published in Statistics and Probability Letters, 78, 190-195, 2008

A NOTE ON ENDOGENOUS CONTROL VARIABLES IN CAUSAL STUDIES

Michael Lechner*
SIAW

Swiss Institute for International Economics and Applied Economic Research

University of St.Gallen

This version: March, 2007
Date this version has been printed: 08 May 2007

Abstract

The issue of potentially endogenous control variables in causal studies based on the assumption of no selection bias conditional on observables (CIA) is discussed. The paper shows that the standard formulation of the CIA obscures the endogeneity problem. It suggests a CIA based on potential variables together with explicit exogeneity conditions which allows a separate assessment of the endogeneity bias and the plausibility of the CIA.

Keywords: Evaluation studies, matching methods, conditional independence assumption, endogeneity

JEL classification: C21, C31

* Address for correspondence: Michael Lechner, Professor of Econometrics, Swiss Institute for International Economics and Applied Economic Research (SIAW), University of St. Gallen, Bodanstr. 8, CH-9000 St. Gallen, Switzerland, Michael.Lechner@unisg.ch, www.siaw.unisg.ch/lechner.
The author is also affiliated with CEPR, London, ZEW, Mannheim, IZA, Bonn, and PSI, London.

1 Introduction

An example of causal studies is the rapidly growing literature on the evaluation of labour market programmes (see for example the surveys by Heckman, LaLonde, and Smith, 1999). Many evaluation studies of labour market programmes take account of selection effects into the programme by collecting large and informative data with many (control) variables that help to explain participation in the programmes as well as outcome variables of interest. Then, they proceed with matching methods (see as examples Sianesi, 2004, and Gerfin and Lechner, 2002, among many others) to estimate the effects of the programmes on the specific sub-groups of participants and nonparticipants. If conditional on the rich control variables, the participation and the outcome variables are independent, the so-called selection on observables, or conditional independence assumption (CIA) holds. Then, a major condition for the consistency of matching estimators is fulfilled (see the survey by Imbens, 2004, for estimation methods).

However, sometimes, the researcher may not be sure whether the control variables are influenced by the treatment. An example is an unclear timing of the control variables, so that some of them may be actually measured after the treatment. Or they are measured prior to the treatment, but the future treatment start is already known and behaviour changes because of this knowledge. Although there is some notion in the literature that conditioning variables should not be influenced by programme participation (not to be confused with the fact that they may be *correlated* with programme participation), which is spelled out for example in Frangakis and Rubin (2002), Heckman, LaLonde, and Smith (1999), Rosenbaum (1984) and Rubin (2004), a clear definition applicable to this case is missing. Strangely, an answer to a

simple question like “Does endogeneity of the control variables matter if the CIA holds?” appears to be missing as well.[1]

This note provides an answer to that question ('it does not matter') and defines the exogeneity conditions required for identification. The main thrust of the paper is that the conditional independence assumption based on observable control variables is sometimes a misleading vehicle for discussing identification of causal effects in a framework based on potential outcomes. Therefore, I suggest a reformulation of the conditional independence assumption together with explicit exogeneity conditions that formally separates the discussion of selection bias coming from a set of control variables that is not rich enough, from the discussion about their potential endogeneity.

The paper is organised as follows: The next section introduces the notation of the causal model based on potential outcomes and defines the parameters of interest. Section 3 considers the endogeneity problem in the standard formulation of the conditional independence assumption (CIA). Sections 4 and 5 reformulate the standard CIA in terms of potential control variables and characterise the bias as well as the exogeneity assumptions required to avoid it.

2 The prototypical binary treatment model

The binary causal model of potential outcomes is currently the workhorse in applied causal analysis. It has its roots in the literature about experimental evaluations in agriculture starting

[1] Since in econometrics *endogeneity* is usually a key issue, it is somewhat surprising that it has not received much general attention in the field of evaluation studies. The papers in this field concentrate fully on the endogeneity problems coming from missing variables that determine the selection process.
Note that we use the term endogeneity (and exogeneity) somewhat casually to mean that the variable is (not) influenced by the treatment, which is not exactly in line with the common use of this language in econometrics (e.g. Engle, Hendry, and Richard, 1983). Alternatively, the exogeneity conditions below may also be termed 'non-causality conditions', although there are different uses for this word as well.

with Neyman (1923).[2] The model is characterised by random variables describing two potential states of the world, also called treatments and denoted by $S \in \{0,1\}$ (e.g. participation or nonparticipation in the programme), the corresponding outcome that would occur if one of the states is realised (Y^1, Y^0) (e.g. employment status that *would* occur in case of participation and nonparticipation), and the observed outcome variables Y.[3] Observed and potential outcome variables are related by the observation rule, $Y = SY^1 + (1-S)Y^0$.

