University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Measuring Treatment Effects Using Semiparametric Models

Zhuo Yu[*]   Mark J. van der Laan[†]

[*]Department of Statistics, University of California, Berkeley, zyu@stat.berkeley.edu

[†]Division of Biostatistics, School of Public Health, University of California, Berkeley, laan@berkeley.edu

Measuring Treatment Effects Using Semiparametric Models

Zhuo Yu and Mark J. van der Laan

Abstract

In order to estimate the causal effect of treatments on an outcome of interest, one has to account for the effect of confounding factors which covary with the treatments and also contribute to the outcome of interest. In this paper, we use the semiparametric regression model to estimate the causal parameters. We assume the causal effect of the treatments can be described by the parametric component of the semiparametric regression model. Following the general methodology which was developed in van der Laan and Robins (2002) we give the orthogonal complement of the nuisance tangent space which identifies all the estimating functions. The estimating function which leads to an estimator given in Newey (1990) is an element of our class of estimating functions. We also give the methods to estimate the influence curve and variance of the resulting estimate. Double protection property is discussed when the nuisance parameters are misspecified. The optimal estimating function or the efficient influence curve is obtained in closed form. A one-step estimator is suggested. If the nuisance parameters in the estimating function are correctly specified, then our estimate is efficient.

1 Introduction

Consider a study in which the goal is to estimate the effect of one or more exposures or treatments on an outcome of interest. Often treatments or exposures may depend on other confounding factors which also contribute to the outcome of interest. In order to estimate the causal effect of treatments or exposures, one has to account for these confounding factors. Typically, parametric regression methods which adjust for both treatment and covariates are used. For example, one assumes $Y = \alpha X + \beta Z + \epsilon$, where X is treatment dose, Z includes all the other confounding factors. Usually we don't have too much knowledge on how the confounding factors contribute to the outcome of interest. In that case, it is more reasonable to adjust for Z additively but nonparametrically. In other words, we assume

$$Y = m_0(X|\alpha) + m_1(Z) + \epsilon, \quad E(\epsilon|X, Z) = 0 \tag{1}$$

the conditional distribution of the error ϵ given X, Z is equal to zero, but has an otherwise non-restricted conditional distribution. Where $m_0(X|\alpha)$ is a known function up to an unknown $p \times 1$ vector α, $m_1(Z)$ is a unknown function of Z. For example, we can choose $m_0(X|\alpha) = \alpha X$. $m_1(Z)$ is an unknown function. Our main goal is to estimate the causal parameter α. A more general model is to also allow m_0 to depend on some subsets $V \in Z$. For example, m_0 could be the linear function of X, V and their interactions. That is,

$$Y = m_0(X, V|\alpha) + m_1(Z) + \epsilon \quad E(\epsilon|X, Z) = 0. \tag{2}$$

In our data analysis, we are concerned with estimating the effect of an activity score on the physical functionality score in a cohort of 1,197 female with age ≥ 55 in the SPARCS (Study of Physical Performance and Age Related Changes in Sonomans, Tager et al. [2000]) project. We will estimate this effect while accounting for the presence of five potential confounding factors that include age, past activity history, smoking history, body mass index, cardiovascular condition. In this example the treatment of interest (activity score) is dichotomous and we assume there is an interaction between treatment and age. That is, we assume that the absolute effect of activity score on physical functionality score depends on a subject's age. In this setting the commonly used approach to estimating the effect of activity score would be to postulate a linear regression model

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_{12} X_i Z_{1,i} + \sum_{k=2}^{K} \alpha_k Z_{k,i} + \epsilon_i, \ E(\epsilon_i|X_i, Z_i) = 0, \tag{3}$$

where $Y_i, X_i, Z_i = (Z_{1,i}, Z_{2,i}, \ldots, Z_{K,i})$ represent physical functionality score, activity score ($X_i = 1$ if engaging moderately vigorous levels of activity and $X_i = 0$

1

otherwise), and values on a vector Z_i of potential confounding factors. We shall assume (Y_i, X_i, Z_i) are independent and identically distributed random vectors.

Suppose we are unwilling to assume that the contribution of the confounders to the outcome Y_i has a parametric functional form. In that case we would generalize model (3) to

$$Y_i = \alpha_1 X_i + \alpha_{12} X_i Z_{1,i} + m_1(Z_i) + \epsilon, \ \ E(\epsilon_i | X_i, Z_i) = 0, \tag{4}$$

where $m_1(Z_i)$ is an unknown real-valued function of the vector Z_i.

To estimate α, we follow the general methodology of constructing estimating functions as presented in van der Laan and Robins [2002]. Consider an idealized experiment where the conditional distribution of treatment given the confounding factors is known because it is under the control of the investigator. For example, X is a simple dichotomous treatment which a subject receives, Z includes all other covariates of the subject. Suppose $P(X = 1|Z)$ satisfies the logistic regression model. That is, $P(X = 1|Z, \beta) = logit^{-1}(\beta_0 + \sum_{i=1}^{K} Z_i)$. In this case, the class of estimating functions turns out to be independent of $m_1(Z)$. The resulting estimate which is obtained by solving the corresponding estimating equation is asymptotically linear. The influence curve and variance of the estimate can be easily calculated. The estimating function which leads to Newey's estimate (1990) is an element of this class of estimating functions.

In case that $G(X|Z)$ is correctly specified, an estimate of $G(X|Z)$, $\hat{G}(X|Z)$, could be used to replace $G(X|Z)$. Since Z is typically high dimensional, nonparametric estimation of G is often not practical. Suppose we assume a parametric or semiparametric model for G. If G is estimated efficiently, we show that the influence curve of the resulting estimate is the influence curve of the estimate which uses the true G projected on the orthogonal complement of the tangent space of the model we choose for G. Thus this estimate is more efficient than the estimate which uses true G. This result can be viewed as a generalization of Theorem A.1, Robins et al. [1992].

