

Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models

BY JAMES ROBINS

*Department of Epidemiology, Harvard School of Public Health, 655 Huntington Avenue,
Boston, Massachusetts 02115, U.S.A.*

robins@hsph.harvard.edu

AND ANDREA ROTNITZKY

*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue,
Boston, Massachusetts 02115, U.S.A.*

andrea@hsph.harvard.edu

SUMMARY

We consider estimation of the received treatment effect on a dichotomous outcome in randomised trials with non-compliance. We explore inference about the parameters of the structural mean models of Robins (1994, 1997) and Robins et al. (1999). We show that, in contrast to the additive and multiplicative structural mean models for continuous and count outcomes, unbiased estimating functions for a nonzero (structural) treatment effect parameter do not exist in the presence of many continuous and discrete baseline covariates, even when the randomisation probabilities are known. The best that can be hoped for are estimators, such as those proposed in this paper, that are guaranteed both to estimate consistently the (null) treatment effect when the null hypothesis of no treatment effect is true and to have small bias when the true treatment effect is close to but not equal to zero.

Some key words: Exclusion restriction; Non-compliance; Structural mean model.

1. INTRODUCTION

In this paper we consider the estimation of the effect of received treatment in randomised clinical trials with non-compliance and a dichotomous outcome using structural mean models. We allow both for the assigned and received treatments to be continuous, categorical or ordinal and for the possibility that the assigned treatment has a direct effect on the outcome through pathways other than the received treatment. We also consider the application of our results to observational studies.

The parameters ψ of a structural mean model measure, on an appropriate scale, how the effect of the received treatment on the treated population varies across levels of pre-treatment covariates. Thus, these models are useful for assessing whether or not the effect of received treatment is modified by baseline covariates.

Robins (1989, 1994) showed that, when the randomisation probabilities are known, both additive and multiplicative structural mean models that respectively impose a linear

or log link function admit unbiased estimating functions for ψ even in the presence of many continuous and discrete baseline covariates. For dichotomous outcomes, additive and multiplicative models cannot generally be used, because these models fail to guarantee that the predicted response probabilities lie in the interval $(0, 1)$.

In this paper we investigate inference in the dichotomous outcome structural mean models for binary responses of Robins et al. (1999) which use as a link function the inverse of a monotone cumulative distribution function, such as the logit or probit link. We prove that no unbiased estimating equations exist for the parameter ψ of dichotomous outcome structural mean models. It is the absence of unbiased estimating functions that is the root cause of the difficulty in constructing satisfactory estimators of the treatment effect in the treated for dichotomous outcomes within levels of high-dimensional baseline covariates. We argue that the best that can be hoped for are estimators, such as those proposed in this paper, that, even in the presence of high-dimensional baseline covariates, are guaranteed (i) to estimate consistently the (null) treatment effect under the joint causal sharp null hypothesis that neither the assigned treatment nor the received treatment has a causal effect on any subject in the population, and (ii) to have small bias when the true treatment effect is close to but not equal to zero. In the presence of noncompliance the possibility of constructing estimators that satisfy (i) and (ii) is one of the most important benefits to be obtained from a randomised experiment, because using such estimators and their estimated standard error we can construct a test that is simultaneously an asymptotically distribution-free α -level test both of the intention-to-treat null hypothesis that the outcome is conditionally independent of assigned treatment given baseline covariates and of the joint causal sharp null hypothesis defined in (i).

Our work was stimulated by a recent article of Vansteelandt & Goetghebeur (2003). These authors assumed that data were available from a two-armed placebo-controlled trial with a dichotomous outcome in which the exclusion restriction that assigned treatment has no direct effect on the outcome was satisfied. By imposing additional parametric modelling restrictions on the law of the observed data these authors succeeded in constructing a simple estimator of the structural mean model parameter ψ which satisfied (i) and (ii). However, in settings in which some patients in the placebo arm take the active treatment, Vansteelandt & Goetghebeur's (2003) additional model restrictions can be incompatible with the structural mean model itself. In this paper, we provide alternative parametric modelling restrictions on the law of the observed data that are guaranteed always to be compatible with the dichotomous outcome structural mean model and, in addition, allow for the possibility that the exclusion restriction does not hold. We use the model restrictions to construct an estimator of the parameter ψ that satisfies (i) and (ii). Our estimators are more complex than the Vansteelandt & Goetghebeur (2003) estimators and so the latter might be preferable when one has data from a two-armed placebo-controlled trial in which no patient in the placebo arm takes the active treatment and in which the exclusion restriction can be assumed to hold.

Our results address the long-standing problem in econometrics and biostatistics, recently reviewed in Angrist (2001), of how to use a randomly assigned treatment to estimate, within levels of pre-treatment baseline covariates, the effect of actual treatment on a dichotomous response variable. In the econometrics literature, a treatment that (A) is randomly assigned and (B) satisfies the exclusion restriction is called an instrument or instrumental variable for the actual treatment. When data on many continuous and discrete baseline covariates have been obtained, to the best of our knowledge, the only

proposed solutions, other than the proposal of Vansteelandt & Goetghebeur (2003), have been in Abadie (2003) and in Hirano et al. (2000) for the special case in which both assigned and received treatments are themselves dichotomous, there is no defier, i.e. there is no subject who takes treatment if assigned to placebo and takes placebo if assigned to treatment, and the estimand of interest is the effect of treatment among those who comply with their assigned treatment. In contrast to the estimators of Vansteelandt & Goetghebeur and of Abadie, those of Hirano et al. fail to satisfy (i) and (ii) above because they do not use knowledge of the randomisation probabilities; see Robins & Ritov (1997).

To help illustrate the ideas in this paper consider a fictitious though not unrealistic six-month two-arm placebo-controlled double-blind randomised trial of the effect of highly active antiretroviral therapy, HAART, on the CD4 lymphocyte count of asymptomatic HIV infected subjects with CD4 counts between 300 and 500 at baseline. As a result of the potential for side effects and for the development of resistant virus, the therapy is only considered a success if a subject's CD4 count increases by 20% or more over the trial. In this example, the assigned treatment variable is the treatment arm indicator and the dichotomous response is equal to one if CD4 count increases by at least 20% and zero otherwise. Suppose that in the trial each patient is tested at weekly clinic visits for the presence of HAART in the serum. Actual treatment is taken to be the total number of weeks a subject tested positive for HAART over the six months. Baseline covariates which may influence the magnitude of the treatment effect include pre-randomisation levels of serum HIV RNA, CD4 count, age, weight, estimated time since HIV infection, gender, ethnicity and recent venereal disease history. Since HAART therapy is associated with characteristic mild side-effects, most subjects can guess whether or not they have been given active treatment, thus effectively preventing the trial from being double blind. As a consequence, a significant fraction of subjects in the placebo arm may violate protocol and choose to obtain HAART outside the trial. Since the trial is not double blind, the exclusion restriction may not hold. For example, subjects who suspect that they have been assigned to the HAART arm may feel protected and so take fewer precautions and be more likely to contract a new venereal infection. If the immune activation caused by such an infection led to the increased killing of CD4 cells by the HIV virus, this would constitute a direct effect of treatment-arm assignment on the outcome through a pathway other than the received treatment.

Suppose it is logistically or ethically impossible to conduct a randomised experiment, but observational data are available, for example, from Health Maintenance Organization records. In that case, under additional assumptions, our methods could still be used to infer causal effects. For example, suppose that we assume, at least as a working hypothesis, that (a) the prescribed dose of a particular drug had no effect on an outcome of interest except through the dose actually taken by the patient, and (b) within levels of the often high-dimensional vector of prognostic factors that both are used by physicians to decide on the need for the therapy and are recorded in the medical chart, the prescribed dose was 'effectively' randomised in the sense of being independent of any unrecorded prognostic factors for the outcome. Note that (a) and (b) are precisely the aforementioned conditions (A) and (B) defining an instrumental variable, so that under (a) and (b) prescribed dose is an instrument for the actual dose. In contrast to randomised experiments, the conditional probability of this instrument would typically depend in a complex fashion on the high-dimensional vector of prognostic factors. Furthermore, unlike in a randomised study, the probability of the instrument, such as prescribed dose, conditional on the vector of

prognostic factors would not be known and would have to be estimated from the data using a dimension-reducing, usually parametric, model. In § 3.5 we will show that in such a case no estimator satisfying (i) and (ii) will exist and the best that can be hoped for is an estimator that is so-called doubly robust under the joint causal sharp null hypothesis. An estimator is doubly robust under this null hypothesis if it is guaranteed to estimate the (null) treatment effect consistently when one has specified dimensional-reducing models for (I) the conditional mean of the instrument and (II) the conditional mean of the binary outcome given the vector of baseline covariates; and at least one, but not necessarily both, of the models for (I) or (II) is correct.

2. MODEL ASSUMPTIONS AND THE INFERENCE PROBLEM

Suppose we observe n independent and identically distributed copies O_i , for $i = 1, \dots, n$, of the vector $O = (L, Z, X, Y)$, where L is a vector recording pre-randomisation covariates, Z records the randomly assigned treatment, X is the treatment actually received, and Y is the dichotomous outcome of interest which we assume takes the value 0 or 1. Both Z and X can be discrete or continuous variables. We assume that O is absolutely continuous with respect to some dominating measure. Throughout, marginal and conditional densities $f(\cdot)$ and $f(\cdot|\cdot)$ refer to the densities of the random variables indicated by their indices; that is, for any pair of random variables V and W , $f(v)$ and $f(v|w)$ denote the marginal density of V and the conditional density of V given $W = w$ evaluated at $V = v$. Throughout, random variables are denoted by capital letters and their supports by the corresponding calligraphic letter.

Until § 3.5, we make the following assumption.

Assumption 1. The randomisation distribution $f(z|l)$ is known by design.

