Ann Inst Stat Math (2011) 63:497–509
DOI 10.1007/s10463-009-0231-7

Representations of efficient score for coarse data problems based on Neumann series expansion

Hua Yun Chen

Received: 1 May 2008 / Revised: 14 October 2008 / Published online: 6 May 2009

Abstract We derive new representations of the efficient score for the coarse data problems based on Neumann series expansion. The representations can be applied to both ignorable and nonignorable coarse data. Approximations to the new representations may be used for computing locally efficient scores in such problems. We show that many of the successive approximation approaches to the computation of the locally efficient score proposed in the literature for such problems can be derived as special cases of the representations. In addition, the representations lead to new algorithms for computing the locally efficient scores for the coarse data problems.

Keywords Auxiliary variable · Adjoint operator · Coarsening at random · Nonparametric information operator · Projection

1 Introduction

Coarse data (Heijtan and Rubin 1991; Heitjan 1994) are commonly occurred in biomedical data. The maximum likelihood approach is often used in analyzing such data. One of the problems with the use of the maximum likelihood approach is that additional high-dimensional nuisance models may need to be specified. Misspecification of those models is of concern in practice. To address this concern, semiparametric models may be used and the efficient estimation procedure may be considered. Estimation of the parameters in semiparametric models with coarse data has been extensively studied. See for examples Bickel et al. (1993), Robins et al. (1994), Rotnitzky and Robins (1997), Scharfstein et al. (1999), and van der Laan and Robins (2003) among

H. Y. Chen (✉)
Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, 1603 West Taylor Street, Chicago, IL, 60612, USA
e-mail: hychen@uic.edu

others. It is well known that finding the semiparametric efficient estimator for coarse data problems is inherently complex. Even when a representation of the semiparametric efficient estimating score is obtained, deriving the locally semiparametric efficient estimator from the score representation can still be computationally challenging.

The approach proposed by Robins and coauthors to find the efficient score for coarse data problems is through augmented weighted estimating equations. Many useful estimating methods were generated from their approach. However, the efficient score represented in their approach usually involves an integral equation which is hard to solve (Nan et al. 2004; Yu and Nan 2006). As a result, successive approximation algorithms were proposed (Robins and Wang 1998). Substantial analytical effort may be required in deriving their algorithm. In this article, we derive new representations of the efficient score, which are simpler to obtain, for the general coarse data problems based on Neumann series expansion. Approximations to the locally efficient score can be naturally obtained from the representations. Furthermore, most of the successive approximation approaches proposed in the literature to finding locally efficient scores with coarsening data can be derived as special cases of the proposed approaches. We concentrate on the general representations of the efficient score in this article and leave the development of theories for estimation and inference based on the approximate score for future studies. Note that Chen (2009) presented such a theory for a special case of the general representations discussed in this paper.

The remainder of the article is organized as follows. In Sect. 2, we introduce the semiparametric estimation problem with observed coarse data. New representations for the semiparametric efficient score based on the Neumann series expansion are derived in Sect. 3. The new representations are then applied to both ignorable and nonignorable missing data problems to obtain the successive approximation algorithms proposed in the literature. Some new algorithms are also derived from the representations. The article is concluded with a discussion on the use of the approximation in practice.

2 The semiparametric problem for coarse data

Let Y be the full data and R be the coarse data variable. Let g be a known vector function defined on R and Y such that $Z = g(R, Y)$ is the observed coarsen data. For missing data problems (Little and Rubin 2002), R is the missing data indicator and $g(R, Y) = \{R, R(Y)\}$, where a component of R takes value 1 if the corresponding component of Y is observed, 0 otherwise, and a component of $R(Y)$ equals the corresponding component of Y if the corresponding component of R takes value 1, and is missing [takes the set value $(-\infty, +\infty)$] if the corresponding component of R takes value 0. For censoring data, R is the censoring variable and $g(R, Y) = \{1_{\{Y\le R\}}, \min(R, Y)\}$.

Let the density of the distribution for (R, Y) with respect to μ, a product of count measures and Lebesgue measures, be $\pi(R|Y, \alpha, \xi) f(Y, \beta, \theta)$, where $(\alpha, \beta) \in \Omega$, $(\xi, \theta) \in \Xi \times \Theta$. (α, β) are the parameters of interest and are usually Euclidean parameters. (ξ, θ) are nuisance parameters, which are usually of infinite dimension. Let $\eta = (\alpha, \beta, \xi, \theta)$ and let $\mathcal{L}$ denote the collection of all the models P_η with $\eta \in \Omega\times\Xi\times\Theta$.

Suppose that each of Ω, Ξ, and Θ is a subset of a Hilbert space. Let

$$H=\left\{h=(h_1,h_2,h_3,h_4)|P_{\eta+h_n/\sqrt{n}}\in\mathcal{L},\, h_n\to h \text{ in the product norm of } \Omega\times\Xi\times\Theta\right\}.$$