When $G(X|Z)$ is completely unspecified, we showed that the class of estimating functions depends on $m_1(Z)$. But the estimating function is unbiased if either $m_1(Z)$ or G is correctly specified. This property is especially attractive in case that we have more precise knowledge of the conditional distribution of X given Z than that of $m_1(Z)$.

Finally we establish a closed form formula for the Optimal estimating function is obtained in closed form. A one-step estimator is suggested. If the nuisance parameters in the estimating function are estimated in appropriate rate, the one-step estimator is efficient under some regularity conditions.

Newey [1995] provides an unbiased estimator for α by modelling G. Methods of confounder control based on an estimated G have also been introduced in Rosenbaum and Rubin [1983]. Robins et al. [1992], Robins and Rotnitzky [2001a] and Robins and Rotnitzky [2001b] propose a class of estimating functions which

2

generalized Newey's method. The current paper provides the optimal estimating function. We carry out a simulation study to show that the proposed estimator is more efficient than Newey and Robins' estimators. We apply the methodology to the SPARCS data to evaluate the activity scores on the physical functioning.

2 The main Theorems

In this section, we provide two theorems which are the basis for our proposed estimators. The proofs are provided in Appendix. (1) and (2) in Theorem 2.1 were also proved in Robins and Rotnitzky [2001b]. We also present them in the Theorem for completeness.

Theorem 2.1. *Assume that $O \equiv (Y, X, Z)$ satisfies (2), where $m_0(X|\alpha)$ is a known function up to an unknown p-dimensional vector α, $m_1(\cdot)$ is an unknown function. The likelihood is given by*

$$L(O) = f_{\epsilon|X,Z}(Y - m_0(X, V|\alpha) - m_1(Z)|X, Z) \times f_{X,Z}(X, Z), \qquad (5)$$

where α is the parameter of interest, $m_1(\cdot)$, $f_{X,Z}(\cdot, \cdot)$ and $f_{\epsilon|X,Z}(\cdot|X, Z)$ are the nuisance parameters. We have

(1) If the conditional distribution of X given Z, $G(X|Z)$, is known, then the orthogonal complement of nuisance tangent space of α in (2) is given by

$$
\begin{aligned}
T_{nuis}^{\perp} &= \left\{ (Y - m_0(X, V|\alpha))h_1(X, Z) + h_2(X, Z) : E(h_i|Z) = 0 \right\} \\
&= \left\{ (Y - m_0(X, V|\alpha))\left(h_1 - E(h_1|Z)\right) + (h_2 - E(h_2|Z)) : h_1, h_2 \right\},
\end{aligned}
$$

where $h_i \equiv (h_{i1}, \ldots, h_{ip})$, $i = 1, 2$ are $p \times 1$ vector valued functions.

(2) If $G(X|Z)$ is completely unspecified, then

$$T_{nuis}^{\perp} = \left\{ (Y - m_0(X|\alpha) - m_1(Z))\left(h(X, Z) - E(h(X, Z)|Z)\right) : h \right\}. \quad (6)$$

(3) For both the case where G is known and unknown, the efficient score for α is

$$IC_{eff}(O|\alpha, G, m_1, w) = (Y - m_0(X, V|\alpha) - m_1(Z)) \times h_{opt}(X, Z), \quad (7)$$

where

$$h_{opt}(X, Z) = u(X, Z) - \frac{E(u(X, Z)|Z)}{E(w(X, Z)|Z)} \times w(X, Z).$$

and $u(X, Z) = d/d\alpha\, m_0(X, V|\alpha) \times w(X, Z)$, $w^{-1}(X, Z) \equiv Var(\epsilon|X, Z)$. Thus, if $Var(\epsilon|X, Z)$ is a constant, we have

$$h_{opt}(X, Z) = \frac{d}{d\alpha} m_0(X, V|\alpha) - E\left(\frac{d}{d\alpha} m_0(X, V|\alpha)|Z\right). \qquad (8)$$

3

If $Var(\epsilon|X,Z) = w^{-1}(Z)$ only depends on Z, we have

$$h_{opt}(X,Z) = w(Z)\left(\frac{d}{d\alpha}m_0(X,V|\alpha) - E(\frac{d}{d\alpha}m_0(X,V|\alpha)|Z)\right). \quad (9)$$

Theorem 2.2. *If G is completely unspecified, then for any $IC(O|\alpha,G,m_1,h) \in T_{nuis}^{\perp}$, where*

$$IC(O|\alpha,G,m_1,h) \equiv (Y - m_0(X|\alpha) - m_1(Z))\big(h(X,Z) - E(h(X,Z)|Z),$$

we have

$$EIC(O|\alpha,G',m_1',h) = 0, \text{ if either } G' = G \text{ or } m_1' = m_1. \quad (10)$$

In particular

$$EIC_{eff}(O|\alpha,G',m_1',w) = 0, \text{ if either } G' = G \text{ or } m_1' = m_1. \quad (11)$$

3 Estimation

According to van der Laan and Robins [2002], the orthogonal complement of the nuisance tangent space for α identifies all the estimating functions. In this section, we discuss the estimation of α based on Theorem 2.1. Section 3.1 discusses the situation where G is known. Section 3.2 discusses the situation where G is correctly modelled. Section 3.3 discusses the situation where G is completely unspecified. Section 3.4 provides the optimal estimating equation.