Assumption 1 holds when, as in a randomised trial, treatment allocation is under the control of the investigator. To define the treatment effect of interest and our model we must first define for each value $z \in \mathcal{Z}$, $x \in \mathcal{X}$ the potential outcome Y_{zx} (Neyman, 1923; Rubin, 1978; Robins, 1986). This is defined as the value of the dichotomous outcome of interest had, possibly contrary to fact, Z been set to z and X set to x . Randomisation guarantees that

$$Z \perp\!\!\!\perp Y_{zx} | L, \quad (1)$$

with probability one for all $z \in \mathcal{Z}$, $x \in \mathcal{X}$, because Y_{zx} can, like genetic make-up, be considered as a generally unobserved baseline characteristic of each subject. Here, for random variables A , B and C , $A \perp\!\!\!\perp B | C$ indicates conditional independence of A and B given C (Dawid, 1979).

We additionally make the consistency assumption that the subject's observed outcome is equal to the potential outcome associated with the subject's assigned and received treatments Z and X . To be specific,

$$Y = Y_{zx} \quad (\text{if } Z = z, X = x). \quad (2)$$

For the subpopulation with covariate level $L = l$, assigned to treatment $Z = z$ and actually treated with $X = x$, let

$$\gamma(l, z, x) = b(zx|z, x, l) - b(z0|z, x, l), \quad (3)$$

where $b(zx|d) := \Phi^{-1}\{E(Y_{zx}|D = d)\}$ and Φ is a known strictly increasing smooth distribution function with support on the real line. For example, if $\Phi(u)$ is the logistic function $e^u/(1 + e^u)$ then

$$\gamma(l, z, x) = \log \left\{ \frac{\text{pr}(Y_{zx} = 1|Z = z, X = x, L = l)}{\text{pr}(Y_{zx} = 0|Z = z, X = x, L = l)} \right\} / \left\{ \frac{\text{pr}(Y_{z0} = 1|Z = z, X = x, L = l)}{\text{pr}(Y_{z0} = 0|Z = z, X = x, L = l)} \right\}$$

is the logarithm of the ratio of the odds of a positive response under the assigned and received treatments over the odds of a positive response had the assigned treatment remained unchanged but the received treatment been withheld. Note that, by definition, $\gamma(l, z, 0) = 0$.

To help interpret $\gamma(l, z, x)$, consider the special case in which $\Phi(u)$ is the logistic function, Z and X are dichotomous and L takes the value 0 with probability one. Define the potential outcome X_z that denotes the treatment that would be received if the assigned treatment were equal to z . Randomisation of assigned treatment implies that $Z \perp\!\!\!\perp (X_z, Y_{zx})|L$; Robins (1989) and Angrist et al. (1996) considered this special setting. Under randomisation, the consistency assumption (2) and the assumption $X_z = X$ if $Z = z$, we have

$$\gamma(l, z, x) = \log \left\{ \frac{\text{pr}(Y_{zx} = 1|X_z = x, L = l)}{\text{pr}(Y_{zx} = 0|X_z = x, L = l)} \right\} / \left\{ \frac{\text{pr}(Y_{z0} = 1|X_z = x, L = l)}{\text{pr}(Y_{z0} = 0|X_z = x, L = l)} \right\}.$$

Angrist et al. (1996) called the subjects with $X_z = z$ for $z = 0, 1$ ‘compliers’, those with $X_z = 1$ for $z = 0, 1$ ‘always takers’, those with $X_z = 0$ for $z = 0, 1$ the ‘never takers’ and those with $X_z = 1 - z$ for $z = 0, 1$ the ‘defiers’, and assumed that there was no defier. Under the Angrist et al. setting, $\gamma(0, 1, 1)$ measures, on the log-odds-ratio scale, the effect of received treatment in the subpopulation consisting of compliers and always takers when they are assigned to treatment $z = 1$. The parameter $\gamma(0, 0, 1)$ measures the received treatment effect in the always takers when they are assigned to treatment $z = 0$. Under the exclusion restriction that for essentially all subjects assigned treatment has no direct effect on the outcome when received treatment is held fixed, i.e.

$$Y_{zx} = Y_x \quad (\text{with probability one, } z \in \mathcal{Z}, x \in \mathcal{X}) \tag{4}$$

when there are neither defiers nor always takers, the parameter $\gamma(0, 1, 1)$ measures the treatment effect in the compliers. Such a parameter is the effect measure studied by Angrist et al. (1996), Abadie (2003) and Hirano et al. (2000).

In § 5 we show that the function $\gamma(l, z, x)$ is not usually nonparametrically identified from the observed data O ; that is, the function $\gamma(l, z, x)$ is not a functional of the joint distribution of the observables O . However, $\gamma(l, z, x)$ is identified provided one imposes sufficient restrictions on it. Therefore, in order to identify $\gamma(l, z, x)$ we shall assume that it follows a parametric model,

$$\gamma(l, z, x) = \gamma(l, z, x; \psi_0), \tag{5}$$

where ψ_0 is an unknown parameter vector of dimension p and $\gamma(l, z, x; \psi)$ is a known function satisfying $\gamma(l, z, 0; \psi) = \gamma(l, z, x; 0) = 0$. Following Robins (1997) and Robins et al. (1999) we call (5) a structural mean model. When $\Phi(u)$ is the logistic function we refer to (5) as a logistic structural mean model. Note that $\gamma(l, z, x; 0) = 0$ implies that if $\psi_0 = 0$ then

$$\text{pr}(Y_{zx} = 1|X_z = x, L = l) = \text{pr}(Y_{z0} = 1|X_z = x, L = l), \tag{6}$$

i.e. that the treatment x has no direct effect, compared to treatment 0, in the subpopulation treated with $X = x$ and assigned to $Z = z$.

To illustrate the structural mean model, we return to the fictitious randomised trial of the introduction. In this trial, for example, the model

$$\gamma(l, z, x; \psi) = x(\psi_1 + \psi_2 z + \psi_3 l_1 + \psi_4 l_2),$$

where l_1 and l_2 record pre-treatment CD4 count and HIV RNA, postulates that the effect on the log odds of a positive response of the actual number of weeks x that HAART therapy was received, as compared to never receiving HAART, among subjects who (a) were randomised to treatment arm z , (b) actually received HAART for x weeks and (c) have values of the pre-treatment covariates equal to l , depends linearly on x but the magnitude of the effect may be modified by treatment arm and the pre-treatment levels of CD4 and HIV RNA. Note that one would in general expect $\gamma(l, z, x; \psi_0)$ to depend on treatment arm z when ψ_0 is nonzero, because, within levels of the pre-treatment covariates L , the subset of the population who take treatment level x when assigned to the placebo arm presumably differs from the subset of the population who take treatment level x when assigned to the HAART arm on many unmeasured factors, some of which may be associated with the magnitude on a logistic scale of response to treatment.

Until § 3.6 we shall assume that we know the direct effect of assigned treatment on the outcome when received treatment is set to zero.

Assumption 2. We assume that $b(z|l) = t(l) + t^*(l, z)$, where $t^*(l, z)$ is a known function satisfying $t^*(l, 0) = 0$ and $t(l)$ is a completely unknown function.

Note that the special choice $t^*(l, z) = 0$, for all z and l , is equivalent to the assumption that $\text{pr}(Y_{z0} = 1|L = l)$ is the same for all z and therefore is implied by, but does not imply, the exclusion assumption (4). Henceforth we refer to $t^*(l, z) = 0$, for all z and l , as the weak exclusion restriction. The function $t^*(\cdot, \cdot)$ would depend on assigned treatment z if, in the fictitious trial of the introduction, the double-blind was effectively subverted and subjects randomised to the active treatment arm took fewer precautions and thereby had lower response rates. Note also that by allowing $t(l, z)$ to depend on baseline covariates l we are allowing for the possibility that these covariates modify the direct effect of treatment assignment. For example, it is possible that men but not women would react to the discovery that they had been assigned to the active treatment by taking fewer precautions, in which case the effect of assigned treatment could be modified by the baseline covariate gender.

In § 3.4, we show that, when the function $t(l)$ is unrestricted, inference about ψ is infeasible because of the curse of dimensionality. As a consequence we will consider estimators of ψ under a dimension-reducing parametric model $t(l; \omega)$ for $t(l)$. For example, in the context of the fictitious HAART trial of the introduction, the model $t(l; \omega) = \omega_1 l_1 + \omega_2 l_2 + \omega_3 l_3$, where L_3 is the gender indicator with $l_3 = 1$ if male, states that the log odds of success, had all subjects in the population been assigned to take placebo and been prevented from having access to HAART, depends linearly on the pre-treatment covariates CD4 count, HIV RNA and gender and on no other baseline covariate. If, in addition, $t^*(l, z) = z(\kappa_1 + \kappa_2 l_3)$ for specified values of κ_1 and κ_2 , regarded as known, then the model additionally postulates that, in the hypothetical scenario in which we could prevent all subjects from having access to HAART, the ratio of the odds of success under

assignment to placebo over the odds of success under assignment to HAART among men and women with the same baseline prognostic factors would be $\exp(\kappa_2)$ times higher for men than for women.

When the weak exclusion assumption is in doubt, one reasonable analytical strategy, the one examined in this paper, is to estimate the parameter ψ_0 regarding $t^*(l, z)$ as known and then to repeat the analysis varying $t^*(l, z)$ over a plausible range as a form of sensitivity analysis. Rather than regarding $t^*(l, z)$ as known one could consider specifying a parametric model $t^*(l, z; \kappa)$ for $t^*(l, z)$ and then jointly estimating the parameters ψ_0 and κ_0 . This strategy would often be infeasible because either ψ_0 and κ_0 will not be jointly identified or they will be so weakly identified that their estimators will be highly correlated and will consequently have very wide associated confidence intervals. In spite of this limitation, for completeness, in § 3 we briefly discuss joint estimation of ψ_0 and κ_0 and in § 5 we provide an example illustrating that ψ_0 and κ_0 are often not jointly identified.

We refer to the semiparametric model defined by data O and the restrictions of Assumption 1, (1), (2), (5) and Assumption 2 as semiparametric model \mathcal{A} .

