

On doubly robust estimation in a semiparametric odds ratio model

BY ERIC J. TCHETGEN TCHETGEN, JAMES M. ROBINS

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

etchetge@hsph.harvard.edu robins@hsph.harvard.edu

AND ANDREA ROTNITZKY

Department of Economics, Universidad Torcuato Di Tella, Sáenz Valiente 1010, 1428 Buenos Aires, Argentina

arotnitzky@utdt.edu

SUMMARY

We consider the doubly robust estimation of the parameters in a semiparametric conditional odds ratio model. Our estimators are consistent and asymptotically normal in a union model that assumes either of two variation independent baseline functions is correctly modelled but not necessarily both. Furthermore, when either outcome has finite support, our estimators are semiparametric efficient in the union model at the intersection submodel where both nuisance functions models are correct. For general outcomes, we obtain doubly robust estimators that are nearly efficient at the intersection submodel. Our methods are easy to implement as they do not require the use of the alternating conditional expectations algorithm of Chen (2007).

Some key words: Doubly robust; Generalized odds ratio; Locally efficient; Semiparametric logistic regression.

1. INTRODUCTION

Given a random vector $O = (Y, A, L)$ the conditional odds ratio function $\gamma(Y, A, y_0, a_0, L)$ between A and Y given L at a given base point (a_0, y_0) is

$$\begin{aligned}\gamma(Y, A, L) &= \frac{f(Y | A, L) f(y_0 | a_0, L)}{f(y_0 | A, L) f(Y | a_0, L)} \\ &= \frac{g(A | Y, L) g(a_0 | y_0, L)}{g(a_0 | Y, L) g(A | y_0, L)},\end{aligned}$$

where the vectors Y and A can take either discrete values, continuous values, or a mixture of both, L is a high-dimensional vector of measured auxiliary covariates, (a_0, y_0) is a user specified point in the sample space and $f(Y | A, L)$, $g(A | Y, L)$ and $h(A, Y | L)$ are, respectively, the conditional densities of Y given A and L , the conditional density of A given Y and L and the joint conditional density of A and Y given L with respect to a dominating measure μ . The odds ratio function is a particularly useful measure of association when Y and A take both discrete and continuous values. For instance, A and Y could each be a mixture of a discrete component encoding, say, the presence or absence of a given bacterium and a continuous component encoding the bacterial counts when it is present. In such a case, as argued by Chen (2007), a complete

characterization of the association between bacterium A and bacterium Y given L would require separate comparisons of the probabilities of absence of one bacterium when the other bacterium is either absent or present at a particular concentration, and of the concentration distribution for one bacterium when the other bacterium is either absent or present at a particular concentration. Instead, the direct estimation of the odds ratio function relating bacterium A to bacterium Y given covariates L provides a unified solution to this problem and obviates the need for separate analyses.

Given n independent and identically distributed copies of O , Chen (2007) proposed a locally efficient iterative estimator of the parameter ψ_0 in a semiparametric model \mathcal{B} that specifies (i) $\gamma(Y, A, L)$ is equal to a known function $\gamma(Y, A, L; \psi)$ evaluated at the unknown true p -dimensional parameter vector ψ_0 , i.e.

$$\gamma(Y, A, L; \psi_0) = \gamma(Y, A, L), \quad (1)$$

where $\gamma(Y, A, L; \psi)$ takes the value 1 if $A = a_0$, $Y = y_0$, or $\psi = 0$, so $\psi_0 = 0$ encodes the null hypothesis that Y and A are conditionally independent given L , and (ii) either but not necessarily both, (a) a given parametric model $f(Y | a_0, L; \theta)$ for $f(Y | a_0, L)$ or (b) a parametric model $g(A | y_0, L; \alpha)$ for $g(A | y_0, L)$ is correct. Model \mathcal{B} is referred to as a union model because it is the union of the model \mathcal{C} that assumes that (i) and (iia) are true and the model \mathcal{D} that assumes that (i) and (iib) are true. An estimator of ψ_0 that is consistent and asymptotically normal under this union model is referred to as doubly robust because, given equation (1), the estimator is consistent and asymptotically normal for ψ_0 if one has succeeded in specifying either a correct model $f(Y | a_0, L)$ or a correct model for $g(A | y_0, L)$, thus giving the data analyst two chances rather than one chance to obtain valid inference for ψ_0 .

An example of a simple parametric model for the odds ratio function is the bilinear log-odds ratio model (Chen, 2003, 2004). It assumes that $\gamma(Y, A, L; \psi_0) = \exp\{\psi_0(Y - y_0) \otimes (A - a_0)\}$, where \otimes is the direct product. This model includes all of the generalized linear regression models with canonical link functions as special cases. In the case of stratified 2×2 tables, it implies homogeneous odds ratios, but is easily extended to the case of nonhomogeneous odds ratios. Other interesting examples of odds ratio models are given by Chen (2007).

Unfortunately, Chen's aforementioned locally efficient doubly robust estimator of ψ_0 under model \mathcal{B} is computationally very demanding, especially when A and Y have multiple continuous components. The main contribution of our paper is to provide novel and highly efficient doubly robust estimators of ψ_0 that are substantially easier to compute than those of Chen.

2. PRELIMINARIES

Before describing our new approach, we briefly summarize Chen's results. He considered the following parametric and semiparametric approaches to the estimation of ψ_0 : a prospective likelihood approach under the model \mathcal{C} that assumes that one has correctly modelled the nuisance baseline function $f(Y | a_0, L)$; a retrospective likelihood approach under the model \mathcal{D} that assumes that one has correctly specified a model for the nuisance baseline function $g(A | y_0, L)$; a joint likelihood approach under the intersection model that assumes that both models \mathcal{C} and \mathcal{D} are correct; and a doubly robust locally semiparametric efficient approach under the union model \mathcal{B} of § 1.