Assume that $H = H_1 \times H_2 \times H_3 \times H_4$, where $H_k = \{h_k|h = (h_1, h_2, h_3, h_4) \in H\}$ is a Hilbert space under the inner product of the corresponding (Hilbert) parameter space, $k = 1, 2, 3, 4$. Denote the inner product in space H_k by $\langle\cdot,\cdot\rangle_{H_k}$, for $k = 1, 2, 3, 4$. Assume that $H_1 \times H_2$ does not change for any $\eta \in \Omega \times \Xi \times \Theta$. Note that both H_3 and H_4 may depend on η. But the inner product, and thus the derived norm, on H_3 and H_4 does not depend on η. Assume further that $\{\pi(R|Y,\alpha,\xi)f(Y,\beta,\theta)\}^{1/2}$ is Fréchet differentiable with respect to η in $L^2(\mu)$. Following the convention (Bickel et al. 1993), let the derivative be denoted by

$$\frac{1}{2}\{\pi(R|Y,\alpha,\xi)f(Y,\beta,\theta)\}^{1/2}\left\{A_{1\eta}(h_1)+A_{2\eta}(h_2)+A_{3\eta}(h_3)+A_{4\eta}(h_4)\right\},$$

where $h_k \in H_k$, for $k = 1, 2, 3, 4$. The score operator at η with (R, Y) observed is defined as the Fréchet derivative times $2\{\pi(R|Y,\alpha,\xi)f(Y,\beta,\theta)\}^{-1/2}$, which is,

$$A_\eta h = A_{1\eta}(h_1) + A_{2\eta}(h_2) + A_{3\eta}(h_3) + A_{4\eta}(h_4).$$

The actual observed data are $Z = g(R, Y)$. The density for the observed data is

$$\int_{\{(r,y)|g(r,y)=z\}} \pi(r|Y,\alpha,\xi)f(Y,\beta,\theta)d\mu(r,y).$$

Suppose that the Fréchet derivative with respect to η in $L^2(\mu)$ exists. This has been verified in missing data and censoring data problems (Bickel et al. 1993). Denote the score operator by

$$B_\eta h = B_{1\eta}(h_1) + B_{2\eta}(h_2) + B_{3\eta}(h_3) + B_{4\eta}(h_4),$$

where $B_{k\eta}h_k = E_\eta\{A_{k\eta}h_k|g(R,Y)\}$, for $k = 1, 2, 3, 4$. The semiparametric efficient score for estimating (α, β) with observed data $Z = g(R, Y)$ can be expressed as

$$\text{Proj}_{P_\eta}[B_{1\eta}(h_1) + B_{2\eta}(h_2)|\left\{B_{3\eta}(H_3) + B_{4\eta}(H_4)\right\}^\perp],$$

where $\{B_{3\eta}(H_3) + B_{4\eta}(H_4)\}^\perp$ denotes the orthogonal complement of the linear space $B_{3\eta}(H_3)+B_{4\eta}(H_4)$ in $L^2(P_\eta)$. Let $\mathcal{B}_{a\eta}(h_1,h_2) = B_{1\eta}(h_1)+B_{2\eta}(h_2)$ and $\mathcal{B}_{b\eta}(h_3,h_4) = B_{3\eta}(h_3)+B_{4\eta}(h_4)$. From the existence of projection, for any fixed $(h_1, h_2) \in H_1 \times H_2$, there exists an $u \in \overline{\mathcal{B}_{b\eta}(H_3 \times H_4)}$, the closure of the linear space $\mathcal{B}_{b\eta}(H_3 \times H_4)$, such that

$$\langle\mathcal{B}_{a\eta}(h_1,h_2) - u, \mathcal{B}_{b\eta}(h_3,h_4)\rangle_{L^2(P_\eta)} = 0,$$

for all $(h_3, h_4) \in H_3 \times H_4$. If we assume further that $\mathcal{B}_{b\eta}(H_3 \times H_4)$ is a closed linear space, there exists an $(h_3^*, h_4^*) \in H_3 \times H_4$ such that $u = \mathcal{B}_{b\eta}(h_3^*, h_4^*)$. It follows from the assumption that

$$\begin{aligned}&\text{Proj}_{P_\eta}[\mathcal{B}_{a\eta}(h_1, h_2)|\{\mathcal{B}_{b\eta}(H_3 \times H_4)\}^\perp] \\ &\quad = \mathcal{B}_{a\eta}(h_1, h_2) - \text{Proj}_{P_\eta}\left[\mathcal{B}_{a\eta}(h_1, h_2)|\mathcal{B}_{b\eta}(H_3 \times H_4)\right],\end{aligned}$$

or equivalently that

$$\langle \mathcal{B}_{a\eta}(h_1, h_2) - \mathcal{B}_{b\eta}\left(h_3^*, h_4^*\right), \mathcal{B}_{b\eta}(h_3, h_4)\rangle_{L^2(P_\eta)} = 0.$$

Note that for any bounded linear operator A mapping $H_3 \times H_4$ to $L^2(P_\eta)$, and any $S \in L^2(P_\eta)$, $\langle S, A(h_3, h_4)\rangle_{L^2(P_\eta)}$ defines a linear functional on $H_3 \times H_4$. By Riesz representation theorem, there exists a unique $(h_3^0, h_4^0) \in H_3 \times H_4$ such that

$$\langle S, A(h_3, h_4)\rangle_{L^2(P_\eta)} = \langle (h_3^0, h_4^0), (h_3, h_4)\rangle_{H_3 \times H_4}.$$