3.1 Estimating α when G is known

When G is known, the class of estimating functions is given by (1) in Theorem 2.1. Given any estimating function

$$IC(O|\alpha,G,h) = (Y - m_0(X,V|\alpha))(h_1 - E(h_1|Z)) + (h_2 - E(h_2|Z))$$

The corresponding estimating equation is

$$\frac{1}{n}\sum_{i=1}^{n} IC(O_i|\alpha,G,h) = 0. \quad (12)$$

Under some regularity conditions the solution of the above estimating equation, α_n^0, is asymptotically linear with influence curve

$$IC(O|\alpha,G,h,c) = c^{-1} \times IC(O|\alpha,G,h). \quad (13)$$

4

where $c = d/d\alpha EIC(O|\alpha, G, h)\big|_{\alpha=\alpha(F_O)}$. That is,

$$\alpha_n^0 - \alpha = \sum_{i=1}^{n} IC(O_i|\alpha, G, h, c) + o_P(\frac{1}{\sqrt{n}}) \tag{14}$$

Example 1: Assume $m_0(X|\alpha) = \alpha X$. Let $h_1(X, Z) = X, h_2(X, Z) = 0$. Then the corresponding estimating equation is

$$\sum_{i=1}^{n}(Y_i - \alpha X_i)(X_i - E(X_i|Z_i)) = 0. \tag{15}$$

The solution of the estimating equation is

$$\alpha_n^0 = \frac{\sum_{i=1}^{n} Y_i(X_i - E(X_i|Z_i))}{\sum_{i=1}^{n} X_i(X_i - E(X_i|Z_i))}. \tag{16}$$

The last estimate is also suggested by Newey [1995], Robins et al. [1992]. By (13), it is easy to show that the influence curve of the last estimator is

$$IC(O|\alpha, G, c) = c^{-1}(Y - \alpha X)(X - E(X|Z)), \tag{17}$$

where $c = Var(X|Z)$. The asymptotic variance of α_n^0 can be easily estimated by

$$\widehat{Var}(\alpha_n^0) = \frac{1}{n}\sum_{i=1}^{n} IC\left(O_i|\alpha_n^0, c\right)^2, \tag{18}$$

Example 2: Assume $m_0(X|\alpha) = \alpha_1 X + \alpha_2 XV$, where V is a subset of Z. $h_1 \equiv (X, XV), h_2 \equiv 0$. Then the corresponding estimating equation is

$$\sum_{i=1}^{n}(Y_i - \alpha_1 X_i - \alpha_2 X_i V_i)(X_i - E(X_i|Z_i)) = 0 \tag{19}$$

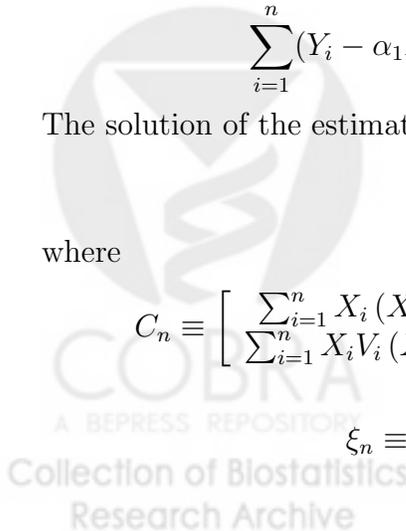$$\sum_{i=1}^{n}(Y_i - \alpha_1 X_i - \alpha_2 X_i V_i)(X_i V_i - V_i E(X_i|Z_i)) = 0. \tag{20}$$

The solution of the estimating equation is

$$\alpha_n^0 = C_n^{-1}\xi_n, \tag{21}$$

where

$$C_n \equiv \left[\begin{array}{cc} \sum_{i=1}^{n} X_i\left(X_i - E(X_i|Z_i)\right) & \sum_{i=1}^{n} X_i V_i\left(X_i - E(X_i|Z_i)\right) \\ \sum_{i=1}^{n} X_i V_i\left(X_i - E(X_i|Z_i)\right) & \sum_{i=1}^{n} X_i V_i^2\left(X_i - E(X_i|Z_i)\right) \end{array} \right]$$

$$\xi_n \equiv \left[\begin{array}{c} \sum_{i=1}^{n} Y_i(X_i - E(X_i|Z_i)) \\ \sum_{i=1}^{n} Y_i V_i(X_i - E(X_i|Z_i)) \end{array} \right]$$

5

3.2 Estimating α when G is correctly modelled

In case that G is unknown, the estimating equation (12) is not feasible since it depends on G. It is natural to replace G by an estimate of G, G_n. That is, one estimates α with solution $\tilde{\alpha}_n^0$ of

$$\frac{1}{n}\sum_{i=1}^{n} IC(O_i|\alpha, G_n, h) = 0, \qquad (22)$$

Example 1: (continued) Estimating equation (15) becomes

$$\sum_{i=1}^{n}(Y_i - \alpha X_i)(X_i - E_{G_n}(X_i|Z_i)) = 0. \qquad (23)$$

The solution of the estimating equation is

$$\tilde{\alpha}_n^0 = \frac{\sum_{i=1}^{n} Y_i(X_i - E_{G_n}(X_i|Z_i))}{\sum_{i=1}^{n} X_i(X_i - E_{G_n}(X_i|Z_i))} \qquad (24)$$

The above estimate was provided in Robins et al. [1992]. Assume that X is a dichotomous treatment and the conditional distribution of X given Z satisfies the linear logistic regression model. If the parameters in the logistic regression are efficiently estimated, Robins et al. [1992] proved that the estimate is asymptotically normal. They also gave a consistent estimate of the asymptotic covariance matrix. The next result generalizes Theorem A.1 in Robins et al. [1992]. This result is an analogue of Theorem 2.4 in van der Laan and Robins [2002].