In his doubly robust approach, Chen establishes that in the semiparametric model \mathcal{A} characterized by the sole restriction (1), the density $h(A, Y | L)$ can be written as $h(A, Y | L; \psi_0)$,

where

$$h(A, Y | L; \psi) = \frac{\gamma(Y, A, L; \psi) f(Y | L, A = a_0) g(A | Y = y_0, L)}{\int \gamma(y, a, L; \psi) f(y | L, A = a_0) g(a | Y = y_0, L) d\mu(a, y)}, \quad (2)$$

$f(y | L, A = a_0)$ and $g(a | Y = y_0, L)$ are the unknown conditional densities that generated the data and are solely restricted by $\int \gamma(y, a, L) f(y | L, A = a_0) g(a | Y = y_0, L) d\mu(a, y) < \infty$ almost everywhere. Then, he specifies parametric models $f(Y | a_0, L; \theta)$ and $g(A | y_0, L; \alpha)$ for the unknown nuisance baseline functions $f(y | a_0, L)$ and $g(a | y_0, L)$, obtains profile estimates $\hat{\theta}(\psi)$ and $\hat{\alpha}(\psi)$ of the nuisance parameters θ and α and calculates the efficient score $\hat{S}_{\text{eff}}(\psi) \equiv S_{\text{eff}}\{\hat{\theta}(\psi), \hat{\alpha}(\psi), \psi\}$ for ψ in the semiparametric model \mathcal{A} evaluated at the law $[\gamma(y, a, l; \psi), f\{y | a_0, l; \hat{\theta}(\psi)\}, g\{a | y_0, l; \hat{\alpha}(\psi)\}]$ indexed by $\{\hat{\theta}(\psi), \hat{\alpha}(\psi), \psi\}$. Next, he estimates ψ_0 with the solution $\hat{\psi}_{\text{eff}}$ to $P_n\{\hat{S}_{\text{eff}}(\psi)\} = 0$, where $P_n(H) = n^{-1} \sum_i H_i$, and proves that $\hat{\psi}_{\text{eff}}$ is regular and asymptotically linear and thus consistent and asymptotically normal under the union model \mathcal{B} . Further general results of Robins and Rotnitzky (2001) imply that $\hat{S}_{\text{eff}}(\psi)$ is also the efficient score for ψ in model \mathcal{B} under the law $[\gamma(y, a, l; \psi), f\{y | a_0, l; \hat{\theta}(\psi)\}, g\{a | y_0, l; \hat{\alpha}(\psi)\}]$. It follows that the estimator $\hat{\psi}_{\text{eff}}$ is locally semiparametric efficient under model \mathcal{B} at the intersection submodel with both nuisance models correct; that is, $\hat{\psi}_{\text{eff}}$ attains the semiparametric efficiency bound for the model \mathcal{B} when both nuisance models happen to hold.

By definition, the efficient score $S_{\text{eff}} = \Pi(S_\psi | \Lambda_{\text{nuis}}^\perp)$ for a parameter ψ in a given model is the projection of the score S_ψ for ψ onto the orthocomplement $\Lambda_{\text{nuis}}^\perp$ to the nuisance tangent space Λ_{nuis} in the Hilbert space $L_2 \equiv L_2(F_O)$ of zero-mean functions of p dimensions, $T \equiv t(A, Y, L) = t(O)$, with inner product $E_{F_O}(T_1^\top T_2) \equiv E(T_1^\top T_2)$, and corresponding squared norm $\|T\|^2 = E(T^\top T)$, where F_O is the distribution function that generated the data. Chen proves that for model \mathcal{A} , the set

$$\Lambda_{\text{nuis}}^\perp = \{v(Y, A, L) : E\{v(Y, A, L) | A, L\} = E\{v(Y, A, L) | Y, L\} = 0\} \cap L_2 \quad (3)$$

contains all functions that have zero-mean conditional on both (A, L) and (Y, L) . When both A and Y contain continuous components and $\psi_0 \neq 0$, Chen (2007) finds that this projection and therefore S_{eff} do not exist in closed form and must be computed using the iterative alternating conditional expectations algorithm. Each iteration requires the evaluation, by numerical integration, of conditional expectations, which seriously limits the practicality of Chen's approach, particularly when A and/or Y have two or more continuous components.

The main contribution of our paper is to show that, even though the projection $\Pi(R | \Lambda_{\text{nuis}}^\perp)$ of a given random variable $R = r(Y, A, L)$ into the orthocomplement $\Lambda_{\text{nuis}}^\perp$ does not exist in closed form when both A and Y contain continuous components, the set $\Lambda_{\text{nuis}}^\perp$ does have a closed-form representation, which appears to be new. We use our representation to obtain doubly robust estimators, i.e. consistent and asymptotically normal estimators of ψ_0 in the union model \mathcal{B} , that are nearly as efficient as $\hat{\psi}_{\text{eff}}$ under the intersection submodel, yet do not require the alternating conditional expectations algorithm. Moreover, our closed-form representation of $\Lambda_{\text{nuis}}^\perp$ is of independent interest, with applications beyond the present paper. For example, Vansteelandt et al. (2008) use our representation to construct multiple robust estimators of the parameter encoding the interaction on an additive and multiplicative scale between two exposures A_1 and A_2 in their effects on an outcome Y .

In the special situation where either Y or A has finite support, Bickel et al. (1993) provide a closed-form expression for $\Pi(R | \Lambda_{\text{nuis}}^\perp)$, which Chen, however, did not use to give a closed-form expression for S_{eff} . We remedy this oversight and obtain doubly robust locally-efficient closed-form estimating functions when Y and/or A has finite support; some emphasis is given to the

important case of dichotomous Y which, incidentally, coincides with the semiparametric logistic regression model.

In the following, for a vector v we write $v^{\otimes 2} = vv^T$. To simplify notation, we suppose $y_0 = 0$ and $a_0 = 0$ throughout, so that $\gamma(Y, 0, L; \psi) = \gamma(0, A, L; \psi) = \gamma(Y, A, L; 0) = 1$. We shall also use the following definition.

DEFINITION 1. *Given conditional densities $f^\dagger(Y | L)$ and $g^\dagger(A | L)$, the density $h^\dagger(Y, A | L) = f^\dagger(Y | L)g^\dagger(A | L)$ that makes A and Y conditionally independent given L is an admissible independence density if the joint law of (Y, A) given L under $h^\dagger(\cdot, \cdot | L)$ is absolutely continuous with respect to the true law of (Y, A) given L with probability one. Furthermore, $E^\dagger(\cdot | \cdot, L)$ denotes conditional expectations with respect to $h^\dagger(Y, A | L)$.*

3. MAIN RESULT

As noted previously, under model \mathcal{A} characterized by restriction (1), Chen showed that $\Lambda_{\text{nuis}}^\perp$ is given by the set (3). We now provide a new closed-form representation of this set. To do so, for a fixed choice of admissible independence density $h^\dagger(Y, A | L) = f^\dagger(Y | L)g^\dagger(A | L)$ and any p -dimensional function d of (Y, A, L) , define the random vector $U(\psi; d, h^\dagger)$ as

$$U(\psi; d, h^\dagger) \equiv u(O; \psi; d, h^\dagger) \equiv \{d(Y, A, L) - d^\dagger(Y, A, L)\} \frac{h^\dagger(Y, A | L)}{h(Y, A | L; \psi)}$$

with $h(Y, A | L; \psi)$ defined in (2) and $d^\dagger(Y, A, L) = E^\dagger(D | A, L) + E^\dagger(D | Y, L) - E^\dagger(D | L)$ for $D \equiv d(Y, A, L)$. The following theorem gives the influence functions of regular asymptotically linear estimators of ψ_0 in model \mathcal{A} and will form the basis for our doubly robust approach.

