Statistica Sinica **14**(2004), 927-944

MARGINAL STRUCTURAL QUANTILE MODELS FOR LONGITUDINAL OBSERVATIONAL STUDIES WITH TIME-VARYING TREATMENT

Joseph W. Hogan and Joo Yeon Lee

Brown University

Abstract: Marginal structural mean (MSM) models for longitudinal data can be used to characterize the causal effect of a time-varying treatment on the mean of an outcome of interest. Several recent applications of MSM models have demonstrated their utility for quantifying the causal effect of new antiviral therapies and treatment regimens in HIV and AIDS. In this paper we describe marginal structural models for quantiles in which potential outcomes distributions corresponding to different treatment histories differ by quantile-specific location shifts. The formulation of marginal structural quantile (MSQ) models is similar in spirit to quantile regression models, and indeed the MSQ model can be estimated using weighted quantile regression routines under certain circumstances. In this paper we describe the formulation of MSQ models for longitudinal data, list the assumptions under which the models can be identified and estimated, and use the methodology to estimate the causal effect of combination antiviral regimens on both CD4 count and HIV viral load using data from an observational cohort study.

Key words and phrases: Causal inference, counterfactuals, highly active antiretroviral therapy, HIV/AIDS, inversely probability weighting, left censoring, longitudinal cohort study, potential outcomes, quantile regressive, selection bias, viral load, weighted estimating equations.

1. Introduction

The problem of estimating treatment effects from observational studies has broad applicability in public health, economics and social sciences. Robins and colleagues (e.g., Robins (1999a, b), Robins, Hernán and Brumback (2000) and Hernán, Brumback and Robins (2002)) have written extensively on the use of marginal structural models for this purpose, with focus primarily on functions of the mean for repeated measures (e.g., Robins (1999b) and Hernán, Brumback and Robins (2002)) and on hazard functions for event histories (e.g., Hernán, Brumback and Robins (2000, 2001)).

In this paper we describe specification and estimation of structural quantile models, wherein the causal effect of treatment can be parameterized in terms of a location shift at one or more specific quantiles of a distribution. For many types

of data, a single measure of central tendency such as the mean or median may not be sufficient to convey the full extent of treatment effects. In HIV and AIDS, for example, CD4 counts are usually positively skewed, increasing the possibility that the mean may be unduly influenced by a heavy right tail. Transformations are useful for correcting skewness but can diminish interpretability. HIV viral load is another important marker with extreme right skewness, and has the added complication of being left censored: most assays have a lower quantification limit of 50 copies/ml, below which it is not possible to measure viral burden. Identifying models of the mean from censored data is not possible without added assumptions about the full distribution (Hughes (2000)).

Formulating a model in terms of quantiles allows the analyst to study the effect of treatment on an entire distribution. Furthermore, the model formulation and the parameter interpretations are similar to those used for quantile regression (Koenker and Bassett (1978)), except that structural models characterize variation in potential outcomes or counterfactuals (Rubin (1974)). Because causal effects are fundamentally nonidentifiable from observed data (cf., Holland (1986)), estimation of structural models rests on a number of unverifiable assumptions.

For purposes of estimation, we adopt a sequential ignorability assumption that implies treatment is randomly received, conditional on a set of observed confounders. This is a longitudinal version of Rosenbaum and Rubin's (1983) weak ignorability assumption (see also Robins, Greenland and Hu (1999)). Under sequential ignorability, a weighted version of the estimating equations used to fit quantile regression models can be used to estimate parameters of a structural quantile model. The weights are inversely proportional to probability of treatment received and, in this sense, we are applying the theoretical results about inverse weighting used in missing data and causal inference problems in mean models (e.g., Robins, Rotnitzky and Zhao (1995) and Robins (1999b)) to structural models of quantiles. Lipsitz, Fitzmaurice, Molenberghs and Zhao (1997) apply similar technology, but not in a causal framework, for estimation of quantile regression models from longitudinal data with monotone missingness under MAR, and Scharfstein, Rotnitzky and Robins (1999, Section 7.1), use inverse weighting techniques to make group comparisons of medians under both MAR and nonignorable dropout mechanisms (using sensitivity analysis for the latter).

