ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Measuring model misspecification: Application to propensity score methods with complex survey data

David Lenis [a,*,1], Benjamin Ackerman [a], Elizabeth A. Stuart [a,b,c]

[a] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St, E3527, Baltimore, MD 21205, USA
[b] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, 624 N. Broadway, 8th Floor, Baltimore, MD 21205, USA
[c] Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, 624. N. Broadway, 4th Floor, Baltimore, MD 21205, USA

ARTICLE INFO

Article history:
Received 14 August 2017
Received in revised form 4 May 2018
Accepted 5 May 2018
Available online xxxx

Keywords:
Model misspecification
Non-experimental study
Propensity score matching
Treatment on the treated weighting
Complex survey data
Causal inference

ABSTRACT

Model misspecification is a potential problem for any parametric-model based analysis. However, the measurement and consequences of model misspecification have not been well formalized in the context of causal inference. A measure of model misspecification is proposed, and the consequences of model misspecification in non-experimental causal inference methods are investigated. The metric is then used to explore which estimators are more sensitive to misspecification of the outcome and/or treatment assignment model. Three frequently used estimators of the treatment effect are considered, all of which rely on the propensity score: (1) full matching, (2) 1:1 nearest neighbor matching, and (3) weighting. The performance of these estimators is evaluated under two different sampling designs: (1) simple random sampling (SRS) and (2) a two-stage stratified survey. As the degree of misspecification of either the propensity score or outcome model increases, so does the bias and the root mean square error, while the coverage decreases. Results are similar for the simple random sample and a complex survey design.

1. Introduction

Model misspecification is a potential problem for nearly all methods that use parametric models, which leads to worries about incorrect inferences from misspecified models. However, to this point there has been relatively little formal investigation of model misspecification, including characterization of the extent of misspecification and how that might impact methods' performance. Here we propose a metric of misspecification and investigate the consequences of model misspecification within the context of causal inference in non-experimental studies, where there have been longstanding debates about whether misspecification of the treatment assignment model or the outcome model is more detrimental to estimation of treatment effects. Even though the measure for the degree of model misspecification presented in this article is used here in the context of causal inference, it can easily be applied to assess the impact of model misspecification in other model-based methods.

Randomized clinical trials (RCTs) are considered the gold standard for estimating causal effects. In an RCT, the researcher knows the treatment assignment mechanisms, allowing unbiased estimators of causal effects. Nevertheless, it is not unusual to find circumstances where a random assignment of the treatment is unfeasible or unethical. When this happens, researchers need to rely on non-experimental data.

* Corresponding author.
E-mail addresses: davidleniz@gmail.com (D. Lenis), backer10@jhu.edu (B. Ackerman), estuart@jhu.edu (E.A. Stuart).
[1] Present Address: Flatiron Health, 200 5th Ave, 8th Floor, New York, NY 10100, USA.

https://doi.org/10.1016/j.csda.2018.05.003
0167-9473/

A main drawback of non-experimental data is that the treatment assignment is not random, therefore there may be confounders that are related to the outcome and differ between treatment and comparison groups. Failure to address confounding will lead to biased estimators of causal effects. One way to mitigate confounding by observed characteristics is using the propensity score, which models the probability of being assigned to the treatment group given the set of confounders.

Non-experimental studies provide a particularly interesting case study for examining model misspecification because model misspecification can be an issue in two ways when using propensity score methods to estimate causal effects: first, in estimating the propensity score, and second, in the outcome model. Since the true treatment assignment mechanism is hardly ever known when working with non-experimental data, different approaches have been suggested to model and estimate the propensity score. While some authors have proposed nonparametric estimation procedures (Hahn, 1998; Imbens, 2004; Ho et al., 2011), it is common practice to estimate the propensity score parametrically via logistic regression.