Theorem 3.1. *Suppose $\alpha_n(G)$ is an estimator of α in the model with the conditional distribution $G(X|Z)$ known. The influence curve of $\alpha_n(G)$ is $IC(O|\alpha, G, h, c)$ as in (17). Let G_n be an estimate of G based on a model for G with Tangent space T_G. Assume that*

$$\alpha_n(G_n) - \alpha = \alpha_n(G) - \alpha + \Phi(G_n) - \Phi(G) + o_P(n^{-\frac{1}{2}}). \qquad (25)$$

for some functional Φ of G_n. Further assume that $\Phi(G_n)$ is an asymptotically efficient estimator of $\Phi(G)$. Then $\alpha_n(G_n)$ is asymptotically linear with influence curve

$$IC_2(O) = IC(O|\alpha, G, h, c) - \Pi(IC(O|\alpha, G, h, c)|T_G). \qquad (26)$$

Proof: By assumption, $\Phi(G_n)$ is an efficient estimate of $\Phi(G)$. Suppose the influence curve of $\Phi(G_n)$ is $IC_{nu}(O)$. By (25), $\alpha_n(G_n)$ is asymptotically linear with influence curve

$$IC_2(O) = IC(O|\alpha, G, h, c) + IC_{nu}(O). \qquad (27)$$

6

Thus
$$\Pi\left(IC_2(O)|T_G\right) = \Pi\left(IC(O|\alpha, G, h, c)|T_G\right) + \Pi\left(IC_{nu}(O)|T_G\right). \qquad (28)$$

Since $IC_2(O)$ is the influence curve of a asymptotically linear estimator, $IC_2(O)$ is orthogonal to the nuisance tangent space. T_G is a subspace of the nuisance tangent space implies that $\Pi(IC_2(O)|T_G) = 0$. But $\Phi(G_n)$ is an efficient estimate of $\Phi(G)$ under the model of G whose tangent space is T_G implies that $IC_{nu}(O) \in T_G$. By (27), $IC_{nu}(O) = -\Pi(IC(O|\alpha, G, h, c)|T_G)$ and the desired result follows. $\square$

The last result teaches us that if G_n is estimated efficiently, the variance of the influence curve for $\alpha_n(G_n)$ is smaller than that of the influence curve for $\alpha_n(G)$. Thus even in case that we know the conditional distribution $G(X|Z)$, if we use G_n instead of G, the resulting estimator will be more efficient.

Specifically, if we choose a parametric model for $G(X|Z)$ with parameter $\beta = (\beta_0, \ldots, \beta_J)$, the efficient estimate of G_n is often available as the maximum likelihood estimator. The last result applies directly. In this case T_2 is just the linear span of $S_\beta = (S_{\beta_0}, S_{\beta_1}, \ldots, S_{\beta_J})$, where S_β is the score function for β. The influence curve of $\alpha(G_n)$, by the last Theorem, is given by

$$IC_2(O) = IC(O|\alpha, G, h, c) - <IC, S_\beta^T> <S_\beta, S_\beta^T>^{-1} S_\beta. \qquad (29)$$

Example 1: (continued) Now we discuss how to estimate the influence curve for $\tilde{\alpha}_n^0$ and its variance. By Theorem 1, estimation of $IC_2(O)$ involves estimation of $IC(O|\alpha, G, c)$ and $\Pi(IC(O|\alpha, G, c)|T_2)$. $IC(O|\alpha, G, c)$ is given by (17). It can be estimated by

$$\widehat{IC}(O_i) \equiv IC(O_i|\tilde{\alpha}_n^0, G_n, c_n) = c_n^{-1}(Y_i - \tilde{\alpha}_n^0 X_i)\left(X_i - E_{G_n}(X_i|Z_i)\right), \qquad (30)$$

where $c_n = 1/n \sum_{i=1}^n (X_i - E_{G_n}(X_i|Z_i))^2$. We have

$$\Pi(IC(O)|T_2) = <IC, S_\beta^T> <S_\beta, S_\beta^T>^{-1} S_\beta(Z) \equiv RQ^{-1}S_\beta, \qquad (31)$$

where $S_\beta(Z)$ is the score function for β in the logistic regression. Let $\hat{\beta}$ be the maximum likelihood estimate of β. The jth component of R, R_j, can be consistently estimated by

$$\hat{R}_j = \frac{1}{n} \sum_{i=1}^n \widehat{IC}(O_i) S_{\hat{\beta}_j}(Z_i).$$

The (i, j) component of Q can be consistently estimated by

$$\hat{Q}_{ij} = \frac{1}{n} \sum_{k=1}^n S_{\hat{\beta}_i}(Z_k) S_{\hat{\beta}_j}(Z_k).$$

7

So the influence curve $IC_2(O_i)$ can be estimated by

$$\widehat{IC}_2(O_i) = \widehat{IC}(O_i) - \hat{R}\hat{Q}^{-1}S_{\hat{\beta}}(Z_i). \tag{32}$$

The asymptotic variance of $\tilde{\alpha}_n^0$ can be estimated by

$$\widehat{Var}(\tilde{\alpha}_n^0) = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{IC}(O_i)\right)^2 - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{R}\hat{Q}^{-1}S_{\hat{\beta}}(Z_i)\right)^2. \tag{33}$$

3.3   Estimating functions when G is unspecified

When G is unknown, the class of estimating functions is given by (2) in Theorem 2.1. In comparison with the case where G is known, we note that the class of estimating functions now depends on both of the nuisance parameters m_1 and G.