The adjoint operator of A, denote by A^*, is defined as the map from $L^2(P_\eta)$ to $H_3 \times H_4$ satisfying $A^*S = (h_3^0, h_4^0)$. By applying this definition of adjoint operator to $\mathcal{B}_{b\eta}$, it follows that

$$\begin{aligned}0 &= \langle \mathcal{B}_{a\eta}(h_1, h_2) - \mathcal{B}_{b\eta}\left(h_3^*, h_4^*\right), \mathcal{B}_{b\eta}(h_3, h_4)\rangle_{L^2(P_\eta)} \\ &= \langle \mathcal{B}_{b\eta}^*\left\{\mathcal{B}_{a\eta}(h_1, h_2) - \mathcal{B}_{b\eta}\left(h_3^*, h_4^*\right)\right\}, (h_3, h_4)\rangle_{H_3 \times H_4},\end{aligned}$$

for all $(h_3, h_4) \in H_3 \times H_4$, where $\mathcal{B}_{b\eta}^*$ is the adjoint operator of $\mathcal{B}_{b\eta}$. It can now be seen that $(h_3^*, h_4^*) \in H_3 \times H_4$ satisfies the normal equation

$$\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}(h_3^*, h_4^*) = \mathcal{B}_{b\eta}^*\mathcal{B}_{a\eta}(h_1, h_2).$$

When $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$, as a map from $H_3 \times H_4$ to itself, is continuously invertible,

$$\left(h_3^*, h_4^*\right) = (\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta})^{-1}\mathcal{B}_{b\eta}^*\mathcal{B}_{a\eta}(h_1, h_2).$$

The projection can now be expressed as

$$\begin{aligned}&\mathcal{B}_{a\eta}(h_1, h_2) - \mathcal{B}_{b\eta}\{(\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta})^{-1}\mathcal{B}_{b\eta}^*\mathcal{B}_{a\eta}(h_1, h_2)\} \\ &\quad = \{I - \mathcal{B}_{b\eta}(\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta})^{-1}\mathcal{B}_{b\eta}^*\}\mathcal{B}_{a\eta}(h_1, h_2).\end{aligned}$$

3 New representations of the semiparametric efficient score

Note that $\mathcal{B}_{b\eta}$ and $\mathcal{B}_{b\eta}^*$ are relatively easy to find. In comparison, the primary difficulty in finding the projection is to find the inverse of $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$. Neumann series expansion can be used to eliminate the necessity of finding the inverse as follows.

For a positive constant c, note that

$$\begin{aligned}&\left\langle (h_3, h_4), \left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)(h_3, h_4)\right\rangle_{H_3\times H_4}\\&= ||(h_3, h_4)||^2_{H_3\times H_4} - \frac{1}{c}\langle (h_3, h_4), \mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}(h_3, h_4)\rangle_{H_3\times H_4}\\&\geq \left\{1 - \frac{||\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}||_{H_3\times H_4}}{c}\right\}||(h_3, h_4)||^2_{H_3\times H_4},\end{aligned}$$

and

$$\leq \left\{1 - \frac{\lambda^{\min}_{\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}}}{c}\right\}||(h_3, h_4)||^2_{H_3\times H_4},$$

where $\lambda^{\min}_{\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}}$ is the minimal spectral point of $\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}$. The constant c can be chosen sufficiently large such that $I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}$ is a positive definite operator with norm in $H_3 \times H_4$ less than 1. Suppose that c has thus been chosen. Note that

$$\left(\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^{-1} = \frac{1}{c}\left\{I - \left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)\right\}^{-1} = \frac{1}{c}\sum_{k=0}^{\infty}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k,$$

and

$$\frac{1}{c}\mathcal{B}_{b\eta}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k\mathcal{B}^*_{b\eta} = \frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}^*_{b\eta}\left(I - \frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}^*_{b\eta}\right)^k.$$

By applying those equalities, it can be seen that the projection operator can be rewritten as

$$\begin{aligned}I - \mathcal{B}_{b\eta}\left(\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^{-1}\mathcal{B}^*_{b\eta} &= I - \frac{1}{c}\mathcal{B}_{b\eta}\sum_{k=0}^{\infty}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k\mathcal{B}^*_{b\eta}\\&= I - \frac{1}{c}\mathcal{B}_{b\eta}\sum_{k=0}^{N-1}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k\mathcal{B}^*_{b\eta}\\&\quad -\frac{1}{c}\mathcal{B}_{b\eta}\sum_{k=N}^{\infty}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k\mathcal{B}^*_{b\eta}.\end{aligned}$$

It can be seen that

$$I - \frac{1}{c}\mathcal{B}_{b\eta}\sum_{k=0}^{N-1}\left(I - \frac{1}{c}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\right)^k\mathcal{B}^*_{b\eta} = \left(I - \frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}^*_{b\eta}\right)^N.$$

Further

$$\left\|\frac{1}{c}\mathcal{B}_{b\eta}\sum_{k=N}^{\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}\right)^{k}\mathcal{B}_{b\eta}^{*}\right\|_{L^2(P_\eta)}$$
$$\leq \frac{1}{c}||\mathcal{B}_{b\eta}||_{H_3\times H_4}\sum_{k=N}^{\infty}||I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}||_{H_3\times H_4}^{N}\left\|\mathcal{B}_{b\eta}^{*}\right\|_{L^2(P_\eta)}$$
$$=\left\|I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}\right\|_{H_3\times H_4}^{N}\frac{\frac{1}{c}||\mathcal{B}_{b\eta}||_{H_3\times H_4}\left\|\mathcal{B}_{b\eta}^{*}\right\|_{L^2(P_\eta)}}{1-\left\|I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}\right\|_{H_3\times H_4}}$$
$$\to 0 \text{ as } N\to\infty.$$

It can now be seen that

$$I-\mathcal{B}_{b\eta}(\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta})^{-1}\mathcal{B}_{b\eta}^{*}=\lim_{N\to\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}\right)^{N},$$

where the limit is in the sense of induced $L^2(P_\eta)$ norm for the operators. This leads to a representation of the semiparametric efficient score as

$$\lim_{N\to\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}\right)^{N}\mathcal{B}_{a\eta}(h_1,h_2) \tag{1}$$

The foregoing derivation requires that c be greater than the norm of $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$ and the inverse of $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$ exist. In fact, neither condition is necessary for the expansion, as we see in the following theorem.