Models are also used in the outcome analysis. Work by Cochran and Rubin (1973), Rubin (1973b), Carpenter (1977), Rubin (1979), Rosenbaum and Rubin (1984), Robins and Rotnitzky (1995), Heckman and Todd (2009), Rubin and Thomas (2000), Glazerman et al. (2003), Imai and Van Dyk (2004), Abadie and Imbens (2006) and Ho et al. (2007) suggested that adjusting for confounders in an outcome model may significantly improve inference on causal effects.

Thus, model assisted estimation of causal effects is a common practice in causal inference. However, there have been relatively few formal investigations of the consequences of model misspecification for different propensity score methods, and whether the degree of misspecification of the treatment assignment model has greater ramifications on the bias or mean squared error of the estimate than that of the outcome model. Previous studies of model misspecification in the context of causal inference have grouped misspecified models in broad ad-hoc categories such as "incorrect model" or "wrong model" (Drake, 1993; Kang and Schafer, 2007; Robins et al., 2007). To our knowledge, this is the first attempt to systematically quantify the degree of model misspecification and evaluate its impact on two of the more commonly used estimation procedures (i.e., propensity score matching and weighting) under different survey designs.

Complex survey designs provide an extra layer of complexity when estimating causal effects. Non-experimental studies often use complex survey data, but there is relatively little guidance on how to incorporate the survey design in propensity score methods. Zanutto et al. (2005) and Zanutto (2006) discussed the use of propensity score subclassification with complex survey data, as illustrated in Hornik et al. (2001). Work by Austin et al. (2016) and Lenis et al. (2017) extended the use of propensity score matching to complex survey data. Similarly, Ridgeway et al. (2015) provided some insight on how to compute inverse probability of treatment weighting (IPTW) estimators using complex survey data; however, it is unclear whether model misspecification would have different implications in the complex survey context.

This paper is organized as follows: in Section 2, we present key definitions and assumptions needed for the estimation of causal effects in the context of non-experimental data. Section 3 reviews the methods implemented in our simulation study. Section 4 contains the details of our simulation study. In Section 5, the main results are presented, followed by the discussion and main conclusions in Section 6.

2. Definitions and assumptions

2.1. The causal inference framework

Traditionally, causal treatment effects are defined using the Rubin Causal Model (RCM) (Rubin, 1974). In the RCM, an individual treatment effect, associated with a binary treatment assignment T, is defined in terms of potential outcomes. For each unit i, $Y_i(t)$ with $t = 0, 1$, represents the outcome that would have been observed if unit i received the treatment t. Thus, the treatment effect for the ith unit is equal to $Y_i(1) - Y_i(0)$. Notice that for any unit i, the pair $(Y_i(0), Y_i(1))$ is not observable — only one of the two potential outcomes is observed. Explicitly, the observed outcome, Y_i, is defined as:

$$Y_i = Y_i(1) \times T_i + Y_i(0) \times (1 - T_i)\,. \tag{1}$$

Eq. (1) is referred as the "consistency of the observed outcome assumption" (Hernan and Robins, 2017). Given that the unit level treatment effects cannot be estimated directly, we are often interested in estimating average treatment effects. At the population level, the most commonly defined average effects are: (1) the population average treatment effect (PATE) and (2) the population average treatment effect on the treated (PATT).

The PATE is defined as average effect across the population:

$$PATE = E\left[Y(1) - Y(0)\right]. \tag{2}$$

Under randomization of the treatment, units in the treated group and the units in the control group have similar distributions of covariates (observed and unobserved) and potential outcomes. In this way, the average outcome computed among the units in the treated group serves as a good counterfactual for the average outcome computed among the units in the control group. The differences between these two averages is an estimator of the population average treatment effect (PATE).

The PATT is defined as the average causal effect, computed only among those units in the population who were actually treated:

$$PATT = E\left[Y(1) - Y(0)|T = 1\right]. \tag{3}$$

Please cite this article in press as: Lenis D., et al., Measuring model misspecification: Application to propensity score methods with complex survey data. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.05.003.