Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Discussion of "Penalized Spline of Propensity Methods for Treatment Comparison" by Zhou, Elliott, and Little

Shu Yang & Donglin Zeng

To cite this article: Shu Yang & Donglin Zeng (2019) Discussion of "Penalized Spline of Propensity Methods for Treatment Comparison" by Zhou, Elliott, and Little, Journal of the American Statistical Association, 114:525, 30-32, DOI: 10.1080/01621459.2018.1537916

To link to this article: https://doi.org/10.1080/01621459.2018.1537916

Published online: 19 Apr 2019.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2019, VOL. 114, NO. 525, 31–32: Applications and Case Studies
https://doi.org/10.1080/01621459.2018.1537916

Discussion of "Penalized Spline of Propensity Methods for Treatment Comparison" by Zhou, Elliott, and Little

Shu Yang[a] and Donglin Zeng[b]

[a]Department of Statistics, North Carolina State University, Raleigh, NC; [b]Department of Biostatistics, University of North Carolina, Chapel Hill, NC

We congratulate the authors on presenting this stimulating article. The article proposes a novel estimation of the average causal effects of the treatment strategies in the presence of time-dependent confounding by indication. Causal inference under the potential outcomes framework can be viewed from a missing data perspective. The article extends the penalized spline of propensity methods in handling missing data to causal inference incorporating a temporal element into consideration.

Under the sequential randomization assumption, Robins and his colleagues have proposed different approaches to handle time-varying confounding, such as g-computation (Robins and Greenland 1992) and marginal structural models (MSMs, Robins 2000) using (augmented) inverse probability weighting (A/IPW). However, g-computation is fully parametric and therefore is sensitive to model assumptions. Weighting requires the positivity assumption that the probability of receiving each treatment is strictly positive for all subjects. In many practices, the weights for some subjects can become extremely large, leading to both bias and large variability. In contrast with the predominantly propensity score weighting approaches, the authors have used the propensity score as the predictors in the outcome mean model in addition to the individual covariates. In particular, the authors adopt penalized splines for propensity scores to provide a flexible model for imputing time-varying potential outcomes; thus, the estimated treatment effects are likely immune to the misspecification of the propensity score when the imputation model is correctly specified. The proposed estimation method successfully avoids the drawbacks of the existing competitors: (i) it does not involve weighting by the inverse of the propensity score and therefore avoids the possibly large variability due to weighting, and (ii) it improves the robustness to model misspecification.

Structural nested models: We would like to bring to the authors' attention another class of structural models that has been proposed in the literature for a while to estimate the treatment effect in longitudinal observational studies, namely the structural nested models (SNMs). SNMs allow modeling time-varying treatment modification effects using the post baseline time-dependent covariates. For example, the structural nested mean model for a continuous outcome specifies the effect of treatments through $\gamma_{\bar{z}_t}(\bar{L}_t;\psi_0) = E(Y_t^{\bar{z}_t} - Y_t^{\bar{0}} \mid \bar{L}_t;\psi_0)$, where $\bar{L}_t = (\bar{X}_t, \bar{Z}_{t-1})$. The g-estimation calculates $H_t(\psi_0) = Y_t - \gamma_{\bar{Z}_t}(\bar{L}_t;\psi_0)$ that mimics the potential outcome $Y_t^{\bar{0}}$, in the sense that $E\{H_t(\psi_0) \mid \bar{L}_t\} = E(Y_t^{\bar{0}} \mid \bar{L}_t)$. As a result, by the sequential moralization assumption, we have $E\{H_t(\psi_0) \mid \bar{L}_t, Z_t\} = E\{H_t(\psi_0) \mid \bar{L}_t\}$, which serves the basis for construction unbiased estimating equations. A general class of estimating functions for ψ_0 is

$$G(\psi_0) = \sum_{t=1}^{T} g(\bar{L}_t)[H_t(\psi_0) - E\{H_t(\psi_0) \mid \bar{L}_t;\beta_0\}] \times \{Z_t - P(Z_t = 1 \mid \bar{L}_t;\alpha_0)\}, \quad (1)$$

for any $g(\bar{L}_t)$. This framework shares the same appealing properties (i) and (ii). Specially, the propensity score enters the estimating equation not in a form of inverse weights. Moreover, $G(\psi_0)$ is unbiased of zero if either $E\{H_t(\psi_0) \mid \bar{L}_t;\beta_0\}$ or $P(Z_t = 1 \mid \bar{L}_t;\alpha_0)$ is correctly specified, but not necessarily both. Although SNMs have substantial promise, their applications in applied research are still relatively unpopular (Vansteelandt and Joffe 2014), partly because of its typically strongly theoretical presentation and challenging implementation. For comparing the aforementioned methods, we continue with the simulation study with the two-time-point treatment scenarios specified in Table 2 in the paper. We obtain the g-estimator by solving the empirical version of (1) with the optimal form of $g(\bar{L}_t)$ in Yang and Lok (2016, 2018). The resulting g-estimator achieves the semiparametric efficiency bound (Bickel et al. 1993) under the treatment effect model and the sequential randomization assumption. Table 1 presents the simulation results from 1000 replicates for sample size of 500. Under Scenarios (A) and (B) when the propensity score model is correctly specified, the g-estimator offers more reduction of the mean square error compared to the proposed estimator, while under Scenario (C) when the propensity score model is misspecified and the prediction model is correctly specified, the g-estimator may produce larger mean square errors. We think this research direction is promising, especially when the time points increase that necessities modeling the treatment effects as raised in the discussion of the paper.