3. ESTIMATION OF THE STRUCTURAL MEAN PARAMETER

3.1. Overview

Though ideally we would like to make inference about ψ under model \mathcal{A} , unfortunately our key result in § 3.3 states that, because of the curse of dimensionality (Huber, 1985; Robins & Ritov, 1997), this is not feasible unless $\psi_0 = 0$ and the weak exclusion restriction holds. To be specific, in § 3.3 we argue that the results of Theorem 1 imply that, when L has two or more continuous components and $\psi_0 \neq 0$ or $t^*(l, z) \neq 0$, impractically large sample sizes would be required for any semiparametric estimator of ψ to be well behaved. Thus, in order to obtain well-behaved estimators of ψ when $\psi_0 \neq 0$, in § 3.4 we propose to estimate ψ under a parametric model for the distribution of the observed data. However, we do not propose to fit the model by maximum likelihood because the maximum likelihood estimator of ψ is not guaranteed either (i) to estimate ψ_0 consistently when the sharp null hypothesis, $Y_{zx} = Y$ with probability one for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$, of no effect of actual or assigned treatment is true, and hence $\psi_0 = 0$ and the weak exclusion restriction holds, or (ii) to have small bias when the true treatment effect ψ is close to but not equal to zero and the weak exclusion restriction holds. The estimators described in § 3.4 satisfy (i) and (ii). To help understand the root of the difficulty for estimating ψ under model \mathcal{A} , we discuss estimation of additive and multiplicative structural mean models in § 3.2.

3.2. Additive and multiplicative structural mean models for continuous and count outcomes

Additive and multiplicative structural mean models differ from our model \mathcal{A} only in that $\Phi(x) = x$ for the additive model and $\Phi(x) = \exp(x)$ for the multiplicative model.

Let models \mathcal{A}_{add} and $\mathcal{A}_{\text{mult}}$ be the models defined by data O , Assumption 1, (1), (2), (5) and Assumption 2 with $\Phi(x) = x$ in \mathcal{A}_{add} and $\Phi(x) = \exp(x)$ in $\mathcal{A}_{\text{mult}}$. In the Appendix we show that, as models for the law of the observed data, models \mathcal{A}_{add} and $\mathcal{A}_{\text{mult}}$ impose on the observed data law only the restriction of Assumption 1 and the conditional mean independence restriction

$$E\{H_{\text{CMI}}(\psi_0, t^*)|Z, L\} = E\{H_{\text{CMI}}(\psi_0, t^*)|L\}, \quad (7)$$

where $H_{\text{CMI}}(\psi_0, t^*) := \Phi\{\Phi^{-1}(Y) - \gamma(L, Z, X; \psi_0) - t^*(L, Z)\}$. It is well known (Chamberlain, 1987) that the set of influence functions of regular asymptotically linear estimators of ψ_0 in any semiparametric conditional mean independence model (7) comprises random variables that are, up to multiplicative matrix constants, equal to

$$h_{\text{CMI}}^*(O; c, d) = r^c(L, Z)H_{\text{CMI}}(\psi_0, t^*) + r^d(L, Z),$$

where, for any function $u(L, Z)$, $r^u(L, Z) := u(L, Z) - E\{u(L, Z)|L\}$, and $c(L, Z)$ and $d(L, Z)$ are arbitrary $p \times 1$ functions. Since, by Assumption 1, $E\{c(L, Z)|L\}$ and $E\{d(L, Z)|L\}$ are known functions of L , it follows that one can obtain consistent and asymptotically normal estimators of ψ by solving the unbiased estimating equation $\sum_i h_{\text{CMI}}^*(O_i; \psi, c, d) = 0$ for arbitrary functions c and d chosen by the investigator, where $h_{\text{CMI}}^*(O; \psi, c, d)$ is defined as $h_{\text{CMI}}^*(O; c, d)$ except that $H_{\text{CMI}}(\psi_0, t^*)$ is replaced by $H_{\text{CMI}}(\psi, t^*)$. The difficulties in the estimation of the parameter ψ in model \mathcal{A} arise precisely because, in contrast to models \mathcal{A}_{add} and $\mathcal{A}_{\text{mult}}$, model \mathcal{A} is not a conditional mean independence model.

3.3. Demonstration that estimation ψ in model \mathcal{A} is infeasible without additional assumptions

In order to state our main result we will need to consider a reparameterisation of the joint density of the observables O ,

$$f(o) = f(y|l, z, x)f(x|l, z)f(z|l)f(l),$$

in terms of the known function $t^*(l, z)$, the known density $f(z|l)$, the unknown finite-dimensional parameter ψ and an infinite-dimensional parameter θ indexing the unknown function $t(l)$, the unknown densities $f(l)$ and $f(x|l, z)$ and the unknown function

$$q(l, z, x) \equiv b(z0|l, z, x) - b(z0|l, z, 0).$$

The latter is completely unrestricted except by the definitional constraint $q(l, z, 0) = 0$. The function $q(l, z, x)$ can be interpreted as a measure of the degree to which non-compliance is nonrandom. To be specific, suppose that, within treatment arm and levels of the baseline covariates L , received dose was determined by a random mechanism whose probabilities depended only on the treatment arm indicator and the values of the baseline covariates. Then we would say that, within levels of the baseline covariates and treatment arm, non-compliance in the trial was random. In such a case, the received dose X would be conditionally independent of Y_{z0} given L and Z and hence $q(l, z, x)$ would be zero for all l, z and x . Thus, the magnitude of the deviation of the function $q(l, z, x)$ from zero is a measure of the degree to which non-compliance is nonrandom.

In § 3.4 we will show that, when the function $q(l, z, x)$ is unrestricted, inference about ψ is infeasible because of the curse of dimensionality. Consequently, we will describe estimators of ψ under a dimension-reducing parametric model $q(l, z, x; \eta)$ for $q(l, z, x)$. For example, in the context of the fictitious HAART trial of the introduction, the model $q(l, z, x; \eta) = x(\eta_1 + \eta_2 z + \eta_3 l_1 + \eta_4 l_2)$ states that, among the subset of the population with baseline covariates values equal to l in treatment arm z , the difference between the subgroup that tested positive for HAART for x weeks, i.e. those with $X = x$, when assigned to arm z and the subgroup that never tested positive for HAART, i.e. those with $X = 0$, when assigned to arm z in the logit of the probability of success under the hypothetical scenario that these two subgroups were assigned to treatment arm z but were prevented from having access to HAART depends linearly on x with the magnitude of the difference modified by treatment arm z and pre-treatment CD4 and HIV RNA levels.

To reparameterise the likelihood, we express $\text{pr}(Y = 1|l, z, x)$ as a function of $\gamma(l, z, x)$, $f(x|l, z)$, $t(l)$ and $q(l, z, x)$. We do so by noting that

$$\Phi^{-1}\{\text{pr}(Y = 1|l, z, x)\} = \gamma(l, z, x) + q(l, z, x) + b(z0|l, z, 0)$$

and that $b(z0|l, z, 0)$ is not an additional parameter because it is, in fact, a function of the parameters listed above. To be specific, $b(z0|l, z, 0)$ is the unique solution $v(l, z)$ to the integral equation

$$\Phi\{t(l) + t^*(l, z)\} = \int \Phi\{q(l, z, x) + v(l, z)\}dF(x|l, z) \tag{8}$$

since $\Phi\{t(l) + t^*(l, z)\} = E(Y_{z0}|l) = E(Y_{z0}|l, z)$ by (2) and

$$\Phi\{q(l, z, x) + b(z0|l, z, 0)\} = E(Y_{z0}|l, z, x).$$

When X is discrete with s levels and Φ is the logit link, equation (8) is a polynomial equation of order s in $u = \exp\{-v(l, z)\}$. In particular, when X is binary the equation is quadratic in u and it has an explicit solution, which we provide in the Appendix. For continuous X , numerical methods would be required to solve (8).

Theorem 1 below gives the influence functions of regular asymptotically linear estimators of ψ in model \mathcal{A} and forms the basis of our argument of why estimation of ψ in model \mathcal{A} is infeasible when L is high-dimensional. To state Theorem 1, it will be convenient to make the following definitions. Let

$$M_1 := \gamma(L, Z, X; \psi_0) + q(L, Z, X) + v(L, Z), \quad M_2 := q(L, Z, X) + v(L, Z),$$

$$M_3 := t(L) + t^*(L, Z), \quad U := \partial\gamma(L, Z, X; \psi)/\partial\psi|_{\psi_0}.$$

Let $\Phi'(\cdot)$ denote the derivative of $\Phi(\cdot)$, that is $\Phi'(u) = \Phi(u)\{1 - \Phi(u)\}$ when $\Phi(u)$ is the logistic function. For any function $d(L, Z)$ define

$$g^d(L, Z) := d(L, Z) - \frac{E\{\Phi'(M_3)d(L, Z)|L\}}{E\{\Phi'(M_3)|L\}}.$$

Also define

$$d_{1,\text{eff}}(L, Z) = d_3(L, Z) - \frac{E\{\Phi'(M_3)d_3(L, Z)|L\}}{E\{\Phi'(M_3)d_4(L, Z)|L\}}d_4(L, Z), \tag{9}$$

where

$$d_3(L, Z) = [E\{\Phi'(M_3)|L\}E\{\Phi'(M_2)U|L, Z\} - E\{\Phi'(M_2)U|L\}\Phi'(M_3)]/d_5(L, Z),$$

$$d_4(L, Z) = \Phi'(M_3)/d_5(L, Z),$$

$$d_5(L, Z) = E\{\Phi'(M_3)|L\} \left(E \left[\frac{\{\Phi'(M_2)\}^2 \Phi(M_1)\{1 - \Phi(M_1)\}}{\{\Phi'(M_1)\}^2} \middle| L, Z \right] + \text{var}\{\Phi(M_2)|L, Z\} \right).$$

Throughout, for any pair of matrices W_1 and W_2 , $W_1 \leq W_2$ means that $W_2 - W_1$ is positive semidefinite and $W_1^{\otimes 2}$ denotes $W_1 W_1^T$.

