Nonparametric Partial and Point Identification of Net or Direct Causal Effects*

Carlos A. Flores† Alfonso Flores-Lagunes‡

December 22, 2008
PRELIMINARY AND INCOMPLETE DRAFT

Abstract

Within the literature on causal statistical inference, an important goal is to examine the causal mechanisms or channels through which the treatment affects the outcome of interest. Net (or direct) causal effects measure the effect of the treatment on the outcome while blocking the effect of the treatment on the variable that represents the mechanism. Hence, net effects are useful in learning about the ways in which the treatment causally affects the outcome. This paper provides sufficient conditions under which net average effects can be partially and point identified without functional form, distributional, or constant-effects assumptions. First, we show that the data usually available to researchers contains information on the relevant potential outcome used in the definition of causal net effects only for a particular subpopulation: those individuals for which the treatment does not affect the mechanism variable. An implication of this result is that estimation of net effects for other subpopulations can only be based on extrapolations involving typically strong assumptions. Second, we show that by imposing a monotonicity condition on the effect of the treatment on the mechanism variable–which is also imposed on methods currently used for estimation of net effects–we can identify a subpopulation with that characteristic. We construct bounds for the average net effect for this subpopulation, and present additional sufficient assumptions under which this parameter is nonparametrically point identified. Finally, we show how the information available in this subpopulation can be used to construct bounds and point identify the average net effect for the entire population.

Key words and phrases: causal inference, net effects, direct effects, nonparametric identification, principal stratification.

JEL classification: C13, C21, C14

*Helpful comments have been provided by seminar participants at the University of Arizona, the 2008 Midwest Econometrics Group, and the 2008 Latin American Meetings of the Econometric Society.

†Department of Economics, University of Miami. Email: caflores@miami.edu

‡Food and Resource Economics Department, Universty of Florida. Email: alfonsofl@ufl.edu

1 Introduction

Within the literature on causal statistical inference, an important goal is to examine the causal mechanisms or channels through which the treatment affects the outcome of interest. This is evidently the case in many fields such as economics (e.g. Currie and Moretti, 2003), education (e.g. Zhang and Rubin, 2003), epidemiology (Petersen et al. 2006), sociology (e.g. Morgan and Winship, 2007), statistics (e.g. Rubin, 2004), among others. Net (or direct) causal effects measure the effect of a treatment on the outcome while blocking the effect of the treatment on a given mechanism (e.g., Pearl, 2000, 2001; Flores and Flores-Lagunes, 2008a). A comparison of the (total) treatment effect and the net effect is thus informative about the causal role that the mechanism has in the impact of the treatment on the outcome.

Identification of net or direct effects, however, is a difficult task since it requires stronger conditions than those necessary to identify total treatment effects (Robins and Greenland, 1992; Petersen et al., 2006; Rubin, 2004). An intuitive reason for this is that identification of net effects implies learning about a different (counterfactual) treatment from the one at hand: an alternative treatment in which the effect of the treatment on the mechanism variable is blocked (Flores and Flores-Lagunes, 2008a; hereafter FF). The assumptions made to estimate net treatment effects in the current literature typically involve (either explicitly or implicitly) a combination of functional form, distributional, and constant treatment effects assumptions (e.g., Robins and Greenland, 1992, Petersen et al., 2006; FF).

In this paper we analyze partial and point identification of net effects under weak assumptions. Employing insights from the analysis of identification of causal effects using instrumental variables in Imbens and Angrist (1994; hereafter IA) and Angrist, Imbens and Rubin (1996; hereafter AIR), we first show that, regardless of the treatment assignment, the typical data contains information on causal net effects only for a particular subpopulation: those individuals for which the treatment does not affect the mechanism variable. Our set up makes clear that this is the only subpopulation in the data for which information is available on the potential outcome when the effect of the treatment on the mechanism is blocked. An important implication of this result is that estimation of net effects for other subpopulations, including the complete population, can only be based on extrapolations involving typically strong assumptions, such as the ones mentioned in the previous paragraph.

Based on this result, we focus on providing sufficient conditions under which average net effects are partially or point identified for that particular subpopulation, subsets of it, and the entire population, allowing for heterogeneous total and net effects. By imposing a monotonicity condition on the effect of the treatment on the mechanism variable–a condition also imposed by current methods for estimation of net effects–enables identification of a subpopulation for which the treatment does not affect the mechanism. We derive bounds for net average treatment

effects in the spirit of Manski (1990), and more recently, Lee (2005) and Zhang et al. (2008). We also show that, under additional assumptions, one can also nonparametrically point identify the net average treatment effect for the subpopulation for which the treatment does not affect the mechanism as well as for the entire population.

There are several definitions in the literature of parameters aimed at capturing the average effect of a treatment on an outcome controlling for variables that are affected by the treatment.[1] Since the intermediate or mechanism variable is affected by the treatment, it is not straightforward to define causal parameters. In this paper, we focus on the Net Average Treatment Effect ($NATE$) defined in FF, which is based on the principal stratification framework developed by Frangakis and Rubin (2002; hereafter FR). It equals the average difference between the potential outcome from an alternative treatment in which the effect of the treatment on the mechanism variable is blocked and the potential outcome under the control treatment, for individuals with the same potential values of the mechanism variable (i.e., principal strata). As discussed in the following section, this parameter is particularly helpful in decomposing the part of the effect of a treatment on an outcome that works through a mechanism and it has a causal interpretation.

A parameter whose definition is close to $NATE$ is the average natural direct effect, or $ANDE$ (Pearl, 2001). Most of the focus on this parameter has been on its point identification (e.g., Robins and Greenland, 1992; Pearl, 2001; Petersen et al., 2006). These approaches require a conditional independence assumption for the selection into potential values of the mechanism variable, along with an assumption regarding the way in which the mechanism variable is allowed to interact with the treatment to affect the outcome. For instance, Robins and Greenland (1992) assume that for all units the effect of the outcome to a change on the treatment does not depend on the level at which the intermediate variable is held.[2] These assumptions are likely strong in many economic applications. Our point identification results do not require "non-interaction" assumptions of this type.

Some recent papers in the econometrics literature–Lee (2005) and Zhang et al. (2008)–focus on interval estimation of the effect of a randomly-assigned training program on wages, considering the fact that wages are only observed for those individuals who are employed. This set up leads to a sample selection problem because employment status may also be affected by the training program. It relates to the present paper since employment status may be regarded as a mechanism through which training affects wages. Zhang et al. (2008) employ a principal stratification approach—as we do—and argue that the relevant average treatment effect (ATE) of training on wages is for the subpopulation of individuals who would be employed whether they

[1] Some examples include the net treatment difference in Rosenbaum (1984), the direct effect in Mealli and Rubin (2003), and the controlled and natural direct effects in Pearl (2001). These parameters will be discussed below.

[2] For a discussion on similar assumptions used in the literature, see Petersen et al. (2006).

received training or not. Similarly, Lee (2005) focuses on the ATE for those individuals who would be employed whether trained or not, presenting estimators for the bounds he develops and deriving their asymptotic distribution. Our partial identification strategy is similar in spirit to those in Zhang et al. (2008) and Lee (2005).[3]

A difference between our general set up and that in Lee (2005) and Zhang et al. (2008) is that in those papers the observability of the outcome (wages) depends on an intermediate variable (employment status), while in ours the outcome is always observed. Another important difference is on the parameters of interest: those papers focus on the ATE, whereas our focus is on $NATE$. Hence, for instance, while the ideal data for them would come from a randomly assigned treatment in which the outcome is always observed (even if the person were unemployed), our ideal data would come from an experiment in which the treatment is the same as the original one but blocks the effect of the treatment on the mechanism variable. Nevertheless, the two effects will be equal for the subpopulation for which the treatment does not affect the mechanism. In this regard, the bounds in those papers are a special case of the ones presented here. Finally, the present paper differs from those in that we also provide bounds for $NATE$ for the entire population, consider point identification, and extend our results to cases when the mechanism variable is multivalued and there is more than one possible mechanism.

Throughout this paper we will assume that the treatment is randomly assigned. This allows us to focus our attention on the specific issues related to estimation of net effects, and it sets the basis for extensions to other treatment assignment mechanisms. Additionally, randomized experiments have gained importance in economics as a way of estimating causal effects, as evidenced in the literatures of program evaluation (e.g., Heckman et al., 1999) and field experiments (e.g., Karlan and List, 2007), among others. The methods developed herein should allow researchers learning further insights from their randomized experiments.

The paper is organized as follows. Section 2 presents the estimands of interest. Section 3 presents the main results in the paper focusing on the case of a binary treatment and a binary mechanism. Subsequently, Section 4 extends the analysis in the previous section to the case of a multi-valued mechanism and to the case where interest lies on learning about the relative causal importance of several mechanisms. To our knowledge, the latter case has not been considered previously in the literature. Finally, section 5 concludes and discusses future extensions.

[3] In addition, Zhang et al. (2008) also look at point estimation of their parameter of interest using a parametric Bayesian model. For instance, they model log potential wages using a standard normal linear regression. As previously mentioned, in the present paper we avoid this kind of parametric assumptions in the identification of our parameters of interest.

2 Definition of Estimands

We start this section by introducing some notation and presenting the general set up, which we base on the potential outcomes framework (Neyman, 1923; Rubin, 1974). Assume we have a random sample of size N from a large population. For each unit i in the sample, let $T_i \in \{0,1\}$ indicate whether the unit received the treatment of interest ($T_i = 1$) or the control treatment ($T_i = 0$). We are interested on the effect of the treatment T on an outcome Y and on analyzing the part of that effect that works through a mechanism variable S. Since S is affected by the treatment, we denote by $S_i(T_i)$ its potential values.[4]

Define the "composite" potential outcomes $Y_i(\tau, \zeta)$, where the first argument refers to one of the treatment arms ($\tau \in \{0,1\}$) and the second argument represents one of the potential values of the post-treatment variable S ($\zeta \in \{S_i(0), S_i(1)\}$). Note that the potential outcomes $Y_i(1, S_i(1))$ and $Y_i(0, S_i(0))$ correspond to the potential outcomes $Y_i(1)$ and $Y_i(0)$ typically used in the literature to define treatment effects. The potential outcome $Y_i(1, S_i(0))$ represents the outcome individual i would receive if she were exposed to the treatment but the effect of the treatment on the mechanism were blocked by keeping the mechanism at $S_i(0)$. This potential outcome plays a crucial role in the definition of net and mechanism effects discussed below.[5] Finally, for each unit i, we observe the vector $\left(T_i, Y_i^{obs}, S_i^{obs}\right)$, where $Y_i^{obs} \equiv T_i Y_i(1, S_i(1)) + (1 - T_i) Y_i(0, S_i(0))$ and $S_i^{obs} = T_i S_i(1) + (1 - T_i) S_i(0)$. As usual in the program evaluation literature, we focus on average causal effects. The population average treatment effect is given by $ATE = E[Y(1, S(1)) - Y(0, S(0))]$.[6]

Defining net effects and attaching a causal interpretation to the definition is not trivial because one has to consider the fact that the mechanism variable is potentially affected by the treatment. Maybe not surprisingly, one can find several definitions of net effects in the literature. In this paper, we focus on the net average treatment effect defined in FF (2008a), which decomposes the ATE into a net and a mechanism effect while having a causal interpretation.

FF based their definition of net effects on the concept of principal stratification introduced to the literature by FR (2002) for defining causal effects in the presence of post-treatment variables. In the potential outcomes framework, a causal effect must be a comparison of potential outcomes for the same group of individuals under treatment and control. The idea in FR is to define the "same group of individuals" based on the potential values of the post-treatment variable.

[4] Note that at this stage S is not restricted to be binary.

[5] Another potential outcome is $Y_i(0, S_i(1))$, the outcome an individual would obtain when the treatment is not given to her but she receives a value of the post-treatment variable equal to $S_i(1)$. A similar decomposition as the one to be presented below is possible using this potential outcome. If interest lies in such decomposition, the results presented in this paper can also be applied there.

[6] We adopt the stable unit treatment value assumption (SUTVA) following Rubin (1980). This assumption is common throughout the literature, and it implies that the treatment effects at the individual level are not affected either by the method used to assign the treatment or by the treatment received by other units. In practice, this assumption rules out general equilibrium effects of the treatment that may impact individuals.