Within the field of public health and epidemiology, causal methods are gaining in importance for research in HIV and AIDS. This is due in large part to the growing amount and availability of natural history data. A prototypical study is the HIV Epidemiology Research Study (HERS) (Smith et al. (1997)), wherein 1,310 women (871 HIV-positive) are followed prospectively for up to seven years on a variety of clinical, behavioral and demographic outcomes related to HIV progression. Like many recent and ongoing natural history studies, the follow-up

period for HERS saw the advent of an important new treatment strategy known as highly-active antiviral therapy, or HAART, which involves the use of multiple antiviral treatments simultaneously (see Carpenter et al. (2000) for details). Although HAART regimens have been tested in the controlled environment of clinical trials, an observational cohort study provides important information about whether the therapy is working in the field. See Ko, Hogan and Mayer (2003) and Hernán, Brumback and Robins (2002) for detailed analyses designed to estimate the causal effects of antiviral treatment on mean CD4 cell counts, and Hernán, Brumback and Robins (2000, 2001) for causal analysis of event histories using proportional hazards models.

The remainder of our paper is organized as follows. The next section defines relevant potential outcomes and describes structural quantile models. Section 3 focuses on parameter identification and estimation under sequential ignorability, and Section 4 illustrates use and interpretation of the models in practice with longitudinal CD4 and viral load data from the HER Study. The objective is to characterize the causal effect of taking HAART on the distribution of CD4 count and viral load. Section 5 concludes.

2. Structural Model for Quantiles

2.1. Potential outcomes

A structural model is by definition a model of potential outcomes (Rubin (1974), Pratt and Schlaifer (1984) and Robins (1999b)). The measurement times are common across individuals, and labeled $t = 1, \ldots, T$. Treatment during the interval $[t-1, t)$ is indexed by $a_t \in \{0, 1\}$, which indicates whether or not HAART was received during the six-month interval leading up to t. Treatment history is denoted by the sequence $\overline{a}_t = \{a_1, \ldots, a_t\}$, which is a member of the set $\mathcal{A}_t = \{0, 1\}^{\otimes t}$. We assume treatment history can be summarized using a known function $g : \mathcal{A}_t \to \mathbb{R}^1$ that maps treatment sequences to a scalar value. Common examples include $g(\overline{a}_t) = a_t$ and $g(\overline{a}_t) = \sum_t a_t$ (cf., Hernán et al. (2002) and Ko et al. (2003)).

For individual i $(= 1, \ldots, n)$ at time t, potential outcomes comprise the set $\{Y_{it}(\overline{a}_t) : \overline{a}_t \in \mathcal{A}_t\}$, where by definition the response of subject i to the treatment sequence $\overline{a}_t$ is $Y_{it}(\overline{a}_t)$. From this set, only one response is observable, namely that corresponding to treatment sequence actually received, denoted by $\overline{A}_{it} = \{A_{i1}, \ldots, A_{it}\}$. The *consistency relation* (Wasserman (1999)) connects the observed response to the potential responses via

$$Y_{it} = \sum_{\overline{a}_t \in \mathcal{A}_t} I(\overline{A}_{it} = \overline{a}_t)\, Y_{it}(\overline{a}_t). \tag{1}$$

2.2. Structural quantile model

Our interest is to identify and estimate parameters indexing a model of the the marginal cumulative distribution function (cdf) $F_{Y_t(\overline{a}_t)}(y)$ of potential outcomes. We adopt a formulation similar to quantile regression for the general location-shift model (Koenker and Bassett (1978)). Valid causal interpretation of estimated parameters will depend on a number of key assumptions, some of which cannot be verified from data. After laying out the model here, we demonstrate in Section 3 how to use the machinery employed in quantile regression problems to estimate parameters of the structural model.