Denote characteristics that influence treatment participation as well as potential outcomes as X (called control variables in econometrics, or confounders in the statistics literature). To address the issue of endogeneity we treat confounders and outcomes in a symmetric way and define random variables for their potential values (X^1, X^0), as well as an observation rule $X = SX^1 + (1-S)X^0$. As usual, we are interested in estimating the average treatment effect on the treated (ATET) $\theta^{10}(1) := E(Y^1 \mid S=1) - E(Y^0 \mid S=1)$, the average treatment effect on the nontreated (ATENT) $\theta^{10}(0) := E(Y^1 \mid S=0) - E(Y^0 \mid S=0)$ and the average treatment effect (ATE) $\theta^{10}(\cdot) := E(Y^1) - E(Y^0)$. Note that $\theta^{10}(\cdot) = \theta^{10}(1)\,\mathrm{P}(S=1) + \theta^{10}(0)[1-P(S=1)]$. $P(S=1)$ denotes the probability of participation. To complete the model, we follow Rubin (1980) and impose the stable unit treatment value assumption (SUTVA). SUTVA implies that treatments on the individual level are well defined and complete and that there is no interference between treated and controls (e.g. no general equilibrium effects for a labour market training programme), and leads to the observation rule. Furthermore, assume that for all values of the observed confounding variables that are of interest $(x \in \chi)$, there is some chance to end up in either of the two states, i.e. $0 < P(S=1 \mid X=x) < 1$ (common support).

[2] In the following we use the terms common in that literature.

[3] We use the binary model for pedagogical reasons only. The extension to a model with multiple treatments along the lines of Imbens (2000) and Lechner (2001, 2002) is straightforward.

With respect to the data available, assume that there is a random sample of the observable variables $\{y_i, s_i, x_i\}_{i=1:N}$, that is large enough so that studying identification in the population instead of the sample is reasonable. Nevertheless, $E(Y^1 \mid S=0)$, $E(Y^0 \mid S=1)$, $E(Y^1)$, and $E(Y^0)$, are not identified without further assumptions.

3 Is there an endogeneity problem when the conditional independence assumption holds?

Assume that 'the data are informative enough, so that controlling for observed variables is sufficient to remove any selection bias' (Rubin, 1974, 1977, 1979). This is the so-called conditional independence (CIA) or selection on observables assumption. It is formalised as:

$$Y^1, Y^0 \perp\!\!\!\perp S \mid X = x, \qquad \forall x \in \chi . \tag{CIA-O}$$

$\perp\!\!\!\perp$ denotes conditional independence as in Dawid (1979). CIA-O implies $F(Y^s \mid S = 1-s, X = x) = F(Y^s \mid S = s, X = x)$, $s \in \{0,1\}$. It suffices to identify the treatment effects because $E(Y^s \mid S = 1-s) = \underset{X \mid S=1-s}{E} E(Y^s \mid S = 1-s, X = x) = \underset{X \mid S=1-s}{E} E(Y^s \mid S = s, X = x) = \underset{X \mid S=1-s}{E} E(Y \mid S = s, X = x)$.

Other than some standard regularity assumptions about the existence of moments, no further assumptions, in particular, *no assumptions* about the exogeneity of the control variables are required. Therefore, the question arises what happens if some of the control variables are influenced by the treatment (i.e. $X^1 \neq X^0$). At this level, the answer is that *it does not matter* at all, as long as (CIA-O) is satisfied. This seems to be odd at first glance, because it is a well known fact that endogenous control variables may lead to tremendous biases of the resulting

estimators. For example, suppose the chosen estimator is a pair matching estimator[4] and the observed outcome variable is included among the control variable, then matched pairs have the same value of the outcome variable and the estimator will always give zero as estimated treatment effect, whatever the true effect may be. Of course, this example is extreme since nobody would actually use outcomes as control variables. However, there are many cases in practice in which there is at least a reasonable suspicion that some of the control variables may actually be a function of the outcome variables. In those cases, qualitatively similar biases may occur.