In FR's terminology, the basic principal stratification with respect to post-treatment variable S is a partition of individuals into groups such that within each group all individuals have the same vector $\{S(0) = s_0, S(1) = s_1\}$, where s_0 and s_1 are generic values of $S(0)$ and $S(1)$, respectively. A principal effect with respect to a principal strata is defined as a comparison of potential outcomes within that strata. Since principal strata are not affected by treatment assignment, individuals in that group are indeed comparable and thus principal effects are causal effects.[7]

In order to causally interpret our parameters of interest, we employ the concept of principal stratification and condition on the principal strata $\{S(0) = s_0, S(1) = s_1\}$. Following FF (2008a), write the ATE controlling for principal strata as

$$ATE = E\{E[Y(1, S(1)) - Y(0, S(0))|S(0) = s_0, S(1) = s_1]\} = E[\tau(s_0, s_1)], \tag{1}$$

where the outer expectation is taken over $S(0)$ and $S(1)$ and we let $\tau(s_0, s_1) = E[Y(1, S(1)) - Y(0, S(0))|S(0) = s_0, S(1) = s_1]$. Then, we can decompose the ATE as:

$$\begin{aligned} ATE &= E\{E[Y(1, S(1)) - Y(1, S(0))|S(0) = s_0, S(1) = s_1]\} \\ &\quad + E\{E[Y(1, S(0)) - Y(0, S(0))|S(0) = s_0, S(1) = s_1]\}. \end{aligned} \tag{2}$$

Define the (causal) net average treatment effect or $NATE$ as:

$$NATE = E\{E[Y(1, S(0)) - Y(0, S(0))|S(0) = s_0, S(1) = s_1]\} \tag{3}$$

and the (causal) mechanism average treatment effect or $MATE$ as:

$$MATE = E\{E[Y(1, S(1)) - Y(1, S(0))|S(0) = s_0, S(1) = s_1]\}. \tag{4}$$

Note that, in the definition of $NATE$ and $MATE$, $Y(1, S(0))$ plays a key role as it is the potential outcome under receipt of treatment when the effect of the treatment on the mechanism is blocked or held at its value under no receipt of treatment.[8] Since by definition $ATE = NATE + MATE$, we focus hereafter on the parameter $NATE$.

An intuitive way to think about $NATE$ is to consider $Y(1, S(0))$ as the potential outcome of an alternative counterfactual experiment in which the treatment is the same as the original one but blocks the effect of T on S by holding S fixed at $S_i(0)$ for each individual i. The $NATE$ for individual i is then the difference between the outcome of this alternative treatment, $Y_i(1, S_i(0))$, and $Y_i(0, S_i(0))$ from the original control treatment.

[7] FR's idea of principal stratification is closely related to the local average treatment effect interpretation of instrumental variables by IA (1994) and AIR (1996). For example, in their terminology, the group of "compliers" is the set of individuals that always comply with their treatment assignment regardless of whether their assignment is to treatment ($T = 1$) or control group ($T = 0$). Therefore, for this group $\{S(0) = 0, S(1) = 1\}$, where S is an indicator of actual treatment reception.

[8] We consider in this section a decomposition of the total effect based on one mechanism of interest. It is possible to extend the decomposition to accommodate more than one mechanism, which we present in Section 4.

An important property of $NATE$ in (3) is that it includes not only the part of the ATE that is totally unrelated to the mechanism variable S, but also the part of the ATE that results from a change *in the way* S affects Y. That is, even though the level of S is held fixed at $S(0)$, the treatment may still affect the way in which S affects the outcome, and this is counted as part of $NATE$. As in FF, we argue that including the effect of T on how S affects Y (i.e., returns to S) in $NATE$ is more relevant from a policy perspective, compared to a different parameter that holds constant the way S affects Y. The reason is that a policy maker typically has some degree of control over S, while very rarely over how S affects Y.[9] Defining $NATE$ in this way is consistent with Holland's (1986) notion of a "treatment" being an intervention that can be potentially applied to each individual. Another important characteristic of $NATE$ is that it has the desirable property that it equals zero when all the effect of the T on Y works through S, and it equals the ATE when none of the effect works through S (either because T does not affect S or S does not affect Y).

Before concluding this section, we briefly discuss the relation of $NATE$ to other parameters in the literature. One of the first parameters considered in the literature was the net treatment difference (NTD) introduced by Rosenbaum (1984), which can be written as $NTD = E\{E[Y(1) - Y(0)|S^{obs}]\}$.[10] However, without further assumptions it has no causal interpretation because it conditions on S^{obs}. Since S^{obs} represents two different potential variables, $S(1)$ and $S(0)$, units with the same value of S^{obs} are generally not comparable. Mealli and Rubin (2003) and Rubin (2004) define the concepts of direct and indirect effects using principal stratification as a comparison of $Y(1)$ and $Y(0)$ within the stratum for which $S(0) = S(1) = s$. Therefore, their concept of direct effect is a special case of NATE defined for that subpopulation, since in this stratum $Y(1) = Y(1, S(0))$. Unless NATE is constant over the population, the direct effect does not equal NATE. Similarly, the parameters considered by Lee (2005) and Zhang et al. (2008) are special cases of $NATE$, since they focus on the ATE of training on wages for those individuals who would be employed whether trained or not. This is a subset (stratum) of the population for which training does not affect employment status.[11]

Other parameters related to $NATE$ are the average controlled direct effect ($ACDE$) and the average natural direct effect ($ANDE$), defined in, e.g., Pearl (2000, 2001), Robins and Greenland (1992), and Petersen et al. (2006). The $ACDE$ at a specific value $\overline{s}$ of S can be written as $ACDE = E[Y(1, S(1) = \overline{s}) - Y(0, S(0) = \overline{s})]$. The $ACDE$ gives the average difference between the counterfactual outcome under the two treatment arms controlling for the value of the mechanism variable at $\overline{s}$. This parameter has some undesirable features such

[9] A case where the policymaker might have some degree of influence on how S affects Y is when general equilibrium effects due to the treatment are present.

[10] Work that implicitly or explicitly estimates the NTD is Black and Smith (2004), Dearden et al. (2002) and Ehrenberg et al. (2007), among others.

[11] In fact, the estimands in Lee (2005) and Zhang et al. (2006) equal the direct effect of training on wages as defined in Mealli and Rubin (2003), with unemployment status as post-treatment variable.

as not decomposing the ATE into a net and a mechanism effect;[12] and that, even if in fact the treatment does not affect the mechanism variable S, the ATE may be different from the $ACDE$ if there is heterogeneity in the effect of T on Y along the values of S. Conversely, the $ANDE$ can be written as $E[Y(1,S(0))-Y(0,S(0))]$. Hence, this parameter is similar to $NATE$ in (3) with the subtle but important difference that $NATE$ conditions on principal strata in order to achieve causal interpretation. This distinction is crucial when stating assumptions for its identification–as shall become clear in the following section.[13]

Finally, Kaufman et al. (2005) and Cai et al. (2007) provide nonparametric bounds for the ACDE. The latter paper extends the former by applying the symbolic Balke-Pearl (1997) linear programming method to derive closed-form formulas for the bounds, and by extending the analysis to the case when the treatment, the intermediate, and the outcome variables are all multivalued. Their approach rests on two important assumptions. First, monotonicity assumptions about the effects of (i) the treatment on the intermediate variable; (ii) the intermediate variable on the outcome; and, (iii) the treatment on the outcome. Second, an assumption that for all units the effect of the outcome to a change on the treatment does not depend on the level at which the intermediate variable is held (i.e. a "no-interaction" assumption). This assumption is likely strong in economic applications. Our bounds below are focused on $NATE$ as opposed to the $ACDE$ for the reasons previously mentioned. In addition, our approach to construction of bounds relaxes some of the assumptions in those papers.

3 Nonparametric Identification

This section presents the main results of the paper. In order to motivate the problem of learning about $NATE$ in (3) from the available data, we start by discussing the two main challenges faced in its identification. The first is that a key potential outcome needed for estimation of $NATE$, $Y_i(1,S_i(0))$, is generally not observed–this is in contrast to the case of estimation of the ATE, where only one of the relevant potential outcomes is missing for every unit. This implies, for instance, that even if all explanatory variables in the regression $Y^{obs}=a+bT+cS^{obs}+d'X+u$ were uncorrelated to the error term u (with X being a set of covariates), b does not equal $NATE$. In this simple example, the coefficient b gives the effect of T on Y holding S fixed at a given value $\overline{s}$ (i.e. the $ACDE$), and not at $S(0)$ as required by $NATE$. The second challenge is that for each unit under study only one of the potential values of the post-treatment variable is observed: S^{obs} represents $S(1)$ for treated units and $S(0)$ for

[12] For example, we could write the ATE as: $ATE = E[Y(1,S(1))-Y(1,S(1)=\overline{s})] + ACDE + E[Y(0,S(0)=\overline{s})-Y(0,S(0))]$. The first term gives the average effect of giving the treatment to the individuals and moving the value of the post-treatment variable from $\overline{s}$ to $S(1)$. The second term represents the average effect of giving the control treatment to the individuals and moving the value of the post-treatment variable from $S(0)$ to $\overline{s}$. These two effects are hard to interpret as mechanism effects of T on Y through S.

[13] Rubin (2005) also emphasizes the importance of defining causal parameters based on principal stratification.

controls units. This implies that the principal strata $\{S(0) = s_0, S(1) = s_1\}$, necessary for a causal interpretation of $NATE$, is not observable.[14]

Some intuition behind these challenges can be gained by thinking about a situation in which none of them are present and thus the estimation of net effects is straightforward. Imagine performing a new experiment in which the treatment is the same as the original one but blocks the effect of the treatment on the mechanism variable. In this hypothetical case, a comparison of the mean outcomes for those receiving this counterfactual treatment and those in the control group would yield $NATE$, and none of the previous challenges would arise. Consequently, we can think of those challenges arising from the desire of learning about a different experiment from the one available.

The approaches currently available in the literature for estimation of $NATE$ or other net effects (e.g., $ACDE$ or $ANDE$) typically involve strong parametric assumptions. For instance, consider one of the estimation procedures in FF. The main idea in this case is to use the potential outcome $Y(1, S(1))$ to learn about $Y(1, S(0))$. More specifically, they model the conditional expectation of $Y(1, S(1))$, $E[Y(1, S(1)) | S(1) = s_1, X = x]$, using a parametric function $f(1, S(1))$ (e.g., OLS), and assume $E[Y(1, S(0)) | S(0) = s_1, X = x] = f(1, S(0))$. In addition to this functional form assumption and assuming T is randomly assigned, this approach requires the strata $\{S(1), S(0)\}$ to be independent of the potential outcomes conditional on the covariates X. Alternatively, in addition to conditional independence, Petersen et al. (2006) employ an assumption regarding the way in which the mechanism variable is allowed to interact with the treatment to affect the outcome, while the implementation of their approach is based on linear regressions. Yet another approach involves the use of a parametric Bayesian model. For instance, following Hirano et al. (2000), Zhang et al. (2008) model the potential outcomes (given the strata and covariates) and the probabilities of the principal strata (given covariates) using a standard normal linear regression and a multinomial logit, respectively, to estimate the ATE of training on wages for the subpopulation of individuals who would be employed whether they received training or not.

Our goal in this section is to analyze how much can be learned about average net effects from the typical data by employing weak assumptions that do not impose functional form or distributional assumptions as the previous studies, while allowing for heterogeneous effects. We start by analyzing the information available in the typical data about the potential outcomes in the definition of $NATE$. Subsequently, we show conditions allowing partial identification of net average treatment effects. Next, we show additional sufficient conditions for their point identification.

[14] Note that S can be regarded as an outcome, and thus the distribution of the principal strata equals the joint distribution of the potential outcomes $\{S(1), S(0)\}$, which is not easily identifiable (e.g., Heckman, Smith and Clements, 1997).

3.1 The Information Contained in Typical Data

Our first result is the observation that the data described in section 2, which is of the kind typically available to researchers, contains information on the key potential outcome $Y(1, S(0))$ only for a particular subpopulation: those for which the treatment does not affect the mechanism. For this subpopulation we have $S_i(1) = S_i(0)$, which implies that $Y_i(1, S_i(0)) = Y_i(1, S_i(1))$ and, hence, $Y_i(1, S_i(0))$ is observed for those receiving treatment. We state this as a result in order to highlight its importance.

