

Statistica Sinica Preprint No: SS-2016-0324.R2

Title	Semiparametric Estimation with Data Missing Not at Random Using an Instrumental Variable
Manuscript ID	SS-2016.0324.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0324
Complete List of Authors	Eric Tchetgen Tchetgen Baoluo Sun Lan Liu Wang Miao Kathleen Wirth and Robin James
Corresponding Author	Eric Tchetgen Tchetgen
E-mail	etchetge@hsph.harvard.edu
Notice: Accepted version subject to English editing.	

the additive, multiplicative or odds ratio scale. However, their approach for estimation is fully parametric and may be sensitive to bias due to model misspecification. Therefore a more robust approach is warranted.

In this paper, we develop a general framework for nonparametric identification of selection models based on an IV. We describe necessary and sufficient conditions for identifiability of the full data distribution with a valid IV. For inference we focus on estimation of an outcome mean, although the proposed methods are easy to adapt to other functionals. We develop semiparametric approaches including inverse probability weighting (IPW) and outcome regression (OR) that extend analogous methods previously developed for missing at random (MAR) settings, and introduce a novel doubly robust (DR) estimation approach. The consistency of each estimator relies on correctly specified models for different parts of the data generating model. We note that IPW in MNAR via calibration weighting (Deville (2000); Kott (2006); Chang and Kott (2008)) has previously been proposed to account for sample nonresponse in survey design settings, and typically requires matching of weighted estimates to population totals for benchmark variables. Besides assuming a correctly specified model for non-response, identification in such settings is made possible by availability of known or estimated population totals, an assumption we do not require.