By Theorem 2.2, if $m_1(\cdot)$ is misspecified, but G is correctly specified, the estimating function remains unbiased for α. So the resulting estimate of α is consistent under some regularity conditions. This property is helpful especially in the case that we have a good knowledge about G. For any estimating function $IC(O|\alpha, G, m_1, h) \in T_{nuis}^{\perp}$, the corresponding estimating equation is

$$\sum IC(O_i|\alpha, G_n, m_{1n}, h) = 0, \tag{34}$$

where G_n and m_{1n} are estimates of G and m_1 respectively. If G_n is an efficient estimate of G under a model whose tangent space is T_G and $m_{1n} \to m_1'$. Under some regularity conditions, the solution of (34), $\hat{\alpha}_n$, is asymptotically linear with influence curve

$$IC(O) = (IC(O|\alpha, G, m_1', h, c) - \Pi(IC(O|\alpha, G, m_1', h, c)|T_G)), \tag{35}$$

where

$$IC(O|\alpha, G, m_1', h, c) = c^{-1}IC(O|\alpha, G, m_1', h)$$

and $c = d/d\alpha E(IC(O|\alpha, G, m_1', h))$. The asymptotic covariance of the estimate can be estimated by

$$\hat{\Sigma} = \frac{1}{n}\sum_i \hat{IC}(O_i)^{\otimes 2} \tag{36}$$

Since it is impractical to take the projection in equation (35), we can estimate the asymptotic covariance of the $\hat{\alpha}_n$ conservatively by

$$\hat{\Sigma} = \frac{1}{n}\sum_i IC(O_i|\alpha, G_n, m_{1n}, h, c_n)^{\otimes 2} \tag{37}$$

Example 2:   Assume $m_0(X|\alpha) = \alpha X$. Let $h(X, Z) = X$. $IC(O|\alpha, G, m_1) = (Y - \alpha X - m_1(Z))(X - E(X|Z)) \in T_{nuis}^{\perp}$. In case that the varince of ϵ given

8

X, Z is a constant, $IC(O|\alpha, G, m_1)$ actually gives the optimal estimating function for estimating α (See Newey [1995], Robins et al. [1992]). The corresponding estimating equation is

$$\sum_{i=1}^{n}(Y_i - \alpha X_i - m_{1n}(Z_i))(X_i - E_{G_n}(X_i|Z_i)) = 0, \tag{38}$$

where $m_{1n}(Z_i)$ is an estimate of $m_1(Z)$. Practically, we can estimate $m_1(Z)$ as follows: suppose $m_1(Z)$ is additive, that is, $m_1(Z) = m_1^1(Z_1) + m_1^2(Z_2) + \ldots + m_1^J(Z_J)$. Regress $Y_i - \alpha_n^0 X_i$ on Z_i using the nonparametric additive model. This can be implemented with the function `gam()` in Splus. The solution of the above estimating equation is

$$\hat{\alpha}_n = \frac{\sum_{i=1}^{n}(Y_i - m_{1n}(Z_i))(X_i - E_{G_n}(X_i|Z_i))}{\sum_{i=1}^{n} X_i(X_i - E_{G_n}(X_i|Z_i))}. \tag{39}$$

The asymptotic variance of $\hat{\alpha}_n$ can be estimated conservatively by

$$\widehat{\Sigma} = \frac{1}{n}\sum_i IC(O_i|\hat{\alpha}_n, G_n, m_{1n}, c_n)^2,$$

where $c_n = 1/n \sum_i (X_i - E_{G_n}(X_i|Z_i))^2$.

3.4 The Optimal estimating equation

The optimal estimating function, the efficient score of α, is given by (3) in Theorem 2.1. Given the optimal estimating function, the corresponding estimating equation is

$$\sum_{i=1}^{n}(Y_i - m_0(X_i, V_i|\alpha) - m_{1n}(Z_i)) \times h_n(X_i, Z_i) = 0, \tag{40}$$

where m_{1n} and h_n are estimates of m_1 and h_{opt}. Next we discuss how to estimate m_1 and h_{opt} in turn.

m_1. Given a initial estiamte α_n^0 , regress $Y - m_0(X|\alpha_n^0)$ on Z based on generalized linear regression or generalized additive model.

h_{opt}. We can estimate $v(X, Z) = w^{-1}(X, Z)$ by regressing the squares of residuals on X and Z. Note that we can calculate $h_{opt}(X, Z)$ based on estimated $w(X, Z)$ and $G(X|Z)$.

We propose the following one-step estimator based on the estimating equation (40). Let α_n^0 be an initial estimate.

$$\alpha_n^1 = \alpha_n^0 + c_n^{-1}\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - m_0(X_i|\alpha_n^0) - m_{1n}(Z_i)\right) \times h_n(X_i, Z_i),$$

9

	Mean	Variance	MSE	Rel. Eff.
$\hat{\alpha}_1$	0.0524	0.6886	0.6893	1.0000
$\hat{\alpha}_2$	0.0028	0.5415	0.5442	0.7793
$\hat{\alpha}_4$	0.0194	0.2493	0.2501	0.3582
$\hat{\alpha}_3$	0.0211	0.2687	0.2696	0.3861
$\hat{\alpha}_5$	0.0146	0.2695	0.2707	0.3877
$\hat{\alpha}_6$	0.0086	0.2723	0.2736	0.3918

Table 1: The relative efficiency of all the estimates

where

$$c_n = \frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\alpha} m_0(X_i|\alpha)|_{\alpha=\alpha_n^0} h_n(X_i, Z_i).$$

If $m_0(X|\alpha) = \alpha X$, the solution of the estimating equation is

$$\alpha_n^1 = \frac{\sum_{i=1}^{n} (Y_i - m_{1n}(Z)) h_n(X, Z)}{\sum_{i=1}^{n} X_i h_n(X, Z)}.$$

If the nuisance parameters G, $m_1(Z)$ and $w(X, Z)$ are correctly specified and consistently estimated, then the resulting estimate is efficient. If either G or $m_1(Z)$ is correctly specified, then the optimal estimating function given by (7) is unbiased. Thus the resulting estimate will be consistent if either G or $m_1(Z)$ is consistently estimated.