Result 1 *The observed data* $\left(T_i, Y_i^{obs}, S_i^{obs}\right)$ *contains information on* $Y(1, S(0))$ *only for those units that receive the treatment and for which the treatment does not affect the mechanism variable* $(S_i(1) = S_i(0))$.

This result does not depend on the assignment mechanism of the treatment, or on whether the treatment or the mechanism variable are binary or continuous. It implies that, under heterogeneous effects, estimation of average net effects for other subpopulations (including the entire population) can only be based on extrapolations of $Y(1, S(0))$ to those units for which the treatment affects the mechanism, since their potential outcome is never observed. This result exemplifies the difficulty of estimating $NATE$ and $MATE$ with the data usually available.

Given this result, we begin investigating the use of weak assumptions for identification of net effects by concentrating on partial identification of the $NATE$ for the subpopulation characterized by $S_i(1) = S_i(0)$. In the spirit of IA (1994), define the local $NATE$ (hereafter $LNATE$) for this subpopulation as:

$$LNATE = E\{E[Y(1, S(0)) - Y(0, S(0))|S(0) = s, S(1) = s]\} \tag{5}$$

There is precedent in the literature on the importance of local average treatment effects (IA, 1994). In fact, the parameter of interest in Lee (2005) and Zhang at al. (2008) is a special case of $LNATE$ since they focus on the ATE for those who would be employed whether trained or not, which equals $LNATE$ for a particular value s. Note that, since in the subpopulation for which $S_i(1) = S_i(0)$ we have $Y_i(1, S_i(0)) = Y_i(1, S_i(1))$, its $LNATE$ equals its ATE. While $LNATE$ pertains to a specific population and does not decompose ATE without further assumptions, its estimation is important in learning if, for at least a subpopulation, the treatment has an effect on the outcome outside of the potential role of S. This has implications regarding testing of exclusion restrictions in the context of triangular simultaneous equations (Flores and Flores-Lagunes, 2008b). In addition, estimating $LNATE$ under mild assumptions clarifies the need for additional assumptions to allow identification of $NATE$, as will be illustrated below.

In estimating $LNATE$, even though we know the data is informative about $Y(1, S(0))$ for the subpopulation for which the treatment does not affect the mechanism, the fact that principal

strata is not observed prevents us from identifying units that belong to this subpopulation without further assumptions. Providing conditions under which this subpopulation is identified from the observed data will be our first task.

We start by considering the case in which the mechanism variable S is binary and the treatment is randomly assigned. This allows us to focus on the general ideas behind the identification results. The next section discusses extensions to the cases when the mechanism variable is multivalued, and when interest lies in analyzing several mechanisms simultaneously. Let the potential values of the mechanism take values $S_i(\tau) = \{0, 1\}$ for $\tau = 0, 1$ and assume the treatment is randomly assigned, which implies that the treatment received by each individual is independent of her potential outcomes and potential values of the post-treatment variable:[15]

Assumption 1 (*Randomly Assigned Treatment).*

$$Y\left(1, S\left(1\right)\right), Y\left(0, S\left(0\right)\right), Y(1, S\left(0\right)), S\left(1\right), S\left(0\right) \perp T.$$

This setup gives rise to four principal strata that are analogous to the "compliance types" of AIR (1996), obtained by combining the values that can be taken by the potential values of S, $S_i(0)$ and $S_i(1)$. In this case we have:

Table 1

		$S_i(0)$ = 0	$S_i(0)$ = 1
$S_i(1)$	0	not-affected at 0 (na^{00})	affected negatively (an)
	1	affected positively (ap)	not-affected at 1 (na^{11})

In this context, the subpopulation for which the treatment does not affect the mechanism is composed of two strata, the not-affected at 0 (na^{00}) and the not-affected at 1 (na^{11}). Based on the concept of principal stratification, we could estimate causal effects by taking averages within principal strata. Unfortunately, the principal strata is not directly observed, but rather T_i and S_i^{obs}, which complicates identification since the observed groups of individuals contain a mix of the principal strata:

Table 2

		T_i = 0	T_i = 1
S_i^{obs}	0	ap & na^{00}	an & na^{00}
	1	na^{11} & an	na^{11} & ap

A common assumption that allows identification of certain principal strata is monotonicity (IA, 1994):

[15] As in Dawid (1979), we write $X \perp Y$ to denote independence of X and Y.

Assumption 2 (*Monotonicity).*

$$S_i(1) \geq S_i(0) \text{ for all } i; \text{ or } S_i(1) \leq S_i(0) \text{ for all } i.$$

Assumption 2 states that the effect of the treatment on the mechanism is monotone for all individuals. IA (1994) employed an assumption of a monotone effect of the instrument on the actual reception of the treatment in the context of identification of average (total) treatment effects using instrumental variables. Here, monotonicity is applied to the effect that the treatment has on the value of the mechanism variable. A similar assumption is also employed by Lee (2005) and Zhang et al. (2008) within their context.

Assumption 2 rules out the existence of the *an* principal stratum, thereby allowing the identification of members of the subpopulations of na^{00} and na^{11}. This application of the work by IA and AIR to the context of net effects leads to the following result. Without loss of generality, let $S_i(1) \geq S_i(0)$ for all i.

Result 2 *Under Assumptions 1-2 (with* $S_i(1) \geq S_i(0)$ *for all* i*), those units with* $(T_i, S_i^{obs}) = (1,0)$ *are* na^{00} *that received the treatment, and those units with* $(T_i, S_i^{obs}) = (0,1)$ *are* na^{11} *that did not received the treatment.*

This result implies that we can identify $E[Y(0, S(0))]$ for na^{11} and $E[Y(1, S(0))]$ for na^{00}. While this is progress, it is not enough to point identify $LNATE$ for any subpopulation, since each expectation is identified for different strata that are in general non-comparable. Point identification will require additional assumptions to construct missing counterfactuals. Conversely, under the current weak assumptions, bounds can be created for the local $NATE$ for na^{00} ($LNATE_{na^{00}}$) and na^{11} ($LNATE_{na^{11}}$), as well as $LNATE$ in (5).

3.2 Partial Identification of Net Average Treatment Effects

3.2.1 Local Net Average Treatment Effects

There is an important body of work in the econometrics of program evaluation that analyzes partial identification of causal effects, such as Manski (1990), Imbens and Manski (2004), Lee (2005), and Zhang et al. (2008), among others. We derive in this subsection bounds on $LNATE$ for the subpopulations of na^{00}, na^{11}, and both, employing the general approach in Manski (1990) and its application to principal stratification in Zhang and Rubin (2003) and Zhang et al. (2008). We also present estimators of these bounds and derive their asymptotic properties based on Lee (2005).

Using the terminology of Table 1, let $\pi_{na^{00}}$, $\pi_{na^{11}}$, π_{ap}, and π_{an} be the population proportions of each of the principal strata na^{00}, na^{11}, ap, and an, respectively. We maintain Assumptions

1 and 2.[16] Under these assumptions $\pi_{an} = 0$ and the observed groups of individuals correspond to those in Table 2 once the *an* have been deleted. We start by constructing bounds on the $LNATE$ for na^{00}, which can be written as:

$$LNATE_{na^{00}} = E[Y(1, S(0))|na^{00}] - E[Y(0, S(0))|na^{00}]. \tag{6}$$

From Result 2, the key counterfactual $Y_i(1, S_i(0))$ is observed for na^{00} in the group with $T_i = 1$ and $S_i^{obs} = 0$, while the group of untreated na^{00} is mixed with ap in the group with $T_i = 0$ and $S_i^{obs} = 0$. Hence, the task is to construct a bound for the second term in (6). Similar to Imbens and Rubin (1997), note that the average outcome for individuals in this group can be written as:

$$E[Y_i^{obs}|T_i = 0, S_i^{obs} = 0] = \frac{\pi_{na^{00}}}{\pi_{na^{00}} + \pi_{ap}} \cdot E[Y_i(0, S_i(0))|na^{00}] + \frac{\pi_{ap}}{\pi_{na^{00}} + \pi_{ap}} \cdot E[Y_i(0, S_i(0))|ap] \tag{7}$$

Under Assumptions 1 and 2, the proportion of na^{00} and ap are identifiable from the data. Let $p_{s|t} = \Pr\left(S^{obs} = s|T_i = t\right)$ for $t, s = 0, 1$. Given a randomly assigned treatment, all four conditional probabilities can be estimated. By Assumption 2 the proportion of na^{00} in the group with $T_i = 0$ and $S_i^{obs} = 0$ can be written as a function of $p_{0|0}$ and $p_{0|1}$. From Table 2, we have that $p_{0|0} = \pi_{ap} + \pi_{na^{00}}$ and $p_{0|1} = \pi_{na^{00}}$, which implies $\pi_{na^{00}}/(\pi_{na^{00}} + \pi_{ap}) = p_{0|1}/p_{0|0}$. Therefore, $E[Y_i(0, S_i(0))|na^{00}]$, which corresponds to the second term in (6), can be bounded from above by the expected value of Y^{obs} in the $p_{0|1}/p_{0|0}$ fraction of *largest values* of Y^{obs} for those in the observed group with $T_i = 0$ and $S_i^{obs} = 0$. Similarly, it can be bounded from below by the expected value of Y^{obs} in the $p_{0|1}/p_{0|0}$ fraction of *smallest values* of Y^{obs} for those in the same observed group.

More formally, let y_r^{ts} be the r-th quantile of Y^{obs} conditional on $T_i = t$ and $S_i^{obs} = s$, or $y_r^{ts} = F^{-1}_{Y^{obs}|T=0,S^{obs}=0}(r)$, with $F.(\cdot)$ the conditional density of Y^{obs} given $T = t$ and $S^{obs} = s$. For example, y_r^{00} is the r-th quantile of Y^{obs} conditional on $T_i = 0$ and $S_i^{obs} = 0$. The bounds for $LNATE_{na^{00}}$ are given in the following proposition.

Proposition 1 *Under Assumptions 1 and 2, the parameter $LNATE_{na^{00}}$ in (6) can be bounded below by $LNATE^{LB}_{na^{00}}$ and above by $LNATE^{UB}_{na^{00}}$, where:*

$$\begin{aligned} LNATE^{LB}_{na^{00}} &= E[Y^{obs}|T_i = 1, S_i^{obs} = 0] - E[Y^{obs}|T_i = 0, S_i^{obs} = 0, Y^{obs} \geq y^{00}_{1-(p_{0|1}/p_{0|0})}] \\ LNATE^{UB}_{na^{00}} &= E[Y^{obs}|T_i = 1, S_i^{obs} = 0] - E[Y^{obs}|T_i = 0, S_i^{obs} = 0, Y^{obs} \leq y^{00}_{(p_{0|1}/p_{0|0})}] \end{aligned}$$

Analogous bounds can be constructed for na^{11} following a similar strategy. Define

$$LNATE_{na^{11}} = E[Y(1, S(0))|na^{11}] - E[Y(0, S(0))|na^{11}]. \tag{8}$$

Then, the corresponding lower and upper bounds are given by:

[16] We note that bounds can be constructed disposing of the monotonicity assumption (Manski, 1990), although they are typically uninformative (Lee, 2005).

Proposition 2 *Under Assumptions 1 and 2, the parameter $LNATE_{na^{11}}$ in (8) can be bounded below by $LNATE^{LB}_{na^{11}}$ and above by $LNATE^{UB}_{na^{11}}$, where:*

$$\begin{aligned} LNATE^{LB}_{na^{11}} &= E[Y^{obs}|T_i=1,S_i^{obs}=1,Y^{obs}\leq y^{11}_{(p_{1|0}/p_{1|1})}] - E[Y^{obs}|T_i=0,S_i^{obs}=1] \\ LNATE^{UB}_{na^{11}} &= E[Y^{obs}|T_i=1,S_i^{obs}=1,Y^{obs}\geq y^{11}_{1-(p_{1|0}/p_{1|1})}] - E[Y^{obs}|T_i=0,S_i^{obs}=1] \end{aligned}$$

The bounds for $LNATE_{na^{11}}$ correspond to those previously derived by Lee (2005) and Zhang et al. (2008) in a different context. Finally, note that we can write $LNATE$ in (5) as:

$$LNATE = [\pi_{na^{00}}/(\pi_{na^{00}}+\pi_{na^{11}})]LNATE_{na^{00}} + [\pi_{na^{11}}/(\pi_{na^{00}}+\pi_{na^{11}})]LNATE_{na^{11}}. \quad (9)$$

Given the bounds previously derived and the fact that the proportions are identifiable, the following proposition presents the bounds for $LNATE$.