Theorem 1 *Suppose that $\mathcal{B}_{b\eta}$ is a bounded linear operator on $H_3\times H_4$, a Hilbert space.*

1. *The projection of $E_\eta\{A_{1\eta}h_1+A_{2\eta}h_2|g(R,Y)\}$ to the null space of $\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}$ is (1), where c is a constant such that $|1-\lambda/c|<1$ for all λ, the nonzero spectral points of $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$.*
2. *When the normal equation $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}h_2=\mathcal{B}_{b\eta}^{*}E_\eta\{A_{1\eta}h_1+A_{2\eta}h_2|g(R,Y)\}$ has at least one solution, (1) is also the projection of $E\{A_{1\eta}h_1+A_{2\eta}h_2|g(R,Y)\}$ to $\mathcal{B}_{b\eta}^{\perp}(H_3\times H_4)$.*
3. *A sufficient condition for the normal equation to have at least one solution is that $\mathcal{B}_{b\eta}(H_3\times H_4)$ is a closed linear space. A sufficient (and necessary) condition for $\mathcal{B}_{b\eta}(H_3\times H_4)$ to be a closed linear space is that $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$ is continuously invertible.*

The proof of the theorem is given in the Appendix. Note that in the foregoing theorem, c needs to be greater than $\frac{1}{2}||\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}||$. To achieve a better convergence rate in the approximation, c should be chosen in such a way that $|1-\lambda/c|<1$ and as small as possible for all nonzero spectral points. If we know the maximal value ($\lambda_{\max}=||\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}||$) and the minimal nonzero value ($\lambda_{\min}$) of the spectral points, the choice of $c = (\lambda_{\max} + \lambda_{\min})/2$ yields the best geometric rate of convergence as $(\lambda_{\max} - \lambda_{\min})/(\lambda_{\max} + \lambda_{\min})$. In many applications, we do not know either $\lambda_{\max}$ or $\lambda_{\min}$. In those cases, an upper bound on the spectral points is enough.

The foregoing theorem can be generalized as follows. For any self-adjoint nonnegative definite operator $\mathcal{W}$ from $H_3 \times H_4$ to itself, $\mathcal{W}^{1/2}$ is defined and is also a self-adjoint nonnegative definite operator from $H_3 \times H_4$ to itself. If the range of $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}$ is the same as the range of $\mathcal{B}_{b\eta}$, i.e., $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(H_3 \times H_4) = \mathcal{B}_{b\eta}(H_3 \times H_4)$, we can redefine $\mathcal{B}_{b\eta}$ as $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}$ and replace $\mathcal{B}^*_{b\eta}$ by $\mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}$ to obtain a useful variation of Theorem 1. Note that c can be absorbed into $\mathcal{W}$.

Theorem 2 *Suppose that $\mathcal{B}_{b\eta}$ is a bounded linear operator on $H_3 \times H_4$, a Hilbert space. Suppose also that $\mathcal{W}$ is a self-adjoint operator from $H_3 \times H_4$ to $H_3 \times H_4$ satisfying $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(H_3 \times H_4) = \mathcal{B}_{b\eta}(H_3 \times H_4)$.*

1. *The projection of $E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$ to the null space of $\mathcal{B}_{b\eta}\mathcal{W}\mathcal{B}^*_{b\eta}$ is*

$$\lim_{N\to\infty}\left(I - \mathcal{B}_{b\eta}\mathcal{W}\mathcal{B}^*_{b\eta}\right)^N \mathcal{B}_{a\eta}(h_1,h_2), \tag{2}$$

   where $|1-\lambda| < 1$ for all λ, the nonzero spectral points of $\mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\mathcal{W}^{1/2}$.
2. *When the normal equation $\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}(h_3,h_4) = \mathcal{B}^*_{b\eta}E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$ has at least one solution, (2) is also the projection of $E\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$ to $\mathcal{B}^\perp_{b\eta}(H_3 \times H_4)$.*

Proof Only the second claim requires a proof. The second claim is proved if we can show that

$$\mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(h^*_3,h^*_4) = \mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$$

has at least one solution in $H_3 \times H_4$ if and only if $\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}(h_3,h_4) = \mathcal{B}^*_{b\eta}E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$ has at least one solution in $H_3 \times H_4$. Note that $\mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(h^*_3,h^*_4) = \mathcal{W}^{1/2}\mathcal{B}^*_{b\eta}E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$ for an $(h^*_3,h^*_4) \in H_3 \times H_4$ implies that

$$\langle \mathcal{B}_{b\eta}\mathcal{W}^{1/2}\left(h^*_3,h^*_4\right) - E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}, \mathcal{B}_{b\eta}\mathcal{W}^{1/2}(h_3,h_4)\rangle_{L^2(P_\eta)} = 0$$

for all $(h_3,h_4) \in H_3 \times H_4$. Since $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(H_3 \times H_4) = \mathcal{B}_{b\eta}(H_3 \times H_4)$, for any $(h_3,h_4) \in H_3 \times H_4$, there exists an $(h_{30},h_{40}) \in H_3 \times H_4$ such that $\mathcal{B}_{b\eta}\mathcal{W}^{1/2}(h_{30},h_{40}) = \mathcal{B}_{b\eta}(h_3,h_4)$. It follows from this and the foregoing displayed equation that

