Proceedings of the American Statistical Association Section on Bayesian Statistical Science 1999, pp. 6 – 10, 2000.

ROBUST ESTIMATION IN SEQUENTIALLY IGNORABLE MISSING DATA AND CAUSAL INFERENCE MODELS

James M. Robins, Harvard School of Public Health
Department of Epidemiology, Harvard School of Public Health,
677 Huntington Avenue, Boston, MA 02115

KEY WORDS Causal inference, longitudinal data, missing data

1. Introduction

The goal of this paper is to (i) construct "doubly robust" augmented inverse probability of treatment weighted (IPTW) estimators and g-estimators in ignorable longitudinal missing data and causal inference models and (ii) show that these estimators can be represented as sequential regression estimators and two stage least squares (2SLS) estimators respectively. In a missing data model an estimator is doubly robust (equivalently doubly protected) if it remains consistent when either the model for missingness mechanism or the model for the distribution of the complete data is correctly specified. In a saturated causal inference model an estimator is doubly robust if it remains consistent when either the model for the treatment assignment mechanism or the model for the remainder of the distribution of the observed data is correctly specified. Scharfstein, Rotnitzky, and Robins (1999, Sec. 3) have discussed the regression representation of doubly robust augmented IPTW estimators in non-longitudinal studies.

2. Monotone Missing Data Models

Let $L = \overline{L}_{K+1} = \left(L_1', \dots L_{K+1}'\right)'$ represent the full data obtained at times $m = 1, \dots, K+1$. Let C be the censoring time such that if $C = m$, then $\overline{L}_m = (L_1, \dots, L_m)$ is observed and $\underline{L}_{m+1} = (L_{m+1}, \dots, L_{K+1})$ is missing. We assume the sample space for C is $\{1, ..., K+1\}$ so L_1 represent the always observed baseline variables. We observe n i.i.d. copies of $O = \left(C, \overline{L}_C\right)$. We assume the data are coarsened at random, i.e.,

$$\lambda\left[m \mid L\right] = \lambda\left[m \mid \overline{L}_m\right], m = 1, ..., K \qquad (1)$$

where $\lambda\left[m \mid \cdot\right] = pr\left[C = m \mid C \geq m, \cdot\right]$ is the discrete hazard of censoring. We assume that *for all m and positive σ*

$$\lambda\left[m \mid \overline{L}_m\right] < 1 - \sigma w.p.\ 1\ . \qquad (2)$$

Our goal is to make inference about the finite dimensional parameter μ in a semiparametric model with likelihood $f(L; \mu, \theta)$ where $\mu \in R^p$ and $\theta \in \Theta$ is an infinite dimensional nuisance parameter. With but little loss of generality, we assume that, in the absence of missing data, the orthogonal complement to the nuisance tangent space is give by the set $\mathcal{D}$ of all $p-$dimensional unbiased estimating functions for μ, i.e., $\mathcal{D} = \{d(L, \mu); E_{\mu,\theta}[d(L, \mu)] = 0\}$. This implies that, in the absence of missing data, any regular asymptotically linear (RAL) estimator of μ must be asymptotically equivalent to the solution to $0 = \sum_i d(L_i; \mu)$ for some $d \in \mathcal{D}$ where the index i will always index the n study subjects. To simplify notation we take μ to be one dimensional. As an example if μ is the mean of L_{K+1} then $d(L, \mu) = c(L_{K+1} - \mu)$ for any constant c.

In the presence of missing data, we will in general not be able to estimate μ without making further modelling assumptions due to the curse of dimensionality. Formally, we are assuming that when there are missing data, the curse of dimensionality appropriate (CODA) information bound of Robins and Ritov (1997) for μ is zero. One approach to reducing the dimension is to assume a parametric submodel $f(L; \mu, \psi)$ where $\psi \in \Psi \subset \Theta$ and Ψ is a finite dimensional space. An alternative approach is to specify a parametric model $\lambda\left[m \mid \overline{L}_m; \alpha\right]$ for $\lambda\left(m \mid \overline{L}_m\right)$.