THEOREM 1. *Given an admissible independence density h^\dagger , an alternative representation of the set $\Lambda_{\text{nuis}}^\perp$ of (3) is $\Lambda_{\text{nuis}}^\perp = \{U(\psi_0; d, h^\dagger) : d \text{ unrestricted}\} \cap L_2$.*

Proof. One can verify by explicit calculation that $\{U(\psi_0; d, h^\dagger) : d\} \cap L_2 \subseteq \Lambda_{\text{nuis}}^\perp$. To show the other direction, take any function $v(A, Y, L)$ in $\Lambda_{\text{nuis}}^\perp$, let $d(Y, A, L) = v(A, Y, L)h(A, Y | L)/h^\dagger(Y, A | L)$. Then $v(A, Y, L) = U(\psi_0; d, h^\dagger)$ since $\int d(y, A, L)f^\dagger(y | L)d\mu(y) = \int v(A, y, L)f(y | A, L)g(A | L)/g^\dagger(A | L)d\mu(y) = E\{v(A, Y, L) | A, L\}g(A | L)/g^\dagger(A | L) = 0$ and $\int d(Y, a, L)g^\dagger(a | L)d\mu(a) = E\{v(A, Y, L) | Y, L\}f(Y | L)/f^\dagger(Y | L) = 0$. \square

Remark. We give an alternative, more abstract, proof of the fact that $U(\psi_0; d, h^\dagger) \equiv \{d(Y, A, L) - d^\dagger(Y, A, L)\}h^\dagger(Y, A | L)/h(Y, A | L; \psi_0) \in \Lambda_{\text{nuis}}^\perp$. Given an admissible independence density h^\dagger , let $\Lambda_{\text{nuis}}^{\perp, \dagger}$ be the set (3) with expectations taken under h^\dagger . It is well known that when, as under h^\dagger , A and Y are conditionally independent given L , $\Lambda_{\text{nuis}}^{\perp, \dagger}$ admits the representation $\{d(Y, A, L) - d^\dagger(Y, A, L) : d\}$. Then $d(Y, A, L) - d^\dagger(Y, A, L) \in \Lambda_{\text{nuis}}^{\perp, \dagger}$ implies $\{d(Y, A, L) - d^\dagger(Y, A, L)\}h^\dagger(Y, A | L)/h(Y, A | L; \psi_0) \in \Lambda_{\text{nuis}}^\perp$, by the Radon–Nikodym theorem.

By standard semiparametric theory (Bickel et al., 1993), Theorem 1 implies that if $\hat{\psi}$ is a regular and asymptotically linear estimator of ψ_0 in model \mathcal{A} , then given any admissible independence density h^\dagger , there exists a p -dimensional function $D \equiv d(O)$ such that $n^{1/2}(\hat{\psi} - \psi_0) = n^{1/2}P_n[E\{\partial U(\psi; d, h^\dagger)/\partial \psi^T |_{\psi=\psi_0}\}^{-1}U(\psi_0; d, h^\dagger)] + o_p(1)$. Furthermore, this also implies that any regular and asymptotically linear estimator of ψ_0 in model \mathcal{A} can be obtained, up to asymptotic equivalence, as the solution to an equation $\sum_{i=1}^n U_i(\psi; d, h^\dagger) = 0$. However, these solutions are infeasible because $h(Y, A | L; \psi)$ depends on the unknown conditional densities $f(y | L, A = 0)$ and $g(a | Y = 0, L)$, which must be estimated from the data. While a

nonparametric smoothing method would, in principle, be the preferred approach to estimate these densities, its finite-sample performance is bound to be poor for continuous L of moderate to high dimension because of the curse of dimensionality. A practical alternative is to proceed as in Chen (2007) and to impose working models of reduced dimension for the unknown baseline functions $f(Y | A = 0, L)$ and $g(A | Y = 0, L)$. Hence, we specify variation independent parametric models $g(A | Y = 0, L; \alpha)$ for $g(A | Y = 0, L)$ and $f(Y | A = 0, L; \theta)$ for $f(Y | A = 0, L)$ with unknown finite-dimensional parameters α and θ . Since we cannot be sure that either $f(Y | A = 0, L; \theta)$ or $g(A | Y = 0, L; \alpha)$ is correctly specified, we shall construct a doubly robust estimator of ψ_0 that is guaranteed to be consistent and asymptotically normal if either, but not necessarily both, of these working models is correct.

To do so, we adopt the notational convention introduced in § 1 that given a function such as $U(\psi_0; d, h^\dagger)$ which depends on the unknown law $h(Y, A | L; \psi_0)$, we let $U(\psi, \theta, \alpha; d, h^\dagger)$ be the function $U(\psi_0; d, h^\dagger)$ evaluated at the law $\{\gamma(y, a, l; \psi), f(y | 0, l; \theta), g(a | 0, l; \alpha)\}$. Then, Theorem 2 shows that, under standard regularity conditions, $\hat{\psi} \equiv \hat{\psi}(d; \hat{h}^\dagger)$ is doubly robust, where $\hat{\psi}(d; \hat{h}^\dagger)$ is the solution to

$$P_n[U\{\psi, \hat{\theta}(\psi), \hat{\alpha}(\psi); d, \hat{h}^\dagger\}] = 0,$$