Proposition 3 *Under Assumptions 1 and 2, the parameter $LNATE$ in (5) can be bounded by:*

$$\begin{aligned} LNATE^{LB} &= \left(\frac{p_{0|1}}{p_{0|1}+p_{1|0}}\right)LNATE^{LB}_{na^{00}} + \left(\frac{p_{1|0}}{p_{0|1}+p_{1|0}}\right)LNATE^{LB}_{na^{11}} \\ LNATE^{UB} &= \left(\frac{p_{0|1}}{p_{0|1}+p_{1|0}}\right)LNATE^{UB}_{na^{00}} + \left(\frac{p_{1|0}}{p_{0|1}+p_{1|0}}\right)LNATE^{UB}_{na^{11}} \end{aligned}$$

Finding estimators for the bounds previously defined is straightforward. We can use sample analogs of the parameters appearing in the bounds of the previous propositions. Let $1(\cdot)$ be the indicator function. Then, for $t,s=0,1$, we use the following estimators for the corresponding unknown objects:

$$\widehat{p}_{s|t} = \frac{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)}{\sum_{i=1}^{n} 1(T_i=t)}$$

$$\widehat{E}[Y^{obs}|T_i=t,S_i^{obs}=s] = \frac{\sum_{i=1}^{n} Y^{obs}\cdot 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)}{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)} \equiv \overline{Y}^{ts}$$

$$\widehat{E}[Y^{obs}|T_i=t,S_i^{obs}=s,Y^{obs}\leq y_r^{ts}] = \frac{\sum_{i=1}^{n} Y^{obs}\cdot 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)\cdot 1\big(Y^{obs}\leq \widehat{y}_r^{ts}\big)}{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)\cdot 1\big(Y^{obs}\leq \widehat{y}_r^{ts}\big)} \equiv \overline{Y}^{ts}_{\leq r}$$

$$\widehat{E}[Y^{obs}|T_i=t,S_i^{obs}=s,Y^{obs}\geq y_r^{ts}] = \frac{\sum_{i=1}^{n} Y^{obs}\cdot 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)\cdot 1\big(Y^{obs}\geq \widehat{y}_r^{ts}\big)}{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)\cdot 1\big(Y^{obs}\geq \widehat{y}_r^{ts}\big)} \equiv \overline{Y}^{ts}_{\geq r}$$

$$\widehat{y}_r^{ts} = \min\left\{ y : \frac{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)\cdot 1\big(Y^{obs}\leq y\big)}{\sum_{i=1}^{n} 1\big(S_i^{obs}=s\big)\cdot 1(T_i=t)} \geq q \right\}$$

Hence, for instance, the corresponding estimators for the bounds of $LNATE_{na^{00}}$ are given by $\widehat{LNATE}^{LB}_{na^{00}} = \overline{Y}^{10} - \overline{Y}^{00}_{\geq(1-(\widehat{p}_{0|1}/\widehat{p}_{0|0}))}$ and $\widehat{LNATE}^{UB}_{na^{00}} = \overline{Y}^{10} - \overline{Y}^{00}_{\leq(\widehat{p}_{0|1}/\widehat{p}_{0|0})}$. The other bound estimators ($\widehat{LNATE}^{LB}_{na^{11}}$, $\widehat{LNATE}^{LB}_{na^{11}}$, $\widehat{LNATE}^{LB}$ and $\widehat{LNATE}^{LB}$) are constructed in a similar way.

To conduct statistical inference, we derive asymptotic properties of the estimators of the bounds. We follow an approach similar to that in Lee (2005) and write our estimators as a solution to a GMM problem. The results then follow by applying standard asymptotic results (e.g., Newey and McFadden, 1994). The proofs of the propositions below are shown in the appendix.[17] In what follows, let an "o" after a parameter bound represent its true value.

Proposition 4 *Let Ω be the parameter space for the bounds. Assume Ω is compact and $E\left[|Y^{obs}|\right] < \infty$. Then, $\widehat{LNATE}^{LB}_{na^{00}}$, $\widehat{LNATE}^{UB}_{na^{00}}$ $\widehat{LNATE}^{LB}_{na^{11}}$, $\widehat{LNATE}^{UB}_{na^{11}}$, $\widehat{LNATE}^{LB}$ and $\widehat{LNATE}^{UB}$ are all consistent.*

We introduce some notation to simplify the expressions for the asymptotic variances of the bound estimators. Let $\mu^{ts}_{\leq r} = E[Y^{obs}|T = t, S^{obs} = s, Y^{obs} \leq y^{ts}_r]$, $\mu^{ts}_{\geq r} = E[Y^{obs}|T = t, S^{obs} = s, Y^{obs} \geq y^{ts}_r]$, $V^{ts}_{\leq r} = Var[Y^{obs}|T = t, S^{obs} = s, Y^{obs} \leq y^{ts}_r]$, $Var^{ts}_{\geq r} = E[Y^{obs}|T = t, S^{obs} = s, Y^{obs} \geq y^{ts}_r]$ and $V^{ts} = Var\left[Y^{obs}|T = t, S = s\right] / E\left[1\left(T = t\right) \cdot 1\left(S^{obs} = s\right)\right]$. In the last expression, the variance is divided by $E\left[1\left(T = t\right) \cdot 1\left(S^{obs} = s\right)\right]$ in order to take into account the fact that the results presented below are scaled by the square root of the total sample size (n), while our estimators are averages over specific subgroups.[18]

Proposition 5 *Assume that the true values of the bounds are in the interior of Ω, that Ω is compact, and that $E\left[|Y^{obs}|^{2+\delta}\right] < \infty$ for some $\delta > 0$. Then,*

$$\begin{aligned}
&\sqrt{n}(\widehat{LNATE}^{LB}_{na^{00}} - LNATE^{LB}_{na^{00}}o) \xrightarrow{d} N\left(0, V^{LB}_{na^{00}}\right), \\
&\sqrt{n}(\widehat{LNATE}^{UB}_{na^{00}} - LNATE^{UB}_{na^{00}}o) \xrightarrow{d} N\left(0, V^{UB}_{na^{00}}\right); \\
&\sqrt{n}(\widehat{LNATE}^{LB}_{na^{11}} - LNATE^{LB}_{na^{11}}o) \xrightarrow{d} N\left(0, V^{LB}_{na^{11}}\right), \\
&\sqrt{n}(\widehat{LNATE}^{UB}_{na^{11}} - LNATE^{UB}_{na^{11}}o) \xrightarrow{d} N\left(0, V^{UB}_{na^{11}}\right); \\
&\sqrt{n}(\widehat{LNATE}^{LB} - LNATE^{LB}o) \xrightarrow{d} N\left(0, V^{LB}\right), \\
&\sqrt{n}(\widehat{LNATE}^{UB} - LNATE^{UB}o) \xrightarrow{d} N\left(0, V^{UB}\right);
\end{aligned}$$

[17] Note that the asymptotic result for the subpopulation of always-takers is equivalent to that previously derived by Lee (2005) since, as previously discussed, the $LNATE$ for this subpopulation equals its local ATE.

[18] For instance, the first term in the bound of $LNATE_{na^{00}}$ corresponds to the sample mean for those in the group with $T_i = 0$ and $S^{obs}_i = 1$.

where

$$
\begin{aligned}
V^{LB}_{na^{00}} &= \frac{\left(p_{0|1}/p_{0|0}\right)^{-1}}{E\left[(1-S^{obs})(1-T)\right]}\left\{V^{00}_{\geq(1-(p_{0|1}/p_{0|0}))}+\left(y^{00}_{(1-(p_{0|1}/p_{0|0}))}-\mu^{00}_{\geq(1-(p_{0|1}/p_{0|0}))}\right)^2\left(1-\frac{p_{0|1}}{p_{0|0}}\right)\right\} \\
&\quad+\left(y^{00}_{(1-(p_{0|1}/p_{0|0}))}-\mu^{00}_{\geq(1-(p_{0|1}/p_{0|0}))}\right)^2\left(\frac{p_{1|1}-\left(1-\frac{p_{0|1}}{p_{0|0}}\right)E\left(T\right)}{E\left[(1-T)\right]p_{0|1}E\left(T\right)}\right)+V^{10}
\end{aligned}
$$

$$
\begin{aligned}
V^{UB}_{na^{00}} &= \frac{\left(p_{0|1}/p_{0|0}\right)^{-1}}{E\left[(1-S^{obs})(1-T)\right]}\left\{V^{00}_{\leq(p_{0|1}/p_{0|0})}+\left(y^{00}_{(p_{0|1}/p_{0|0})}-\mu^{00}_{\leq(p_{0|1}/p_{0|0})}\right)^2\left(1-\frac{p_{0|1}}{p_{0|0}}\right)\right\} \\
&\quad+\left(y^{00}_{(p_{0|1}/p_{0|0})}-\mu^{00}_{\leq(p_{0|1}/p_{0|0})}\right)^2\left(\frac{p_{1|1}-\left(1-\frac{p_{0|1}}{p_{0|0}}\right)E\left(T\right)}{E\left[(1-T)\right]p_{0|1}E\left(T\right)}\right)+V^{10}
\end{aligned}
$$

$$
\begin{aligned}
V^{LB}_{na^{11}} &= \frac{1}{E\left[S^{obs}T\right]\left(p_{1|0}/p_{1|1}\right)}\left\{V^{11}_{\leq(p_{1|0}/p_{1|1})}+\left(y^{11}_{(p_{1|0}/p_{1|1})}-\mu^{11}_{\leq(p_{1|0}/p_{1|1})}\right)^2\left(1-\frac{p_{1|0}}{p_{1|1}}\right)\right\} \\
&\quad+\left(y^{11}_{(p_{1|0}/p_{1|1})}-\mu^{11}_{\leq(p_{1|0}/p_{1|1})}\right)^2\left(\frac{p_{1|1}-\left(1-\frac{p_{0|1}}{p_{0|0}}\right)E\left(T\right)}{E\left[(1-T)\right]p_{0|1}E\left(T\right)}\right)+V^{01}
\end{aligned}
$$

$$
\begin{aligned}
V^{UB}_{na^{11}} &= \frac{1}{E\left[S^{obs}T\right]\left(p_{1|0}/p_{1|1}\right)}\left\{V^{11}_{\geq(1-(p_{1|0}/p_{1|1}))}+\left(y^{11}_{(1-(p_{1|0}/p_{1|1}))}-\mu^{11}_{\geq(1-(p_{1|0}/p_{1|1}))}\right)^2\left(1-\frac{p_{1|0}}{p_{1|1}}\right)\right\} \\
&\quad+\left(y^{11}_{(1-(p_{1|0}/p_{1|1}))}-\mu^{11}_{\geq(1-(p_{1|0}/p_{1|1}))}\right)^2\left(\frac{p_{1|1}-\left(1-\frac{p_{0|1}}{p_{0|0}}\right)E\left(T\right)}{E\left[(1-T)\right]p_{0|1}E\left(T\right)}\right)+V^{01}
\end{aligned}
$$

$$
\begin{aligned}
V^{LB} &= \left(\frac{p_{0|1}}{p_{0|1}+p_{1|0}}\right)^2 V^{LB}_{na^{00}}+\left(\frac{p_{1|0}}{p_{0|1}+p_{1|0}}\right)^2 V^{LB}_{na^{11}} \\
V^{UB} &= \left(\frac{p_{0|1}}{p_{0|1}+p_{1|0}}\right)^2 V^{UB}_{na^{00}}+\left(\frac{p_{1|0}}{p_{0|1}+p_{1|0}}\right)^2 V^{UB}_{na^{11}}
\end{aligned}
$$

To provide intuition behind the variances of the bound estimators, consider the variance of the estimator of the lower bound for $LNATE_{na^{00}}$ in (6), which contains four terms. The last term, V^{10}, corresponds to the variance of the estimator of the second term in (6), which is directly estimable from the data as the usual variance of a sample mean, except that here it is divided by $E\left[T(1-S^{obs})\right]$ to take into account the scaling by the square root of the total sample size. The remaining terms in $V^{LB}_{na^{00}}$ correspond to the asymptotic variance of $\widehat{E}[Y^{obs}|T_i=0,S^{obs}_i=0,Y^{obs}\geq y^{00}_{1-(p_{0|1}/p_{0|0})}]$, which is uncorrelated with $\widehat{E}[Y^{obs}|T_i=1,S^{obs}_i=0]$. Note that estimating the second term in $LNATE^{LB}_{na^{00}}$ involves three unknowns: the mean, the $\left(1-(p_{0|1}/p_{0|0})\right)$-th quantile (i.e., $y^{00}_{1-(p_{0|1}/p_{0|0})}$) and the proportion probability

$1-(p_{0|1}/p_{0|0})$. The first three terms in $V^{LB}_{na^{00}}$ correspond to estimators of these objects. For instance, the first two terms are equal to the asymptotic variance of the trimmed mean when the probabilities $p_{0|1}$ and $p_{0|0}$ are known. Again, note the division the first two terms by $E\left[(1-S^{obs})(1-T)\right]\left(p_{0|1}/p_{0|0}\right)$ to employ the right scaling in obtaining the asymptotic result.