4   Simulation

In the simulation study, we choose $m_0(X|\alpha) = \alpha X$ ($\alpha = 10$) and $m_1(Z) = 6 + 3Z_1 + 2Z_1^2 + 3Z_2^2$. Z_1, Z_2 are sampled from exponential distribution with mean one. Given X, Z, ϵ is sampled from $N(0, SD = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2)$, where $((\gamma_0, \gamma_1, \gamma_2) = (1.5, 1.5, 1.5))$. Given Z, X is sampled from Binomial$(1, p = logit^{-1}(\beta_0 + \beta_1 Z_1 + \beta_2 Z_2))$, where $(\beta_0, \beta_1, \beta_2) = (0.015, 0.02, 0.015)$. The simulation results are based on 200 repetitions. The finite sample Bias, variance and MSE are summarized in Table 1.

We have that $\hat{\alpha}_1$ is based on Newey's method defined by (24), $\hat{\alpha}_2$ is based on Robins approach which was obtained by solving estimating equation (38), $\hat{\alpha}_3$ - $\hat{\alpha}_6$ are all obtained by solving estimating equation (40). For $\hat{\alpha}_3$, we use the true nuisance parameters in the estimating equation. For $\hat{\alpha}_4$, all the nuisance parameters are estimated with correctly specified models (say, we regress X on Z using linear logistic regression). For $\hat{\alpha}_5$, we correctly specify G and estimate the other nuisance parameters using a nonparametric additive model. For $\hat{\alpha}_6$, we estimate all the nuisance parameters nonparametrically.

	LS	Newey	Robins	Eff.
Treat	-0.211	-0.126	-0.199	-0.069
	(0.114)		(0.115)	(0.113)
Treat:Age	0.0033	0.0021	0.0031	0.0013
	(0.0016)		(0.0017)	(0.0016)

Table 2: Results of different estimates

From the simulation results, we can see that the finite sample biases of $\hat{\alpha}_3$ - $\hat{\alpha}_6$ are much smaller than those of the Newey and Robins's estimates. The finite sample MSEs of $\hat{\alpha}_4$ - $\hat{\alpha}_6$ are very close to those of $\hat{\alpha}_3$ which is obtained by plugging in the true nuisance parameters.

5 Data Analyses

The data in this analyses is from the SPARCS (Study of Physical Performance and Age Related Changes in Sonomans) project. SPARCS is a community based longitudinal study of physical activity and fitness in people ≥ 55 years of age who live in the city and environs of Sonoma, California. The objective of the current analysis is to estimate the causal effects of activity scores on physical functional score (NRB). The range of the activity score is $\{1, 2, 3, 4\}$. Let $Moderate \equiv I(GSM\_ORD \geq 3)$ be the indicator of engaging in moderately vigorous levels of activity based on GSM_ORD. $X = Moderate$ is used as treatment variable in our analyses. $Y = NRB$ is the outcome of interest. Other variables $Z = (Age, Habitual, TeamDecline, Bmass)$ are considered to be confounders which predict both the Treatment (Activity score) and the outcome. Let $V = Age$. We assume $m_0(X, V) = \alpha_1 X + \alpha_2 XV$.

As we see in Section 3.4, we need to estimate the conditional distribution of X given Z and the conditional variance of ϵ given (X, Z) to solve the optimal estimating equation (40). We use a linear logistic regression to model the conditional distribution of X given Z. We use a linear regression to model the residuals on (X, Z) to obtain the h_{opt} given in Section 3.4. The following table summarizes the coefficient estimates of treatment and interaction of treatment and age based on different methods (standard errors are reported in the parenthesis). The first column is the LS estimate. The second column is Newey's estimate (24). The 3rd column is Robins' estimate obtained by solving estimating equation (38). The last colum is obtained by solving the optimal estimating equation (40).

11

References

W.K. Newey. Semiparametric efficiency bounds. *Journal of applied econometrics*, 1(4):335–341, 1995. ISSN 1350-7265.

J.M. Robins, S.D Mark, and W.K. Newey. Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48: 479–495, 1992.

J.M. Robins and A. Rotnitzky. Comment on "inference for semiparametric models: some questions and an answer" by bickel, j. and kwon, jaimyoung. *Statistica Sinica*, 11:920–935, 2001a.

J.M. Robins and A. Rotnitzky. Double robustness in statistical models. *Scandinavian Journal of Statistics*, To appear, 2001b.

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

I. Tager, T. Haight, Hollenberg, and Satariano. Physical functioning and mortality in elderly females. Unpublished manual, University of California, Berkeley, School of Public Health., 2000.

M.J. van der Laan and J. M. Robins. *Unified methods for censored longitudinal data and causality.* Springer-Verlag, New York, 2002.

Appendix

Proof of Theorem 2.1: We first prove (2). The joint density of (Y,X,Z) is $f_{\epsilon|X,Z}(Y - m_0(X, V|\alpha) - m_1(Z)|X, Z) \times f_{X,Z}(X, Z)$, where $f_{\epsilon|X,Z}(\cdot|X, Z)$ and $f_{X,Z}(\cdot, \cdot)$ denote the conditional density of ϵ given X, Z and joint density of X, Z respectively. There are three nuisance parameters: m_1, $f_{\epsilon|X,Z}(\cdot|X, Z)$ and $f_{X,Z}(\cdot, \cdot)$. We have $T_{nuis} = T_1 \oplus T_2$, where T_1 is the tangent space for $f_{\epsilon|X,Z}(Y - m_0(X, V|\alpha) - m_1(Z)|X, Z)$ and T_2 is the tangent space for $f_{X,Z}(X, Z)$. Since we did not assume anything about the joint density $f_{X,Z}(\cdot, \cdot)$,

$$T_2 = L_0^2(X, Z) \equiv \left\{ h(X, Z) : Eh(X, Z) = 0, Var(h(X, Z)) < \infty \right\}.$$

