inf is not attained), and second where the inf is attained. First consider the case where the inf is not attained. By construction, the event $[U(\omega) = p(z), D_z(\omega) = 0]$ implies $D_z(\omega) = 1$ for any z^* s.t. $p(z^*) > p(z)$. Thus, for any such z^*, $p(z^*) \geq p(z) + r\mathbf{P}[\omega \in C]$, and thus $(p(z), p(z) + r\mathbf{P}[\omega \in C]) \bigcap \mathcal{P} = \emptyset$. But then $U(\omega) \neq \inf\limits_{z \in \mathcal{Z}_1(\omega)} p(z)$, a contradiction. Now consider the case where the inf is attained. Then there exists z' s.t. $p(z') = p(z)$ and $D_{z'} = 1$ and $D_z = 0$. The independence and monotonicity assumptions (assumptions (L-1) and (L-2)) immediately imply that this is a zero probability event.

B Taking the Distribution of U to be Absolutely Continuous With Respect to Lebesgue Measure

TO BE MADE AVAILABLE UPON REQUEST FROM THE AUTHOR

I now show that for any selection model, there will exist on some probability space an alternative selection model with the distribution of U being absolutely continuous with respect to Lebesgue measure and implying the same distribution of $(\{D_z\}, Y_0, Y_1, Z)$.

Theorem 2 *For any selection model consistent with assumption (S-1), there exists on some probability space an alternative selection model with the following three properties: (i) the alternative selection model implies the same distribution of* $(\{D_z\}, Y_0, Y_1, Z)$ *as the original selection model; (ii) the alternative selection model is consistent with assumption (S-1); (iii) in the alternative selection model, the distribution of* U *is absolutely continuous with respect to Lebesgue measure.*

Proof: Let G denote any given cumulative distribution function for (U, Y_1, Y_0, Z) consistent with assumption S-1. By assumption, for any $z \in \mathcal{Z}$, $D_z = 1[\nu(z) \geq U]$, and thus the distribution G of (U, Y_1, Y_0, Z) induces a distribution of $(\{D_z\}, Y_1, Y_0, Z)$.

Now consider an alternative distribution function. Define the functions

$$S(t) \equiv \inf\{s \in \text{Supp}(G_U(U) : s \leq G_U(t)\},$$

$$H(u^*|u) = \begin{cases} 0 & \text{if } u^* < S(u), \\ \frac{u^* - S(u)}{G_U(u) - S(u)} & \text{if } S(u) \leq u^* < G_U(u) \;\;\&\;\; S(u) \neq G_U(u) \\ 1 & \text{if } G_U(u) \leq u^*, \end{cases}$$

and

$$\tilde{G}(u^*, u, y_1, y_0, z) = \int_{-\infty}^{u} H(u^*|t) G_{Y_1, Y_0, Z|U}(y_1, y_0, z|t) dG_U(t).$$

We then have that there exists on some probability space a random vector $(\tilde{U}^*, \tilde{U}, \tilde{Y}_1, \tilde{Y}_0, \tilde{Z})$ having distribution function $\tilde{G}$. (This follows, for example, by a straightforward modification of Theorem 20.4 of Billingsley, 1995). Note that, by construction, the marginal distribution of $(\tilde{U}, \tilde{Y}_1, \tilde{Y}_0, \tilde{Z})$ equals the distribution of (U, Y_1, Y_0, Z),

$$\lim_{u^* \to \infty} \tilde{G}(u^*, u, y_1, y_0, z) = G(u, y_1, y_0, z).$$

Thus, defining $\tilde{D}_z = 1[\nu(z) \geq \tilde{U}]$ for each $z \in \mathcal{Z}$ (recall, by construction, Supp($\tilde{Z}$)=Supp(Z)), we have that the distribution of $(\{\tilde{D}_z\}, \tilde{Y}_1, \tilde{Y}_0, \tilde{Z})$ equals the distribution of $(\{D_z\}, Y_1, Y_0, Z)$.