THEOREM 1. Let θ_0 be the value of θ at the true law of the observed data.

(i) If $\hat{\psi}$ is a regular asymptotically linear estimator of ψ in model \mathcal{A} then there exist $p \times 1$ vector functions $d_1(L, Z)$ and $d_2(L, Z)$ such that $\hat{\psi}$ has influence function

$$h(O; d_1, d_2; \psi_0, \theta_0) = [E\{h^*(O; d_1, d_2; \psi_0, \theta_0)h^*(O; d_{1,\text{eff}}, 0; \psi_0, \theta_0)^\top\}]^{-1} \\ \times h^*(O; d_1, d_2; \psi_0, \theta_0),$$

where

$$h^*(O; d_1, d_2; \psi_0, \theta_0) = g^{d_1}(L, Z)H(\psi_0, \theta_0; t^*) + r^{d_2}(L, Z), \\ H(\psi_0, \theta_0; t^*) = \frac{\Phi'(M_2)}{\Phi'(M_1)} \{Y - \Phi(M_1)\} + [\Phi(M_2) - E\{\Phi(M_2)|L, Z\}].$$

Equivalently, as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\psi} - \psi_0) \rightarrow N(0, \Sigma)$$

in distribution, where $\Sigma = E\{h(O; d_1, d_2; \psi_0, \theta_0)^{\otimes 2}\}$.

(ii) The function $h(O; d_{1,\text{eff}}, 0; \psi_0, \theta_0)$ is the efficient influence function for ψ in model \mathcal{A} , so that, for any $h(O; d_1, d_2; \psi_0, \theta_0)$,

$$E\{h(O; d_{1,\text{eff}}, 0; \psi_0, \theta_0)^{\otimes 2}\} \leq E\{h(O; d_1, d_2; \psi_0, \theta_0)^{\otimes 2}\}.$$

The form of the influence functions stated in Theorem 1 implies the assertions in the following two remarks. As we argue next, these remarks imply that well-behaved estimators of ψ in model \mathcal{A} when $\psi_0 \neq 0$ or $t^*(L, Z) \neq 0$ with the moderate sample sizes found in practice are not feasible.

Remark 1. The influence function $h(O; d_1, d_2; \psi_0, \theta_0)$ depends on θ_0 through the unknown functions $t(l)$ and $q(l, z, x)$ and conditional expectations of functions of X given $L = l$ and $Z = z$.

Remark 2. When $\psi_0 \neq 0$ or $t^*(L, Z) \neq 0$, $h(O; d_1, d_2; \psi_0, \theta)$ does not have mean zero if any of the functions $t(l)$ or $q(l, z, x)$ or the required conditional expectations of functions of X given L and Z associated with θ are not the true ones.

Remarks 1 and 2 imply that, under $\psi_0 \neq 0$ or $t^*(L, Z) \neq 0$, estimation of ψ under model \mathcal{A} would require preliminary consistent estimation of the functions $t(l)$ and $q(l, z, x)$ and the required conditional expectations of functions of X given $L = l$ and $Z = z$. However, when L is a vector with two or more continuous components, these functions would not be well estimated with the moderate sample sizes found in practice essentially because no two units would have values of L close enough to each other to allow the borrowing of information needed for smoothing. Hence, unrealistically large sample sizes would be required for any estimator of ψ to have an approximately centred normal sampling distribution with variance small enough to be of substantive use. This phenomenon is often referred to as the curse of dimensionality. We emphasise that inference about ψ is infeasible in spite of the fact that the randomisation distribution $f(z|l)$ is known.

3.4. Feasible estimators of ψ that are guaranteed to be consistent under the null hypothesis

The following lemma is the key to our proposed approach to estimation of ψ .

LEMMA 1. If $t^*(L, Z) = 0$ and $\psi_0 = 0$, then $h(O; d_1, d_2; \psi_0, \theta)$ has mean zero even if evaluated at functions $t(l)$, $f(x|l, z)$ and $q(l, z, x)$ different from the true ones.

Proof. Under $\psi_0 = 0$, $M_1 = M_2$ and therefore

$$H(\psi_0, \theta_0; t^*) = Y - E\{\Phi(M_2)|L, Z\}.$$

Furthermore, $t^*(L, Z) = 0$ implies that $E\{\Phi(M_2)|L, Z\} = \Phi\{t(L)\}$ and $M_3 = t(L)$, so that

$$g^{d_1}(Z, L)H(\psi_0, \theta_0; t^* = 0) + r^{d_2}(Z, L) = r^{d_1}(Z, L)[Y - \Phi\{t(L)\}] + r^{d_2}(Z, L),$$

which has mean zero even when $t(L)$ is not the true function, because $\psi_0 = 0$ and $t^*(L, Z) = 0$ imply the intention-to-treat null hypothesis $E(Y|Z, L) = E(Y|L)$. \square

Thus, when $t^*(L, Z) = 0$, there exist regular asymptotically linear estimators of $\psi_0 = 0$ in model \mathcal{A} that are robust to model misspecification of $f(x|l, z)$, $t(l)$ and $q(l, z, x)$. To obtain such robust estimators we therefore propose estimating ψ by initially postulating working parametric models for the functions $f(x|l, z)$, $t(l)$ and $q(l, z, x)$ and then solving an estimating equation for ψ which has a solution $\hat{\psi}$ such that, when $t^*(L, Z) = 0$ and $\psi_0 = 0$, $\hat{\psi}$ has the following properties.

Property 1. The estimator $\hat{\psi}$ is a regular asymptotically linear estimator of $\psi_0 = 0$ in model \mathcal{A} regardless of whether the working parametric models are correct or not.

Property 2. The asymptotic variance of $\hat{\psi}$ attains the semiparametric variance bound for model \mathcal{A} when the working models are correct.

To be specific, our proposed estimator $\hat{\psi}$ of ψ is computed by the following two-stage estimation procedure.

Step 1. Postulate working parametric models $f(x|l, z; \alpha)$, $t(l, \omega)$ and $q(l, z, x; \eta)$ for the unknowns $f(x|l, z)$, $t(l)$ and $q(l, z, x)$. For each fixed value of ψ , compute estimators $\hat{\alpha}(\psi)$, $\hat{\omega}(\psi)$ and $\hat{\eta}(\psi)$ of α , ω and η , that are \sqrt{n} -consistent under $\psi_0 = \psi$. Such estimators can be obtained by maximising the parametric likelihood

$$\prod_i P(Y_i|L_i, Z_i, X_i; \psi, \eta, \alpha, \omega)f(X_i|L_i, Z_i; \alpha),$$

where $P(Y_i|L_i, Z_i, X_i; \psi, \eta, \alpha, \omega)$ is the conditional probability that the outcome takes the value Y_i given L_i , Z_i and X_i under the parameter values $(\psi, \eta, \alpha, \omega)$.

Alternatively, one can compute $\hat{\alpha}(\psi) = \hat{\alpha}$ by maximising $\prod_i f(X_i|L_i, Z_i; \alpha)$ and then compute $\hat{\eta}(\psi)$ and $\hat{\omega}(\psi)$ by maximising the parametric conditional likelihood

$$\prod_i P(Y_i|L_i, Z_i, X_i; \psi, \eta, \hat{\alpha}, \omega)$$

over η and ω with $\hat{\alpha}$ and ψ held fixed.

Step 2. Compute $\hat{H}(\psi; t^*)$ and $\hat{g}^{d_1, \text{eff}}(L, Z)$, which are equal to $H(\psi; t^*)$ and $g^{d_1, \text{eff}}(L, Z)$ evaluated at the law determined by $\alpha = \hat{\alpha}(\psi)$, $\omega = \hat{\omega}(\psi)$, $\eta = \hat{\eta}(\psi)$ and ψ . Finally, estimate ψ_0 with the solution $\hat{\psi}$ to the estimating equation

$$\sum_i \hat{H}_{\text{eff}, i}(\psi) = 0$$

with

$$\hat{H}_{\text{eff}, i}(\psi) = \hat{g}^{d_1, \text{eff}}(L_i, Z_i)\hat{H}(\psi; t^*).$$

It follows from Lemma 1 and a Taylor expansion that the estimator $\hat{\psi}$ satisfies Properties 1 and 2 mentioned above.

If the working models are correctly specified then, under mild regularity conditions,

$$\hat{\Sigma} = n \left\{ \sum_{i=1}^n \frac{\partial}{\partial \psi^T} \hat{H}_{\text{eff},i}(\psi) \Big|_{\psi=\hat{\psi}} \right\}^{-1} \left\{ \sum_{i=1}^n \hat{H}_{\text{eff},i}(\hat{\psi}) \hat{H}_{\text{eff},i}(\hat{\psi})^T \right\} \left\{ \sum_{i=1}^n \frac{\partial}{\partial \psi} \hat{H}_{\text{eff},i}(\psi)^T \Big|_{\psi=\hat{\psi}} \right\}^{-1}$$

is a consistent estimator of the asymptotic variance of $\hat{\psi}$, appropriately standardised. However, $\hat{\Sigma}$ is no longer a consistent variance estimator if the working models are misspecified because $\hat{H}_{\text{eff},i}(\psi)$ does not converge to a random vector which is orthogonal to the nuisance tangent space for ψ . To derive an analytical expression for a variance estimator for $\hat{\psi}$ that would be consistent even under misspecification of the working models one would need to derive Taylor expansions of the joint estimating functions for $(\psi, \eta, \alpha, \omega)$. Alternatively, under mild regularity conditions, the bootstrap estimator of the variance of $\hat{\psi}$ is a consistent estimator because the estimator $\hat{\psi}$ is a regular asymptotically linear estimator of its probability limit.