Spline versus kernel: To avoid the wrong imputation when the propensity score model is misspecified, the authors propose a penalized spline for the derived propensity scores. However, the

CONTACT Donglin Zeng dzeng@email.unc.edu Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599.

Table 1. Ratios of empirical RMSE from g-estimation to RMSE of PENCOMP under (A) correctly specified propensity and prediction models; (B) a correctly specified propensity score model only; (C) a correctly specified prediction model only.

		Δ_{11}			Δ_{10}			Δ_{01}		
		Low	Mod	High	Low	Mod	High	Low	Mod	High
		Linear outcome								
Ratio (×100)	A	54	55	57	52	48	41	55	57	58
	B	48	45	48	49	34	29	39	41	48
	C	65	137	214	274	371	389	180	130	120
		Nonlinear outcome								
Ratio (×100)	A	68	68	73	72	81	91	95	92	95
	B	53	50	47	57	63	46	53	85	135
	C	78	80	83	80	87	96	113	107	113

choice of the splines and knots can be a challenge in practice, especially when the estimated propensity scores are distributed unevenly, which is a common situation when some predictors are discrete. Alternatively, kernel based estimation can be useful to maintain nonparametric relationship between the imputed outcome and propensity scores. In other words, the imputation model can be built locally around each fixed propensity score, say x, where the local region is specified by assigning weights to each observed propensity score, say x_i, and the weights are determined by $a_n^{-1}K\{(x - x_i)/a_n\}$. The choice of $K(\cdot)$ includes the Gaussian kernel or Epanichikov kernel if the bounded support is desired. The bandwidth a_n for the kernel can be flexible to be different for different x so we allow the support of the propensity scores to have both continuous and discrete ranges. The kernel-based imputation was used in Zeng (2004) and Zeng and Chen (2010) for missing data. Furthermore, in these papers, the imputation for the missing outcomes was the mean imputation using a two-dimensional local kernel weights around not only the propensity score but also the prediction score derived from the posited model for the missing outcome. They showed that the derived estimators also possess double robust property and are demonstrated to be numerically more reliable than weighted estimators or their augmentation version. It would be interested that the proposed penalized spline approach can be compared with the kernel based approach.

Machine learning for prediction: The models for the propensity score and missing potential outcomes are parametric, although the latter is partially nonparametric for the included propensity score. These models are likely to be misspecified when the dimension of the observed confounders is large and the number of the stages increases. On the other hand, many algorithms from the machine learning field are powerful to produce fairly accurate prediction using complex and high-dimensional data. For example, random forest, support vector machine and deep learning algorithms are widely used for producing scores for predicting binary response (the treatment indicator in this paper) and continuous outcomes (the potential outcomes in this paper). It will be very interesting to see that the proposed method can incorporate different machine learning algorithms to improve the accuracy of propensity score estimation and imputation. Finally, for all these algorithms including the parametric models used in the paper, how to tune algorithms in machine learning or retain important predictors in the parametric models will remain to be a challenging issue, given that the eventual goal is to obtain a valid and precise estimate of the average treatment effects.

Summary: Many different modeling and algorithms exist for predicting propensity scores and missing outcomes, and different approaches can be adopted to incorporate the propensity score (weighting, covariate adjustment and g-estimation). It is natural to ask which model or method should be chosen for a typical application. We would like to hear the thoughts from the authors. We congratulate the authors again on this nice work.

Funding

Dr. Yang is partially supported by NSF grant DMS 1811245, NCI grant P01 CA142538, and (ORAU) Oak Ridge Associated Universities. Dr. Zeng is partially supported by NIH funding: NIH GM124104.

References

Bickel, P. J., Klaassen, C., Ritov, Y., and Wellner, J. (1993), *Efficient and Adaptive Inference in Semiparametric Models*, Baltimore: Johns Hopkins University Press. [31]

Robins, J. M. (2000), "Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, eds. M. E. Halloran and D. Berry, New York: Springer, pp. 95–133. [31]

Robins, J. M., and Greenland, S. (1992), "Identifiability and Exchangeability for Direct and Indirect Effects," *Epidemiology*, 3, 143–155. [31]

Vansteelandt, S., and Joffe, M. (2014), "Structural Nested Models and g-Estimation: The Partially Realized Promise," *Statistical Science*, 29, 707–731. [31]

Yang, S., and Lok, J. J. (2016), "A Goodness-of-Fit Test for Structural Nested Mean Models," *Biometrika*, 103, 734–741. [31]

——— (2018), "Sensitivity Analysis for Unmeasured Confounding in Coarse Structural Nested Mean Models," *Statistica Sinica*, 28, 1703–1723. [31]

Zeng, D. (2004), "Estimating Marginal Survival Function by Adjusting for Dependent Censoring Using Many Covariates," *The Annals of Statistics*, 32, 1533–1555. [32]

Zeng, D., and Chen, Q. (2010), "Adjustment for Informative Missingness Using Auxiliary Information in Semiparametric Regression, *Biometrics*, 66, 115–122. [32]