$$\langle \mathcal{B}_{b\eta}\mathcal{W}^{1/2}(h^*_3,h^*_4) - E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}, \mathcal{B}_{b\eta}(h_3,h_4)\rangle_{L^2(P_\eta)} = 0$$

for all $(h_3,h_4) \in H_3 \times H_4$. This implies that

$$\mathcal{B}^*_{b\eta}\mathcal{B}_{b\eta}(h_3,h_4) = \mathcal{B}^*_{b\eta}E_\eta\{A_{1\eta}h_1 + A_{2\eta}h_2|g(R,Y)\}$$

has at least one solution, $\mathcal{W}(h^*_3,h^*_4)$, in $H_3 \times H_4$. The reverse is easily seen.

Note that $\mathcal{B}_{b\eta}(h_3, h_4) = E_\eta\{A_{3\eta}h_3 + A_{4\eta}h_4|g(R, Y)\}$. It follows that $\mathcal{B}^*_{b\eta} = (A^*_{3\eta}, A^*_{4\eta}k^*_\eta)$ where k_η is the operator of taking the conditional expectation given the observed data $g(R, Y)$ under P_η and $k^*_\eta s\{g(R, Y)\} = E_\eta(s|Y)$ is the adjoint of k_η defined on square-integrable functions of the full data. Let $m_\eta = k^*_\eta k_\eta = E_\eta[E_\eta\{\cdot|g(R, Y)\}|Y]$, which maps square integrable functions of Y in $L^2(P_\eta)$ to square integrable functions of Y in $L^2(P_\eta)$. Let $M_\eta\{s(R, Y)\} = E_\eta[E_\eta\{s(R, Y)|g(R, Y)\}|Y]$. It follows that m_η is M_η restricted on square integrable functions of Y in $L^2(P_\eta)$. The efficient score representation can be rewritten as

$$\lim_{N\to\infty} E\left[\left\{I - (A_{3\eta}, A_{4\eta})\mathcal{W}\begin{pmatrix} A^*_{3\eta}k_\eta \\ A^*_{4\eta}M_\eta \end{pmatrix}\right\}^N (A_1h_1 + A_2h_2)\middle| g(R, Y)\right].$$

□

4 Applications of the new representations

The efficient score represented as the limit of a sequence can be used in practice by approximation with a finite N as $E\{\mathcal{U}_N|g(R, Y)\}$, where

$$\mathcal{U}_N = \left\{I - (A_{3\eta}, A_{4\eta})\mathcal{W}\begin{pmatrix} A^*_{3\eta}k_\eta \\ A^*_{4\eta}M_\eta \end{pmatrix}\right\}^N (A_1h_1 + A_2h_2).$$

It is easy to convert the expression into successive approximation as

$$\mathcal{U}_N = \left\{I - (A_{3\eta}, A_{4\eta})\mathcal{W}\begin{pmatrix} A^*_{3\eta}k_\eta \\ A^*_{4\eta}M_\eta \end{pmatrix}\right\}\mathcal{U}_{N-1}.$$

Different algorithms for computing the efficient score in coarse data problems can be obtained from the approximation with different choices of $\mathcal{W}$. We show in the following that many successive approximation algorithms proposed in the literature for coarse data problems can be derived from this approximation expression.

The optimal $\mathcal{W}$ is the inverse of

$$\begin{pmatrix} A^*_{3\eta}k_\eta \\ A^*_{4\eta}M_\eta \end{pmatrix}(A_{3\eta}, A_{4\eta})$$

when the inverse exists or an appropriate generalized inverse when the inverse does not exist. With the optimal choice, the approximation becomes exact for any $N \geq 1$. In practice, however, the optimal choice is usually unavailable, which is primarily the reason for considering the approximation. We therefore aim at choosing a $\mathcal{W}$ which is convenient to use and in the same time as close to the optimal choice as possible. Let $\mathcal{P}_j = A_{j\eta}(A^*_{j\eta}A_{j\eta})^{-1}A^*_{j\eta}$ for $j = 1, 2, 3, 4$ and $\mathcal{W} = \text{diag}\{\frac{1}{c_3}(A^*_{3\eta}A_{3\eta})^{-1}, \frac{1}{c_4}(A_{4\eta}A^*_{4\eta})^{-1}\}$. (2) can be rewritten as

$$\lim_{N\to\infty} E_\eta\left\{\left(I - \frac{1}{c_3}\mathcal{P}_3k_\eta - \frac{1}{c_4}\mathcal{P}_4M_\eta\right)^N (A_{1\eta}h_1 + A_{2\eta}h_2)\middle| g(R, Y)\right\}. \quad (3)$$

It can be seen that the choice of $c_3 = 1$ and $c_4 = 1$ always works though it is not an optimal choice in terms of the rate of convergence of the series.