We can use $\hat{\psi}$ to construct an asymptotically distribution-free test of the sharp null hypothesis of no effect of assigned or received treatment:

$$H_0: Y_{zx} = Y$$

for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. To be specific, the test that rejects if

$$\left\{ \sum_i \hat{H}_{\text{eff},i}(0) \right\}^T \left\{ \sum_i \hat{H}_{\text{eff},i}(0) \hat{H}_{\text{eff},i}(0)^T \right\}^{-1} \sum_i \hat{H}_{\text{eff},i}(0)$$

exceeds the $(1 - \alpha)$ -quantile of a χ_p^2 distribution is an asymptotically distribution free α -level test of the intent-to-treat null hypothesis that Y and Z are independent given L and thus of the sharp null hypothesis of no effect of assigned or received treatment. When ψ is one-dimensional this test is the locally most powerful asymptotically distribution free test of the intent to treat null against Pitman alternatives which have $\psi_0 \neq 0$ and are contained in the models $f(x|l, z; \alpha)$, $t(l; \omega)$ and $q(l, z, x; \eta)$.

Remark 3. Note that even after specifying parametric models for $f(x|l, z)$, $t(l)$ or $q(l, z, x)$ we do not propose to estimate ψ by the easier-to-compute solution of the estimating equation

$$\sum_i (Y_i - \Phi[M_1\{\psi, \hat{\eta}(\psi), \hat{\alpha}(\psi), \hat{\omega}(\psi)\}])m(L_i, Z_i, X_i) = 0, \quad (10)$$

where $M_1\{\psi, \hat{\eta}(\psi), \hat{\alpha}(\psi), \hat{\omega}(\psi)\}$ is defined like M_1 except that it is evaluated at the laws determined by $\eta = \hat{\eta}(\psi)$, $\alpha = \hat{\alpha}(\psi)$, $\omega = \hat{\omega}(\psi)$ and ψ , and $m(L, Z, X)$ is any user-supplied function. The reason is that our proposed two-stage estimator $\hat{\psi}$ of ψ is consistent when $\psi_0 = 0$ and $t^*(L, Z) = 0$ regardless of whether or not the models for $f(x|l, z)$, $t(l)$ and $q(l, z, x)$ are correctly specified and its large-sample bias is small under model misspecification when ψ_0 is close to zero. Neither maximum likelihood nor the solutions of equation (10) satisfy this robustness property.

3.5. Estimation in non-randomised studies

Consider now the situation described in the introduction in which a randomised experiment was not conducted but Z was believed to be an instrument conditional on L . In such a case $f(z|l)$ would be unknown and in practice, because of the curse of dimensionality, with moderate sample sizes it would have to be estimated from the data using a model $f(z|l; \lambda)$ with λ an unknown finite-dimensional parameter. In such setting we propose using

the two-stage estimation procedure described above for estimating ψ_0 with the only modification that $f(z|l)$ is replaced with $f(z|l; \hat{\lambda})$, where $\hat{\lambda}$ is the maximum likelihood estimator of λ . Such a procedure gives an estimator of ψ that is doubly robust at $\psi_0 = 0$ and $t^*(L, Z) = 0$ in the sense that it is consistent for ψ at $\psi_0 = 0$ and $t^*(L, Z) = 0$ if either, but not necessarily both, the model $f(z|l; \lambda)$ for $f(z|l)$ or the model $t(l; \omega)$ for $t(l)$ is correctly specified. For additive or multiplicative structural mean models, doubly robust estimators of the structural parameters at nonnull values are also available (Robins, 2000).

3.6. Estimation of the direct effects of assigned treatment

Finally, suppose that, rather than regarding $t^*(l, z)$ as known, we assume a parametric model for $t^*(l, z)$; that is,

$$t^*(l, z) = t^*(l, z; \kappa_0), \tag{11}$$

with $t^*(l, z; \kappa)$ a known function and κ_0 an unknown p^* -dimensional parameter. In the Appendix we show that, when model \mathcal{A} is redefined by the restrictions defining the previous model \mathcal{A} and the additional assumption, Theorem 1 remains true provided ψ is replaced by the parameter vector $\rho = (\psi, \kappa)$, $d_1(L, Z)$ and $d_2(L, Z)$ are $(p + p^*) \times 1$ vector functions and U is redefined as the $(p + p^*) \times 1$ vector

$$U = \begin{pmatrix} \partial\gamma(L, Z, X; \psi)/\partial\psi|_{\psi_0} \\ [\partial t^*(L, Z; \kappa)/\partial\kappa]|_{\kappa_0} \Phi'(M_3)E\{\Phi'(M_2)|L, Z\}^{-1} \end{pmatrix}.$$

Furthermore, it remains true that, when $t^*(L, Z) = 0$ and $\psi_0 = 0$, $h(O; d_1, d_2; \psi_0, \theta)$ has mean zero even if evaluated at functions $t(l), f(x|l, z)$ and $q(l, z, x)$ different from the true ones. Thus, to estimate ρ we propose the two-stage estimation procedure given in § 3.4 with the only modifications that $\hat{\psi}$ is replaced by $\hat{\rho} = (\hat{\psi}, \hat{\kappa})$, ψ and ψ_0 are replaced by ρ and ρ_0 respectively, and $d_{1,\text{eff}}(L, Z)$ is now a $(p + p^*)$ -dimensional function defined as in (9) but with U defined as in this section.

4. IDENTIFICATION

In this section we will elaborate on the identifiability of the functions $\gamma(l, z, x)$ and $t^*(l, z)$. The simple special case in which Z, X and Y are all dichotomous suffices to illustrate the essence of why the functions $\gamma(l, z, x)$ and $t^*(l, z)$ are generally not identified if assumed unrestricted and why they are identified under a sufficiently stringent structural mean model.

Suppose that L is a constant with probability one so that we can write $\gamma(l, z, x) = \gamma(z, x)$ and $t^*(l, z) = t^*(z)$ because L can be ignored. Since by definition $\gamma(z, 0) = t^*(0) = 0$ for $z = 0, 1$, then the functions $\gamma(\cdot, \cdot)$ and $t^*(\cdot)$ are defined by the three numbers $\gamma(0, 1), \gamma(1, 1)$ and $t^*(1)$. Consider first the model for the joint distribution of $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, Y, Z, X)$, which we will denote by $\mathcal{B}(1)$, defined by the consistency restriction (2) and the randomisation restriction (1) and with $\gamma(0, 1), \gamma(1, 1)$ and $t^*(1)$ unrestricted. We will now show that, for most distributions of the observed data $O = (Y, X, Z)$, the conditions (2) and (1) do not restrict the distribution of $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, Y, Z, X)$ enough so as to determine uniquely any of the components of the triplet $(\gamma(0, 1), \gamma(1, 1), t^*(1))$. To be specific, by definition,

$$\begin{aligned} \gamma(0, 1) &= \Phi^{-1}\{E(Y_{01}|Z = 0, X = 1)\} - \Phi^{-1}\{E(Y_{00}|Z = 0, X = 1)\}, \\ \gamma(1, 1) &= \Phi^{-1}\{E(Y_{11}|Z = 1, X = 1)\} - \Phi^{-1}\{E(Y_{10}|Z = 0, X = 1)\}. \end{aligned}$$

Under the consistency restriction (2) we then have

$$\begin{aligned}\gamma(0, 1) &= \Phi^{-1}\{E(Y|Z=0, X=1)\} - \Phi^{-1}\{E(Y_{00}|Z=0, X=1)\}, \\ \gamma(1, 1) &= \Phi^{-1}\{E(Y|Z=1, X=1)\} - \Phi^{-1}\{E(Y_{10}|Z=1, X=1)\}.\end{aligned}$$

Therefore, for any fixed observed data distribution for O , under the consistency restriction (2), the numbers $\gamma(0, 1)$ and $\gamma(1, 1)$ are one-one functions of $E(Y_{00}|Z=0, X=1)$ and $E(Y_{10}|Z=1, X=1)$ respectively. We will now see that the second restriction defining model $\mathcal{B}(1)$, that is the randomisation restriction (1), does not uniquely determine these two conditional expectations. To be specific, for $z=1$ and $x=1$, assumption (1) is equivalent to

$$\begin{aligned}E(Y_{10}|Z=1, X=1) \text{pr}(X=1|Z=1) + E(Y_{10}|Z=1, X=0) \text{pr}(X=0|Z=1) \\ = E(Y_{10}|Z=0, X=1) \text{pr}(X=1|Z=0) + E(Y_{10}|Z=0, X=0) \text{pr}(X=0|Z=0)\end{aligned}$$

and, in turn, under the consistency assumption (2), the last display is equivalent to

$$\begin{aligned}E(Y_{10}|Z=1, X=1) \text{pr}(X=1|Z=1) + E(Y|Z=1, X=0) \text{pr}(X=0|Z=1) \\ = E(Y_{10}|Z=0, X=1) \text{pr}(X=1|Z=0) + E(Y_{10}|Z=0, X=0) \text{pr}(X=0|Z=0).\end{aligned}$$

This last equation is the only restriction imposed by model $\mathcal{B}(1)$ on the conditional expectation of Y_{10} given (Z, X) . Put another way, the assumptions of model $\mathcal{B}(1)$ only allow us to learn one restriction about the conditional mean function $E(Y_{10}|Z, X)$ and this is given by the equation in the last display. Unless $\text{pr}(Z=1, X=1)=0$, that is unless no subject in the active treatment arm receives active treatment, in which case $E(Y_{10}|Z=1, X=1)$ is undefined, the last display is a linear equation in three unknowns, or two unknowns if either $\text{pr}(X=1|Z=0)$ or $\text{pr}(X=0|Z=0)$ is equal to 0, one of which is the conditional expectation $E(Y_{10}|Z=1, X=1)$ which, as we have argued previously, uniquely determines $\gamma(0, 1)$. We therefore conclude that there exist an infinite number of values for $\gamma(0, 1)$ that are compatible with the distribution of the observed data under model $\mathcal{B}(1)$. Arguing identically we conclude that there exist infinitely many values for $\gamma(1, 1)$ compatible with the observed data under model $\mathcal{B}(1)$. Therefore, neither $\gamma(0, 1)$ nor $\gamma(1, 1)$ is identified under model $\mathcal{B}(1)$. Incidentally, note that the values of $\gamma(0, 1)$ and $\gamma(1, 1)$ are variation independent in the sense that the range of plausible values for $\gamma(1, 1)$ remains the same regardless of the value of $\gamma(0, 1)$.