Now consider the alternative selection model $\tilde{D}^*_z = 1[p(z) \geq \tilde{U}^*]$ for each $z \in \mathcal{Z}$. Note that $\text{Supp}(p(Z)) = \text{Supp}(G_U(\nu(Z)) \subseteq \text{Range}(G_U) = \text{Supp}(G_U(U))$. Since $1[\tilde{U} \leq t] = 1[\tilde{U}^* \leq t]$ w.p.1. for any $t \in \text{Supp}(G_U(U)$, and $\text{Supp}(p(Z)) \subseteq \text{Supp}(G_U(U))$, we have that $\tilde{D}_z = \tilde{D}^*_z$ w.p.1 for each $z \in \mathcal{Z}$. Thus, the distribution of $(\{\tilde{D}^*_z\}, \tilde{Y}_1, \tilde{Y}_0, \tilde{Z})$ equals the distribution $(\{\tilde{D}_z\}, \tilde{Y}_1, \tilde{Y}_0, \tilde{Z})$, and thus equals the distribution of $(\{D_z\}, Y_1, Y_0, Z)$. Thus, the alternative selection model implies the same distribution of $(\{D_z\}, Y_1, Y_0, Z)$ as the original selection model. Since $\nu(z)$ is a measurable, nontrivial function of z, and since $\tilde{Z}$ is independent of $(\tilde{U}^*, \tilde{Y}_1, \tilde{Y}_0)$ by construction, we have that this alternative selection model satisfies the conditions of (S-1). Finally, note that the marginal distribution of $\tilde{U}^*$ is unit uniform, and thus absolutely continuous with respect to Lebesgue measure.

C Measurability with $\mathcal{Z}$ Uncountable

TO BE MADE AVAILABLE UPON REQUEST FROM THE AUTHOR

I now show that the analysis of the note holds for the case where $\mathcal{Z}$ is uncountable. The difficulty for the case where $\mathcal{Z}$ is uncountable is that A, N, and C need not be measurable sets, and U need not be a random variable, since they are defined by uncountable intersections and unions of sets.

We proceed as follows. We redefine the sets A, N, and C using an intersection involving only one element of $\{z \in \mathcal{Z} : p(z) = t\}$ for each $t \in \mathcal{P}$.[13] Given this construction, we show that A, N, C are measurable sets, and that U is a measurable function. The proof that Z is independent of (U, Y_1, Y_0) and the proof that $\tilde{D}_z(\omega) = D_z(\omega)$ for all $z \in \mathcal{Z}$ w.p.1. then follow with only trivial modifications. Thus, even though we will define the selection model representation using only a subset of $\mathcal{Z}$, the equivalence will still hold w.p.1. for each $z \in \mathcal{Z}$.

We now redefine the sets A, N, and C. Let $\Psi \subset \mathcal{Z}$ denote a set such that, for each $t \in \mathcal{P}$, there exists one and only one $z \in \Psi$ s.t. $p(z) = t$. Define

$$\begin{aligned} N &\equiv \bigcap_{z \in \Psi} D_z^{-1}(0) \\ A &\equiv \bigcap_{z \in \Psi} D_z^{-1}(1) \\ C &\equiv \left(N \bigcup A \right)^c \end{aligned}$$

The following lemma shows that the sets A, N and C are measurable.

[13] As is made clear by the following analysis, the central feature of this construction is not that we are defining A, N and C using only one element of $\{z \in \mathcal{Z} : p(z) = t\}$ for each $t \in \mathcal{Z}$, but that we are using a finite or countable number of elements of $\{z \in \mathcal{Z} : p(z) = t\}$ for each $t \in \mathcal{Z}$. All analysis trivially extends to a construction using a countable subset of $\{z \in \mathcal{Z} : p(z) = t\}$ for each t.