Our goal will be to construct an estimator $\widehat{\mu}_{robust}$ of μ that is RAL in the semiparametric *union* model that assumes that the data are CAR, the semiparametric model $f(L; \mu, \theta), \theta \in \Theta$ is true, and at least one of the lower-dimensional models $f(L; \mu, \psi), \psi \in \Psi$ or $\lambda\left(m \mid \overline{L}_m; \alpha\right)$ is correct. Further when both parametric models are correct and the model $f(L; \mu, \theta)$ does not restrict the distribution of L (i.e. it is nonparametric), the estimator should attain the semiparametric variance bound for the union model. An sequential regression estimator that satisfies this goal is constructed as follows.

1. Compute the restricted MLE $\widehat{\psi}=\widehat{\psi}(\mu)$ of ψ with μ held fixed and the MLE $\widehat{\alpha}$ of α from the observed data.
2. Select a particular $d(L,\mu)$ from $\mathcal{D}$. (The choice can only affect efficiency .)
3. Let $\widehat{B}_m(\mu) = \Phi\left\{E_{\mu,\widehat{\psi}}\left[d(L,\mu) \mid \overline{L}_m\right]\right\}$, where Φ^{-1} is the canonical link function of a given generalized linear model (GLM). (This expectation will generally have to be computed numerically or by simulation)
4. Set $\widehat{H}_{K+1}(\mu) = d(L,\mu)$. For $m = K+1,\ldots,2$, for subjects with $C \geq m-1$, let $\widehat{H}_{m-1}(\mu) = \Phi^{-1}\left\{\widehat{B}_{m-1}(\mu) + \widehat{\phi}_{m-1}(\mu)\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})\right\}$ be the predicted value from the iteratively reweighted least squares (IRLS) fit of the regression model $\Phi\{E\left[\widehat{H}_m(\mu) \mid C \geq m, \overline{L}_{m-1}\right]\} = \widehat{B}_{m-1}(\mu) + \phi_{m-1}\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})$ where ϕ_{m-1} is the sole unknown parameter in the regression model, $\overline{\pi}_m(\widehat{\alpha}) = \prod_{j=1}^{m}\left(1-\lambda\left[j \mid \overline{L}_j;\widehat{\alpha}\right]\right)$, and $\widehat{\phi}_{m-1}(\mu)$ is the IRLS estimator of ϕ_{m-1}. This means that $\widehat{\phi}_{m-1}(\mu)$ satisfies $0 = \widetilde{E}\{I(C \geq m)[\widehat{H}_m(\mu) - \Phi^{-1}$ $\{\widehat{B}_{m-1}(\mu) + \widehat{\phi}_{m-1}(\mu)\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})\}]\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})\}$ where $\widetilde{E}\{V\} = n^{-1}\sum_{i=1}^{n} V_i$.
5. $\widehat{\mu}_{robust}$ solves $0 = \sum_i \widehat{H}_{1i}(\mu)$.

The argument at the end of the following section shows the sequential regression estimator $\widehat{\mu}_{robust}$ is also an augmented IPTW estimator which implies that it satisfies our goals.

3. Marginal structural models

In this section, the temporally ordered observed data are now $O = (L_1, A_1, L_2, A_2, \ldots L_K, A_K, L_{K+1})$ where A_k is a treatment given at time k and L_k are other variables measured just prior to treatment. Associated with each history $\overline{a} = (a_1, \ldots, a_K)$ there is a counterfactual random variable $L_{\overline{a}} = \overline{L}_{\overline{a},K+1}$ satisfying the consistency assumption $\overline{L}_{\overline{a},m} = \overline{L}_m$ if $\overline{A}_{m-1} = \overline{a}_{m-1}$. We impose the assumption of sequential ignorability (i.e., no unmeasured confounders) that *for all* $\overline{a}$ *and* m

$$L_{\overline{a}} \coprod A_m \mid \overline{L}_m, \overline{A}_{m-1} = \overline{a}_{m-1}. \qquad (3)$$

Further we assume that for all a_m in the support of A_m

$$\text{If } f\left(\overline{A}_{m-1}, \overline{L}_m\right) > 0, \qquad (4)$$
$$\text{then } f\left[a_m \mid \overline{A}_{m-1}, \overline{L}_m\right] > 0.$$

