CAMBRIDGE
UNIVERSITY PRESS

Violating Ignorability of Treatment by Controlling for Too Many Factors

Author(s): Jeffrey M. Wooldridge

Source: *Econometric Theory*, Oct., 2005, Vol. 21, No. 5 (Oct., 2005), pp. 1026-1028

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/3533523

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Cambridge University Press is collaborating with JSTOR to digitize, preserve and extend access to *Econometric Theory*

JSTOR

This content downloaded from
71.195.40.232 on Tue, 18 May 2021 20:09:56 UTC
All use subject to https://about.jstor.org/terms

Econometric Theory, **21**, 2005, 1026–1028. Printed in the United States of America.
DOI: 10.1017/S0266466605050516

VIOLATING IGNORABILITY OF TREATMENT BY CONTROLLING FOR TOO MANY FACTORS

JEFFREY M. WOOLDRIDGE
Michigan State University

This problem shows how the key ignorability-of-treatment assumption used in estimating treatment effects can be violated when certain factors are included among the covariates. The case considered is when there are $J + 1$ treatment levels, treatment is randomized with respect to potential outcomes, but the distribution of included covariates differs across treatment levels.

1. MOTIVATION AND RESULTS

In introductory econometrics classes, students are sometimes warned of the perils of controlling for too many factors in multiple regression analysis. In particular, it may not make sense to hold fixed certain factors when estimating the partial effect of the variable of interest. Overcontrolling is often a result of using goodness-of-fit measures or statistical tests to decide whether to include certain covariates and forgetting about the ceteris paribus interpretation of regression coefficients.

The counterfactual setting of modern treatment effect analysis provides an ideal setting in which to illustrate how one can easily include too many factors. Plus, propensity-score-based methods of estimating treatment effects are becoming rather popular, and one sees a tendency to include many factors in estimating the propensity scores (the probabilities of the various treatment levels). This exercise provides a simple demonstration of how including certain controls—specifically, those that are themselves affected by treatment—generally violates the key ignorability assumption.

To formalize this conclusion, consider the counterfactual treatment setup where w is a discrete treatment outcome taking values in $\{0,1,2,\ldots,J\}$, so that there are $J + 1$ treatment levels. The corresponding counterfactual outcomes are $(y_0, y_1, \ldots, y_J)$; that is, for a generic population unit, y_j is the outcome or response if the treatment level is $w = j$. (For the binary case, see, e.g., Rosenbaum and Rubin, 1983; Heckman, Ichimura, and Todd, 1997.) We do not observe each counterfactual outcome; rather, we observe only the outcome associated

Address correspondence to Jeffrey M. Wooldridge, Department of Economics, Michigan State University, East Lansing, MI 48824-1038, USA; e-mail: wooldri1@msu.edu

 © 2005 Cambridge University Press 0266-4666/05 $12.00

This content downloaded from
71.195.40.232 on Tue, 18 May 2021 20:09:56 UTC
All use subject to https://about.jstor.org/terms

with the treatment level w. Write the observed outcome in terms of the treatment level and counterfactual outcomes as

$$y \equiv 1[w = 0]\cdot y_0 + 1[w = 1]\cdot y_1 + \cdots + 1[w = J]\cdot y_J$$
$$\equiv w_0 y_0 + w_1 y_1 + \cdots + w_J y_J, \tag{1}$$

where $w_j \equiv 1[w = j]$ and $1[\cdot]$ is the indicator function. Let $\mathbf{x}$ be a K-vector of observed covariates thought to be "related to" the treatment level and also the counterfactual outcomes. A key assumption in the treatment effect literature is *ignorability of treatment*: conditional on $\mathbf{x}$, w and $(y_0, y_1, \ldots, y_J)$ are independent. A weaker version of the assumption—which often suffices for consistency of propensity-score or regression-adjustment methods—is conditional mean independence:

$$E(y_j|w,\mathbf{x}) = E(y_j|\mathbf{x}), \qquad j = 0,1,\ldots,J. \tag{2}$$

Typically, we think (2) is more plausible the richer is $\mathbf{x}$; that is, the more factors that might predict treatment status, the better. Nevertheless, as in standard regression contexts, it is possible to include too many factors in $\mathbf{x}$. In particular, one might inadvertently include in $\mathbf{x}$ variables (other than y, of course) that are themselves influenced by the treatment level. For example, in evaluating the effects of drug courts on recidivism, should we include in $\mathbf{x}$ a measure of postsentencing education levels or employment status? Almost certainly not. Differences in postsentencing education and employment, depending on whether the juvenile was sentenced in regular court or a drug court, are themselves responses to the policy treatment.

To obtain an unambiguous result, consider the extreme case where treatment is actually randomized with respect to $(y_0, y_1, \ldots, y_J)$. That is, assume that w and $(y_0, y_1, \ldots, y_J)$ are independent. Letting $D(\cdot)$ denote distribution, show that if $D(\mathbf{x}|w) \neq D(\mathbf{x})$—so that treatment is *not* randomized with respect to $\mathbf{x}$—then the ignorability assumption in (2) *cannot* hold unless $E(y_j|\mathbf{x}) = E(y_j)$ for all j.

2. PROOF AND DISCUSSION

The proof that (2) cannot hold under the stated assumptions is a simple application of the law of iterated expectations. We derive a contradiction by assuming that (2) holds, that w and $(y_0, y_1, \ldots, y_J)$ are independent, and that $D(\mathbf{x}|w) \neq D(\mathbf{x})$. First, by iterated expectations,

$$E(y_j|w) = E[E(y_j|w,\mathbf{x})|w]. \tag{3}$$

But, because w is independent of y_j, the left-hand side does not depend on w, and $E(y_j|w,\mathbf{x})$ does not depend on w if (2) is supposed to hold. Write $\mu_j(\mathbf{x}) \equiv E(y_j|\mathbf{x})$. Then (3) and the assumptions imply

$$E(y_j) = E[\mu_j(\mathbf{x})|w], \tag{4}$$

This content downloaded from
71.195.40.232 on Tue, 18 May 2021 20:09:56 UTC
All use subject to https://about.jstor.org/terms

which is impossible if the right-hand side depends on w. If $\mu_j(\mathbf{x})$ depends on $\mathbf{x}$ and $D(\mathbf{x}|w) \neq D(\mathbf{x})$, then the right-hand side of (4) generally depends on w.

Intuitively, it may seem clear that one should not include in $\mathbf{x}$ variables that are themselves responses to treatment, but this can get lost, especially when applying propensity-score methods. In a first stage, the propensity scores, $p_j(\mathbf{x}) \equiv P(w = j|\mathbf{x})$, are estimated, and then there are various ways to use these in a second stage (see, e.g., Wooldridge, 2002, Ch. 18). The initial focus is on getting "good" estimates of the probabilities of each treatment class, and, if "good" is taken to mean "best" fit, it is tempting to include anything in $\mathbf{x}$ that helps predict w.

REFERENCES

Heckman, J.J., H. Ichimura, & P. Todd (1997) Matching as an econometric evaluation estimator. *Review of Economic Studies* 65, 261–294.

Rosenbaum, P.R. & D.B. Rubin (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

This content downloaded from
71.195.40.232 on Tue, 18 May 2021 20:09:56 UTC
All use subject to https://about.jstor.org/terms