$d(Y, A, L)$ is a user-supplied function,

$$\hat{\alpha}(\psi) = \arg \max_{\alpha} \sum_{i=1}^n \log\{g(A_i | Y_i, L_i; \psi, \alpha)\}$$

is the profile maximum likelihood estimator of α at a fixed ψ , $g(A | Y, L; \psi, \alpha) = \gamma(Y, A, L; \psi)g(A | Y = 0, L; \alpha) / \int g(a | Y = 0, L; \alpha)\gamma(Y, a, L; \psi)d\mu(a)$,

$$\hat{\theta}(\psi) = \arg \max_{\theta} \sum_{i=1}^n \log\{f(Y_i | A_i, L_i; \psi, \theta)\}$$

is the profile maximum likelihood estimator of θ at a fixed ψ , $f(Y | A, L; \psi, \theta) = \gamma(Y, A, L; \psi)f(Y | A = 0, L; \theta) / \int \gamma(y, A, L; \psi)f(y | A = 0, L; \theta)d\mu(y)$, and $\hat{h}^\dagger(Y, A | L) \equiv f^\dagger(Y | L; \hat{\omega}_f)g^\dagger(A | L, \hat{\omega}_g)$. Here $f^\dagger(Y | L; \hat{\omega}_f)$ is a user-specified density when $\hat{\omega}_f$ is chosen to be nonrandom, and $f^\dagger(Y | L; \hat{\omega}_f)$ is a user-supplied parametric model $f^\dagger(Y | L; \omega_f)$ for the density of $Y | L$ evaluated at $\hat{\omega}_f$ maximizing $\prod_{i=1}^n f^\dagger(Y_i | L_i; \omega_f)$, otherwise. Similarly, $g^\dagger(A | L, \hat{\omega}_g)$ is a user-specified density when $\hat{\omega}_g$ is chosen nonrandom and $g^\dagger(A | L, \hat{\omega}_g)$ is a user-supplied parametric model $g^\dagger(A | L; \omega_g)$ for the density of $A | L$, evaluated at $\hat{\omega}_g$ maximizing $\prod_{i=1}^n f^\dagger(A_i | L_i; \omega_g)$, otherwise.

THEOREM 2. *Suppose $\hat{h}^\dagger(Y, A | L)$ converges in probability to an admissible independence density $h^\dagger(Y, A | L)$. Then subject to the regularity conditions given in the Appendix, under the union model \mathcal{B} characterized by (1) and the assumption that either the model $f(y | L, A = 0; \theta)$ or $g(a | Y = 0, L; \alpha)$ is correct, $n^{1/2}(\hat{\psi} - \psi_0)$ is regular and asymptotically linear, with influence function*

$$E \left[\frac{\partial}{\partial \psi} M\{\psi, \theta^*(\psi_0), \alpha^*(\psi_0); d, h^\dagger\} \Big|_{\psi=\psi_0} \right]^{-1} M\{\psi_0, \theta^*(\psi_0), \alpha^*(\psi_0); d, h^\dagger\} \quad (4)$$

and thus converges in distribution to $N(0, \Sigma)$, where

$$\Sigma = E \left\{ \left(E \left[\frac{\partial}{\partial \psi} M\{\psi, \theta^*(\psi_0), \alpha^*(\psi_0); d, h^\dagger\} \Big|_{\psi=\psi_0} \right]^{-1} M\{\psi_0, \theta^*(\psi_0), \alpha^*(\psi_0); d, h^\dagger\} \right)^{\otimes 2} \right\}$$

with $\theta^*(\psi)$ and $\alpha^*(\psi)$ denoting the probability limits of $\hat{\theta}(\psi)$ and $\hat{\alpha}(\psi)$, respectively, and

$$M(\psi, \theta, \alpha; d, h^\dagger) = U(\psi, \theta, \alpha; d, h^\dagger) - E \left\{ \frac{\partial}{\partial \theta} U(\psi, \theta, \alpha; d, h^\dagger) \right\} E \left\{ \frac{\partial}{\partial \theta} C(\psi, \theta) \right\}^{-1} C(\psi, \theta) - E \left\{ \frac{\partial}{\partial \alpha} U(\psi, \theta, \alpha; d, h^\dagger) \right\} E \left\{ \frac{\partial}{\partial \alpha} B(\psi, \alpha) \right\}^{-1} B(\psi, \alpha), \tag{5}$$

where $C(\psi, \theta) = \frac{\partial}{\partial \theta} \log f(Y | A, L; \psi, \theta)$ and $B(\psi, \alpha) = \frac{\partial}{\partial \alpha} \log \{g(A | Y, L; \psi, \alpha)\}$ are the scores for θ and α , respectively.

A consistent estimator of Σ is

$$\hat{\Sigma} = n^{-1} \sum_{i=1}^n \left(\left[n^{-1} \sum_{j=1}^n \frac{\partial}{\partial \psi} \hat{M}_j \{ \psi, \hat{\theta}(\hat{\psi}), \hat{\alpha}(\hat{\psi}); d, \hat{h}^\dagger \} |_{\psi=\hat{\psi}} \right]^{-1} \hat{M}_i \{ \hat{\psi}, \hat{\theta}(\hat{\psi}), \hat{\alpha}(\hat{\psi}); d, \hat{h}^\dagger \} \right)^{\otimes 2},$$

where \hat{M} is defined as M but with expectations replaced by their empirical version. Thus, $\hat{\Sigma}$ can easily be used to obtain Wald-type confidence intervals for components of ψ_0 .

Remark. When $\hat{\omega}_f$ and/or $\hat{\omega}_g$ are random, the asymptotic distribution of $\hat{\psi}(d; \hat{h}^\dagger)$ is equal to that of $\hat{\psi}(d; h^\dagger)$ with $h^\dagger = f^\dagger \times g^\dagger = f^\dagger(Y | L) \times g^\dagger(A | L)$ the probability limit of $\hat{f}^\dagger \times \hat{g}^\dagger = f^\dagger(Y | L; \hat{\omega}_f) \times g^\dagger(A | L, \hat{\omega}_g)$. In practice, it will be convenient to use an estimated density \hat{h}^\dagger rather than a fixed choice h^\dagger .