Lemma 4 *Given L-1 and L-2, the sets N, A, and C are measurable.*

Proof: Consider the set $N = \bigcap_{z\in\Psi} D_z^{-1}(0)$. Consider the case where there does exist $z \in \Psi$ s.t. $p(z) = \sup \mathcal{P}$ (the supremum is attained). In this case, by the monotonicity and independence conditions,

$$\bigcap_{z\in\Psi} D_z^{-1}(0) = D_{z^*}^{-1}(0)$$

for the unique $z^* \in \Psi$ s.t. $p(z^*) = \sup \mathcal{P}$. Since $D_z^{-1}(0)$ is measurable for any $z \in \Psi \subseteq \mathcal{Z}$, we have that $\bigcap_{z\in\Psi} D_z^{-1}(0)$ is measurable for the case where the supremum is attained. Now consider the case where there does not exist $z \in \Psi$ s.t. $p(z) = \sup \mathcal{P}$ (the supremum is not attained). Let $\Psi^0 \subset \Psi$ denote a countable set s.t. $\{p(z) : z \in \Psi^0\}$ is dense in $\mathcal{P}$. Since $\Psi \supseteq \Psi^0$, we have

$$\bigcap_{z\in\Psi} D_z^{-1}(0) \subseteq \bigcap_{z\in\Psi^0} D_z^{-1}(0).$$

Since the supremum is not attained and $\{p(z) : z \in \Psi^0\}$ is dense in $\mathcal{P}$, we have that for any $z \in \Psi$ there exists $\tilde{z} \in \Psi^0$ s.t. $p(\tilde{z}) > p(z)$ and thus $D_{\tilde{z}}^{-1}(0) \subset D_z^{-1}(0)$. Thus,

$$\bigcap_{z\in\Psi} D_z^{-1}(0) \supseteq \bigcap_{z\in\Psi^0} D_z^{-1}(0),$$

and thus

$$\bigcap_{z\in\Psi} D_z^{-1}(0) = \bigcap_{z\in\Psi^0} D_z^{-1}(0)$$

for the case where the supremum is not attained. Conclude that N is a measurable set. The parallel argument shows that A is a measurable set, and thus that C is a measurable set.

For $\omega \in C$, define

$$\Psi_1(\omega) = \{z \in \Psi : D_z(\omega) = 1\},$$

$$U(\omega) = \begin{cases} 1 & \text{if } \omega \in N, \\ 0 & \text{if } \omega \in A, \\ \inf_{z \in \Psi_1(\omega)} p(z) & \text{if } \omega \in C, \end{cases}$$

The following lemma shows that U is a random variable.

Lemma 5 *Given L-1 and L-2, U is a random variable, i.e. the function U, $U : \Omega \mapsto \Re$, is measurable $\mathcal{F}$.*

Proof: Let $\mathcal{P}^c = [0,1] \setminus \mathcal{P}$. Let $\mathcal{P}_0$ denote a countable set that is dense in $\mathcal{P}$. For $j \in \mathbb{N}$, define $\mathcal{P}_j = \{p \in \mathcal{P} : (p, p + \frac{1}{j}) \subset \mathcal{P}^c\}$. $\mathcal{P}_j$ is countable for each j, so that $\bigcup_{j=0}^{\infty} \mathcal{P}_j$ is countable. Let $\Lambda \subseteq \Psi$ denote the countable set such that $\{p(z) : z \in \Lambda\} = \bigcup_{j=0}^{\infty} \mathcal{P}_j$. Define $\Lambda_1(\omega) = \{z \in \Lambda : D_z(\omega) = 1\}$.