In general we allow the distribution of potential outcomes to depend on observed covariates. Let $\mathbf{x}_{it} = (x_{1it}, \ldots, x_{Kit})$ denote a K-vector of *exogenous* covariates (defined below) observed up to time t, and let $\mathbf{x}^*_{it}(\overline{a}_t)$ represent a $1 \times P$ vector that is a known function of $\mathbf{x}_{it}$ and $g(\overline{a}_t)$. For example, if interest is in estimating causal effects conditional on age at time t, we might set $\mathbf{x}_{it} = (1,\ \text{age}_{it})$ and $\mathbf{x}^*_{it} = (1,\ \text{age}_{it},\ g(\overline{a}_t)\ )$. For a continuously-valued random variable Y, let $Q_\theta(Y)$ denote the θ-th quantile of its distribution, where $0 < \theta < 1$; i.e., $\text{pr}\{Y \leq Q_\theta(Y)\} = \theta$. For $t = 1, \ldots, T$, we assume the potential outcomes are related to treatment $\overline{a}_t$ and covariates $\mathbf{x}_{it}$ via the marginal structural quantile (MSQ) model

$$Q_\theta\{Y_{it}(\overline{a}_t) \mid \mathbf{x}_{it}\} = \mathbf{x}^*_{it}(\overline{a}_t)\boldsymbol{\beta}_{\theta t}, \tag{2}$$

where $\boldsymbol{\beta}_{\theta t}$ is a $P \times 1$ vector of unknown parameters. Alternately we can write

$$Y_{it}(\overline{a}_t) = \mathbf{x}^*_{it}(\overline{a}_t)\boldsymbol{\beta}_{\theta t} + U_{\theta it}(\overline{a}_t), \tag{3}$$

where $U_{\theta it}(\overline{a}_t)$ is an error term having unspecified distribution but satisfying $Q_\theta\{U_{\theta it}(\overline{a}_t)\} = 0$ for all $\overline{a}_t \in \mathcal{A}_t$. Hence the MSQ model can be represented as the location model

$$\text{pr}\{Y_{it}(\overline{a}_t) \leq y \mid \mathbf{x}_{it}\} = F_{U_{\theta t}(\overline{a}_t)}(y - \mathbf{x}^*_{it}(\overline{a}_t)\boldsymbol{\beta}_{\theta t})$$

(Basset and Koenker (1982) and Buchinsky (1998)). Covariates $\mathbf{x}_{it}$ are said to be exogenous if, for every t and for each $\overline{a}_t \in \mathcal{A}_t$, $\mathbf{x}_{it}$ is independent of $U_{\theta it}(\overline{a}_t)$.

Because (2) models features of the potential outcomes, the components of $\boldsymbol{\beta}_{\theta t}$ related to treatment have a causal interpretation, and in particular they can be used to contrast, for two different treatment regimes $\overline{a}_t$ and $\overline{a}'_t$, the marginal population quantiles $Q_\theta\{Y_t(\overline{a}_t)\}$ and $Q_\theta\{Y_t(\overline{a}'_t)\}$. Importantly, treatment effect in a MSQ model is *not* the θth quantile of individual treatment differences. A simple example illustrates more clearly. Let $g(\overline{a}_t) = a_t \in \{0, 1\}$, an indicator of whether treatment was received in the previous interval, and assume $\theta = 0.5$. A simple MSQ model (here a marginal structural median model) is

$$Q_{0.5}\{Y_{it}(\overline{a}_t)\} = \alpha_{(0.5)t} + a_t\beta_{0.5} \quad t = 1, \ldots, T,$$

where the $\{\alpha_{(0.5)t}\}$ are time-specific intercepts, and $\beta_{0.5}$ is the difference in the median of two potential-outcomes distributions: one that corresponds to no one receiving treatment ($a_t = 0$) and the other corresponding to everyone receiving treatment ($a_t = 1$).