When parameter ξ is not involved in the model formulation, such as when the coarsening mechanism is modeled by a parametric model, $A_{3\eta} = 0$ in (3). The representation can be simplified as

$$\lim_{N\to\infty} E_\eta \left\{ \left(I - \frac{1}{c_4}\mathcal{P}_4 M_\eta \right)^N (A_{1\eta}h_1 + A_{2\eta}h_2) \middle| g(R,Y) \right\}.$$

Note further that

$$A_{1\eta}h_1 = m_\eta^{-1}\{M_\eta(A_{1\eta}h_1)\} + [A_{1\eta}h_1 - m_\eta^{-1}\{M_\eta(A_{1\eta}h_1)\}]$$

and M_η maps the second term to zero. The projection can be rewritten as $E[A_{1\eta}h_1 - m_\eta^{-1}\{M_\eta(A_{1\eta}h_1)\}|g(R,Y)]$ plus

$$\lim_{N\to\infty} E_\eta \left\{ \left(I - \frac{1}{c_4}\mathcal{P}_4 m_\eta \right)^N \left(m_\eta^{-1}\{M_\eta(A_{1\eta}h_1)\} + A_{2\eta}h_2 \right) \middle| g(R,Y) \right\}.$$

When $c_4 = 1$,

$$\left\{ I - \mathcal{P}_4 m_\eta \right\}^N = \{I - \mathcal{P}_4\} + \mathcal{P}_4(I - m_\eta)\left\{ I - \mathcal{P}_4 m_\eta \right\}^{N-1}. \tag{4}$$

It can be seen that

$$\mathcal{U}_N = \{I - \mathcal{P}_4\}\left(m_\eta^{-1}\{M_\eta(A_{1\eta}h_1)\} + A_{2\eta}h_2 \right) + \mathcal{P}_4(I - m_\eta)\mathcal{U}_{N-1}.$$

For nonignorable missing data, the foregoing recursive formula is the successive approximation algorithm proposed in Rotnitzky and Robins (1997, Theorem A1.7). Note that the successive approximation derived from the original representation, i.e.,

$$\mathcal{U}_N = (I - \mathcal{P}_4 M_\eta)^N \left(A_{1\eta}h_1 + A_{2\eta}h_2 \right) = \left(I - \mathcal{P}_4 M_\eta \right) \mathcal{U}_{N-1}$$

with $\mathcal{U}_0 = A_{1\eta}h_1 + A_{2\eta}h_2$ is simpler to use than that of Rotnitzky and Robins (1997), especially when m^{-1} is not easy to obtain.

When the parameter in the coarsening data mechanism is not a part of the interest, we usually set $A_{1\eta} = 0$. The representation then reduces to

$$\lim_{N\to\infty} E_\eta \left\{ \left(I - \frac{1}{c_3}\mathcal{P}_3 k_\eta - \frac{1}{c_4}\mathcal{P}_4 M_\eta \right)^N A_{2\eta}h_2 \middle| g(R,Y) \right\}. \tag{5}$$

When m^{-1} is not difficult to obtain, we can apply the decomposition

$$A_{3\eta}h_3 = m_\eta^{-1}\{M_\eta(A_{3\eta}h_3)\} + [A_{3\eta}h_3 - m_\eta^{-1}\{M_\eta(A_{3\eta}h_3)\}]$$

to reduce the formula. Since both $A_{2\eta}h_2$ and $A_{4\eta}h_4$ are functions of Y only, it follows that

$$\begin{aligned}\langle A_{3\eta}h_3, k_\eta A_{j\eta}h_j\rangle &= \langle k_\eta A_{3\eta}h_3, k_\eta A_{j\eta}h_j\rangle \\ &= \langle k_\eta m_\eta^{-1}\{M_\eta(A_{3\eta}h_3)\}, k_\eta A_{j\eta}h_j\rangle \\ &\quad + \langle k_\eta[A_{3\eta}h_3 - m_\eta^{-1}\{M_\eta(A_{3\eta}h_3)\}], k_\eta A_{j\eta}h_j\rangle \\ &= \langle k_\eta m_\eta^{-1}\{M_\eta(A_{3\eta}h_3)\}, k_\eta A_{j\eta}h_j\rangle \\ &= \langle M_\eta\{A_{3\eta}(h_3)\}, A_{j\eta}h_j\rangle,\end{aligned}$$

for $j = 2, 4$. If we define $\mathcal{P}_3^{\text{new}}$ as the projection to the space generated by $M_\eta\{A_{3\eta}(H_3)\}$, (5) can be rewritten as

$$\lim_{N\to\infty} E_\eta\{\left(I - \mathcal{P}_3^{\text{new}} - \mathcal{P}_4 m_\eta\right)^N A_{2\eta}h_2 \,|g(R, Y)\} \tag{6}$$

when $c_3 = c_4 = 1$. In the special case where the coarse data are coarsening at random, $M_\eta\{A_{3\eta}(H_3)\} = \{0\}$. It follows that $\mathcal{P}_3^{\text{new}} = 0$. Equation (6) can be simplified to

$$\lim_{N\to\infty} E\{(I - \mathcal{P}_4 m_\eta)^N A_{2\eta}h_2|g(R, Y)\},$$

when $c_3 = c_4 = 1$. Notice that $(I - \mathcal{P}_4)A_2h_2 = S^{F,\text{eff}}(h_2)$, where $S^{F,\text{eff}}$ denote the efficient score for β under the full data model. By applying (4), we obtain that the efficient score is $\lim_{N\to\infty} E\{\mathcal{U}_N|g(R, Y)\}$, where $\mathcal{U}_0 = S^{F,\text{eff}}(h_2)$ and

$$\mathcal{U}_N = S^{F,\text{eff}}(h_2) + \mathcal{P}_4(I - m_\eta)\mathcal{U}_{N-1},$$

which is the same as the successive approximation algorithm proposed in Robins et al. (1994, Proposition 8.1e) for missing data problems with MAR missing data.