We can construct estimators of the asymptotic variances in Proposition 5 by using sample analogs of the unknown terms in the same way it was done to construct estimators of the bounds. These estimators can then be used in the construction of confidence intervals for our parameters. There exist a growing literature on inference in partially identified models.[19] Part of this literature focuses on deriving confidence intervals that cover the entire identification region with a fixed probability (e.g., Horowitz and Manski, 2000; Chernozhukov, et al., 2007). Alternatively, Imbens and Manski (2004) introduced confidence intervals (CIs) that cover the true value of the parameter of interest with a fixed probability. This latter view is analogous to the one commonly used in the construction of CIs for point identified parameters. Both types of CIs can be constructed in our setting. For instance, consider the CIs proposed in Imbens and Manski (2004), which are shown to converge uniformly across different values for the width of the identification region. Let $k=\{na^{00}, na^{11},\}$, and $\widehat{V}^{LB}_k$ and $\widehat{V}^{UB}_k$ be the estimators of the corresponding asymptotic variances in Proposition 5. Then, an α-th CI for $LNATE_k$ can be constructed as

$$\overline{CI}_\alpha = [\widehat{LNATE}^{LB}_k - \overline{C}_n \cdot (\widehat{V}^{LB}_k/n)^{1/2},\ \widehat{LNATE}^{UB}_k - \overline{C}_n \cdot (\widehat{V}^{UB}_k/n)^{1/2}]$$

where $\overline{C}_n$ satisfies

$$\Phi\left(\overline{C}_n + \sqrt{n}\frac{\widehat{LNATE}^{UB}_k - \widehat{LNATE}^{LB}_k}{\max\left((\widehat{V}^{LB}_k/n)^{1/2}, (\widehat{V}^{UB}_k/n)^{1/2}\right)}\right) - \Phi\left(-\overline{C}_n\right) = \alpha,$$

with $\Phi(\cdot)$ the cdf of a standard normal distribution.

3.2.2 Net Average Treatment Effect

In order to point identify the $NATE$ for the entire population and thus decompose the ATE, additional assumptions are required. An assumption allowing the point identification of $NATE$ after any of the $LNATE$ parameters have been identified with the propositions presented above is a constant net treatment effect assumption. Under this assumption, $NATE$ is equal to any of the $LNATE$ parameters above, and thus the part of the ATE that is due to the mechanism S is given by $MATE = ATE - NATE$. This kind of assumption has also been pointed out in FF. Two observations are in order. First, this assumption is similar to that of a constant average treatment effect when estimating ATE using instrumental variables as in IA (1994).

[19] See, for instance, Horowitz and Manski (2000), Imbens and Manski (2004), Chernozhukov, et al. (2007), Beresteanu and Molinari (2008), and Romano and Shaikh (2008).

In that case we can only identify $LATE$ for the group of individuals who change treatment status in response to a change in the instrumental variable (i.e. the compliers); but under the assumption of a constant ATE we have that $LATE = ATE$. Second, this assumption is weaker than assuming a constant ATE. It allows for heterogeneous effects of the treatment on the outcome variable, but such heterogeneity is restricted to work through the mechanism S (i.e. through $MATE$). Nevertheless, this assumption may still be too strong in empirical settings.

In this subsection we present partial identification results for $NATE$. We show that $NATE$ in (3) can be bounded by adding to Assumptions 1-2 above two stochastic dominance assumptions. Without loss of generality, we continue working with the case for which $S_i(1) \geq S_i(0)$ for all i in Assumption 2.[20] Note that, under Assumption 1, the second term in (3) can be estimated from observed data on those in the control group, but the first term is not identifiable. By noting that under Assumption 2 $\pi_{na^{00}} + \pi_{na^{11}} + \pi_{ap} = 1$, the first term can be written as

$$E[Y(1,S(0))|na^{00}, na^{11}, ap] = \pi_{na^{00}} E[Y(1,S(0))|na^{00}] + \pi_{na^{11}} E[Y(1,S(0))|na^{11}] + \pi_{ap} E[Y(1,S(0))|ap]. \tag{10}$$

As in the previous section, the proportions are identified from the data since $\pi_{na^{00}} = p_{0|1}$, $\pi_{na^{11}} = p_{1|0}$ and $\pi_{ap} = 1 - p_{0|1} - p_{1|0}$. The first expectation is point-identified from the data (Result 1) and the second expectation can be bounded (Proposition 2). The non-identified term is the third expectation, unless additional assumptions that relate it to identifiable terms are made. One possible assumption is stochastic dominance, which has been previously employed in the literature in different settings (e.g., Manski, 1995; Zhang and Rubin, 2003; Mattei and Mealli, 2007; Zhang, et al, 2008).[21]

Assumption 3 (*Stochastic Dominance I).* For any real number q,

$$P(Y(1,S(0)) \leq q|ap) \leq P(Y(1,S(0)) \leq q|na^{00}).$$

Assumption 3 implies that $E[Y(1,S(0))|ap] \geq E[Y(1,S(0))|na^{00}]$, i.e., that in expectation, the counterfactual $Y(1,S(0))$ for the group ap is at least as large as that of the group na^{00}. This assumption will be appropriate when units for which the treatment affects the value of S (ap) are expected to have better outcomes than those not affected at 0 (na^{00}). Alternatively, this assumption is satisfied if the outcome for ap is expected to be better than that of na^{00} independently of the mechanism S. The validity of this assumption in particular applications can be gauged in light of these remarks.

[20] If it is the case that $S_i(1) \leq S_i(0)$ for all i, the discussion below is still valid but the subpopulations used, the inequalities in the assumptions, and the lower and upper bounds employed have to be modified accordingly.

[21] Zhang et al. (2008) employ an stochastic dominance condition in order to tighten their bounds on the effect of training on wages for the subpopulation of individuals who would be employed whether they received training or not. Here, this assumption is used to derive bounds for the population $NATE$. Note that the bounds on $LNATE$ obtained in the previous subsection could also be tightened by adding a stochastic dominance assumption.

Under Assumptions 1-3 the lower bound for $NATE$ can be constructed and is given in Proposition 6 below (which uses the lower bound for $E[Y(1,S(0))|na^{00}]$ in the previous section). For the upper bound, we follow a similar approach:

Assumption 4 (*Stochastic Dominance II).* For any real number q,

$$P(Y(1,S(0)) \leq q \left| na^{11} \right|) \leq P(Y(1,S(0)) \leq q|ap).$$

Similar to Assumption 3, it implies that $E[Y(1,S(0))|na^{11}] \geq E[Y(1,S(0))|ap]$, i.e., that in expectation, the counterfactual $Y(1,S(0))$ for the group na^{11} is at least as large as that of the group ap. Under this assumption, the upper bound of $E[Y(1,S(0))|na^{11}]$ is also an upper bound for $E[Y(1,S(0))|ap]$. We summarize these arguments in the following proposition.

Proposition 6 *Under Assumptions 1 through 4, $NATE$ in (3) can be bounded below by $NATE^{LB}$ and above by $NATE^{UB}$*, where:

$$\begin{aligned} NATE^{LB} &= \left(1-p_{1|0}\right) E[Y^{obs}|T_i=1,S_i^{obs}=0] \\ +p_{1|0}E[Y^{obs}|T_i &= 1,S_i^{obs}=1,Y^{obs} \leq y^{11}_{(p_{1|0}/p_{1|1})}] \\ -E[Y^{obs}|T_i &= 0] \end{aligned}$$

$$\begin{aligned} NATE^{UB} &= p_{0|1}E[Y^{obs}|T_i=1,S_i^{obs}=0] \\ +\left(1-p_{0|1}\right) E[Y^{obs}|T_i &= 1,S_i^{obs}=1,Y^{obs} \geq y^{11}_{1-(p_{1|0}/p_{1|1})}] \\ -E[Y^{obs}|T_i &= 0] \end{aligned}$$

As in the previous section, we can construct estimators for the bounds in Proposition 6, estimate their variance, and construct confidence intervals for $NATE$. Finally, we point out that $MATE$ can also be bounded once the bounds for $NATE$ have been obtained. In this case, given bounds on $NATE$, we can bound $MATE$ by $MATE^{LB} = ATE - NATE^{UB}$ and $MATE^{UB} = ATE - NATE^{LB}$.

Lastly, the bounds of Proposition 6 can be tightened by adding a monotonicity assumption of the effect of S on Y. More precisely, for the case of a non-negative effect, assume $Y_i(1,S(1)) \geq Y_i(1,S(0))$ for all i, which implies that $E[Y(1,S(1))|ap] \geq E[Y(1,S(0))|ap]$ and allows using $E[Y(1,S(1))|ap]$ as an upper bound for the previously unidentified $E[Y(1,S(0))|ap]$. In this case, the upper bound for (10) equals $E[Y(1,S(1))|n^{00},n^{11},ap]$ since for both $n^{00}$and n^{11} $E[Y(1,S(0))] = E[Y(1,S(1))]$. Therefore, the upper bound for $NATE$ is equal to the ATE, while the lower bound remains as in Proposition 6.

3.3 Point Identification of Net Average Treatment Effects

3.3.1 Local Net Average Treatment Effects

In order to go beyond partial identification of $LNATE$, additional assumptions are necessary. Recall that the reason the previous assumptions are not enough to point identify average treatment effects is that only one of the two potential outcomes needed is identifiable for any given strata. More specifically, Result 2 states that we can only identify $E\left[Y\left(0,S\left(0\right)\right)\right]$ for na^{11} and $E\left[Y\left(1,S\left(0\right)\right)\right]$ for na^{00}. In this subsection we present assumptions that allow estimation of counterfactual outcomes that in turn allow point identification of net average treatment effects.

A common approach to construct missing counterfactuals is to assume conditional independence or unconfoundedness. In the present context, an analogous approach is to assume that, conditional on a set of covariates X, the principal strata of interest are independent of the potential outcomes.

Assumption 5 *(Unconfounded Strata)*

$$Y\left(1,S(1)\right),Y\left(0,S(0)\right),Y(1,S(0))\perp\{S\left(1\right),S\left(0\right)\}|X.$$

This assumption implies that individuals in different strata are comparable once we condition on a set of covariates X, ruling out the existence of variables not included in X that simultaneously affect the principal strata an individual belongs to and her potential outcomes. Assumptions 1 and 5 together imply that $Y\left(1,S(1)\right),Y\left(0,S(0)\right),Y(1,S(0))\perp\{T,S\left(1\right),S\left(0\right)\}|X$, so that control and treated units in different strata but with the same values of covariates are comparable.[22]

Based on Result 2 and Assumption 5 applied to the relevant strata, we can nonparametrically point identify $NATE$ for different subpopulations, including the entire population. Consider the following proposition that pertains to identification of $LNATE_{na^{00}}$ in (6), where the missing counterfactual is the term $E[Y(0,S\left(0\right))|na^{00}]$.