4. LOCAL EFFICIENCY

We first consider the case in which both A and Y contain continuous components. Chen’s estimator $\hat{\psi}_{\text{eff}}$ solving $P_n\{\hat{S}_{\text{eff}}(\psi)\} = 0$ is locally efficient in model \mathcal{B} at the intersection submodel. However, as previously noted, the estimated efficient score $\hat{S}_{\text{eff}}(\psi)$ does not exist in closed form when $\psi \neq 0$ and the alternating conditional expectations algorithm is needed to compute $\hat{S}_{\text{eff}}(\psi)$ and thus $\hat{\psi}_{\text{eff}}$. In this section, we propose estimators that exploit our representation of the set (3) and thus are easier to compute than $\hat{\psi}_{\text{eff}}$, and yet are nearly locally efficient, i.e. have asymptotic variance almost equal to that of $\hat{\psi}_{\text{eff}}$ at the intersection submodel. The first estimator $\hat{\psi}(d_{\text{ind}}, \hat{h}_{\text{ind}}^\dagger)$ is the easiest to compute, although its asymptotic variance is close to that of $\hat{\psi}_{\text{eff}}$ only when all components of ψ_0 are close to zero; nonetheless $\hat{\psi}(d_{\text{ind}}, \hat{h}_{\text{ind}}^\dagger)$ will be useful in practice, because in many epidemiologic studies, the investigator will know from previously published results that all components of ψ_0 are small. Specifically, we set $d_{\text{ind}}(Y, A, L) \equiv [\partial \log \{\gamma^\top(Y, A, L; \psi)\} / \partial \psi] |_{\psi=0}$ and $\hat{h}_{\text{ind}}^\dagger(Y, A | L) = \hat{f}_{\text{ind}}^\dagger(Y | L) \hat{g}_{\text{ind}}^\dagger(A | L)$, with $\hat{f}_{\text{ind}}^\dagger(Y | L) \equiv f\{Y | L, A; \psi = 0, \hat{\theta}(\psi = 0)\}$ and $\hat{g}_{\text{ind}}^\dagger(A | L) \equiv g\{A | Y, L; \psi = 0, \hat{\alpha}(\psi = 0)\}$. When the true parameter ψ_0 is 0 and thus A and Y are independent given L , $\hat{\psi}_{\text{eff}}$ and $\hat{\psi}(d_{\text{ind}}, \hat{h}_{\text{ind}}^\dagger)$ have identical limiting distributions under the union model B . This result follows from the fact that, when $\psi_0 = 0$, $S_{\text{eff}}(\psi) = d_{\text{ind}}(Y, A, L) - d_{\text{ind}}^\dagger(Y, A, L)$ with $h^\dagger(Y, A | L)$ equal to the true density $h(Y, A | L)$ (Chen, 2007). By continuity, the asymptotic variances of $\hat{\psi}(d_{\text{ind}}, \hat{h}_{\text{ind}}^\dagger)$ and $\hat{\psi}_{\text{eff}}$ will be close, whenever ψ_0 is nearly zero.

When ψ_0 is not known to be nearly zero, we adopt a general approach proposed by Newey (1993). We take a basis system $\phi_j(A, Y, L)$ ($j = 1, \dots$) of functions dense in L_2 , such as tensor products of trigonometric, wavelets or polynomial bases when the components of A, Y and L are all continuous. For some finite $K > \dim(\psi)$, we form the K -dimensional

vector $U(\psi; \tilde{\phi}_K, h^\dagger)$ with $\tilde{\phi}_K$ the vector of the first K basis functions and let $\hat{W}_K(\psi) \equiv U\{\psi, \hat{\theta}(\psi), \hat{\alpha}(\psi); \tilde{\phi}_K, \hat{h}^\dagger\}$, and $\hat{\Gamma}_K(\tilde{\psi}) = \sum_{i=1}^n \hat{W}_{K,i}(\tilde{\psi}) \hat{W}_{K,i}^\top(\tilde{\psi})$, where $\tilde{\psi}$ is any preliminary doubly robust estimator of ψ_0 . Let $\tilde{\psi}_{K,\text{eff}} \equiv \tilde{\psi}_{K,\text{eff}}(\tilde{\phi}_K, \hat{h}^\dagger)$ be the minimizer of the quadratic form $\{\sum_{i=1}^n \hat{W}_{K,i}(\psi)\}^\top \{\hat{\Gamma}_K(\tilde{\psi})\}^- \{\sum_{i=1}^n \hat{W}_{K,i}(\psi)\}$ with $\{\hat{\Gamma}_K(\tilde{\psi})\}^-$ a generalized inverse of $\hat{\Gamma}_K(\tilde{\psi})$. Then, $\tilde{\psi}_{K,\text{eff}} \equiv \tilde{\psi}_{K,\text{eff}}(\tilde{\phi}_K, \hat{h}^\dagger)$ is consistent and asymptotically normal in the semiparametric union model \mathcal{B} ; furthermore, with K chosen sufficiently large, the asymptotic variance of $n^{1/2}(\tilde{\psi}_{K,\text{eff}} - \psi_0)$ nearly attains the semiparametric efficiency bound for the union model at the intersection submodel with both nuisance models correct. In particular, the inverse of the asymptotic variance of $\tilde{\psi}_{K,\text{eff}}$ at the intersection submodel is

$$\begin{aligned} \Omega_K &= \left[E \left\{ \frac{\partial}{\partial \psi^\top} U(\psi; \tilde{\phi}_K, h^\dagger) \Big|_{\psi=\psi_0} \right\} \right]^\top \Gamma_K^- E \left\{ \frac{\partial}{\partial \psi^\top} U(\psi; \tilde{\phi}_K, h^\dagger) \Big|_{\psi=\psi_0} \right\} \\ &= E \{ S_\psi U^\top(\psi_0; \tilde{\phi}_K, h^\dagger) \} \Gamma_K^- \{ E \{ S_\psi U^\top(\psi_0; \tilde{\phi}_K, h^\dagger) \} \}^\top \end{aligned}$$

where Γ_K^- is a generalized inverse of $\Gamma_K = E\{U^\top(\psi_0; \tilde{\phi}_K, h^\dagger)U(\psi_0; \tilde{\phi}_K, h^\dagger)\}$. Thus, Ω_K is the variance of the population least squares regression of S_ψ on the linear span of $U(\psi_0; \tilde{\phi}_K, h^\dagger)$. By $\tilde{\phi}_K$ dense in L_2 , as $K \rightarrow \infty$, the components of $U(\psi_0; \tilde{\phi}_K, h^\dagger)$ become dense in $\Lambda_{\text{nuis}}^\perp$ so that $\Omega_{K \rightarrow \infty} \rightarrow \|\Pi(S_\psi | \Lambda_{\text{nuis}}^\perp)\|^2 = \text{var}(S_{\psi,\text{eff}})$, the semiparametric information bound for estimating ψ_0 under model \mathcal{B} .

Neither of these two strategies is needed if Y or A have finite support as an explicit form for the efficient score in this case was given by Bickel et al. (1993). Without loss of generality, assume Y has finite support say $\{y_0, y_1, \dots, y_{M-1}\}$, with $y_0 = 0$. In the following, we use the result obtained by Bickel et al. (1993) to construct a doubly robust locally-efficient estimating function in model \mathcal{B} . We then demonstrate that this estimating function is in fact a particular member of the class of estimating functions in § 3. For clarity of exposition, this demonstration is restricted to the case of dichotomous Y , but can be easily extended to Y with arbitrary finite support.