We yet need to calculate T_1. Let T_{11} be the tangent space of $f_{\epsilon|X,Z}(Y - m_0(X, V|\alpha) - m_1(Z)|X, Z)$ for fixed m_1 (i.e., tangent space obtained by varying $f_{\epsilon|X,Z}(\cdot|X, Z)$) and T_{12} be the tangent space of $f_{\epsilon|X,Z}(Y - m_0(X, V|\alpha) - m_1(Z))$

for fixed $f_{\epsilon|X,Z}(\cdot|X,Z)$ (i.e., tangent space obtained by varyint m_1). Consider one dimensional model $m_1^\delta(\cdot) = m_1(\cdot) + \delta h(\cdot)$, We have

$$T_{11} = \left\{ \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} h(Z) : h(Z) \right\}. \tag{41}$$

Consider one-dimensional model $f_{\epsilon|X,Z}^\delta(\cdot|X,Z) = (1 + \delta h(\cdot|X,Z)) f_{\epsilon|X,Z}(\cdot|X,Z)$, where $h(\cdot|X,Z)$ satisfies $E\big(h(\epsilon|X,Z)|X,Z\big) = E\big(\epsilon h(\epsilon|X,Z)|X,Z\big) = 0$, We have

$$T_{12} = \left\{ h(\epsilon|X,Z) : E\big(h(\epsilon|X,Z)|X,Z\big) = E\big(\epsilon h(\epsilon|X,Z)|X,Z\big) = 0 \right\}. \tag{42}$$

It is known that $T_1 = T_{11} + T_{12}$. So $T_{nuis} = T_1 \oplus T_2 = T_{11} + (T_{12} \oplus T_2)$. It is straightforward to show $T_{nuis}^\perp \supset \big\{ \epsilon\big(h(X,Z) - E(h|Z)\big) : h(X,Z) \big\}$. Now we prove the other direction. We have $T_{nuis}^\perp = T_{11}^\perp \cap (T_{12} \oplus T_2)^\perp$. It is easily checked that $(T_{12} \oplus T_2)^\perp = \big\{ \epsilon h(X,Z) : h \big\}$. So for any $D \in T_{nuis}^\perp$, $D = \epsilon h_1(X,Z)$ for some h_1 and $D \perp T_{11}$. Thus,

$$
\begin{aligned}
0 &= < \epsilon h_1(X,Z), \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} h_2(Z) > \qquad \text{for any } h_2(\cdot) \\
&= E\left(h_1(X,Z) h_2(Z) E\left(\epsilon \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} | X,Z \right) \right) \\
&= E\left(h_2(Z) E\left(h_1(X,Z)|Z \right) \right).
\end{aligned}
$$

The second equality uses the fact that $\int t f'(t) dt = 1$ for any density function f with mean zero. The last equality is true for any h_2 implies $E(h_1(X,Z)|Z) = 0$. We thus conclude $T_{nuis}^\perp \subset \big\{ \epsilon\big(h(X,Z) - E(h|Z)\big) : h(X,Z) \big\}$ and the desired result follows.

Assume that $G(X|Z)$ is known. The nuisance tangent space in this case is $T_{nuis} = T_1 \oplus L_0^2(Z) = T_{11} + (T_{12} \oplus L_0^2(Z))$. So $T_{nuis}^\perp = T_{11}^\perp \cap (T_{12} \oplus L_0^2(Z))^\perp$. We first show

$$(T_{12} \oplus L_0^2(Z))^\perp = \big\{ \epsilon h_1(X,Z) + \big(h_2(X,Z) - E(h_2|Z) \big) : h_1, h_2 \big\}. \tag{43}$$

It is straightforward to check that "$\supset$" holds. The "$\subset$" part follows by the following projection formula. For any $V \in L_0^2(Y,X,Z)$,

$$
\begin{aligned}
\Pi(V|(T_{12} \oplus L_0^2(Z))^\perp) &= V - \Pi(V|T_{12}) - \Pi(V|L_0^2(Z)) \\
&= E(V|X,Z) - E(V|Z) + \frac{E\big((V - E(V|X,Z))\epsilon|X,Z\big)}{Var(\epsilon|X,Z)} \epsilon.
\end{aligned}
$$

The above calculation uses the projection formula $\Pi(V|L_0^2(Z)) = E(V|Z)$ and
$\Pi(V|T_{12}) = V - E(V|X,Z) - \dfrac{E\big((V - E(V|X,Z))\epsilon|X,Z\big)}{Var(\epsilon|X,Z)} \epsilon.$

13

Next we will prove (1). We need to show

$$T_{nuis}^{\perp} = \left\{ \epsilon\big(h_1(X,Z) - E(h_1|Z)\big) + \big(h_2(X,Z) - E(h_2|Z)\big) : h_1, h_2 \right\}. \qquad (44)$$

Again "⊃" part is easy. Let's prove the other direction. Given $D \in T_{nuis}^{\perp} = T_{11}^{\perp} \cap (T_{12} \oplus L_0^2(Z))^{\perp}$. By (43), $D = \epsilon h_1(X,Z) + \big(h_2(X,Z) - E(h_2|Z)\big)$, for some h_1, h_2 and $D \perp T_{11}$. Thus

$$< \epsilon h_1(X,Z) + \big(h_2(X,Z) - E(h_2|Z)\big), \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} h_3(Z) >= 0, \text{ for any } h_3(Z).$$
$$(45)$$

Using the fact that $\int f'(t)dt = 0$, it is easy to show

$$< \big(h_2(X,Z) - E(h_2(X,Z)|Z)\big), \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} h_3(Z) >= 0. \qquad (46)$$

So

$$< \epsilon h_1(X,Z), \frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} h_3(Z) >= 0, \text{ for any } h_3(Z). \qquad (47)$$

By the same argument as before we conclude $E\left(h_1(X,Z)|Z\right) = 0$. Thus $D \in \left\{ \epsilon\big(h_1(X,Z) - E(h_1|Z)\big) + \big(h_2(X,Z) - E(h_2|Z)\big) : h_1, h_2 \right\}$ and the desired result follows.

