Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 1281–1301
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Inverse probability weighted estimation for general missing data problems

Jeffrey M. Wooldridge*

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

Available online 9 March 2007

Abstract

I study inverse probability weighted M-estimation under a general missing data scheme. Examples include M-estimation with missing data due to a censored survival time, propensity score estimation of the average treatment effect in the linear exponential family, and variable probability sampling with observed retention frequencies. I extend an important result known to hold in special cases: estimating the selection probabilities is generally more efficient than if the known selection probabilities could be used in estimation. For the treatment effect case, the setup allows a general characterization of a "double robustness" result due to Scharfstein et al. [1999. Rejoinder. Journal of the American Statistical Association 94, 1135–1146].

JEL classification: C13; C21; C23

Keywords: Inverse probability weighting; Sample selection; M-estimator; Censored duration; Average treatment effect

1. Introduction

In this paper I extend earlier work on inverse probability weighted (IPW) M-estimation along several dimensions. One important extension is that I allow the selection probabilities to depend on selection predictors that are not fully observed. In Wooldridge (2002a), building on the framework of Robins and Rotnitzky (1995) for attrition in regression, I assumed that the variables determining selection were always observed and

*Tel.: +1 517 353 5972; fax: +1 517 432 1068.
E-mail address: wooldri1@msu.edu.

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2007.02.002

that the selection probabilities were estimated by binary response maximum likelihood. These assumptions exclude some interesting cases, including: (i) variable probability (VP) sampling with known retention frequencies; (ii) a censored response variable with varying censoring times, as in Koul et al. (1981); (iii) unobservability of a response variable due to censoring of a second variable, as in Lin (2000).

Extending previous results to allow more general selection mechanisms is fairly routine when interest lies in consistent estimation. My goal here is to expand the scope of a result that has appeared in a variety of settings with missing data: estimating the selection probabilities generally leads to a more efficient weighted estimator than if the known probabilities could be used. A few examples include Imbens (1992) for choice-based sampling, Robins and Rotnitzky (1995) for IPW estimation of nonlinear regression models, and Wooldridge (2002a) for general M-estimation under the Robins and Rotnitzky (1995) sampling scheme.

Having a unified setting where asymptotic efficiency is improved by using estimated selection probabilities has several advantages. First, knowing that an estimator produces narrower asymptotic confidence intervals has obvious benefits. Second, the proof of relative efficiency leads to a computationally simple estimator of the asymptotic variance for a broad class of estimation problems, including popular nonlinear models. For example, Koul et al. (1981) and Lin (2000) treat only the linear regression case, and the formulas are almost prohibitively complicated. A third benefit is that I expand the scope of models and estimation methods where one can obtain conservative inference by ignoring the first-stage estimation of the selection probabilities.

Another innovation in this paper is my treatment of exogenous selection when some feature of a conditional distribution is correctly specified. Namely, I study the properties of the IPW M-estimator when the selection probability model is possibly misspecified. Among other things, allowing misspecified selection probabilities in the exogenous selection case leads to key insights for more robust estimation of average treatment effects (ATEs).

The remainder of the paper is organized as follows. In Section 2, I briefly introduce the underlying population minimization problem. In Section 3, I describe the selection problem and propose a class of conditional likelihoods for estimating the selection probabilities; obtain the asymptotic variance of the IPW M-estimator; show that it is more efficient to use estimated probabilities than to use the known probabilities; and provide a simple estimator of the efficient asymptotic variance matrix. Section 4 covers the case of exogenous selection, allowing the selection probability model to be misspecified. In Section 5, I provide a general discussion of the considerations when deciding whether or not to use inverse-probability weighting. I cover three examples in Section 6: (i) estimating a conditional mean function when the response variable is missing due to a censored duration; (ii) estimating an ATE with a possibly misspecified conditional mean function; and (iii) VP sampling with observed retention frequencies.

2. The population optimization problem and random sampling

The starting point is a population optimization problem, which essentially defines the parameters of interest. Let w be an $M \times 1$ random vector taking values in $W \subset \Re^M$. Some aspect of the distribution of w depends on a $P \times 1$ parameter vector, θ, contained in a parameter space $\Theta \subset \Re^P$. Let $q(w, \theta)$ denote an objective function.