Proposition 7 *Suppose Assumptions 1 and 2 hold, as well as Assumption 5 for the strata* $\{S(0)=0,S(1)=0\}$ *and* $\{S(0)=0,S(1)=1\}$. *Then,*

$$\begin{aligned} LNATE_{na^{00}} &= E_X\Big\{E[Y^{obs}|T=1,S^{obs}=0,X=x]- \\ &\quad E\left[Y^{obs}|T=0,S^{obs}=0,X=x\right]|S(0)=S(1)=0\Big\} \end{aligned}$$

Where we have used the fact that ap and na^{00} are comparable conditional on X (per Assumption 5). Estimation under this proposition is easily accomplished by first isolating the

[22] Moreover, Assumption 5 implies that, conditional on X, the potential outcomes are independent of any function of T, $S\left(1\right)$ and $S\left(0\right)$, such as S^{obs}.

strata of na^{00} ($T_i = 1, S_i^{obs} = 0$) and those units with $T_i = 0, S_i^{obs} = 0$, and then applying to these two groups any of the available methods for estimating the effect of T on Y under unconfoundedness (e.g., Imbens, 2004). Interestingly, in this particular case of $LNATE_{na^{00}}$ we do not need an overlap condition since by random assignment the probability of finding comparable individuals in the control group is one. Also, note that Proposition 7 achieves non-parametric identification of $LNATE_{na^{00}}$ without the use of functional form or constant effect assumptions. Of course, the availability of relevant covariates that validate the unconfoundedness assumption is required.

Analogous results as in Proposition 7 for the point identification of $LNATE$ for other subpopulations are not straightforward. The reason relates to Result 2: the only units in the sample for which we know the value of $Y(1, S(0))$ are those na^{00} with $T_i = 1$ (even though the data contains this information also for na^{11}, they are mixed with ap for which $Y^{obs} = Y(1, S(1)) \neq Y(1, S(0))$). Thus, in order to nonparametrically point identify net effects for any subpopulation (including the entire population), we need to use the na^{00} with $T_i = 1$ to construct the missing counterfactual $E[Y(1, S(0))]$. Intuitively, these are the only units we observe that received the counterfactual treatment that blocks the effect of T on S, so they are our "counterfactual-treated group". The following proposition illustrates the point by giving the details for the non-parametric point identification of $LNATE_{na^{11}}$.

Proposition 8 *Suppose Assumptions 1 and 2 hold, as well as Assumption 5 for the strata* $\{S(0) = 1, S(1) = 1\}$ *and* $\{S(0) = 0, S(1) = 0\}$. *Also, suppose that* $0 < \Pr(T = 1, S(0) = 0, S(1) = 0|X)$. *Then,*

$$\begin{aligned} LNATE_{na^{11}} &= E_X \Big\{ E[Y^{obs}|T = 1, S^{obs} = 0, X = x] - \\ &\quad E\left[Y^{obs}|T = 0, S^{obs} = 1, X = x\right] |S(0) = S(1) = 1 \Big\} \end{aligned}$$

Note that the conditions in Proposition 8 are stronger than those in Proposition 7, as an overlap condition is required. For completeness, we state the nonparametric identification of $LNATE$ in (5).

Proposition 9 *Suppose Assumptions 1 and 2 hold, as well as Assumption 5 for all strata. Also, suppose that* $0 < \Pr(T = 1, S(0) = 0, S(1) = 0|X)$. *Then,*

$$\begin{aligned} LNATE &= E_X \Big\{ E[Y^{obs}|T = 1, S^{obs} = 0, X = x] - \\ &\quad E\left[Y^{obs}|T = 0, X = x\right] |S(0) = s_0, S(1) = s_1 \Big\} \end{aligned}$$

3.3.2 Net Average Treatment Effect

Following the same logic as in the point identification of $LNATE$ in the previous subsection, the following proposition provides details for the non-parametric point identification of $NATE$.

Proposition 10 *Suppose Assumptions 1 and 2 hold, as well as Assumption 5 for all strata. Also, suppose that* $0 < \Pr(T = 1, S(0) = 0, S(1) = 0|X)$. *Then,*

$$\begin{aligned} NATE &= E_X \Big\{ E[Y^{obs}|T = 1, S^{obs} = 0, X = x] - \\ &\quad E\left[Y^{obs}|T = 0, X = x\right] \Big\} \end{aligned}$$

Intuitively, we can think of Proposition 10 as if for each member in the control group we were to find someone in the ($T = 1$, $S^{obs} = 0$) group that were comparable in terms of X. The overlap condition ensures that in infinite samples we are able to compare treated and control individuals for all values of X.

To end this section, we mention a literature that provides alternative nonparametric identification results for the $ANDE$ parameter discussed in Section 2, which is closely related to $NATE$. Although the identification results in this literature are in principle nonparametric, estimation is typically suggested using parametric models.[23] In general, the approach is based on writing the $ANDE$ as a function of the $ACDE$ (discussed in section 2), which is identifiable under the approach's assumptions. The two key assumptions are a conditional independence assumption similar to Assumption 5 plus a form of a "non-interaction" assumption. The latter assumption specifies the way in which the mechanism variable is allowed to interact with the treatment to affect the outcome. For example, Robins and Greenland (1992) assume that the intermediate variable does not interact with the treatment at all to affect the outcome. Hence, for example, it implies that if T and S are both randomly assigned, then $E[Y|T, S]$ is an additive function of T and S.

Arguably, the weakest non-interaction assumption is in Petersen et al. (2006). Using our notation, their assumption may be written as: $Y(1, \overline{s}) - Y(0, \overline{s}) \perp S(0)|X$. It states that, conditional on X, the *controlled* direct effect is independent of the potential value $S(0)$ of the intermediate variable *at all levels of* $\overline{s}$. The $ANDE$ is then identified from the $ACDE$. The assumption, however, is likely strong in economic settings. It would imply, for instance, that if we were to analyze the effect of veteran status on earnings having schooling as an intermediate variable, the effect of being a veteran holding the level of schooling fixed *at any level* is independent of the level of schooling under the control treatment, conditional on covariates.[24] Contrary

[23] Examples of this approach are contained in Robins and Greenland (1992), Pearl (2001), and Petersen et al. (2006), among others.

[24] To illustrate the use of this interaction assumption to identify $ANDE$, we can write:

$$\begin{aligned} ANDE &= E[Y(1, S(0)) - Y(0, S(0))] \\ &= E_{S(0),X}\{E[Y(1, S(0)) - Y(0, S(0))|S(0) = s, X = x]\} \\ &= E_{S(0),X}\{E[Y(1, s) - Y(0, s)|S(0) = s, X = x]\} \\ &= E_{S(0),X}\{E[Y(1, s) - Y(0, s)|X = x]\} \end{aligned}$$

where the assumption $Y(1, \overline{s}) - Y(0, \overline{s}) \perp S(0)|X$ has been used in the last line. The term inside the outer expectation equals the $ACDE$ given X. Petersen et al. (2006) implement this approach by first running OLS on

to this literature, by focusing on the information contained in the data about $Y(1, S(0))$, we are able to avoid using non-interaction assumptions.

4 Extensions

This section presents two extensions of the previous results to the case of a multivalued mechanism and the case of multiple mechanisms. These extensions share a common issue in that the number of principal strata increases. Two consequences are that nonparametric identification is based on a subpopulation that is more specific, and that estimation is more difficult with a fixed sample size as less observations are available for each subpopulation. In the case of partial identification, the latter consequence can potentially result in uninformative bounds.

4.1 Multivalued Mechanism

In order to focus on the main ideas, we start by considering a mechanism that can take up three values: $S_i \in \{0, 1, 2\}$. In this case there are nine strata:

Table 3

		$S_i(0)$ = 0	$S_i(0)$ = 1	$S_i(0)$ = 2
$S_i(1)$	0	na^{00}	an^{10}	an^{20}
	1	ap^{01}	na^{11}	an^{21}
	2	ap^{02}	ap^{12}	na^{22}

where we use a similar notation as in the previous section with the superscript "$s_0 s_1$" referring to potential values of the multivalued mechanism S under control and treatment states, respectively. By Result 1, we know the data contains information on $Y(1, S(0))$ only for the groups na^{00}, na^{11} and na^{22}. If Assumption 1 (randomly assigned treatment) and Assumption 2 (monotonicity) are imposed, then the three strata corresponding to an are eliminated and the observed groups based on S_i^{obs} and T_i are composed of the following strata:

Table 4

		T_i = 0	T_i = 1
S_i^{obs}	0	$na^{00}, ap^{01}, ap^{02}$	na^{00}
	1	na^{11}, ap^{12}	na^{11}, ap^{01}
	2	na^{22}	$na^{22}, ap^{12}, ap^{02}$

Hence, similar to Result 2, we have that under Assumptions 1 and 2 those units with $(T_i, S_i^{obs}) = (1, 0)$ are non affected with $S_1(1) = S_1(0) = 0$ that received the treatment, and

$Y = a + bT + cTX + dTS + eTSX + gX + hS$ to obtain $Y(1, \overline{s})$. Using the estimated coefficients they calculate $CDE = Y(1, \overline{s}) - Y(0, \overline{s}) = b + cX + dS + eSX$. Finally, they get estimates of $E(X)$, $E(S(0))$ and $E(S(0)X)$ and plug them into CDE. The outcome is their estimate of the $ANDE$.

those units with $(T_i, S_i^{obs}) = (0, 2)$ are non affected with $S_1(1) = S_1(0) = 2$ that did not receive the treatment. Clearly, this will hold in the case in which S can take more than 3 values, for which we state the following result.

Result 3 *Let S take values on a set $\{l, l+1, \ldots, u\}$. Under Assumptions 1-2, treated units with $(T_i, S_i^{obs}) = (1, l)$ belong to the strata $S_1(1) = S_1(0) = l$, and controls units with $(T_i, S_i^{obs}) = (0, u)$ belong to the strata $S_1(1) = S_1(0) = u$.*

It is important to note that in the multivalued mechanism case, information on the non affected units with values of S strictly between l and u (e.g., na^{11} strata in Table 3) is not identifiable from the observed groups. The reason is that these strata are mixed with other strata under both treatment and control group, and their proportions are not identifiable from the data.[25] This prevents us from exploiting the information available in these strata without stronger assumptions (e.g., ruling out the existence of more strata to identify such probabilities). As a result, the multivalued case reduces to the binary-mechanism case considered in Section 3 with the lowest and highest values of the mechanism. Consequently, the approaches for partial and point identification in that section can be employed based on the strata na^{ll} and na^{uu}.

To illustrate, consider construction of bounds in the three-value case for the $LNATE$ for the stratum of non affected with $S(0) = S(1) = 0$, or

$$LNATE_{na^{00}} = E[Y(1, S(0))|na^{00}] - E[Y(0, S(0))|na^{00}]. \tag{11}$$

Using similar notation as in Section 3, we have that $\pi_{na^{00}} = p_{0|1}$ and $\pi_{na^{00}} + \pi_{ap^{01}} + \pi_{ap^{02}} = p_{0|0}$. The average outcome for the observed group with $T_i = 0$ and $S_i^{obs} = 0$ can be written as

$$\begin{aligned} E[Y_i^{obs}|T_i \;=\; 0, S_i^{obs} = 0] &= \frac{\pi_{na^{00}}}{\pi_{na^{00}} + \pi_{ap^{01}} + \pi_{ap^{02}}} \cdot E[Y_i(0, S_i(0))\,|na^{00}] \\ &+ \frac{\pi_{ap^{01}} + \pi_{ap^{02}}}{\pi_{na^{00}} + \pi_{ap^{01}} + \pi_{ap^{02}}} \cdot E[Y_i(0, S_i(0))\,|ap^{01} \text{ or } ap^{02}] \end{aligned}$$

and thus $E[Y_i(0, S_i(0))\,|na^{00}]$ can be bounded from below by $\mu^{00}_{\leq (p_{0|1}/p_{0|0})}$ and from above by $\mu^{00}_{\geq 1-(p_{0|1}/p_{0|0})}$. Thus, the resulting bounds are given by the expressions in Proposition 1. In general, the nonparametric partial and point identification results can be stated as follows. Let A^C denote the complement of set A.