To resolve this puzzle, note that with an endogenous variable included among the control variables, the untestable assumption CIA may become hard to defend on a priori grounds. Using the observation rule for the confounders, we see the reason: CIA implies $F(Y^s \mid S = 1-s, X^{1-s} = x) = F(Y^s \mid S = s, X^s = x)$. In this case the randomization interpretation of CIA ('if we observe all variables jointly influencing the outcomes and selection, then, conditional on the same value of these variables, selection is random') does not hold any longer, because the comparison relates to the same value, but for different control variables. Since we cannot observe X^1 for nonparticipants and vice versa, we must condition on the observed *X*. However, conditioning on the observed *X* implies conditioning on X^1 for the participants, and conditioning on X^0 for the nonparticipants. This could become a very serious problem when X^1 and X^0 differ.

To conclude, endogeneity of the control variables does not play any role once the CIA is assumed, but it makes the CIA unlikely to hold. Thus, the endogeneity problem is obscured when stipulating the CIA in the usual way as in CIA-O.[5]

---

[4] See Imbens (2004) for an extensive survey of relevant estimation methods.

[5] Simonsen and Skipper (2004) are concerned with endogeneity problems in matching studies as well. However, they assume that CIA is based on exogenous variables only. The thrust of their

4 The explicit treatment of endogeneity

This section discusses potential biases due to endogeneity of the control variables (i.e. they are influenced by the treatment) and gives conditions under which no such biases occur. Such variables may contain valuable information about the selection process as considered in Rosenbaum (1984), they may be in the 'causal pathway' of the treatment, as in Rubin (2005), or they may be intermediate outcomes of a dynamic treatment not yet completed (Lechner and Miquel, 2005).

First, I restate the conditional independence assumption conditioning on variables that are by definition exogenous, namely the potential confounders:[6]

$$Y^1, Y^0 \coprod S \mid X^s = x, \qquad s \in \{0,1\}, \qquad \forall x \in \chi. \tag{CIA-P}$$

CIA-P implies $F(Y^s \mid S = 1-s, X^s = x) = F(Y^s \mid S = s, X^s = x)$. Since this condition is a contrast for the same conditioning variables, the requirement is that if all variables influencing outcomes (Y^1, Y^0) and selection (D) are included in X^s, CIA-P must hold. Note that conditioning on X^1 *or* X^0 is sufficient, whichever appears to be more plausible in the intended application.

CIA-P has the virtue of using conditioning variables that are by definition exogenous. However, it does not allow identification of the effects, because there is only data on X^s for those observations participating in state S=s. Thus, additional assumptions are required linking the potential X to the observed X. For the strongest of such assumptions, i.e. that S has no influ-

---

paper seems to be to uncover effect heterogeneity driven by endogenous variables as well as uncovering the effect of the endogenous variables together with S on the outcomes, which is a different issue than the topic discussed here.

[6] Note that *potential* confounders are in principle the same type of variables as potential outcomes. However, the difference is that we are not interested in the direct effect of the treatment on potential confounders.

ence on X, $X = X^1 = X^0$, CIA-P and CIA-O are identical.[7] In fact, they amount to what Rosenbaum and Rubin (1983) call the assumption of *strongly ignorable treatment assignment*.

4.1 Bias

Next, consider the potential asymptotic bias of a matching estimator due to the endogeneity of the control variables for the counterfactuals $E(Y^s \mid S = 1-s)$. This bias can be characterised by using the observation rule for outcomes and control variables together with CIA-P. The characterisation depends on whether the potential outcome and potential confounders used in CIA-P relate to the same or a different potential state. For the same state, we obtain:

$$\begin{aligned}
\underbrace{E(Y^s \mid S=1-s) - \underset{X\mid S=1-s}{E} E(Y \mid S=s, X=x)}_{Bias} &= \underset{X^s\mid S=1-s}{E} E(Y^s \mid S=1-s, X^s=x) - \underset{X^{1-s}\mid S=1-s}{E} E(Y \mid S=s, X=x) \\
&= \underset{X^s\mid S=1-s}{E} E(Y \mid S=s, X^s=x) \quad - \underset{X^{1-s}\mid S=1-s}{E} E(Y \mid S=s, X=x) \\
&= \underset{X^s\mid S=1-s}{E} E(Y \mid S=s, X=x) \quad - \underset{X^{1-s}\mid S=1-s}{E} E(Y \mid S=s, X=x).
\end{aligned}$$