Robins (1999) discusses the use of MSMs to estimate the effect of $\overline{a}$ on $L_{\overline{a}}$. With little loss of generality, he shows that the statistical problem which arises from estimation of MSMs under sequential ignorability often reduces to the estimation of the finite dimensional parameter of μ in a semiparametric model with likelihood $f(O;\mu,\theta,\rho) = \prod_{m=1}^{K+1} f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1};\mu,\theta\right] \times \prod_{m=1}^{K} f\left[A_m \mid \overline{A}_{m-1}, \overline{L}_m;\rho\right]$, where $f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1};\mu,\theta\right]$ and $f\left[A_m \mid \overline{A}_{m-1}, \overline{L}_m;\rho\right]$ are densities with respect to dominating measures ν_l and ν_a respectively, (μ,θ) and ρ are variation-independent, and θ and ρ are infinite-dimensional nuisance parameters, that is characterized by the restriction for all functions $s = s(\cdot,\cdot,\cdot)$ in a given set $\mathcal{S}$

$$E\left[s\left(\overline{L}_{K+1}, \overline{A}_K;\mu\right)/\overline{\pi}_K\right] = 0 \qquad (5)$$

where now $\overline{\pi}_m = \prod_{j=1}^{m} f\left[A_j \mid \overline{L}_j, \overline{A}_{j-1}\right]$. Specifically, if we specify a marginal structural model characterized by the restriction that $E\{s(L_{\overline{a}},\overline{a};\mu)\} = 0$ for $s \in \mathcal{S}$, then (5) will hold under (3) and (4). As an example, the marginal structural mean model $E[L_{\overline{a},K+1}] = g(\overline{a},\mu)$ with $g(\cdot,\cdot)$ a known function is equivalent to the model $E\{s(L_{\overline{a}},\overline{a};\mu)\} = 0$ with $\mathcal{S} = \{s(L_{\overline{a}},\overline{a};\mu) = s^*(\overline{a})[L_{\overline{a},K+1} - g(\overline{a},\mu)]; s^*(\cdot)$ *arbitrary*$\}$. If the assumption that $E\{s(L_{\overline{a}},\overline{a};\mu)\} = 0$ for $s \in \mathcal{S}$ does not restrict the distribution of the $L_{\overline{a}}$, we say our MSM is saturated.

In order to reduce dimensionality, we consider two parametric submodels, $f\left(a_m \mid \overline{l}_m, \overline{a}_{m-1};\alpha\right)$ and $f\left(l_m \mid \overline{l}_{m-1}, \overline{a}_{m-1};\mu,\psi\right)$ where ψ and α are finite-dimensional parameters. Our goal will be to construct a RAL estimators of μ in the semiparametric *union* model that assumes (3), (4), the model $f(O;\mu,\theta,\rho)$ (so (5) holds), and at least one of the two finite-dimensional submodels is correct as well. Further, when both parametric models are correct and our MSM is saturated, the estimator should attain the semiparametric variance bound for the union model. For simplicity we take μ to be one dimensional. In that case $\mathcal{S}$ will be a one dimensional vector space when the MSM is saturated. A sequential regression estimator that satisfies our goal can be constructed as follows.

1. Compute the restricted MLE $\widehat{\psi}=\widehat{\psi}(\mu)$ of ψ with μ held fixed and the MLE $\widehat{\alpha}$ of α from the observed data.

2. Select a particular s from $\mathcal{S}$. (The choice can only affect efficiency.)

3. Let $\widehat{B}_m(\mu) = \Phi\{E_{\mu,\widehat{\psi}}[s(\overline{L}_{K+1}, \overline{A}_K, \mu) / \underline{\pi}_{m+1} \mid \overline{L}_m, \overline{A}_m]\} =$
$$\Phi\{\int \cdots \int s(\overline{L}_{K+1}, \overline{A}_K, \mu) \prod_{j=m+1}^{K+1} f\left[L_j \mid \overline{L}_{j-1}, \overline{A}_{j-1}; \mu, \widehat{\psi}\right] d\nu_l(L_j)\, d\nu_a(A_j)\}$$
where for $m < K+1$, $\underline{\pi}_m = \prod_{j=m}^{K} f\left[A_j \mid \overline{L}_j, \overline{A}_{j-1}\right]$, $\underline{\pi}_{K+1} = 1$, and Φ^{-1} is the canonical link function of a given generalized linear model. This expectation will generally have to be computed numerically or by simulation. [The connection of this definition to that in the previous section becomes clear upon noting that, under (3), $\widehat{B}_m(\mu) = \widehat{b}_m(\overline{L}_m, \overline{A}_{m,}\mu)$ where $\widehat{b}_m(\overline{l}_m, \overline{a}_{m,}\mu) = \int \cdots \int E_{\mu,\widehat{\psi}}\left[s(L_{\overline{a}}, \overline{a}; \mu) \mid \overline{L_{\overline{a},m}} = \overline{l}_m\right] \prod_{j=m+1}^{K} d\nu_a(a_j)$.]