3. Parameter Identification and Estimation

In this section we describe procedures for estimating the parameters $\boldsymbol{\beta}_\theta$ in (2) from observed data. Because $\boldsymbol{\beta}_\theta$ is a causal parameter, it cannot be identified from observable data without reliance on unverifiable assumptions. We rely on the assumption of *treatment ignorability* (Rosenbaum and Rubin (1983)), which characterizes the mechanism by which individuals receive treatment at each time point. Other approaches include instrumental variables (for quantile models, see Abadie, Angrist and Imbens (1998)) and principal stratification (Frangakis and Rubin (2002)).

3.1. Sequential treatment ignorability

Consider first the cross-sectional case with binary treatment, and suppose inference is conditional on covariates $\mathbf{x}_i$. The ignorability assumption states that receipt of treatment is independent of potential outcomes, i.e., $A_i \perp\!\!\!\perp Y_i(a) \mid \mathbf{x}_i$, for $a = 0, 1$. Here and throughout we maintain *weak* ignorability assumptions; *strong* ignorability states that the potential outcomes are jointly independent of treatment receipt (see Rosenbaum and Rubin (1983) for details). The longitudinal-data analogue of the ignorability assumption is as follows (recall that $\overline{A}_t$ is treatment history up to but not including t):

A0: Sequential Ignorability. Receipt of treatment at time t is sequentially ignorable given $\mathbf{x}_{it}$ if for all t and for all $\overline{a}_t \in \mathcal{A}_t$, $A_{it} \perp\!\!\!\perp Y_{it}(\overline{a}_t) \mid \mathbf{x}_{it}$.

(Robins (1986), Robins et al. (1999), see also Rubin (1976)). Assumption (A0) can be interpreted to mean that receipt of treatment is sequentially randomized, conditional on model covariates $\mathbf{x}_{it}$. It implies that A_{it} is exogenous given $\mathbf{x}_{it}$, and therefore standard estimation methods that assume exogenous treatment (such as those used in quantile regression software) can be used to estimate the causal parameter $\boldsymbol{\beta}_\theta$; however, (A0) cannot generally be expected to hold for observational studies. In HIV/AIDS for example, the decision to administer treatment may depend on variables other than $\mathbf{x}_{it}$, including previous realizations of the outcome of interest (e.g., CD4 and viral load). Furthermore, receipt of treatment may even depend on the potential outcomes themselves, as would be the case when doctors preferentially treat patients who would be sicker under

no treatment (whereby A_{it} depends on $Y_{it}(0)$), or patients who are expected to respond favorably (whereby A_{it} depends on $Y_{it}(1) - Y_{it}(0)$).

Although (A0) will not generally hold for observational studies, it may be possible to identify a covariate process $\{\mathbf{c}_{it}\}$ that fully or partially explains the conditional dependence between the potential outcomes and treatment receipt, given $\mathbf{x}_{it}$. A covariate that partially explains this dependence is a confounder. (See Greenland, Robins and Pearl (1999) and Geng, Guo and Fung (2002) for a detailed treatment.) We are assuming here that primary interest lies in modeling the conditional distribution of $Y_t(\overline{a}_t)$ given $\mathbf{x}_t$, and that there may exist important confounders that are not part of $\mathbf{x}_t$.) Let $\overline{\mathbf{c}}_{it}$ represent the accumulated history, up to and including t, of $\{\mathbf{c}_{it}\}$. If treatment is sequentially ignorable conditional on $\overline{\mathbf{c}}_{it}$ (and $\mathbf{x}_{it}$), then $\overline{\mathbf{c}}_{it}$ is said to be sufficient for control of confounding (Greenland, Robins and Pearl (1999)). In practice, it will not be possible to know whether a given set of confounders is sufficient to induce ignorability, but in many settings the determinants of treatment administration and clinical outcomes are reasonably well understood. In HIV/AIDS for example, if we are interested in estimating the effect of time-varying treatment on CD4 count or viral load, candidate confounders may include previous measures of CD4 and viral load, previous treatment history, whether progression to AIDS has occurred, and duration of HIV infection. The availability of confounders leads to specialization of Assumption (A0) (see Hernán et al. (2002) for example).