We will now show that, for most distributions of the observed data O , the value of $t^*(1)$ is also not uniquely determined under model $\mathcal{B}(1)$. To be specific, by definition,

$$t^*(1) = E(Y_{10}) - E(Y_{00}).$$

Under the randomisation assumption (1),

$$t^*(1) = E(Y_{10}|Z=1) - E(Y_{00}|Z=0)$$

and, under the consistency assumption (2), the last display is equivalent to

$$\begin{aligned}t^*(1) &= E(Y_{10}|Z=1, X=1) \text{pr}(X=1|Z=1) + E(Y|Z=1, X=0) \text{pr}(X=0|Z=1) \\ &\quad - E(Y_{00}|Z=0, X=1) \text{pr}(X=1|Z=0) - E(Y|Z=0, X=0) \text{pr}(X=0|Z=0).\end{aligned}\tag{12}$$

The last equation is the only restriction imposed on $t^*(1)$ by model $\mathcal{B}(1)$ and the observed data distribution. Thus we conclude that, unless $\text{pr}(X = 0|Z = 1) = \text{pr}(X = 0|Z = 0) = 1$, that is unless no subject in either arm takes active treatment, $t^*(1)$ is a nontrivial function of $E(Y_{10}|Z = 1, X = 1)$ and $E(Y_{00}|Z = 0, X = 1)$ under model $\mathcal{B}(1)$. Since we have already seen that neither $E(Y_{10}|Z = 1, X = 1)$ nor $E(Y_{00}|Z = 0, X = 1)$ is uniquely determined under model $\mathcal{B}(1)$ we conclude that $t^*(1)$ is not identified under model $\mathcal{B}(1)$ unless $\text{pr}(X = 0|Z = 1) = \text{pr}(X = 0|Z = 0) = 1$.

The previous argument shows that the function $t^*(1)$ is indeed uniquely determined by the values of $\gamma(0, 1)$ and $\gamma(1, 1)$, since these are in one-one correspondence with $E(Y_{10}|Z = 1, X = 1)$ and $E(Y_{00}|Z = 0, X = 1)$. However this, in turn, shows that, if our structural mean model was so restrictive that indeed it had a priori assumed the values of $\gamma(0, 1)$ and $\gamma(1, 1)$, then the value of $t^*(1)$ would be identified. This, of course, is an uninteresting model, but it nevertheless serves to make the point that, with a sufficiently restrictive structural mean model, the function $t^*(l, z)$ is identified.

Now consider model $\mathcal{A}(1)$ which is defined like $\mathcal{B}(1)$ except that the value of $t^*(1)$ is assumed known. Model $\mathcal{A}(1)$ is model \mathcal{A} of § 2 in which $\gamma(z, x; \psi) = \psi_1 x(1 - z) + \psi_2 xz$ with ψ_1 and ψ_2 arbitrary. Note that the model $\gamma(z, x; \psi)$ for $\gamma(z, x)$ is saturated because it places no restriction on the function $\gamma(z, x)$ beyond the definitional restriction that $\gamma(z, 0) = 0$. Suppose first that $\text{pr}(X = 0|Z = 1) \neq 1$ and $\text{pr}(X = 0|Z = 0) \neq 1$ so that $\gamma(0, 1)$ and $\gamma(1, 1)$ are well defined. Since, under $\mathcal{A}(1)$, $t^*(1)$ is given, equation (12) imposes one linear constraint on the values of $E(Y_{10}|Z = 1, X = 1)$ and $E(Y_{00}|Z = 0, X = 1)$. Thus, under model $\mathcal{A}(1)$, $\gamma(0, 1)$ and $\gamma(1, 1)$ are no longer variation independent. However, the unique constraint (12) does not suffice to identify the values of $\gamma(0, 1)$ and $\gamma(1, 1)$ because we still have infinitely many choices for one of them. Next, suppose that $\text{pr}(X = 0|Z = 1) = 1$ but $\text{pr}(X = 0|Z = 0) \neq 1$. In this case $\gamma(1, 1)$ is not defined and $\gamma(0, 1)$ is uniquely determined because $E(Y_{00}|Z = 0, X = 1)$ is determined from equation (12). Similarly, $\gamma(0, 1)$ is not defined and $\gamma(1, 1)$ is uniquely determined when $\text{pr}(X = 0|Z = 1) \neq 1$ but $\text{pr}(X = 0|Z = 0) = 1$. Finally, if $\text{pr}(X = 0|Z = 1) = \text{pr}(X = 0|Z = 0) = 1$, then neither $\gamma(0, 1)$ nor $\gamma(1, 1)$ is defined. Interestingly, in this case equation (12) reduces to

$$t^*(1) = E(Y|Z = 1, X = 0) - E(Y|Z = 0, X = 0),$$

which shows that indeed model $\mathcal{A}(1)$ places restrictions on the observed data distribution. In conclusion, assuming a saturated structural mean model but a known function $t^*(1)$ does not identify the function $\gamma(z, x)$ even though the model is not a nonparametric model for the observed data distribution; Pearl (2000, Ch. 8) makes a similar remark.

The previous discussion suggests that, in order to identify the function $\gamma(z, x)$ when $t^*(1)$ is assumed known, we must place one additional restriction. For example, suppose that we assume model $\mathcal{A}(2)$ which is defined like $\mathcal{A}(1)$ except that $\gamma(z, x; \psi) = \psi x(1 - z)$. This model postulates a priori that $\gamma(1, 1) = 0$. Since $\gamma(1, 1)$ uniquely determines $E(Y_{10}|Z = 1, X = 1)$, then now in the linear equation (12) the unique unknown is $E(Y_{00}|Z = 0, X = 1)$ which is therefore uniquely determined. This value, in turn, uniquely determines the value of $\gamma(0, 1)$. Thus $\gamma(0, 1)$ is identified under model $\mathcal{A}(2)$. Alternatively, we could have assumed for example model $\mathcal{A}(3)$ defined like $\mathcal{A}(1)$ except that $\gamma(z, x; \psi) = \psi x$. This model postulates a priori that $\gamma(0, 1) = \gamma(1, 1)$. This last identity together with (12) determine a system of two equations in the unknowns $E(Y_{00}|Z = 0, X = 1)$ and $E(Y_{10}|Z = 1, X = 1)$. The values of $E(Y_{00}|Z = 0, X = 1)$ and $E(Y_{10}|Z = 1, X = 1)$, and

hence $\gamma(0, 1)$ and $\gamma(1, 1)$, are therefore now uniquely determined if the system has a solution. Hence $\gamma(0, 1)$ and $\gamma(1, 1)$ are identified under model $\mathcal{A}(3)$. Note also that for some observed data distributions the system does not have a solution. For example, it can be easily checked that if

$$\begin{aligned} \text{pr}(X = 1|Z = 0) &\geq \text{pr}(X = 1|Z = 1) > 0, \\ t^*(1) - [E(ZY|X = 0) - E\{(1 - Z)Y|X = 0\}] &\geq 0 \end{aligned}$$

then a necessary condition for the system to have a solution is that $E(Y|Z = 1, X = 1) \geq E(Y|Z = 0, X = 1)$. Thus, model $\mathcal{A}(3)$ is not a nonparametric model for the observed data law.

In conclusion, we have shown that $\gamma(z, x)$ is not identified if it is left unrestricted even if $t^*(z)$ is assumed known. However, $\gamma(z, x)$ is identified if, as in models $\mathcal{A}(2)$ or $\mathcal{A}(3)$, we place restrictions on it. Of course, care must be taken in placing restrictions on the function $\gamma(z, x)$, as some restrictions may be tantamount to unrealistic assumptions. For example, suppose that we are analysing data from a trial in which the double-blind protocol was not violated and the strong exclusion restriction (4) held. Assuming model $\mathcal{A}(3)$ when Φ^{-1} is the logistic function is tantamount to assuming that

$$\begin{aligned} &\left\{ \frac{\text{pr}(Y_{01} = 1|Z = 1, X = 1)}{\text{pr}(Y_{01} = 0|Z = 1, X = 1)} \right\} / \left\{ \frac{\text{pr}(Y_{00} = 1|Z = 1, X = 1)}{\text{pr}(Y_{00} = 0|Z = 1, X = 1)} \right\} \\ &= \left\{ \frac{\text{pr}(Y_{01} = 1|Z = 0, X = 1)}{\text{pr}(Y_{01} = 0|Z = 0, X = 1)} \right\} / \left\{ \frac{\text{pr}(Y_{00} = 1|Z = 0, X = 1)}{\text{pr}(Y_{00} = 0|Z = 0, X = 1)} \right\}, \end{aligned}$$

i.e. that the effect of received treatment is the same in the subset of the population who receive active treatment when assigned to active treatment as in the subset of the population who, when assigned to placebo, actually switch to active treatment. This assumption is unrealistic because these two subpopulations are likely to be quite different with regard to modifiers of the effect of active treatment on the outcome of interest.

Finally, we note that physical randomisation implies not only our assumption (1) but also the stronger assumption,

$$Z \perp\!\!\!\perp \{Y_{zx}, z \in \mathcal{Z}, x \in \mathcal{X}\} | L,$$

with probability one, of joint independence of Z with the counterfactuals. Surprisingly, it follows from results given in Pearl (2000) that, under certain unusual laws for (Z, X, Y) , imposing joint randomisation allows identification of some of the parameters of the saturated model $\mathcal{A}(1)$ provided we impose the stronger exclusion assumption (4) rather than the weaker weak exclusion assumption $t^*(1) = 0$. To see this suppose that $\text{pr}(Y = 1|X = x, Z = z) = 1$ if $x = z = 1$ and is 0 otherwise, and that $\text{pr}(X = 1|Z = 1) = 1 - \text{pr}(X = 1|Z = 0) = 0.55$ and we impose the exclusion restriction (4). Then it is easy to show that under our strengthened assumptions

$$\begin{aligned} \text{pr}(Y_{11}|Z = 1, X = 1) &= 1, \\ \text{pr}(Y_{10}|Z = 1, X = 1) &= \text{pr}(Y_{10}|Z = 0, X = 1) = \text{pr}(Y_{11}|Z = 0, X = 1) = 0. \end{aligned}$$

Thus $\psi_2 = \infty$ and $\psi_1 = \log 0/0$ is undefined.