4. Set $\widehat{T}_{K+1}(\mu) = s(\overline{L}_{K+1}, \overline{A}_K, \mu)$.
For $m = K+1, \ldots, 2$, let
$\widehat{H}_{m-1}(\mu) \equiv \widehat{h}_{m-1}(\overline{L}_{m-1}, \overline{A}_{m-1,}\mu) = \Phi^{-1}\left\{\widehat{B}_{m-1}(\mu) + \widehat{\phi}_{m-1}(\mu)\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})\right\}$
be the predicted value from the iteratively reweighted least squares (IRLS) fit of the regression model $\Phi\{E\left[\widehat{T}_m(\mu) \mid \overline{A}_{m-1,}, \overline{L}_{m-1}\right]\} = \widehat{B}_{m-1}(\mu) + \phi_{m-1}\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})$ where ϕ_{m-1} is the sole unknown parameter in the regression model, $\overline{\pi}_m(\widehat{\alpha})$ is the estimated version of $\overline{\pi}_m$, and $\widehat{\phi}_{m-1}(\mu)$ is the IRLS estimator of ϕ_{m-1}. Here for $m = K, \ldots, 1$, $\widehat{T}_m(\mu) = \int \widehat{h}_m(\overline{L}_m, \overline{A}_{m,}\mu)\, d\nu_a(A_m)$.

5. $\widehat{\mu}_{robust}$ solves $0 = \sum_i \widehat{T}_{1i}(\mu)$.

Now, $\widehat{\mu}_{robust}$ is precisely the augmented IPTW estimator solving $0 = \sum_i U_{IPTW,i}(\mu)$; because $\sum_i U_{IPTW,i}(\mu) = \sum_i \widehat{T}_{1i}(\mu)$ for all μ, where $U_{IPTW}(\mu) = s(\overline{L}_{K+1}, \overline{A}_K, \mu) / \overline{\pi}_K(\widehat{\alpha}) - \sum_{m=0}^{K}\{c(m, \overline{L}_m, \overline{A}_m, \mu) - E_{\widehat{\alpha}}[c(m, \overline{L}_m, \overline{A}_m, \mu) \mid \overline{L}_m, \overline{A}_{m-1}]\}$ is an augmented IPTW estimating function with $c(m, \overline{L}_m, \overline{A}_m, \mu) = \widehat{H}_m(\mu) / \overline{\pi}_m(\widehat{\alpha})$.

To see this, note $E_{\widehat{\alpha}}\left[c(m, \overline{L}_m, \overline{A}_m, \mu) \mid \overline{L}_m, \overline{A}_{m-1}\right] = \widehat{T}_m(\mu)\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})$. Hence the identity $\sum_i U_{IPTW,i}(\mu) = \sum_i \widehat{T}_{1i}(\mu)$ is a consequence of the fact that the inclusion of the term $\phi_{m-1}\overline{\pi}_{m-1}^{-1}(\widehat{\alpha})$ in the canonical link GLM in step 4 guarantees that the sample averages of $\widehat{H}_m(\mu)/\overline{\pi}_m(\widehat{\alpha})$ and $\widehat{T}_{m+1}(\mu)\overline{\pi}_m^{-1}(\widehat{\alpha})$ are equal for $m = 1, \ldots, K$.

It follows from Robins (1999) that, because it is an augmented IPTW estimator, $\widehat{\mu}_{robust}$ is RAL when the model $f(a_m \mid \overline{l}_m, \overline{a}_{m-1}; \alpha)$ is correct. The key step in showing that $\widehat{\mu}_{robust}$ is RAL when model $f_\cdot(l_m \mid \overline{l}_{m-1}, \overline{a}_{m-1}; \mu, \psi)$ is correct is the observation that then $\widehat{\phi}_m$ is converging to 0 for each m and hence that $n^{-1}\sum_i \widehat{T}_{1i}(\mu)$ is converging to $E\{s(\overline{L}_{K+1}, \overline{A}_K; \mu)/\overline{\pi}_K\}$. The efficiency result follows from the fact that, if the MSM is saturated, then at laws where both parametric models are true, the tangent space (i.e. the closed linear span of scores for correctly specified regular parametric submodels) for the union model is all random variables with finite variance. This implies that all regular estimators have the same efficient influence function.