A1: Sequential Ignorability given confounders. Receipt of treatment at time t is sequentially ignorable given confounders $\overline{\mathbf{c}}_{it}$ and past treatment history $\overline{A}_{i,t-1}$ if, for all t and for all $\overline{a}_t \in \mathcal{A}_t$, $Y_{it}(\overline{a}_t) \perp\!\!\!\perp A_{it} \mid (\overline{A}_{i,t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it})$.

This assumption also is referred to as the 'no unmeasured confounders' assumption (e.g., Robins, Greenland and Hu (1999) and Hernán et al. (2002)) because it states that within levels of $\mathbf{x}_t$, the confounders and the past treatment history contain all relevant information about $Y_t(\overline{a}_t)$ that leads to observed treatment A_t.

3.2. Estimation of structural quantile models

Estimation of model parameters from observed data can be most clearly motivated as a missing data problem, where responses under treatments other than the actual treatment received are viewed as missing data (Rubin (1976)). In missing data problems, it is useful to cast the estimation problem in terms of a model for the *full* data − in this case the structural model of all potential outcomes − and then to impose assumptions or constraints that allow estimation of the full-data parameters from the partially observed data.

Against this backdrop, we begin by considering estimation in the full-data setting. Recall that at each time point t, the full data for individual i comprise potential outcomes $\{Y_{it}(\overline{a}_t) : \overline{a}_t \in \mathcal{A}_t\}$, exogenous covariates $\mathbf{x}_{it}$, treatment history $\overline{A}_{i,t-1}$, and confounder history $\overline{\mathbf{c}}_{it}$. Our objective is consistent estimation of $\boldsymbol{\beta}_\theta$ in (2). For treatment history $\overline{a}_t$ and for a fixed $\theta \in (0,1)$, define the $P \times 1$ moment function

$$\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t), \boldsymbol{\beta}_\theta) = \mathbf{x}^*_{it}(\overline{a}_t)^{\mathrm{T}} \, [\, I\{Y_{it}(\overline{a}_t) \geq \mathbf{x}^*_{it}(\overline{a}_t)\boldsymbol{\beta}_\theta\} - \theta \,].$$

For simplicity we write $\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t), \boldsymbol{\beta}_\theta) = \psi_{it}(\overline{a}_t, \boldsymbol{\beta}_\theta)$. Under a correctly specified model for $Q_\theta\{Y_{it}(\overline{a}_t)\}$ and assuming suitable regularity conditions, $E\{\psi_{it}(\overline{a}_t, \boldsymbol{\beta}_\theta)\} = \mathbf{0}$ (Buchinsky (1998)). Hence, if the full data were available, the solution $\widehat{\boldsymbol{\beta}}_\theta$ to the $P \times 1$ system of estimating equations

$$\mathbf{U}(\boldsymbol{\beta}_\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{\overline{a}_t \in \mathcal{A}_t} \psi_{it}(\overline{a}_t, \boldsymbol{\beta}_\theta) = \mathbf{0}$$

would be consistent for $\boldsymbol{\beta}_\theta$ and asymptotically normal. The moment-based estimator can also be characterized as the solution to a linear programming problem and $\widehat{\boldsymbol{\beta}}_\theta$ can be computed using appropriate techniques (Basset and Koenker (1982)).

In reality, the potential outcomes are not all observable. Receipt of specific levels of treatment can be viewed as a selection process that acts on the potential outcomes and yields the observed outcome Y_{it} via (1). If all confounders are observed in the sense that (A1) holds, and if the probability of treatment receipt as a function of confounders is known or can be consistently estimated, then consistent estimates of $\boldsymbol{\beta}_\theta$ can be obtained from a system of weighted estimating equations, where the weight for the contribution of subject i at time t is inversely proportional to the (estimated) probability of receiving the observed treatment $\overline{A}_{it}$. The probability of receiving treatment sequence $\overline{a}_t$, given $\overline{\mathbf{c}}_{it}$ and $\mathbf{x}_{it}$, is