When either $(A_{3\eta}A_{3\eta}^*)^{-1}$ or $(A_{4\eta}A_{4\eta}^*)^{-1}$ or both are difficult to obtain, we can respectively use $\mathcal{W} = \text{diag}(\frac{1}{c_3}I, \frac{1}{c_4}(A_{4\eta}A_{4\eta}^*)^{-1})$, $\text{diag}(\frac{1}{c_3}(A_{3\eta}A_{3\eta}^*)^{-1}, \frac{1}{c_4}I)$, and diag $(\frac{1}{c_3}I, \frac{1}{c_4}I)$ in the representation. In the last case, (2) becomes

$$\lim_{N\to\infty} E_\eta\left\{\left(I - \frac{1}{c_3}A_{3\eta}A_{3\eta}^*k_\eta - \frac{1}{c_4}A_{4\eta}A_{4\eta}^*M_\eta\right)^N \left(A_{1\eta}h_1 + A_{2\eta}h_2\right)\Big|g(R, Y)\right\}, \tag{7}$$

where c_3 and c_4 may respectively be taken as the norms of $A_{3\eta}^*A_{3\eta}$ and $A_{4\eta}^*A_{4\eta}$. From the foregoing discussion, it can also be seen that, when data are coarsening at random and there is no α involved, which implies $A_{1\eta}$ is zero, the representation reduces to

$$\lim_{N\to\infty} E_\eta\left\{\left(I - \frac{1}{c_4}A_{4\eta}A_{4\eta}^*m_\eta\right)^N A_{2\eta}h_2\,|g(R, Y)\right\}. \tag{8}$$

5 Discussion

We proposed a general approximation approach to finding the locally semiparametric efficient score in coarse data problems, including ignorable and nonignorable missing data and censoring data problems. Algorithms derived from the proposed representation cover most existing successive approximation approaches. New algorithms for computing locally efficient scores are also generated from the proposed representations. Further work is needed to study the inferential properties of the proposed Neumann approximations in practice.

Appendix: Proof of Theorem 1

Proof Note that $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$ is a self-adjoint operator. Let $\mathcal{P}_t$ denote the projection operator in $L^2(P_\eta)$ that projects scores to the null space of operator $\psi_t(\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}) = \lim_{n\to\infty}\phi_{nt}(\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta})$, where $\phi_n(u)$, $n \geq 1$ are polynomials approximating $\psi_t(u) = (u-t)1_{(u>t)}$. The self-adjoint operator $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$ has an integral representation (Kantorovich and Akilov 1982, pp 258–274) which, for the nonnegative operator, appears as

$$\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta} = \int_0^M t d\mathcal{P}_t,$$

where M is the norm of $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$ under $L^2(P_\eta)$. Since $(I - \frac{1}{c}\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta})^N$ is a polynomial of $\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}$, it has the following integral representation

$$\left(I - \frac{1}{c}\mathcal{B}_{b\eta}^*\mathcal{B}_{b\eta}\right)^N = \int_0^M \left(1 - \frac{t}{c}\right)^N d\mathcal{P}_t.$$

Because for any fixed $\delta > 0$, $\lim_{N\to+\infty}(1-t/c)^N = 0$ uniformly on $[\delta, M]$, it follows that

$$\lim_{N\to+\infty}\int_\delta^M \left(1 - \frac{t}{c}\right)^N d\mathcal{P}_t = 0.$$

On the other hand, because

$$\mathcal{P}_0 \leq \int_0^\delta \left(1 - \frac{t}{c}\right)^N d\mathcal{P}_t \leq \int_0^\delta d\mathcal{P}_t = \mathcal{P}_\delta,$$

and the projection operator $\mathcal{P}_t$ is right continuous, it is seen that

$$\lim_{\delta\to 0}\int_0^\delta \left(1 - \frac{t}{c}\right)^N d\mathcal{P}_t = \mathcal{P}_0,$$

where the inequalities between operators are in the sense of positive definite of the difference operator. It follows that

$$\lim_{N\to+\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}\right)^{N}=\lim_{N\to+\infty}\int_{0}^{M}\left(1-\frac{t}{c}\right)^{N}d\mathcal{P}_{t}=\mathcal{P}_{0},$$

which is the projection onto the null space of the operator $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$. Similarly, it can be shown that $\lim_{N\to+\infty}(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*})^{N}=\mathcal{P}_{0}^{*}$, the projection onto the null space of the operator $\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}$.

To prove (2), note that for any $U\{g(R,Y)\}$ and any $(h_3,h_4)\in H_3\times H_4$,

$$\begin{aligned}
&\left\langle \mathcal{B}_{b\eta}(h_3,h_4),\lim_{N\to\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}\right)^{N}U\right\rangle_{L^2(P_\eta)}\\
&=\left\langle (h_3,h_4),\mathcal{B}_{b\eta}^{*}\lim_{N\to\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}\right)^{N}U\right\rangle_{H_3\times H_4}\\
&=\left\langle (h_3,h_4),\lim_{N\to\infty}\left(I-\frac{1}{c}\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}\right)^{N}\mathcal{B}_{b\eta}^{*}U\right\rangle_{H_3\times H_4}\\
&=\langle (h_3,h_4),\mathcal{P}_0\mathcal{B}_{b\eta}^{*}U\rangle_{H_3\times H_4}.
\end{aligned}$$