Consider the vector $\{I(Y = y_1), \dots, I(Y = y_{M-1})\}$ which we again denote by Y . Next, let $\Psi(A, L; \psi_0) = E\{\epsilon(\psi_0)^{\otimes 2} | A, L\}$ and $k \mapsto \tilde{U}(\psi_0; k) = [k(A, L) - \tilde{E}\{k(A, L) | L; \psi_0\}] \times \epsilon(\psi_0)$ be a function that maps the space of $p \times M - 1$ matrix functions of A and L into L_2 , where $\tilde{E}\{k(A, L) | L; \psi_0\} = E\{k(A, L) \times \Psi(A, L; \psi_0) | L\} \times E\{\Psi(A, L; \psi_0) | L\}^{-1}$ and $\epsilon(\psi_0) = Y - E(Y | A, L; \psi_0)$. Then, by Theorem A.4.5 of Bickel et al. (1993), the closed linear set $\{\tilde{U}(\psi_0; k) : k = k(A, L) \text{ unrestricted}\} \cap L_2$ as k varies over the set of all $p \times (M - 1)$ -dimensional functions of A and L is equal to the set $\Lambda_{\text{nuis}}^\perp$ for model \mathcal{A} .

Furthermore, Bickel et al. (1993) show that $\tilde{U}\{\psi_0; k_{\text{eff}}(\psi_0)\}$ is the efficient score function of ψ in model \mathcal{A} , where $k_{\text{eff}}(\psi_0)$ equals $k_{\text{eff}}(\psi_0) = [\partial \log\{\rho^\top(A, L; \psi)\} / \partial \psi] |_{\psi=\psi_0}$, with $\rho(A, L; \psi)$ defined to be the $(M - 1) \times 1$ vector with the j th component equal to $\gamma(y_j, A, L, \psi)$, $j = 1, \dots, M - 1$. Robins and Rotnitzky (2001) prove that the efficient score in models \mathcal{A} and \mathcal{B} is identical at the intersection submodel. Therefore, a doubly robust, locally efficient at the intersection submodel, estimator of ψ_0 in model \mathcal{B} is obtained by solving either $\sum_{i=1}^n \tilde{U}_i\{\psi; k_{\text{eff}}(\psi), \hat{\theta}(\psi), \hat{\alpha}(\psi)\} = 0$, or $\sum_{i=1}^n \tilde{U}_i\{\psi; k_{\text{eff}}(\hat{\psi}_{\text{mle}}), \hat{\theta}(\psi), \hat{\alpha}(\psi)\} = 0$, where $\tilde{U}(\psi; k_{\text{eff}}, \theta, \alpha)$ is equal to the function $\tilde{U}(\psi; k_{\text{eff}})$ evaluated at the law $\{\gamma(y, a, l; \psi), f(y | A = 0, l; \theta), g(a | Y = 0, l; \alpha)\}$ and $(\hat{\psi}_{\text{mle}}, \hat{\theta}_{\text{mle}}, \hat{\alpha}_{\text{mle}})$ is the maximum likelihood estimator in the parametric model $h(A, Y | L; \psi, \alpha, \theta)$ for $h(A, Y | L)$.

We next derive a doubly robust locally-efficient estimating function $U(\psi, \theta, \alpha; d_{\text{eff}}, h^\dagger)$ in our class that equals $\tilde{U}(\psi; k_{\text{eff}}, \theta, \alpha)$, in the special case where Y is dichotomous. This case is of particular interest as model \mathcal{A} is then equivalent to the familiar semiparametric logistic regression

model

$$\text{logit}\{\Pr(Y = 1 \mid A, L; \psi_0) = \log\{\gamma(1, A, L; \psi_0)\} + \eta(L)$$

with $y_1 = 1$ and $\eta(L) = \log[\Pr(Y = 1 \mid A = 0, L)/\{1 - \Pr(Y = 1 \mid A = 0, L)\}]$ is an unrestricted function of L . Since Y is binary, any function $d(Y, A, L)$ may be written as $Ym(A, L) + n(A, L)$ with $m(A, L) = d(1, A, L) - d(0, A, L)$ and $n(A, L) = d(0, A, L)$. Given an admissible independence density $h^\dagger(Y, A \mid L) = f^\dagger(Y \mid L)g^\dagger(A \mid L)$, let $r \mapsto V(\psi_0; r, h^\dagger) = \{r(A, L) - r^\dagger(L)\} \times (-1)^{1-Y}g^\dagger(A \mid L)/h(Y, A \mid L; \psi_0)$, be a function that maps the space of p -dimensional functions of A and L into L_2 , where $r^\dagger(L) \equiv E^\dagger\{r(A, L) \mid L\}$. For a given choice of h^\dagger and $d(Y, A, L)$, $U(\psi_0; d, h^\dagger)$ simplifies to $V(\psi_0; r, h^\dagger)$ with $r(A, L) = m(A, L)f^\dagger(1 \mid L)\{1 - f^\dagger(1 \mid L)\}$.

Furthermore, by

$$\frac{(-1)^{1-Y}}{h(Y, A \mid L; \psi_0)} = \frac{\{Y - \Pr(Y = 1 \mid A, L; \psi_0)\}}{\text{var}(Y \mid A, L; \psi_0) \int h(y, A \mid L; \psi_0) d\mu(y)},$$

we have that

$$V(\psi_0; r, h^\dagger) = \frac{\{r(A, L) - r^\dagger(L)\}g^\dagger(A \mid L)}{\text{var}(Y \mid A, L; \psi_0) \int h(y, A \mid L) d\mu(y)} \times \epsilon(\psi_0).$$

Thus, since $S_{\text{eff}} = \tilde{U}(k_{\text{eff}})$ is the efficient score, we conclude that $V\{\psi_0; r_{\text{eff}}(h^\dagger; \psi_0), h^\dagger\} = S_{\text{eff}}$ with

$$r_{\text{eff}}(h^\dagger; \psi_0) \equiv \frac{g(A \mid L)}{g^\dagger(A \mid L)} \times \text{var}(Y \mid A, L; \psi_0) \times [k_{\text{eff}}(\psi_0) - E\{k_{\text{eff}}(\psi_0) \times \text{var}(Y \mid A, L; \psi_0) \mid L\} \times E\{\text{var}(Y \mid A, L; \psi_0) \mid L\}^{-1}].$$

Therefore, the solution to either of the following estimating equations is doubly robust locally semiparametric efficient $\sum_{i=1}^n U_i[\psi, \hat{\theta}(\psi), \hat{\alpha}(\psi); d_{\text{eff}}\{\psi, \hat{\theta}(\psi), \hat{\alpha}(\psi), \hat{h}^\dagger\}, \hat{h}^\dagger] = 0$, or $\sum_{i=1}^n U_i\{\psi, \hat{\theta}(\psi), \hat{\alpha}(\psi); d_{\text{eff}}(\hat{\psi}_{\text{mle}}, \hat{\theta}_{\text{mle}}, \hat{\alpha}_{\text{mle}}, \hat{h}^\dagger), \hat{h}^\dagger\} = 0$ where $d_{\text{eff}}(Y, A, L; \psi, \theta, \alpha, \hat{h}^\dagger) = Yr_{\text{eff}}(\hat{h}^\dagger; \psi, \theta, \alpha)$, $\hat{\alpha}(\psi)$ as defined earlier and $\hat{\theta}(\psi) = \arg \max_{\theta} \sum_{i=1}^n [Y_i \log\{b(A_i, L_i; \psi, \theta)\} + (1 - Y_i) \log\{1 - b(A_i, L_i; \psi, \theta)\}]$ with $\text{logit}\{b(A, L; \psi, \theta)\} = \log\{\gamma(1, A, L; \psi)\} + \eta(L; \theta)$. More precisely, each solution is regular and asymptotically linear under model \mathcal{B} and attains the semiparametric efficiency bound for the model at the intersection submodel.