Proposition 11 *Let S take values on a set $\{l, l+1, \ldots, u\}$, and let Assumption 1 hold. Also, without loss of generality, consider the case in Assumption 2 where $S_i(1) \geq S_i(0)$ for all i.*

[25] For instance, using similar notation as in Section 3 it is straightforward to show in the three-value example above that there is no unique solution for $\pi_{n^{11}}$ from the identifiable probabilities $p_{0|0}$, $p_{0|1}$, $p_{1|0}$, $p_{1|1}$, $p_{2|0}$ and $p_{2|1}$.

(i) *Let* $LNATE_l = E[Y(1,S(0)) - Y(0,S(0)) | S(0)=l, S(1)=l]$, $LNATE_u = E[Y(1,S(0)) - Y(0,S(0)) | S(0)=u, S(1)=u]$ *and* $LNATE_{lu} = \left(\frac{p_{0|1}}{p_{0|1}+p_{1|0}}\right) LNATE_l + \left(\frac{p_{1|0}}{p_{0|1}+p_{1|0}}\right) LNATE_u$. *Then, the bounds for* $LNATE_l$, $LNATE_u$ *and* $LNATE_{lu}$ *equal those for* $LNATE_n$, $LNATE_a$ *and* $LNATE$ *from Propositions 1-3, respectively.*

(ii) *Assume for any real number* q, $P(Y(1,S(0)) \leq q | \{n^{ll}, n^{uu}\}^C) \leq P(Y(1,S(0)) \leq q | n^{ll})$, *and* $Y_i(1,S(1)) \geq Y_i(1,S(0))$ *for all* i. *Then, the bounds for* $NATE$ *are given by those in Proposition 6.*

(iii) *Suppose Assumption 5 holds for the strata* $\{S(0)=l, S(1)=l\}$ *and* $\{S(0)=0, S(1)=s\}$ *with* $s = \{l+1, l+2, \ldots, u\}$. *Then,* $LNATE_l = E_X\{E[Y^{obs}|T=1, S^{obs}=l, X=x] - E[Y^{obs}|T=0, S^{obs}=l, X=x] | S(0)=S(1)=l\}$.

(iv) *Suppose Assumption 5 holds for all strata and* $0 < \Pr(T=1, S(0)=l, S(1)=l|X)$. *Then,* $NATE = E\left\{E[Y^{obs}|T=1, S^{obs}=l, X=x] - E[Y^{obs}|T=0, X=x]\right\}$.

The fact that we only employ the lowest and highest values of the mechanism has important implications. First, in contrast to the binary case, combining the $LNATE$ for the highest and lowest values of the mechanism (say, $LNATE_{na^{ll}}$ and $LNATE_{na^{uu}}$) does not yield bounds for all the the not-affected because the strata for those not-affected with values between l and u are not identified. Second, to the extent that the proportion of not-affected units with the lowest and highest values of S in the population is small, the bounds for $NATE$ are more likely to become uninformative. Finally, given that the only strata for which we know $Y(1,S(0))$ is na^{ll}, our nonparametric point identification results base the construction of the missing counterfactual $E[Y(1,S(0))]$ on a smaller subpopulation and, as a result, the overlap condition will become a stronger assumption. In summary, nonparametric identification of net effects when the mechanism variable is multivalued becomes more difficult. In such a case, additional assumptions (e.g. functional form or distributional) may become necessary for extrapolation.[26]

4.2 Multiple Mechanisms

We focus on the case of two different binary mechanism variables S and R in order to highlight the salient issues. Following similar notation for R as we have employed for S, composite potential outcomes equal to $Y_i(1)$ and $Y_i(0)$ are $Y_i(1,S_i(1),R_i(1))$ and $Y_i(0,S_i(0),R_i(0))$, respectively. Other composite potential outcomes are $Y_i(1,S_i(0),R_i(0))$ which represents the part of individual i's effect that does not work through mechanisms S or R; $Y_i(1,S_i(0),R_i(1))$ which represents the part of the effect that does not work through S; and $Y_i(1,S_i(1),R_i(0))$ which represents the part of the effect that does not work through R. As a result, there are

[26] In this case, the methods in FF (2008a) could be employed. That paper also provides empirical illustrations where the mechanism is multivalued or continuous.

multiple decompositions of ATE that may be of interest. Table 5 shows several ATE decompositions along with the definitions of different net and mechanism effects relative to S and R, where, for simplicity, the expectation operator $E[\cdot]$ and the conditioning on principal strata $\{S(0) = s_0, S(1) = s_1, R(0) = r_0, R(1) = r_1\}$ are left implicit.

Table 5

Mechanisms	Definitions	KC = Key Counterfactual
S and R	$ATE = MATE_{SR} + NATE_{SR}$ $MATE_{SR} = Y_i(1, S_i(1), R_i(1)) - KC$ $NATE_{SR} = KC - Y_i(0, S_i(0), R_i(0))$	$Y_i(1, S_i(0), R_i(0))$
S	$ATE = MATE_S + NATE_S$ $MATE_S = Y_i(1, S_i(1), R_i(1)) - KC$ $NATE_S = KC - Y_i(0, S_i(0), R_i(0))$	$Y_i(1, S_i(0), R_i(1))$
R	$ATE = MATE_R + NATE_R$ $MATE_R = Y_i(1, S_i(1), R_i(1)) - KC$ $NATE_R = KC - Y_i(0, S_i(0), R_i(0))$	$Y_i(1, S_i(1), R_i(0))$
Interaction	$MATE_{intSR} = MATE_{SR} - MATE_S - MATE_R$	

The last term in Table 5, labeled "Interaction", measures the part of the effect of T on Y due to the interaction from a change in both S and R due to T. Hence, $MATE_{intSR}$ equals the total mechanism effect of S and R combined ($MATE_{SR}$) minus each one of the individual mechanism effects of S and R (i.e., $MATE_S$ and $MATE_R$). Suppose there is no interaction effect between S and R, so that the mechanism effect for S and R equals the sum of the separate mechanism effects of S and R. This implies that $MATE_{SR} = MATE_S + MATE_R$, $MATE_{intSR} = 0$, and thus

$$ATE = MATE_S + MATE_R + NATE_{SR}. \tag{12}$$

If there exists a non-zero interaction term, a more general decomposition of ATE is as follows:

$$ATE = MATE_S + MATE_R + MATE_{intSR} + NATE_{SR} = MATE_{SR} + NATE_{RS} \tag{13}$$

where the second equality follows from the definition of $MATE_{intSR}$ in the last row of Table 5.

We now discuss nonparametric identification of the parameters in Table 5. For each unit i, we observe the vector $\left(T_i, Y_i^{obs}, S_i^{obs}, R_i^{obs}\right)$, where Y_i^{obs} and S_i^{obs} are as before, and $R_i^{obs} = T_i R_i(1) + (1 - T_i) R_i(0)$. In this setup where each T, S, and R are binary, there are sixteen principal strata.[27] An expanded assumption of monotonicity (Assumption 2) that includes also R deletes 7 strata of an. As before, without loss of generality we assume the effects of T on R and S are non-negative for all units. Using na^{00}, ap, and na^{11} as in Section 3, and a subscript denoting the corresponding mechanism, the remaining principal strata are: $na_s^{00}na_r^{00}$, $ap_s na_r^{00}$,

[27] In general, with m mechanisms there are $2^{(2\times m)}$ strata.

$na_s^{11}na_r^{00}$, $na_s^{00}ap_r$, ap_sap_r, $na_s^{11}ap_r$, $na_s^{00}na_r^{11}$, $ap_sna_r^{11}$, and $na_s^{11}na_r^{11}$. Assuming a randomly assigned treatment (Assumption 1), the observed groups are composed of:

Table 6

		$T_i=0$		$T_i=1$	
		R_i^{obs}		R_i^{obs}	
		0	1	0	1
S_i^{obs}	0	$na_s^{00}na_r^{00}$ $na_s^{00}ap_r$ $ap_sna_r^{00}$ ap_sap_r	$na_s^{00}na_r^{11}$ $ap_sna_r^{11}$	$na_s^{00}na_r^{00}$	$na_s^{00}na_r^{11}$ $na_s^{00}ap_r$
	1	$na_s^{11}na_r^{00}$ $na_s^{11}ap_r$	$na_s^{11}na_r^{11}$	$na_s^{11}na_r^{00}$ $ap_sna_r^{00}$	$na_s^{11}na_r^{11}$ $na_s^{11}ap_r$ $ap_sna_r^{11}$ ap_sap_r

where it becomes evident the fact that the number of observed groups increases geometrically with the number of mechanisms considered, resulting in less observations available per cell for any fixed sample.

Consider the first decomposition in Table 5: $ATE = MATE_{SR} + NATE_{SR}$. Analogous to Result 1, note that in this case the data contains information on $Y(1, S(0), R(0))$ only for the groups $n_s^{00}n_r^{00}$ and $n_s^{11}n_r^{11}$. Moreover, similar to Result 2, we have that those treated units with $(T_i, S_i^{obs}, R_i^{obs}) = (1, 0, 0)$ are n^{00} with respect to both mechanisms; and those control units with $(T_i, S_i^{obs}, R_i^{obs}) = (0, 1, 1)$ are n^{11} with respect to both mechanisms. Hence, we can apply the methods in Section 3 by focusing on these two groups. Let $p_{sr|t} = \Pr\left(S^{obs} = s, R^{obs} = r | T = t\right)$. Then, we state the following

Proposition 12 *Let Assumption 1 hold. Also, assume $S_i(1) \geq S_i(0)$ and $R_i(1) \geq R_i(0)$ for all i.*

(i) *Let $LNATE_{SR}^{na_r^{00}na_s^{00}} = E[Y(1, S(0), R(0)) - Y(0) | na_r^{00}na_s^{00}]$, $LNATE_{SR}^{na_r^{11}na_s^{11}} = E[Y(1, S(0), R(0)) - Y(0) | na_r^{11}na_s^{11}]$ and*

$$LNATE_{SR} = E\left\{E[Y(1, S(0), R(0)) - Y(0) | S(0) = S(1) = s, R(0) = R(1) = r]\right\}$$

Then, the bounds for $LNATE_{SR}^{na_r^{00}na_s^{00}}$, $LNATE_{SR}^{na_r^{11}na_s^{11}}$ and $LNATE_{SR}$ equal those for $LNATE_{na^{00}}$, $LNATE_{na^{11}}$ and $LNATE$ from Propositions 1-3, respectively; with $p_{0|0}$, $p_{1|0}$, $p_{0|1}$ and $p_{1|1}$ replaced by $p_{00|0}$, $p_{11|0}$, $p_{00|1}$and $p_{11|1}$; and replacing the groups $(T_i, S_i^{obs}) = \{(0,0), (0,1), (1,0), (1,1)\}$ with the corresponding groups $(T_i, S_i^{obs}, R_i^{obs}) = \{(0,0,0), (0,1,1), (1,0,0), (1,1,1)\}$.

(ii) *Assume for any real number q, $P(Y(1, S(0), R(0)) \leq q | \left\{na_r^{00}na_s^{00}, na_r^{11}na_s^{11}\right\}^C) \leq P(Y(1, S(0), R(0)) \leq q | na_r^{00}na_s^{00})$, and $Y_i(1, S(1), R(1)) \geq Y_i(1, S(0), R(0))$ for all*

i. Then, the bounds for $NATE_{SR}$ are given by those in Proposition 6, using the corresponding proportions and observed groups from (i).

(iii) *Suppose* $Y(1), Y(0), Y(1, S(0), R(0)) \perp \{S(1), S(0), R(1), R(0)\}|X$ *holds for the strata* $na_r^{00}na_s^{00}$, $na_s^{00}ap_r$, $ap_sna_r^{00}$ and ap_sap_r. *Then,* $LNATE_{SR}^{na_r^{00}na_s^{00}} = E_X\{E[Y^{obs}|T = 1, S^{obs} = 0, R^{obs} = 1, X = x] - E[Y^{obs}|T = 0, S^{obs} = 0, R^{obs} = 0, X = x]|S(0) = S(1) = 0, R(0) = R(1) = 0\}$.

(iv) *Suppose* $Y(1), Y(0), Y(1, S(0), R(0)) \perp \{S(1), S(0), R(1), R(0)\}|X$ *holds for all strata and* $0 < \Pr(T = 1, S(0) = 0, S(1) = 0, R(0) = 0, R(1) = 0|X)$. *Then,* $NATE = E\{E[Y^{obs}|T = 1, S^{obs} = 0, R^{obs} = 0, X = x] - E[Y^{obs}|T = 0, X = x]\}$.