Proof of (3): The score for α is

$$S_{\alpha} = -\frac{f'_{\epsilon|X,Z}(\epsilon|X,Z)}{f_{\epsilon|X,Z}(\epsilon|X,Z)} \times \frac{d}{d\alpha} m_0(X|\alpha)$$

The rest of the proof is a straightforward calculation by (1), (2) of Theorem 2.1 and Lemma 5.1. □

Proof of Theorem 2.2: The proof is straightforward and omitted. □

Lemma 5.1. *Suppose the conditional distribution of X given Z is unspecified. For any $V \in L_0^2(Y, X, Z)$, the projection of V onto the orthogonal complement of the nuisance tangent space is given by*

$$\Pi(V|T_{nuis}^{\perp}) = \epsilon\left(u(X,Z) - \frac{E(u(X,Z)|Z)}{E\left(w(X,Z)|Z\right)} \times w(X,Z)\right), \qquad (48)$$

where $w^{-1}(X,Z) \equiv v(X,Z) \equiv Var(\epsilon|X,Z), u(X,Z) \equiv \dfrac{E(\epsilon V|X,Z)}{w(X,Z)}$. If the conditional distribution of X given Z, $G(X|Z)$, is known, then the projection of V onto the orthogonal complement of the nuisance tangent space is given by

$$\Pi(V|T_{nuis}^{\perp}) = \epsilon\left(u_1(X,Z) - \frac{E(u_1(X,Z)|Z)}{E\left(w(X,Z)|Z\right)} \times w(X,Z)\right)$$
$$+ (u_2(X,Z) - E(u_2(X,Z)|Z)),$$

where $u_1(X,Z) \equiv w(X,Z)E(\epsilon V|X,Z), u_2(X,Z) \equiv E(V - \epsilon u_1(X,Z)|X,Z)$.

14

Proof: If $G(X|Z)$ is unspecified, $T_{nuis}^{\perp} = \{\epsilon h(X, Z) : E(h(X, Z)|Z) = 0\} \subset H \equiv \{\epsilon h(X, Z) : h\}$. It is easy to check $\Pi(V|H) = \epsilon u(X, Z)$. Thus, $\Pi(V|T_{nuis}^{\perp}) = \Pi\left(\Pi(V|H)|T_{nuis}^{\perp}\right) = \Pi\left(\epsilon u(X, Z)|T_{nuis}^{\perp}\right)$. We need to find $\epsilon u_0(X, Z) \in T_{nuis}^{\perp}$ such that $E(\epsilon u(X, Z) - \epsilon u_0(X, Z))^2 = Ev(X, Z)(u(X, Z) - u_0(X, Z))^2$ is minimized. To show

$$u_0(X, Z) = u(X, Z) - \frac{E(u(X, Z)|Z)}{E(w(X, Z)|Z)} \times w(X, Z). \tag{49}$$

Consider Hilbert space: $\tilde{H} = \{h(X) : Ew(X)h^2(X) < \infty\}$ equipped with inner product $< f, g >_{\tilde{H}} \equiv Ev(X)f(X)g(X)$. Let $\tilde{H}_0 = \{h_0(X) \in \tilde{H} : Eh_0(X) = < \frac{1}{v}, h_0 >_{\tilde{H}} = 0\}$.

$$\begin{aligned}
\Pi(h|\tilde{H}_0) &= \Pi\left(h\Big|\left[\frac{1}{v}\right]^{\perp}\right) = h - \Pi\left(h\Big|\left[\frac{1}{v}\right]\right) \\
&= h - \frac{< h, \frac{1}{v} >_{\tilde{H}}}{< \frac{1}{v}, \frac{1}{v} >_{\tilde{H}}} \times \frac{1}{v} = h - \frac{Eh(X)}{E\frac{1}{v(X)}} \times \frac{1}{v(X)}.
\end{aligned}$$

If $G(X|Z)$ is specified, consider

$$\begin{aligned}
H &= \{\epsilon h_1(X, Z) + h_2(X, Z) : h_1, h_2\} \\
&\supset T_{nuis}^{\perp} = \{\epsilon h_1(X, Z) + h_2(X, Z) : E(h_1|Z) = 0, E(h_2|Z) = 0\}.
\end{aligned}$$

It is easy to show $\Pi(V|H) = \epsilon u_1(X, Z) + u_2(X, Z)$. To calculate $\Pi(\epsilon u_1(X, Z) + u_2(X, Z)|T_{nuis}^{\perp})$, we need to find $\epsilon u_{10}(X, Z) + u_{20}(X, Z) \in T_{nuis}^{\perp}$, such that

$$\begin{aligned}
&E\left(\epsilon u_1(X, Z) + u_2(X, Z) - \epsilon u_{10}(X, Z) - u_{20}(X, Z)\right)^2 \\
&= E\epsilon^2\left(u_1(X, Z) - u_{10}(X, Z)\right)^2 + E\left(u_2(X, Z) - u_{20}(X, Z)\right)^2.
\end{aligned}$$

Is minimized. (Note that the cross terms in the last formula disappear since $E(\epsilon|X, Z) = 0$). By the same argument as before, we have

$$u_{10} = u_1(X, Z) - \frac{E(u_1(X, Z)|Z)}{E(w(X, Z)|Z)} \times w(X, Z)$$

$$u_{20} = u_2(X, Z) - E(u_2(X, Z)|Z)$$

which conclude (2). $\square$

15