5. DISCUSSION

Although the common variation independent parameterization of $h(A, Y \mid L)f_L(L)$ under model \mathcal{A} with Y binary is (ψ, f, g, f_L) with $f_L = f_L(l)$, $f = f(y \mid l, A = 0)$ and $g = g(a \mid l)$, we instead used the parameterization of Chen (2007) that has $g = g(a \mid Y = 0, l)$ rather than $g = g(a \mid l)$. Our use of Chen's parameterization was the key to our obtaining the doubly robust estimating functions for ψ and hence doubly robust estimators. Formally, following Robins and Rotnitzky (2001), a function $S(\psi, f^*, g^*) = s(O; \psi, f^*, g^*)$ of a single subject's data O is said to be doubly robust for ψ under a particular parameterization for model \mathcal{A} if, when either, but not necessarily both $f = f^*$ or $g = g^*$, (i) $E_{\psi, f, g, f_L}\{S(\psi; f^*, g^*)\} = 0$ and $\text{var}_{\psi, f, g, f_L}\{S(\psi; f^*, g^*)\} < \infty$ for all ψ and (ii) $\partial[E_{\psi^*, f, g, f_L}\{S(\psi; f^*, g^*)\}/\partial\psi]_{\psi=\psi^*} \neq 0$ for all ψ^*, f, g, f_L . Part (ii) guarantees power against local alternatives. As shown in the Appendix, $U(\psi; \theta, \alpha, d, h^\dagger)$ and $\tilde{U}(\psi; k_{\text{eff}}, \theta, \alpha)$ satisfy this definition under Chen's parameterization with $f^*(y \mid l, A = 0) = f(y \mid l, A = 0; \theta)$ and $g^*(a \mid Y = 0, l) = g(a \mid Y = 0, l; \alpha)$. In contrast, no

doubly robust estimating function for ψ exists under the common parameterization. In fact, the following result holds.

THEOREM 3. *Under the common parameterization by (ψ, f, g, f_L) with $f_L = f_L(l)$, $f = f(l) = f(y | l, A = 0)$ and $g = g(a | l)$, there does not exist a doubly robust estimating function $S(\psi, f^*, g^*) = s(O; \psi, f^*, g^*)$ in model \mathcal{A} with Y binary characterized by the sole restriction (1).*

In the Appendix, we prove this result for discrete A , thereby avoiding technicalities that arise in the continuous case.

ACKNOWLEDGEMENT

Andrea Rotnitzky and James Robins were funded by grants from the U.S. National Institutes of Health. The authors wish to thank the reviewers for helpful comments. Andrea Rotnitzky is also affiliated with the Harvard School of Public Health.

APPENDIX

Proof of Theorem 2. We assume that the regularity conditions of Theorem 1A of Robins et al. (1992) hold for $U(\psi_0; \theta, \alpha, h^\dagger)$, $C(\psi_0, \theta)$ and $B(\psi_0, \alpha)$ and that $E[\partial M\{\psi, \theta^*(\psi), \alpha^*(\psi); d, h^\dagger\} / \partial \psi |_{\psi=\psi_0}]$ is nonsingular. We first show that $E\{U(\psi_0; \theta^*(\psi_0), \alpha^*(\psi_0), h^\dagger)\} = 0$ when the data are generated under either $f(Y | A, L; \psi_0, \theta_0)$ or $f(A | Y, L; \psi_0, \alpha_0)$. By symmetry, it is enough to consider the case where the data were generated under $f(Y | A, L; \psi_0, \theta_0)$. Under standard conditions guaranteeing the consistency of the maximum likelihood estimator, $\theta^*(\psi_0) = \theta_0$. Now, under $f(Y | A, L; \psi_0, \theta_0)$,

$$\begin{aligned} & E[U\{\psi_0; \theta_0, \alpha^*(\psi_0), h^\dagger\}] \\ &= E[f^\dagger(Y | L)g^\dagger(A | L)\{d(Y, A, L) - d^\dagger(Y, A, L)\} / h\{Y, A | L; \psi_0, \theta_0, \alpha^*(\psi_0)\}] \\ &= E\left(\frac{g^\dagger(A | L)}{\int h\{u, A | L; \psi_0, \theta_0, \alpha^*(\psi_0)\}d\mu(u)} E\left[\frac{f^\dagger(Y | L)}{h(Y | A, L; \psi_0, \theta_0)}\{d(Y, A, L) - d^\dagger(Y, A, L) | A, L\}\right]\right) \\ &= E\left[\frac{g^\dagger(A | L)}{\int h\{u, A | L; \psi_0, \theta_0, \alpha^*(\psi_0)\}d\mu(u)} \int f^\dagger(y | L)\{d(y, A, L) - d^\dagger(y, A, L)\}d\mu(y)\right] = 0, \end{aligned}$$

since

$$\begin{aligned} & \int f^\dagger(y | L)\{d(y, A, L) - d^\dagger(y, A, L)\}d\mu(y) \\ &= \int f^\dagger(y | L)d(y, A, L)d\mu(y) - \int d(y, A, L)f^\dagger(y | L)d\mu(y) \\ &\quad - \int d(y, a, L)f^\dagger(y | L)g^\dagger(a | L)d\mu(a, y) + \int d(y, a, L)g^\dagger(a | L)f^\dagger(y | L)d\mu(y, a) = 0. \end{aligned}$$

Then, under the assumed regularity conditions the formulae (4) and (5) follow from standard Taylor series arguments, whenever $E[\partial M\{\psi, \theta^*(\psi), \alpha^*(\psi); d, h^\dagger\} / \partial \psi |_{\psi=\psi_0}]$ is nonsingular. The asymptotic normality result follows from the standard application of Slutsky's theorem and the central limit theorem. \square