The bias arises because X^1 and X^0 may have different distributions in the population of interest (*1-s*). If $F(X^1 \mid S=1-s) = F(X^0 \mid S=1-s)$ $[= F(X \mid S=1-s)]$, the bias is zero. This is an exogeneity condition, because it stipulates that the distribution of the controls is not affected by the treatment in a specific subpopulation. For example, if interest is in ATET ($\theta^{10}(1)$) and conditioning is on X^0, then this is all that is needed for identification.

When the potential confounder related to the alternative state, then we can rewrite the bias differently, such that a second interpretation becomes obvious:

[7] See for example Assumption A6 in Heckman and Vytlacil (2005).

$$E(Y^s \mid S=1-s) - \underset{X|S=1-s}{E} E(Y \mid S=s, X=x) = \underset{X^{1-s}|S=1-s}{E} E(Y^s \mid S=1-s, X^{1-s}=x) - \underset{X|S=1-s}{E} E(Y \mid S=s, X=x)$$
$$= \underset{X^{1-s}|S=1-s}{E} E(Y \mid S=s, X^{1-s}=x) - \underset{X|S=1-s}{E} E(Y \mid S=s, X=x)$$
$$= \underset{X|S=1-s}{E} [E(Y \mid S=s, X^{1-s}=x) - E(Y \mid S=s, X=x)].$$

The bias arises because we should evaluate the first conditional expectation of Y as a function of the counterfactual conditioning variable (X^{1-s}), instead of the observed one (X^s). Thus, if both expectations coincide, i.e. $E(Y \mid S=s, X^{1-s}=x) = E(Y \mid S=s, X^s=x)$, then no bias arises. This is again an exogeneity condition in a specific subpopulation, although weaker than the previous one. It stipulates that the X may be influenced by the treatment, but these influences should not matter for the outcomes.

4.2 Identification under explicit exogeneity conditions

The previous considerations suggest explicitly introducing two exogeneity conditions for X:

$$F(X^s \mid S=1-s) = F(X^{1-s} \mid S=1-s); \tag{E.1}$$

$$E(Y \mid S=s, X^{1-s}=x) = E(Y \mid S=s, X^s=x). \tag{E.2}$$

If interest is in $\theta^{1-s,s}(1-s)$, then *CIA-P* (for Y^s only) conditioning on X^s in combination with *E.1*, or *CIA-P* (for Y^s only) conditioning on X^{1-s} in combination with *E.2*, is sufficient.[8] If interest is in the average treatment effect, $\theta^{1-s,s}(\cdot)$, then CIA-P conditioning on X^s or X^{1-s} in combination with E.1 and E.2, or CIA-P conditioning on X^s and CIA-P conditioning on X^{1-s} in combination with either E.1 or E.2 are sufficient. The proofs follow directly from the expressions of the biases given above. If the treatment does not influence the control variables at all ($X^1 = X^0$), E.1 and E.2 hold trivially anyway. Note that together with the two

[8] By an obvious change of indices, a similar result is obtained for $\theta^{1-s,s}(s)$.

exogeneity conditions, CIA-P implies CIA-O, and vice versa. However, without those exogeneity conditions, neither implies the other.

It is interesting to compare the exogeneity conditions suggested above to conditions already considered by Heckman, LaLonde, and Smith (1999). They require $F(X \mid Y^1, Y^0, S = s) = F(X \mid Y^1, Y^0)$, which implies $F(X^s \mid Y^1, Y^0, S = s) = F(X^{1-s} \mid Y^1, Y^0, S = 1-s)$. Although this condition is similar to E.1 and E.2, the two equations characterising the bias presented in section 4.1 show that this does not solve the problem. When used together with CIA based on observables (which is of course unnecessary for identification), it has however intuitive appeal, because it rules out some dependence of the observed *X* on *S*.