Let $h_0=\mathcal{P}_0\mathcal{B}_{b\eta}^{*}U$. It follows from the definition of $\mathcal{P}_0$ that $\mathcal{B}_{b}^{*}\mathcal{B}_{b\eta}h_0=0$. When the normal equation $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}(h)=\mathcal{B}_{b\eta}^{*}U$ has at least one solution,

$$\begin{aligned}
0&=\left\langle \mathcal{P}_0\mathcal{B}_{b\eta}^{*}U,\mathcal{B}_{b\eta}^{*}U-\mathcal{P}_0\mathcal{B}_{b\eta}^{*}U\right\rangle_{(h_3,h_4)}=\left\langle h_0,\mathcal{B}_{b\eta}^{*}U-h_0\right\rangle_{H_3\times H_4}\\
&=\left\langle h_0,\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}(h_3,h_4)\right\rangle_{H_3\times H_4}-||h_0||_{H_3\times H_4}^{2}\\
&=\left\langle \mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}h_0,(h_3,h_4)\right\rangle_{H_3\times H_4}-||h_0||_{H_3\times H_4}^{2}=-||h_0||_{H_3\times H_4}^{2},
\end{aligned}$$

which implies that $h_0=0$. Thus, $\lim_{N\to\infty}(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*})^{N}U\in\mathcal{B}_{b\eta}^{\perp}(H_3\times H_4)$. Further, since $U-(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*})^{N}U=\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*}\Delta_N U\in\mathcal{B}_{b\eta}(H_3\times H_4)$ for a bounded linear operator Δ_N, it follows that $U-\lim_{N\to+\infty}(I-\frac{1}{c}\mathcal{B}_{b\eta}\mathcal{B}_{b\eta}^{*})^{N}U\in\overline{\mathcal{B}_{b\eta}(H_3\times H_4)}$. Hence, (1) is the projection of $U=E_\eta\{\mathcal{B}_{a\eta}(h_1,h_2)|g(R,Y)\}$ to $\mathcal{B}_{b\eta}^{\perp}(H_3\times H_4)$.

When $\mathcal{B}_{b\eta}(H_3\times H_4)$ is closed, the projection onto $\mathcal{B}_{b\eta}(H_3\times H_4)$ exists. Let $\mathcal{B}_{b\eta}(h_3,h_4)$ be the projection of score S onto $\mathcal{B}_{b\eta}(H_3\times H_4)$. It then follows that

$$\begin{aligned}
0&=\langle \mathcal{B}_{b\eta}(h_3,h_4)-S,\mathcal{B}_{b\eta}(h_3,h_4)\rangle_{L^2(P_\eta)}\\
&=\langle \mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}(h_3,h_4)-\mathcal{B}_{b\eta}^{*}S,(h_3,h_4)\rangle_{H_3\times H_4}.
\end{aligned}$$

Therefore, (h_3,h_4) satisfies the normal equation. From the proof, we conclude that $\mathcal{P}_0^{*}$ is the projection onto $\mathcal{B}_{b\eta}^{\perp}(H_3\times H_4)$ when $\mathcal{B}_{b\eta}(H_3\times H_4)$ is closed. When $\mathcal{B}_{b\eta}^{*}\mathcal{B}_{b\eta}$ is invertible, for any $g_n = \mathcal{B}_{b\eta}(H_3 \times H_4)$ such that $g_n \to g_0$, $(h_{3n}, h_{4n}) = (\mathcal{B}_{b\eta}^* \mathcal{B}_{b\eta})^{-1} \mathcal{B}_{b\eta}^* g_n \to (\mathcal{B}_{b\eta}^* \mathcal{B}_{b\eta})^{-1} \mathcal{B}_{b\eta}^* g_0 \in H_3 \times H_4$. Hence, $g_0 = \mathcal{B}_{b\eta}(\mathcal{B}_{b\eta}^* \mathcal{B}_{b\eta})^{-1} \mathcal{B}_{b\eta}^* g_0$, which implies that $g_0 \in \mathcal{B}_{b\eta}(H_3 \times H_4)$. Hence $\mathcal{B}_{b\eta}(H_3 \times H_4)$ is closed. □

References

Bickel, P., Klassen, C., Ritov, Y., Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models*. Baltimore: John Hopkins University Press.

Chen, H. Y. (2009). Estimation and inference based on Neumann series approximation to locally efficient score in missing data problems. *Scadinavian Journal of Statistics*, in press.

Heitjan, D. F. (1994). Ignorability in general incomplete data models. *Biometrika, 94*, 701–708.

Heijtan, D. F., Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics, 19*, 2244–2253.

Kantorovich, L. V., Akilov, G. P. (1982). *Functional analysis* (2nd ed.). Oxford: Pergamon Press.

Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Nan, B., Emond, M. J., Wellner, J. A. (2004). Information bounds for Cox regression models with missing data. *Annals of Statistics, 32*, 723–735.

Robins, J. M., Wang, N. (1998). Discussion on the papers by Forster and Smith and Clayton et al. *Journal of Royal Statistical Society, Ser. B, 60*, 91–93.

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *The Journal of American Statistical Association, 89*, 846–866.

Rotnitzky, A., Robins, J. M. (1997). Analysis of semiparametric models with nonignorable nonresponses. *Statistics in Medicine, 16*, 81–102.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Scharfstein, D. O., Rotnitzky, A., Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion). *The Journal of American Statistical Association, 94*, 1096–1120.

Van der Laan, M. J., Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. New York: Springer.

Yu, M., Nan, B. (2006). Semiparametric regression models with missing data: the mathematical review and a new application. *Statistica Sinica, 16*, 1193–1212.