Proof of Theorem 3. The proof is by contradiction: if $S(\psi, f^*, g^*)$ were doubly robust, then, for every f^* , $S(\psi, f^*, g^*)$ would be an unbiased estimating function for ψ with power against local alternatives in the submodel \mathcal{A}_{g^*} of model \mathcal{A} in which $g = g^*$ is known a priori. Hence, it suffices to prove that model \mathcal{A}_{g^*} does not admit such unbiased estimating functions. Noting that model \mathcal{A}_{g^*} can be parameterized by (ψ, f, f_L) , we need to prove there is no function $Q(\psi) = q(O; \psi)$ such that $E_{\psi, f, f_L}\{Q(\psi)\} = 0$,

$\text{var}_{\psi, f, f_L} \{Q(\psi)\} < \infty$ and $\partial[E_{\psi^*, f, f_L} \{Q(\psi)\}] / \partial \psi|_{\psi=\psi^*} \neq 0$ for all ψ^*, f, f_L . Now Bickel et al. (2003) proved that an unbiased estimating function for a parameter ψ lies in the orthocomplement to the nuisance tangent space for the model. For model \mathcal{A}_{g^*} , it is straightforward to show that the orthocomplement to the nuisance tangent space at law (ψ, f, f_L) , say $\Lambda_{\text{nuis}, g^*}^\perp(\psi, f)$, does not depend on f_L and is the direct sum of the orthocomplement for model \mathcal{A} plus the space of functions $v = v(A, L)$ of (A, L) with zero-mean given L . Thus,

$$\Lambda_{\text{nuis}, g^*}^\perp(\psi, f) = [T(k, v; \psi, f) = \tilde{U}(\psi, f; k) + v(A, L); k \text{ unrestricted, } v \text{ with } E\{v(A, L) | L\} = 0] \cap L_2,$$

where $\tilde{U}(\psi, f; k) = \epsilon(\psi, f)[k(A, L) - \tilde{E}\{k(A, L) | L; \psi, f\}]$ is $\tilde{U}(\psi; k)$ defined in §4 with the dependence on f now made explicit. Suppose $Q(\psi)$ existed. Then $Q(\psi) = T(k_f, v_f; \psi, f) \in \Lambda_{\text{nuis}, g^*}^\perp(\psi, f)$ holds for each f , where k_f and v_f are the particular functions k and v associated with a given f . Thus, $T(k_f, v_f; \psi, f) = T(k_{f^*}, v_{f^*}; \psi, f^*)$ for any f, f^* . Noting $\epsilon(\psi, f) \equiv Y - \text{pr}(Y = 1 | A, L; \psi, f)$, the previous equality implies that $Y\{k_f(A, L) - \tilde{E}\{k_f(A, L) | L; \psi, f\} - k_{f^*}(A, L) - \tilde{E}\{k_{f^*}(A, L) | L; \psi, f^*\}]$ is equal to a function that does not depend on Y . Hence, it must be that

$$k_f(A, L) - \tilde{E}\{k_f(A, L) | L; \psi, f\} = k_{f^*}(A, L) - \tilde{E}\{k_{f^*}(A, L) | L; \psi, f^*\}.$$

Thus, for a function $r(L)$, $k_{f^*}(A, L) = k_f(A, L) + r(L)$. Substituting for $k_{f^*}(A, L)$ in the last display we obtain $k_f(A, L) - \tilde{E}\{k_f(A, L) | L; \psi, f\} = k_f(A, L) + r(L) - \tilde{E}\{k_f(A, L) | L; \psi, f^*\} - r(L)$ and hence $\tilde{E}\{k_f(A, L) | L; \psi, f\} = \tilde{E}\{k_f(A, L) | L; \psi, f^*\}$ with probability one for all f, f^*, ψ which, as shown in the next paragraph, implies $k_f(A, L)$ is not a function of A , i.e. $k_f(A, L) = \tilde{k}(L)$ with probability one for some $\tilde{k}(L)$. But $k_f(A, L)$ not a function of A implies $\partial[E_{\psi^*, f, f_L} \{Q(\psi)\}] / \partial \psi|_{\psi=\psi^*} = 0$, which is a contradiction. We conclude that no unbiased estimating function $Q(\psi)$ with power against local alternatives exists. We show that, for A binary, $\tilde{E}\{h(A, L) | L; \psi, f\}$ depends on f on a set of nonzero probability whenever the conditional odds ratio function $\gamma(1, 1, L; \psi) \neq 1$ with probability one, and $h(A, L) = h_1(L)A + h_0(L)$ depends on A , i.e. whenever $h_1(L)$ is nonzero with positive probability. Let $f(l)$ denote $f(1 | A = 0, l)$. When $\gamma(1, 1, L; \psi) \neq 1$ with probability one, $\tilde{E}[A | L; \psi, f] = \tilde{E}[A | L; \psi, f] = [1 + \{g(A = 0 | L) / g(A = 1 | L)\} \{1 - f(L) + \gamma(1, 1, L) f(L)\}^2 / \gamma(1, 1, L)]^{-1}$ obviously depends on $f(L)$ with probability one which implies $\tilde{E}\{h(A, L) | L; \psi, f\}$ depends on f on the set where $h_1(L)$ is nonzero. The proof for arbitrary discrete A is identical except that extra bookkeeping is required. \square

REFERENCES

- BICKEL, P., KLASSEN, C., RITOV, Y. & WELLNER, J. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer.
- CHEN, H. Y. (2003). A note on prospective analysis of outcome-dependent samples. *J. R. Statist. Soc. B* **65**, 575–84.
- CHEN, H. Y. (2004). Nonparametric and semiparametric models for missing covariates in parametric regression. *J. Am. Statist. Assoc.* **99**, 1176–89.
- CHEN, H. Y. (2007). A semiparametric odds ratio model for measuring association. *Biometrics* **63**, 413–21.
- NEWBY, W. (1993). Efficient estimation of models with conditional moment restrictions. In *Handbook of Statistics, IV*, Ed. G. S. Maddala, C. R. Rao and H. Vinod, pp. 427–61. Amsterdam: Elsevier Science.
- ROBINS, J. M., MARK, S. D. & NEWBY, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48**, 479–95.
- ROBINS, J. M. & ROTNITZKY, A. (2001). Comment on the Bickel and Kwon article, ‘Inference for semiparametric models: some questions and an answer’. *Statist. Sinica* **11**, 920–36.
- VANSTEELENDT, S., VANDERWEELE, T., TCHETGEN, E. J. & ROBINS, J. M. (2008). Semiparametric inference for statistical interactions. *J. Am. Statist. Assoc.* **103**, 1693–704.

[Received September 2008. Revised May 2009]