The results in the previous section on missing data models are a special case of those in this section. To see this define $A_m = 1$ if $C > m$ and $A_m = 0$ otherwise and $s(\overline{L}_{K+1}, \overline{A}_K, \mu) = I(\overline{A}_K = \mathbf{1})\, d(L, \mu)$, where $\mathbf{1}$ is the vector with all components equal to 1. Note in this special case $\widehat{H}_m(\mu) = \widehat{T}_m(\mu)$.

4. Structural Nested Models

The set-up in this section is as in the previous. Our goal is to represent a g-estimator of an additive structural nested mean model (SNMM) as a sequential two-stage least squares (2SLS) estimator. For pedagogical purposes, we treat the simplest case. Specifically, we impose the additive SNMM

$$E\left[\overline{L}_{(\overline{A}_m,0),K+1} - \overline{L}_{(\overline{A}_{m-1},0),K+1} | \overline{L}_m, \overline{A}_m\right] = \gamma\left(\overline{L}_m, \overline{A}_m, \mu\right) \quad (6)$$

where $\left(\overline{A}_m, 0\right)$ is the treatment history that is $\overline{A}_m$ through time m and zero thereafter, $\gamma\left(\overline{L}_m, \overline{A}_m, \mu\right)$ is a known function, and μ is an unknown parameter. The statistical problem which arises in estimation of SNMMs under (3) and (4) reduces to the estimation of the finite dimensional parameter μ in a semiparametric model with likelihood $f(O; \mu, \theta, \rho) = \prod_{m=1}^{K+1} f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}; \mu, \theta\right] \times \prod_{m=1}^{K} f\left[A_m \mid \overline{A}_{m-1}, \overline{L}_m; \rho\right]$, where $f\left[L_m \mid \overline{L}_{m-1}, \overline{A}_{m-1}; \mu, \theta\right]$ and $f\left[A_m \mid \overline{A}_{m-1}, \overline{L}_m; \rho\right]$ are densities with respect to dominating measures ν_l and ν_a respectively, (μ, θ) and ρ are variation-independent, θ and ρ are infinite-dimensional nuisance parameters, that is characterized by the restriction that

$$E\left[H_m(\mu) | \overline{L}_m, \overline{A}_m\right] = E\left[H_m(\mu) | \overline{L}_m, \overline{A}_{m-1}\right] \quad (7)$$

where $H_m(\mu) = \overline{L}_{K+1} - \sum_{j=m}^{K} \gamma\left(\overline{L}_j, \overline{A}_j, \mu\right)$.

In order to reduce dimensionality, we consider two parametric submodels, $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}; \alpha\right)$ and $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}; \mu, \psi\right)$ where ψ and α are finite-dimensional parameters. Our goal will be to construct an estimator $\widehat{\mu}_{robust}$ of μ that is RAL in the semiparametric *union* model which assumes that the model $f(O; \mu, \theta, \rho)$ is correct (so (7) holds) and at least one of the two finite-dimensional submodels is correct as well. Let $\widehat{\psi}, \widehat{\mu}, \widehat{\alpha}$ be the MLEs when both submodels are imposed. Let $\widehat{B}_m = E_{\widehat{\psi},\widehat{\mu}}\left[H_m(\widehat{\mu}) | \overline{L}_m, \overline{A}_{m-1}\right]$ and $\widehat{G}_m = E_{\widehat{\alpha}}\left[A_m | \overline{L}_m, \overline{A}_{m-1}\right]$. Define $U(\mu, \phi_1, ..., \phi_K) = \sum_{m=1}^{K}\left(H_m(\mu) - \widehat{B}_m - \phi_m \widehat{G}_m\right) \left(A_m, 0, ..., 0, \widehat{G}_m, 0, ..0\right)'$ where $\left(A_m, 0, ..., \widehat{G}_m, 0, ..0\right)'$ is of dimension $K+1$ with $\widehat{G}_m$ the $m+1^{st}$ component. Then $\left(\widehat{\mu}_{robust}, \widehat{\phi}_1, ..., \widehat{\phi}_K\right)$ is the 2SLS estimator solving $0 = \sum_i U_i(\mu, \phi_1, ..., \phi_K)$. Furthermore, $\widehat{\mu}_{robust}$ is also the g-estimator solving $0 = \sum_i U_i^*(\mu)$ where $U^*(\mu) = \sum_{m=1}^{K}\left(H_m(\mu) - \widehat{B}_m - \widehat{\phi}_m(\mu)\widehat{G}_m\right)\left(A_m - \widehat{G}_m\right)$, $\widehat{\phi}_m(\mu)$ solves $0 = \sum_i U_i^{**}(\mu, \phi_m)$ with μ held fixed, and $U^{**}(\mu, \phi_m) = \sum_{m=1}^{K}\left(H_m(\mu) - \widehat{B}_m - \phi_m \widehat{G}_m\right)\widehat{G}_m$. This representation as a g-estimator implies $\widehat{\mu}_{robust}$ is RAL when the model $f\left(a_m \mid \overline{\ell}_m, \overline{a}_{m-1}; \alpha\right)$ is correct (Robins et al.,1999). When $f\left(\ell_m \mid \overline{\ell}_{m-1}, \overline{a}_{m-1}; \mu, \psi\right)$ is correct, $\widehat{\phi}_m(\mu)$ converges to 0 so the estimator will be RAL.