ACKNOWLEDGEMENT

This work was partially completed while Andrea Rotnitzky was visiting the Department of Economics at Di Tella University. She would like to thank the authorities of the school for providing a stimulating environment. James Robins and Andrea Rotnitzky were partially funded by grants from the National Institutes of Health.

APPENDIX

Technical details

Explicit solution to the integral equation (8). For fixed (l, z) , let $u = \exp \{-v(l, z)\}$. Then, when X is binary and Φ is the logit link, equation (8) is the same as

$$k_1 = (1 + u)^{-1}p + (1 + k_2u)^{-1}(1 - p),$$

where $k_1 = \Phi\{t(l) + t^*(l, z)\}$ and $k_2 = \exp \{-q(l, z, 1)\}$. This is the quadratic equation in u , $au^2 + bu + c = 0$, where $a = k_1k_2$ and $c = k_1 - 1$. Since by definition $u > 0$, we must choose the positive root of the equation. This corresponds to $u = \{-b + (b^2 - 4ac)^{1/2}\}/(2a)$ because $c < 0$ and $a > 0$.

The additive and multiplicative structural mean models are conditional mean independence models. Consider first the additive model for which $\Phi(u) = u$. By definition of $\gamma(l, z, x)$,

$$E(Y|Z = z, X, L) - \gamma(L, z, X) = E(Y_{z0}|Z = z, X, L). \tag{A1}$$

Thus,

$$E\{Y - \gamma(L, z, X)|Z = z, L\} = E(Y_{z0}|Z = z, L) = E(Y_{z0}|L) = t(L) + t^*(z, L),$$

where the first equality follows by taking conditional expectations with respect to X on both sides of (A1), the second is by assumption (1) and the third is by Assumption 2. The last equality gives restriction (7) because $t(L)$ is unrestricted. Consider next the multiplicative model for which $\Phi(u) = \exp(u)$. By definition of $\gamma(l, z, x)$,

$$E(Y|Z = z, X, L) \exp \{-\gamma(L, z, X)\} = E(Y_{z0}|Z = z, X, L).$$

Thus,

$$\begin{aligned} E[Y \exp \{-\gamma(L, z, X)\}|Z = z, L] &= E(Y_{z0}|Z = z, L) = E(Y_{z0}|L) \\ &= \exp \{t(L) + t^*(z, L)\}, \end{aligned}$$

where the first equality follows by taking conditional expectations with respect to X on both sides of (A1), the second is by assumption (1) and the third is by Assumption 2. The last equality gives restriction (7) because $t(L)$ is unrestricted.

We have shown that models \mathcal{A}_{add} and $\mathcal{A}_{\text{mult}}$ impose restriction (7) on the law of the observed data O . To show that this is the only restriction, other than Assumption 1, imposed on the observed data law, we must exhibit for any law of O satisfying (7) and Assumption 1 a joint law of O and $(Y_{zx})_{x \in \mathcal{X}, z \in \mathcal{Z}}$ satisfying the restrictions (1) and Assumption 2, and with the potential outcomes satisfying (2). We construct such a random vector $(Y_{zx})_{x \in \mathcal{X}, z \in \mathcal{Z}}$ as follows. Given $(Z = z^*, X = x^*, L = l, Y = y)$ with $x^* \neq 0$, we define $Y_{z^*x^*} = y$. This guarantees that condition (2) is satisfied. Next we set $f(Y_{z^*x^*}|Z = z^*, X \neq x^*, L = l, Y = y)$ to be an arbitrary density. This construction, together with the law of O , determines $f(Y_{z^*x^*}|Z = z^*, L = l)$. For any $z \neq z^*$ and any x , we define $f(Y_{zx^*}|Z = z, X = x, L = l, Y = y)$ to be any density such that $f(Y_{zx^*}|Z = z, L = l) = f(Y_{z^*x^*}|Z = z^*, L = l)$. This guarantees that condition (1) is satisfied. Finally, for $z^* = 0$, we define

$$f(Y_{z^*0}|Z = z^*, X = x, L = l, Y = y) = f(Y_{z^*0}|Z = z^*, X = x, L = l)$$

to be any density such that its mean is equal to $E(Y|Z = z^*, X = x, L = l) - \gamma(l, z^*, x)$. This, together with (7), guarantees that Assumption 2 holds. We complete the construction by imposing that the random variables $(Y_{zx})_{x \in \mathcal{X}, z \in \mathcal{Z}}$ are conditionally independent given O .

Sketch of the proof of Theorem 1 and the assertions in § 3.6. The nuisance tangent space Λ for model \mathcal{A} is the closed linear span of the sum of the infinite-dimensional parameter-specific tangent sets $\Lambda_l, \Lambda_x, \Lambda_t$ and Λ_q comprising respectively products of constant conformable matrices of p rows times the scores for the parameters in any regular parametric submodel for $f(l), f(x|l, z), t(l)$ and $q(l, z, x)$; for a precise definition of the nuisance tangent space, see for example Newey (1990). The sets $\Lambda_l, \Lambda_x, \Lambda_t$ and Λ_q are linear subspaces of the Hilbert space of p -dimensional zero-mean random vectors with covariance inner product. Throughout, a space superscribed with \perp denotes the orthogonal complement of that space and $\Pi(\cdot)$ denotes the projection operator. In model \mathcal{A} , the set of influence functions of regular asymptotically linear estimators of the parameter ψ_0 is the set $\{E(AS_{\text{eff}}^T)^{-1}A : A \in \Lambda^\perp\}$, where S_{eff} is the efficient score for ψ . To derive the set of all influence functions we will first derive the restrictions defining the elements of each of the spaces $\Lambda_l, \Lambda_x, \Lambda_t$ and Λ_q , and then use the fact that $\Lambda^\perp = \Lambda_l^\perp \cap \Lambda_x^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp$.

To derive the restrictions on the elements of each of the sets $\Lambda_l, \Lambda_x, \Lambda_t$ and Λ_q consider arbitrary correctly specified parametric submodels $f(l; \phi), f(x|l, z; \alpha), t(l; \omega)$ and $q(l, z, x; \eta)$. Let $\Gamma(\psi) = \gamma(L, Z, X; \psi), T^* = t^*(L, Z), T(\omega) = t(L; \omega), Q(\eta) = q(L, Z, X; \eta)$ and $V(\eta, \omega, \alpha) = v(L, Z; \eta, \omega, \alpha)$, where $V(\eta, \omega, \alpha)$ is the unique solution to the equation (8) with $f(L), f(x|L, Z), t(L)$ and $q(L, Z, x)$ replaced by $f(L; \phi), f(x|L, Z; \alpha), t(L; \omega)$ and $q(L, Z, x; \eta)$. Let $M_j (j = 1, \dots, 4)$ and U be defined as in § 3.

Let S_ϕ, S_x, S_ω and S_η denote the scores for the parameters ϕ, α, ω and η respectively, evaluated at the truth. To derive the restrictions that the scores S_ϕ, S_x, S_ω and S_η must satisfy at the truth we consider the loglikelihood for a single observation O ,

$$\begin{aligned} \mathcal{L}(\phi, \alpha, \omega, \eta, \psi) &= Y \log \Phi\{\Gamma(\psi) + Q(\eta) + V(\eta, \omega, \alpha)\} \\ &+ (1 - Y) \log [1 - \Phi\{\Gamma(\psi) + Q(\eta) + V(\eta, \omega, \alpha)\}] + \log f(X|L, Z; \alpha) + \log f(L; \phi). \end{aligned}$$

Let $G = \{Y - \Phi(M_1)\} \Phi'(M_1) / [\Phi(M_1)\{1 - \Phi(M_1)\}]$. By (8),

$$\begin{aligned} S_\eta &= G\{\partial Q(\eta_0)/\partial \eta + \partial V(\eta_0, \omega_0, \alpha_0)/\partial \eta\} \\ &= G \left[\partial Q(\eta_0)/\partial \eta - \frac{E\{\Phi'(M_2)\partial Q(\eta_0)/\partial \eta | Z, L\}}{E\{\Phi'(M_2) | Z, L\}} \right], \\ S_\omega &= G\partial V(\eta_0, \omega_0, \alpha_0)/\partial \omega = \frac{G\Phi'(M_3)\{\partial T(\omega_0)/\partial \omega\}}{E\{\Phi'(M_2) | L, Z\}}, \\ S_x &= -\frac{GE\{\Phi(M_2)S_{x,\text{part}} | L, Z\}}{E\{\Phi'(M_2) | L, Z\}} + S_{x,\text{part}}, \end{aligned}$$

where $S_{x,\text{part}} = \partial \log f(X|L, Z; \alpha_0)/\partial \alpha$ and $S_\phi = \partial \log f(L; \phi_0)/\partial \phi$. From these expressions it can be easily shown that

$$\Lambda_q = \left\{ G \left[B - \frac{E\{\Phi'(M_2)B | Z, L\}}{E\{\Phi'(M_2) | Z, L\}} \right]; B = b(L, Z, X) \text{ is arbitrary} \right\};$$

we arrive at this set after noting that the set Λ_q with $b(L, Z, X)$ arbitrary is the same as the set Λ_q with $b(L, Z, X)$ satisfying $b(L, Z, 0) = 0$, which is the only restriction on $\partial Q(\eta_0)/\partial \eta$. Also,

$$\begin{aligned} \Lambda_t &= \left\{ G \left[\frac{\Phi'(M_3)e(L)}{E\{\Phi'(M_2) | Z, L\}} \right]; e(L) \text{ is arbitrary} \right\}, \\ \Lambda_x &= \left\{ -G \left[\frac{E\{\Phi(M_2)c(L, Z, X) | Z, L\}}{E\{\Phi'(M_2) | Z, L\}} \right] + c(L, Z, X); E\{c(L, Z, X) | L, Z\} = 0 \right\}, \\ \Lambda_l &= \{a(L); E\{a(L)\} = 0\}. \end{aligned}$$