Regarding the decompositions in the second and third rows of Table 5, these are exactly the decompositions analyzed in Section 3 using separate mechanisms S and R. Hence, those results are directly applicable here.

The more complicated decomposition is that in equation (13), where the total mechanism effect $MATE_{SR}$ from the first decomposition in Table 5 is broken up into each of the individual mechanism effects S, R, and their interaction. Note that in an ideal situation, to estimate the terms in (13), we would perform an experiment in which units are randomized into one of five possible treatments: the control treatment, the original treatment, and a treatment corresponding to each one of the key potential outcomes in Table 5.[28] Hence, estimating the terms in that decomposition is more difficult as we are implicitly trying to learn about five different treatments from the two at hand.

First, consider using the $LNATE$s identified in Proposition 12 and the identified effects for each separate mechanism in rows 2 and 3 of Table 5. Since by definition all those subpopulations do not have mechanism effects, it is not possible to use them to identify the terms in (13). A possible way to do so, though, is to impose constant treatment effect assumptions to relate them to their mechanism effects.[29] The notes about this assumption in section 3.2.2 apply here as well. A second approach is to use our partial identification results about $NATE$ in Proposition 6 and the propositions in section 3.2.1 to identify the terms in (13). The bounds derived for $NATE_{SR}$, $NATE_S$ and $NATE_R$ can be used to construct bounds for the corresponding mechanism effects $MATE_{SR}$, $MATE_S$ and $MATE_R$. Although valid, given how much we want to learn from the limited data available, combining all these intervals may result in uninformative bounds for the decomposition in (13) for a given sample.

Alternatively, we can use the nonparametric point identification results for $NATE_{SR}$, $NATE_S$, and $NATE_R$. In this case, we use those values of Y^{obs} that we know correspond to each one of the three potential outcomes $Y_i(1, S_i(0), R_i(0))$, $Y_i(1, S_i(1), R_i(0))$ and $Y_i(1, S_i(0), R_i(1))$,

[28] In this case $MATE_{intSR}$ would be obtained as the difference of the rest of the terms in equation (13).

[29] This assumption would state that the individual effects $Y_i(1, S_i(0), R_i(0)) - Y_i(0)$, $Y_i(1, S_i(1), R_i(0)) - Y_i(0)$ and $Y_i(1, S_i(0), R_i(1)) - Y_i(0)$ are constant for all units in the population.

to construct the missing counterfactuals, i.e., the values of Y^{obs} from the observed groups $\left(T_i = 1, S_i^{obs} = 0, R_i^{obs} = 0\right)$, $\left(T_i = 1, R_i^{obs} = 0\right)$ and $\left(T_i = 1, S_i^{obs} = 0\right)$, respectively. In principle, since we are implicitly comparing five different treatments, three overlap conditions $0 < \Pr\left(T = 1, S\left(0\right) = 0, S\left(1\right) = 0, R\left(0\right) = 0, R\left(1\right) = 0|X\right)$, $0 < \Pr\left(T = 1, S\left(0\right) = 0, S\left(1\right) = 0|X\right)$ and $0 < \Pr\left(T = 1, R\left(0\right) = 0, R\left(1\right) = 0|X\right)$ must hold simultaneously to be able to find comparable individuals in all treatments at the same time.[30] In the current situation, however, the unconfounded strata assumption and overlap condition in Proposition 12 imply the ones in Section 3, so they are sufficient to nonparametrically point identify the decomposition in (13). Finally, note that imposing the condition $MATE_{intSR} = 0$ and estimating all the terms in the right hand side of (12) allows testing whether their sum equals the ATE.

5 Conclusions

This paper has analyzed nonparametric partial and point identification of net and mechanism effects under weak assumptions, allowing for heterogeneous effects. Employing insights from the seminal analysis in Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996), we clarify that the typical data contains information on the key potential outcome used in the definition of net effects (i.e., $Y\left(1, S\left(0\right)\right)$) only for the subpopulation for which the treatment does not affect the mechanism. The main implication of this result is that estimation of net effects for other subpopulations–including the entire population–can only be based on extrapolations involving typically strong conditions such as constant-effect, parametric and/or conditional independence assumptions.

Following this result, we provide identification conditions for the case in which the treatment assignment is random and the mechanism variable is binary. Our partial identification results for average net effects for the subpopulation for which the data contains information on $Y\left(1, S\left(0\right)\right)$ rely only on a monotonicity condition for the effect of the treatment on the mechanism. By adding a stochastic dominance condition and an additional monotonicity condition for the effect of the mechanism on the outcome, we derive bounds for the net average treatment effect for the entire population. As with any partial identification results, estimated bounds from a given sample may turn out to be uninformative, in which case making additional assumptions will be required. Sufficient assumptions for point identification of average net effects for different subpopulations, including the whole population, are then presented. Finally, Section 4 discussed extensions of the identification strategies to the cases of a multivalued mechanism variable and multiple mechanisms.

Several extensions of the results contained here are ongoing. So far we have developed results

[30] This would ensure that all ATE, $MATE_{SR}$, $MATE_S$ and $MATE_R$ are estimated for the same population. For a discussion on different ways to implement the overlap condition in a multiple treatment setting see, for instance, Flores and Mitnik (2008).

for the situation in which the treatment is assigned at random. Extensions to the case in which the treatment is not randomly assigned are possible for some of our identification strategies. For example, Flores and Flores-Lagunes (2008a) provide such extensions for situations in which assumptions such as unconfoundedness, functional form or constant effects are tenable. Further extensions to non-randomly assigned treatments under milder conditions, as well as the use of instrumental variables in this context, are at the top of our research agenda.

6 Appendix

To be completed ...

References

[1] Angrist, J., Imbens, G., and Rubin, D. (1996) "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, 91, 444–472.

[2] Balke, A., and Pearl, J. (1997) "Bounds on Treatment Effects from Studies with Imperfect Compliance" *Journal of the American Statistical Association*, 92, 1172–1176.

[3] Beresteanu, A. and Molinari, F. (2008) "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, 76(4), 763-814.

[4] Black, D. and Smith, J. (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*, 121, 99-124.

[5] Cai, Z., Kuroki, M., Pearl, J. and Tian, J. (2007) "Bounds on Direct Effects in the Presence of Confounded Intermediate Variables", *Biometrics*, forthcoming.

[6] Chernozhukov, V., Hong, H. and Tamer, E. (2007) "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, 75 (5), 1243-1284.

[7] Currie, J. and Moretti, E. (2003), "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings", *Quarterly Journal of Economics*, 118(4), 1495-1532.

[8] Dawid, A. (1979) "Conditional Independence in Statistical Theory (with Discussion)" *Journal of the Royal Statistical Society, Series B*, 41, 1-31.

[9] Dearden, L. Ferri, J. and Meguir, C. (2002), "The Effect of School Quality on Educational Attainment and Wages." *Review of Economics and Statistics*, 84, 1-20.

[10] Ehrenberg, R., Jakubson, G., Groen, J., So, E., and Price, J. (2007), "Inside the Black Box of Doctoral Education: What Program Characteristics Influence Doctoral Students' Attrition and Graduation Probabilities?" *Educational Evaluation and Policy Analysis* (June).

[11] Flores, C. A. and Flores-Lagunes, A. (2008a), "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment", Working Paper, Department of Economics, University of Miami.

[12] Flores, C. A. and Flores-Lagunes, A. (2008b), "Testing Implications of the Exclusion Restriction Assumption in Just-Identified Instrumental Variable Models", mimeo, Department of Economics, University of Miami.

[13] Flores, C. A. and Mitnik, O. (2008), "An Evaluation of Non-Experimental Estimators of Multiple Treatments", Working Paper, Department of Economics, University of Miami.

[14] Frangakis,C.E. and Rubin D. (2002) "Principal Stratification in Causal Inference", *Biometrics*, 58, 21-29.

[15] Heckman, J., Smith, J. and Clements, N. (1997), "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, 64(3), 487-535.

[16] Heckman, J., LaLonde, R. and Smith, J. (1999) "The Economics and Econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*. Elsevier Science North Holland, 1865-2097.

[17] Hirano, K., Imbens, G. W., Rubin, D. and Zhou, X. (2000), "Assesing the Effect of an Influenza Vaccine in an Encouragement Design", *Biostatistics*, 1 (1), 69-88.

[18] Holland, P. (1986) "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-70.

[19] Horowitz, J. and Manski, C. (2000) "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data", *Journal of the American Statistical Association*, 95, 77-84.

[20] Imbens, G. (2004) "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review" *Review of Economics and Statistics*, 84, 4-29.

[21] Imbens, G. W. and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62 (2), 467-475.

[22] Imbens, G. W. and Manski, C. F. (2004), "Confidence Intervals for Partially Identified Parameters", *Econometrica*, 72 (6), 1845-1857.

[23] Imbens, G. W. and Rubin, D. (1997), "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance", *The Annals of Statistics*, 25 (1), 305-327.

[24] Karlan, D., and List, J. (2007), "Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment", *American Economic Review* 97(5), 1774-1793.

[25] Kaufman, S., Kaufman, J., MacLenose, R., Greenland, S., and Poole, C. (2005), "Improved Estimation of Controlled Direct Effects in the Presence of Unmeasured Confounding of Intermediate Variables" *Statistics in Medicine* 24, 1683-1702. (Correction, 25, 3228).

[26] Lee, D.S. (2005) "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", National Bureau of Economic Research Working Paper #11721.

[27] Manski, C. (1990) "Nonparametric Bounds on Treatment Effects" *American Economic Review Papers and Proceedings*, 80, 319-23.

[28] Manski, C. (1990) *Identification Problems in the Social Sciences*. Harvard University Press.

[29] Mattei, A. and Mealli, F. (2007) "Application of the Principal Stratification Approach to the Faenza Randomized Experiment on Breast Self-Examination" *Biometrics*, 63, 437-446.

[30] Mealli, F. and Rubin, D. (2003) "Assumptions Allowing the Estimation of Direct Causal Effects" *Journal of Econometrics*, 112, 79-87.

[31] Morgan, S.L. and Winship, C. (2007), *Counterfactuals and Causal Inference*. Cambridge University Press.

[32] Newey, W., and McFadden, D. (1994) "Large Sample Estimation and Hypothesis Testing" in R. Engle and D. McFadden (eds.) *Handbook of Econometrics*. Elsevier Science North Holland, 2111-2245.

[33] Neyman, J. (1923), "On the Application of Probability Theory to Agricultural Experiments: Essays on Principles" Translated in *Statistical Science*, 5, 465-80.

[34] Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

[35] Pearl, J. (2001), "Direct and Indirect Effects" *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 411-20.

[36] Petersen, M., Sinisi, S., and van der Laan, M. (2006) "Estimation of Direct Causal Effects" *Epidemiology*, 17, 276-284.

[37] Robins, J. and Greenland, S. (1992) "Identifiability and Exchangeability for Direct and Indirect Effects" *Epidemiology*, 3, 143-155.

[38] Romano, J. and Shaikh, A. (2008), "Inference for Identifiable Parameters in Partially Identified Econometric Models", *Journal of Statistical Planning and Inference*, forthcoming.

[39] Rosenbaum, P. (1984) "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment" *Journal of the Royal Statistical Society, Series A*, 147, 656-66.

[40] Rubin, D. (1974) "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

[41] Rubin, D. (1980) "Discussion of 'Randomization Analysis of Experimental Data in the Fisher Randomization Test' by Basu" *Journal of the American Statistical Association*, 75, 591-93.

[42] Rubin, D. (2004) "Direct and Indirect Causal Effects via Potential Outcomes" *Scandinavian Journal of Statistics*, 31, 161-70.

[43] Rubin, D. (2005) "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions" *Journal of the American Statistical Association*, 100, 322-331.

[44] Zhang, J.L. and Rubin, D. (2003) "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by 'Death'" *Journal of Educational and Behavioral Statistics*, 28, 353-68.

[45] Zhang, J.L., Rubin, D. and Mealli, F. (2008) "Evaluating the Effects of Job Training Programs on Wages Through Principal Stratification" in D. Millimet et al. (eds) *Advances in Econometrics vol XXI*, Elsevier.