The early paper by Rosenbaum (1984) considers a formal framework in which the endogeneity comes from some variables affected by the treatment that are, however, not necessary as control variables for the CIA. Thus, his question is about the potential damage when such variables are included in the conditioning set. His sufficient conditions for no selection bias imposes $X^1 = X^0$. Thus, it seems to be stronger than necessary.[9]

The idea to condition on potential confounders is similar to an idea recently advanced in papers by Frangakis and Rubin (2002) and Rubin (2004). The aim of these authors is to arrive at valid results for the effects based on matching estimation despite the endogeneity problem of the controls. They define what they call *principle strata* of the data. These strata are defined by the values of the potential confounding variables. They 'solve' the problem that such strata are inherently unobservable by, firstly, imputing the values for the potentially endogenous control variables. Then, they use the estimated values of the potential control variables as control variables in the final estimation. This method that shows some similarity to ideas underlying the old parametric two stage least squares estimator in econometrics is

computationally and conceptionally attractive, but it hinges on additional strong assumptions, which pose a new identification problem that may be as difficult to solve - convincingly - as the original one.

5 Conclusion

This paper showed that the fact that some control variables may be influenced by the treatment does not matter as long as the usual formulation of the conditional independence assumption holds. In other words, the usual formulation of the CIA obscures this endogeneity problem. Therefore, an alternative formulation of the conditional independence assumption together with explicit exogeneity conditions is proposed. The new conditions allow separating the discussion of eliminating selection bias by conditioning on the necessary control variables from the discussion about their potential endogeneity with respect to outcome and selection variables. Thus, these conditions should provide useful and in fact better guidance for applied work, in particular for the assessment of the plausibility of the untestable assumptions that are always required for the identification of causal effects.

Acknowledgments

I thank Don Rubin, Jeff Smith, Markus Frölich, and an anonymous referee for helpful comments on previous versions of this paper. The usual disclaimer applies.

References

Dawid, A. P. (1979): "Conditional Independence in Statistical Theory", *The Journal of the Royal Statistical Society, Series B*, 41, 1-31.

Engle R., D. F. Hendry, and J.-F. Richard (1983): "Exogeneity", *Econometrica*, 51, 277-304.

[9] Due to the specific framework he uses, he must impose a further condition that amounts to a 'no selection bias conditional on X assumption' for the affected conditioning variables as well.

Frangakis, C. E., and D. B. Rubin (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

Gerfin, M., and M. Lechner (2002): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland", *The Economic Journal*, 112, 854-893.

Heckman, J., R. LaLonde and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Vol. 3, 1865-2097, Amsterdam: North-Holland.

Heckman, James J, and Edward Vytlacil (2005): "Causal Parameters, Structural Equations, Treatment Effects and Randomized Evaluation of Social Programs", *Econometrica*.

Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *The Review of Economics and Statistics*, 86, 4–29.

Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.

Lechner, M. (2002): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", *Review of Economics and Statistics*, 84, 205-220.

Lechner, M., and R. Miquel (2005): "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions", Discussion Paper, Department of Economics, University of St. Gallen, 2002, revised 2005.

Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essays on Principles. Section 9," translated in *Social Science*, 5, 465-480.

Rosenbaum, P. R. (1984): "The Consequences of Adjustment for a Concominat Variable That Has Been Affected by the Treatment", *The Journal of the Royal Statistical Society, Series A*, 147, 656-666.

Rosenbaum, P.R., and D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrica*, 70, 41-50.

Robins, J. M. (1986): “A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect.” *Mathematical Modelling*, 7, 1393-1512,

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

Rubin, D. B. (1977): "Assignment to Treatment Group on the Basis of a Covariate", *Journal of Educational* Statistics, 2 (1), 1-26.

Rubin, D. B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association*, 74, 318-328.

Rubin, D. B. (1980): "Discussion of 'Randomization analysis of experimental data in the Fisher randomisation test by Basu", *Journal of the American Statistical Association*, 75, 591-593.

Rubin, D. B. (2004): "Direct and Indirect Causal Effects via Potential Outcomes", *Scandinavian Journal of Statistics*, 31, 161-170.

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

Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s", The *Review of Economics and Statistics*, 86, 133–155.

Simonsen, M., and L. Skipper (2004): "The Cost of Motherhood: An Analysis Using Matching Estimators", *mimeo*.