It can be easily checked that the orthogonal complement of the spaces Λ_t , Λ_r and Λ_q are given by

$$\begin{aligned} \Lambda_r^\perp &= \{Gb(L, Z, X) + c(L, Z, X) + d(L, Z); b(L, Z, X) \text{ arbitrary}, \\ &E\{c(L, Z, X)|L, Z\} = E\{d(L, Z)|L\} = 0\}, \\ \Lambda_t^\perp &= \{Gb_1(L, Z, X) + b_2(L, Z, X); b_2(L, Z, X) \text{ arbitrary}, \\ &E[E(G^2|Z, L, X)\Phi'(M_3)E\{\Phi'(M_2)|Z, L\}^{-1}b_1(L, Z, X)|L] = 0\}, \\ \Lambda_q^\perp &= \{GRd(L, Z) + b(L, Z, X): d(L, Z) \text{ and } b(L, Z, X) \text{ arbitrary}\}, \end{aligned}$$

where $R = \Phi'(M_2)/E(G^2|Z, L, X)$. Thus,

$$\begin{aligned} \Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp &= \{GRd_1(L, Z) + c(L, Z, X) + d_2(L, Z); E\{\Phi'(M_3)d_1(L, Z)|L\} = 0, \\ &E\{c(L, Z, X)|L, Z\} = E\{d_2(L, Z)|L\} = 0\}. \end{aligned}$$

Now, A is an element of $\Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp \cap \Lambda_x^\perp$ if and only if $A \in \Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp$ and A is orthogonal to all the elements of Λ_x ; that is, A satisfies

$$\begin{aligned} 0 &= E \left\{ A \left(b(L, Z, X) - G \left[\frac{E\{\Phi(M_2)b(L, Z, X)|Z, L\}}{E\{\Phi'(M_2)|Z, L\}} \right] \right) \right\} \\ &\quad - E \left\{ A \left(E\{b(L, Z, X)|L, Z\} - G \left[\frac{E\{b(L, Z, X)|L, Z\}E\{\Phi(M_2)|Z, L\}}{E\{\Phi'(M_2)|Z, L\}} \right] \right) \right\} \end{aligned}$$

for all $b(L, Z, X)$. This last equation is equivalent to

$$\begin{aligned} 0 &= E\{E(A|L, Z, X)b(L, Z, X)\} - E \left[E(AG|L, Z) \frac{\Phi(M_2)b(L, Z, X)}{E\{\Phi'(M_2)|Z, L\}} \right] \\ &\quad - E\{E(A|L, Z)b(L, Z, X)\} + E \left(E(AG|L, Z) \frac{E[\{\Phi(M_2)|Z, L\}]b(L, Z, X)}{E\{\Phi'(M_2)|Z, L\}} \right) \end{aligned}$$

for all $b(L, Z, X)$, which in turn is equivalent to

$$E(A|L, Z, X) - E(A|L, Z) - E(AG|L, Z) \frac{\Phi(M_2) - E\{\Phi(M_2)|Z, L\}}{E\{\Phi'(M_2)|Z, L\}} = 0. \tag{A2}$$

However, for $A \in \Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp$, we have

$$\begin{aligned} E(A|L, Z) &= d_2(L, Z), \quad E(A|L, Z, X) = d_2(L, Z) + c(L, Z, X), \\ E(AG|L, Z) &= d_1(L, Z)E\{\Phi'(M_2)|L, Z\}, \end{aligned}$$

where $E\{c(L, Z, X)|L, Z\} = 0$ and $E\{\Phi'(M_3)d_1(L, Z)|L\} = 0$. Thus, (A2) is equivalent to

$$c(L, Z, X) = d_1(L, Z)[\Phi(M_2) - E\{\Phi(M_2)|Z, L\}],$$

which is therefore the only restriction that $A \in \Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp$ must satisfy to be also in Λ_x^\perp . Hence, we conclude that

$$\begin{aligned} \Lambda^\perp &= \Lambda_r^\perp \cap \Lambda_t^\perp \cap \Lambda_q^\perp \cap \Lambda_x^\perp \tag{A3} \\ &= \{(GR + [\Phi(M_2) - E\{\Phi(M_2)|Z, L\}])d_1(L, Z) + d_2(L, Z); \\ &E\{\Phi'(M_3)d_1(L, Z)|L\} = 0, E\{d_2(L, Z)|L\} = 0\}. \end{aligned}$$

Note that $(GR + [\Phi(M_2) - E\{\Phi(M_2)|Z, L\}])d_1(L, Z) + d_2(L, Z)$ is the same as $h^*(O; d_1, d_2; \psi_0, \eta_0)$.

The score for ψ is $S_\psi = GU$ and therefore $S_{\text{eff}} = \prod(GU|\Lambda^\perp)$. It can be easily checked that

$$\prod(GU|\Lambda^\perp) = (GR + [\Phi(M_2) - E\{\Phi(M_2)|Z, L\}])d_{1,\text{eff}}(L, Z)$$

with

$$d_{1,\text{eff}}(L, Z) = d_3(L, Z) - \frac{E\{\Phi'(M_3)d_3(L, Z)|L\}}{E\{\Phi'(M_3)d_4(L, Z)|L\}}d_4(L, Z),$$

where

$$d_3(L, Z) = [E\{\Phi'(M_3)|L\}E\{\Phi'(M_2)U|L, Z\} - E\{\Phi'(M_2)U|L\}\Phi'(M_3)]/d_5(L, Z),$$

$$d_4(L, Z) = \Phi'(M_3)/d_5(L, Z),$$

$$d_5(L, Z) = E\{\Phi'(M_3)|L\}[E\{\Phi'(M_2)R|L, Z\} + \text{var}\{\Phi(M_2)|L, Z\}].$$

That the conjectured quantity is indeed the efficient score can be shown by simply checking that

$$E[\{GU - (GR + [\Phi(M_2) - E\{\Phi(M_2)|Z, L\}])d_{1,\text{eff}}(L, Z)\}^T A] = 0$$

for all $A \in \Lambda^\perp$. Thus, $S_{\text{eff}} = h(O; d_{1,\text{eff}}, 0; \psi_0, \theta_0)$. This result together with the set (A3) completes the proof of Theorem 1.

If model \mathcal{A} is re-defined as in § 3.6, and the parameter of interest is ρ , the nuisance parameters remain the same and hence Λ^\perp remains the same except that the functions d_1 and d_2 are column vectors of the same dimension as ρ . Finally, the score S_ρ for ρ is equal to HU with U redefined as in § 3.6 and hence the efficient score for ρ is the same as S_{eff} defined above but with U redefined as in § 3.6.

REFERENCES

- ABADIE, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *J. Economet.* **113**, 231–63.
- ANGRIST, J. (2001). Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. *J. Bus. Econ. Statist.* **19**, 1–16.
- ANGRIST, J., IMBENS, G. & RUBIN, D. B. (1996). Identification of causal effects using instrumental variables. *J. Am. Statist. Assoc.* **91**, 444–55.
- CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Economet.* **34**, 305–34.
- DAWID, A. P. (1979). Conditional independence in statistical theory (with Discussion). *J. R. Statist. Soc. B* **41**, 1–31.
- HIRANO, K., IMBENS, G. W., RUBIN, D. B. & ZHOU, X.-H. (2000). Assessing the effects of influenza vaccine in an encouragement design. *Biostatistics* **1**, 69–88.
- HUBER, P. (1985). Projection pursuit (with Discussion). *Ann. Statist.* **13**, 435–525.
- NEWWEY, W. (1990). Semiparametric efficiency bounds. *J. Appl. Econ.* **5**, 99–135.
- NEYMAN, J. (1923). Sur les applications de la thar des probabilitiés aux expériences Agaricales: Essay des principes. Excerpts reprinted (1990) in English (D. Dabrowska and T. Speed, Trans.) in *Statist. Sci.* **5**, 463–72.
- PEARL, J. (2000). *Causality, Models, Reasoning and Inference*. Cambridge University Press.
- ROBINS, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods—Application to control of the healthy worker survivor effect. *Math. Mod.* **7**, 1393–512.
- ROBINS, J. M. (1989). The analysis of randomised and non-randomised AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, Ed. L. Sechrest, H. Freeman and A. Mulley, pp. 113–59. Washington, D.C.: U.S. Public Health Service, National Center for Health Services Research.
- ROBINS, J. M. (1994). Correcting for non-compliance in randomised trials using structural nested mean models. *Commun. Statist. A* **23**, 2379–412.
- ROBINS, J. M. (1997). Causal inference from complex longitudinal data. Latent variable modeling and applications to causality. In *Lecture Notes in Statistics*, **120**, Ed. M. Berkane, pp. 69–117. New York: Springer Verlag.

- ROBINS, J. M. (2000). Robust estimation in sequentially ignorable missing data and causal inference models. In *Proc. Bayesian Statist. Sect., Am. Statist. Assoc.*, pp. 6–10. Washington D.C: Am. Statist. Assoc.
- ROBINS, J. & RITOV, Y. (1997). Toward a curse of dimensionality appropriate (coda) asymptotic theory for semiparametric models. *Statist. Med.* **16**, 185–319.
- ROBINS, J. M., ROTNITZKY, A. & SCHARFSTEIN, D. (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*, IMA Volume 116, Ed. M. E. Halloran and D. Berry, pp. 1–92. New York: Springer-Verlag.
- RUBIN, D. B. (1978). Bayesian inference for causal effects: the role of randomisation. *Ann. Statist.* **6**, 34–58.
- VANSTEELENDT, S. & GOETGHEBEUR, E. (2003). Causal inference with generalized structural mean models. *J. R. Statist. Soc. B* **65**, 817–35.

[Received August 2002. Revised February 2004]