Research in Economics 70 (2016) 232–237

Contents lists available at ScienceDirect

Research in Economics

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

Should instrumental variables be used as matching variables?☆

Jeffrey M. Wooldridge*

Department of Economics, Michigan State University, East Lansing, MI 48824-1038, United States

ARTICLE INFO

Article history:
Received 25 November 2015
Accepted 9 January 2016
Available online 15 February 2016

Keywords:
Matching
Instrumental variable
Inconsistency
Treatment effect

ABSTRACT

I show that for a linear model and estimating a coefficient on an endogenous explanatory variable, adding covariates that satisfy instrumental variables assumptions increases the amount of inconsistency. A special case is an endogenous binary treatment and estimating a constant treatment effect when matching on covariates that satisfy instrumental variables, rather than ignorability, assumptions. I also establish a general result that implies that regression adjustment using the propensity score based on instrumental variables actually maximizes the inconsistency among regression-type estimators.

1. Introduction

In the context of instrumental variables estimation, it is fairly well known that covariates satisfying proxy variable assumptions make poor instruments: proxy variables are supposed to be highly correlated with unobservables, which is why proxy variables should be included as controls, not used as instruments. What seems to be less well understood is that, in the context of matching-type estimators, we should not match on covariates that satisfy instrumental variable assumptions. The current paper is motivated by the work of Heckman and Navarro-Lozano (2004) (HN-L), a paper I discussed in an invited session at the 2005 ASSA meetings in Philadelphia. In their paper, Heckman and Navarro-Lozano present simulation results for matching estimators when the key ignorability (or unconfoundedness) assumption used in matching fails. In particular, the authors assume a setup with self-selection into treatment – the kind of self-selection that can be, under certain assumptions, solved by instrumental variables (IV) or control function methods. Within this context the authors study, via simulations, the performance of matching estimators that match on the basis of the instrumental variables.

At the time I read the HN-L paper, it struck me that matching on instrumental variables – rather than on covariates that have predictive power for unobservables affecting the response – was not a good idea, and that few empirical researchers would use such a strategy. But were HN-L really setting up a straw man? The literature has been somewhat vague on the kinds of covariates that make sense in matching estimators. HN-L state this explicitly: "The method of matching offers no guidance as to which variables to include or exclude in conditioning sets" (p. 30).

It is pretty well known that covariates that are influenced by the treatment can cause the ignorability assumption to be violated and lead to larger biases when they are included. Rosenbaum (1984) characterizes the bias that can occur when posttreatment outcomes are included in the covariates, and the simulations in Heckman and Navarro-Lozano (2004) effectively make this point via simulations. Wooldridge (2005) formally shows that if treatment is randomized with respect

☆ This is a revised version of an unpublished paper that began circulating in late 2006.
* Tel.: +1 517 353 5972; fax: +1 517 432 1068.
E-mail address: wooldri1@msu.edu

http://dx.doi.org/10.1016/j.rie.2016.01.001
1090-9443/

to the counterfactual outcomes but not with respect to the covariates, ignorability is generally violated. But what about covariates that affect treatment without having a partial effect on the response?

In this paper, I formalize the notion that when the goal is to estimate a treatment (or causal) effect, matching on instrumental variables is an especially bad idea when treatment is endogenous and cannot be made ignorable by conditioning on covariates. The result for matching is a special case of a more general result for linear regression: including in a regression analysis any functions of instrumental variables, along with an endogenous explanatory variable and other covariates, leads to more asymptotic bias than excluding the instrumental variables. The exception is when there is no bias in the short regression to begin with, in which case including instrumental variables among the covariates reduces precision.

After I posted a version of this paper in 2008, I was made aware of independent, related work by Bhattacharya and Vogt (2012) (BV). The basic message in the BV paper is the same as the one here, although there are some notable differences in the approaches. First, I consider a general regression framework that allows non-binary treatments, with binary treatments as a special case. Second, I explicitly allow for conditioning on covariates that satisfy the usual ignorability assumptions. BV allow for heterogeneous treatment effects but confine their attention to the binary treatment case without extra covariates. The two papers are complementary in that they reach the same basic conclusion using different settings.

The rest of the paper is organized as follows. Section 2 begins with the simple case where all available covariates satisfy instrumental variables assumptions. Section 3 extends to the more realistic case where some covariates do not satisfy instrumental variables assumptions and are included as controls, as is very standard in practice. An important result is that it is always worse to include instrumental variables in the matching covariates than to exclude those covariates that satisfy IV assumptions.

2. The problem in a simple setting

Consider the simple model

$$y = \alpha + \beta w + u, \tag{2.1}$$

where all quantities are scalars and w is thought to be endogenous (correlated with the error, u). The nature of w is unrestricted – it can be continuous, discrete, or exhibit both properties. The parameter of interest is β.

Given a random sample of size N, $\{(y_i, w_i): i = 1, \ldots, N\}$, the probability limit of the slope estimator $\hat{\beta}$ from the regression

$$y_i \text{ on } 1,\ w_i, \quad i = 1, \ldots, N, \tag{2.2}$$

is well known:

$$plim(\hat{\beta}) = \beta + Cov(w,u)/Var(w). \tag{2.3}$$

That $\hat{\beta}$ is inconsistent for β when $Cov(w,u) \neq 0$ motivates the search for different estimators. Broadly speaking, in a cross-sectional environment, there are two possibilities for obtaining estimators with less "asymptotic bias." If we have a K-vector of extra controls, say $\mathbf{x}$, that satisfies

$$E(u|w, \mathbf{x}) = E(u|\mathbf{x}) \equiv g(\mathbf{x}), \tag{2.4}$$

then we can consistently estimate β by adding the function $g(\mathbf{x})$ to the regression. In practice, we tend to approximate $g(\mathbf{x})$ using parametric functions, particularly those linear in parameters. In (2.4), we call $\mathbf{x}$ a set of *proxy variables* for the unobservables, u. Wooldridge (2010, Section 4.3.2) contains further discussion. Even if (2.4) does not hold, it could be that adding a function of $\mathbf{x}$ to the regression can reduce the asymptotic bias. In the treatment effect literature, adding functions of $\mathbf{x}$ to a regression is typically called "regression adjustment."

Alternatively, we might have another set of variables, say, an L-vector $\mathbf{z}$, that satisfies a very different assumption:

$$E(u|\mathbf{z}) = 0. \tag{2.5}$$

In (2.5), we say $\mathbf{z}$ is a set of *instrumental variables* candidates. (For the purposes of this paper, we state the exogeneity assumption in terms of a zero conditional mean, rather than zero correlation. The main reason for this choice is to allow general nonlinear functions of covariates and instruments within a regression framework.)

We can easily show that if w is correlated with u – so that w is endogenous in Eq. (2.1) – including IVs as regressors is always worse than using the simple regression estimator.

Proposition 2.1. *Let $h(\mathbf{z})$ be any function of $\mathbf{z}$ and let $\tilde{\beta}$ be the coefficient on w_i from the regression*

$$y_i \text{ on } 1,\ w_i,\ h(\mathbf{z}_i), \quad i = 1, \ldots, N. \tag{2.6}$$

Then, assuming standard moment assumptions such that the law of large numbers can be applied,

$$|plim(\tilde{\beta}) - \beta| \geq |plim(\hat{\beta}) - \beta|, \tag{2.7}$$

with strict inequality whenever $Cov(w,u) \neq 0$ and $Cov[h(\mathbf{z}), w] \neq 0$.