5. A Correction

Because this paper extends results in Scharfstein et.al. (1999), it will be useful here to point out an error in that paper. Specifically the formula for $\widehat{b}_{eff}(O; \psi)$ given at the end of their Sec 3.1 is incorrect unless the data are CAR (i.e., unless $\alpha = 0$ in their model (3)). If we let $\widehat{b}_{eff}(O; \psi)$ be the incorrect formula as given in their text, the correct formula is $\widehat{b}_{correct,eff}(O; \psi) = \widehat{b}_{eff}(O; \psi) - (0, c)' \widehat{b}_{eff,1}(O; \psi)$ where $\widehat{b}_{eff,1}(O; \psi)$ is the first component of $\widehat{b}_{eff}(O; \psi)$ and $c = E[\{\pi^{-1}(L; \rho_{mis}) - 1\} \{\pi(L; \rho_{mis})(1, V, Y) - E_{\rho}^*[\pi(L; \rho_{mis})(1, V, Y) | V]\} \{Y - \mu\}]$. An exactly analogous correction is required to the formula for $\widehat{b}_{eff}$ in the second to last paragraph of their section 3.4 when their model (4) is not CAR (ie when $r(L, \beta_0)$ is not identically zero).

6. Discussion

We have acted as if it were no problem to specify and maximize fully parametric MSM and SNMM models with μ satisfying the restriction (5) for MSMs and (7) for SNMs. This is not the case as discussed in Sec. 8 of Robins et al. (1999) where variation independent parameterization of the observed data distribution are given for SNMMs and MSMs and are shown to be quite complex. An alternative parametrization would be to specify a parametric model depending on (μ, ψ) for the joint distribution of the counterfactuals $L_{\overline{a}}$, but this leads to intractable high or even infinite dimensional integrals. The message is that the analysis of fully parametric MSMs and SNMMs is anything but straightforward. .Thus the method of estimator construction given above may be more theoretical than practical because of the difficulty with fitting the required parametric models. In practice, we would replace $\widehat{B}_m(\mu)$ and $\widehat{B}_m$ by $\varsigma_m B_m$,where B_m is a known vector function of $(\overline{L}_m, \overline{A}_m)$ in Section 3 and of $(\overline{L}_m, \overline{A}_{m-1})$ in Section 4 and ς_m is an unknown parameter (row) vector. In both Sections 3 and 4, we estimate ς_m analogously to and simultaneously with ϕ_m.

References:

1. Robins J.M. (1999). Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In: **Statistical Models in Epidemiology: The Environment and Clinical Trials.** Halloran, M.E. and Berry, D., eds. NY: Springer-Verlag, pp. 95-134.

2. Robins J.M. and Ritov Y. (1997). A curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine,* 16:285-319.

3. Robins J.M., Scharfstein D., Rotnitzky A. (1999). Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In: **Statistical Models in Epidemiology: The Environment and Clinical Trials.** Halloran M.E. and Berry D., eds. NY: Springer-Verlag, pp. 1-94 (to appear).

4. Scharfstein D.O., Rotnitzky A., and Robins J.M. (1999). Rejoinder to Adjusting for non-ignorable drop-out using semiparametric non-response models. *Journal of the American Statistical Association, 94:1135-1146.*