We now show that $\inf_{z \in \Psi_1(\omega)} p(z) = \inf_{z \in \Lambda_1(\omega)} p(z)$ for each ω. Fix some ω. Note that $\Psi_1(\omega) \supseteq \Lambda_1(\omega)$ so that $\inf_{z \in \Psi_1(\omega)} p(z) \leq \inf_{z \in \Lambda_1(\omega)} p(z)$. Thus, we need only establish the converse, that $\inf_{z \in \Psi_1(\omega)} p(z) \geq \inf_{z \in \Lambda_1(\omega)} p(z)$. First assume $\inf_{z \in \Psi_1(\omega)} p(z)$ is in the closure of $\{p(z) : z \in \Psi, p(z) > \inf_{z \in \Psi_1(\omega)} p(z)\}$. Using that $\{p(z) : z \in \Lambda\}$ is dense in $\mathcal{P}$, we have that $\inf_{z \in \Psi_1(\omega)} p(z)$ is in the closure of $\{p(z) : z \in \Lambda, p(z) > \inf_{z \in \Psi_1(\omega)} p(z)\}$. By the monotonicity and independence conditions, we have $\{p(z) : z \in \Lambda, p(z) > \inf_{z \in \Psi_1(\omega)} p(z)\} \subseteq \{p(z) : z \in \Lambda, D_z(\omega) = 1\}$, so that $\inf_{z \in \Psi_1(\omega)} p(z)$ is in the closure of $\{p(z) : z \in \Lambda, D_z(\omega) = 1\}$. Thus, $\inf_{z \in \Psi_1(\omega)} p(z) \geq \inf_{z \in \Lambda_1(\omega)} p(z)$. Now assume that $\inf_{z \in \Psi_1(\omega)} p(z)$ is not in the closure of $\{p(z) : z \in \Psi, p(z) > \inf_{z \in \Psi_1(\omega)} p(z)\}$. Then there exists a $z^* \in \Psi$ such that $p(z^*) = \inf_{z \in \Psi_1(\omega)} p(z)$ and $D_{z^*}(\omega) = 1$. In this case, we have that there exists $j \in \mathbb{N}$ s.t. $p(z^*) \in \mathcal{P}_j$ and thus $p(z^*) \in \bigcup_{j=0}^{\infty} \mathcal{P}_j$ so that $z^* \in \Lambda$ and $\inf_{z \in \Lambda_1(\omega)} p(z) \leq p(z^*)$. Thus, $\inf_{z \in \Psi_1(\omega)} p(z) \geq \inf_{z \in \Lambda_1(\omega)} p(z)$. Conclude that $\inf_{z \in \Psi_1(\omega)} p(z) = \inf_{z \in \Lambda_1(\omega)} p(z)$ for each ω.

We now show that $U(\omega)$ is a measurable function. For any $t \in [0,1]$, $\{\omega \in C : \inf_{z \in \Psi_1(\omega)} p(z) < t\} = \{\omega \in C : \inf_{z \in \Lambda_1(\omega)} p(z) < t\} = \bigcup_{z \in \Lambda(t)} \left[C \bigcap D_z^{-1}(1)\right]$, where $\Lambda(t) = \{z \in \Lambda : p(z) < t\}$. Since $D_z^{-1}(1)$ is a measurable set for any $z \in \Lambda$ and Λ is countable, we have that $\bigcup_{z \in \Lambda(t)} \left[C \bigcap D_z^{-1}(1)\right]$ is measurable $C \bigcap \mathcal{F}$ for any $t \in [0,1]$. Thus, the restriction of U to C is measurable. The restrictions of U to N and to A are trivially measurable. Finally the sets N, A, and C are measurable and form a partition of Ω, and we thus have that $U(\omega)$ is a random variable, i.e., the function U, $U : \Omega \mapsto \Re$, is measurable $\mathcal{F}$.

The proof that Z is independent of (U, Y_1, Y_0) (Lemma 3) is valid without modification. The proof that $\tilde{D}_z(\omega) = \mathbf{1}[p(z) \geq U(\omega)]$ equals $D_z(\omega)$ w.p.1 for all $z \in \mathcal{Z}$ follows with only trivial modifications.[14]

[14]For example, note that the above construction of N implies $D_z(\omega) = 0$ for all $\omega \in N$, for all $z \in \Psi$, but does not imply $D_z(\omega) = 0$ for all $\omega \in N$, for all $z \in \mathcal{Z}$. However, the construction of N, combined with the independence and monotonicity assumptions of LATE, imply that $D_z(\omega) = 0$ for all $\omega \in N$ outside of a set of probability zero, for all $z \in \mathcal{Z}$.