$$\pi_{it}(\overline{a}_t \mid \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}) = \prod_{j=1}^{t} \mathrm{pr}(A_{ij} = a_j \mid \overline{A}_{i,j-1} = \overline{a}_{j-1}, \overline{\mathbf{c}}_{ij}, \mathbf{x}_{ij}), \tag{4}$$

where we define $A_{i0} = 0$ for all i. For binary treatment, $\pi_{it}(\overline{a}_t \mid \overline{\mathbf{c}}_{it}, \mathbf{x}_{it})$ can be consistently estimated if we assume the conditional probability terms under the product sign have a known functional form. For example, let $\lambda_{it}(\overline{a}_{t-1}) = \mathrm{pr}(A_{it} = 1 \mid \overline{A}_{i,t-1} = \overline{a}_{t-1}, \overline{\mathbf{c}}_{it}, \mathbf{x}_{it})$ and assume

$$\mathrm{logit}\{\lambda_{it}(\overline{a}_{t-1}, \boldsymbol{\gamma})\} = f(\overline{a}_{t-1}, \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma}), \tag{5}$$

where f is a known linear function of $\boldsymbol{\gamma}$ and the covariates, and $\boldsymbol{\gamma}$ is an unknown, finite-dimensional parameter. We can then fit the model $\text{logit}\{\lambda_{it}(\overline{A}_{i,t-1}, \boldsymbol{\gamma})\} = f(\overline{A}_{t-1}, \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})$ to observed treatment data in order to estimate $\boldsymbol{\gamma}$. From this point forward, we assume that $\pi_{it}(\overline{a}_t \mid \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}) = \pi_{it}(\overline{a}_t \mid \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})$ is known up to $\boldsymbol{\gamma}$ and furthermore that it can be consistently estimated.

For consistent estimation of $\boldsymbol{\beta}_\theta$, we also require that at time t, each individual must have strictly positive probability of receiving treatment (the positivity assumption; see Hernán, Brumback and Robins (2002, p.1694)). At the true value of $\boldsymbol{\gamma}$, the root $\widehat{\boldsymbol{\beta}}_\theta$ of the estimating function

$$\begin{aligned}\mathbf{U}(\boldsymbol{\beta}_\theta) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{\overline{a}_t \in \mathcal{A}_t} \frac{I(\overline{A}_{it} = \overline{a}_t)}{\pi_{it}(\overline{a}_t \mid \mathbf{c}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})}\, \psi_{it}(\overline{a}_t, \boldsymbol{\beta}_\theta) \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T}\{\pi_{it}(\overline{A}_{it} \mid \mathbf{c}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})\}^{-1}\, \psi_{it}(\overline{A}_{it}, \boldsymbol{\beta}_\theta) \qquad (6)\end{aligned}$$

is consistent for $\boldsymbol{\beta}_\theta$ (see Appendix for details). This remains true when $\pi_{it}(\overline{A}_{it} \mid \mathbf{c}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})$ is replaced by $\pi_{it}(\overline{A}_{it} \mid \mathbf{c}_{it}, \mathbf{x}_{it}; \widehat{\boldsymbol{\gamma}})$, so long as it is an $n^{1/4}$-consistent estimator (Robins and Rotnitzky (1995)). Based on (4) and (5), we use

$$\widehat{\pi}_{it}(\overline{A}_{it} \mid \overline{\mathbf{c}}_{it}, \mathbf{x}_{it}, \widehat{\boldsymbol{\gamma}}) = \prod_{j=1}^{t}[A_{it}\lambda_{it}(\overline{A}_{i,t-1}, \widehat{\boldsymbol{\gamma}}) + (1 - A_{it})\{1 - \lambda_{it}(\overline{A}_{i,t-1}, \widehat{\boldsymbol{\gamma}})\}]. \qquad (7)$$

The form of $\text{Var}\,(\widehat{\boldsymbol{\beta}}_\theta)$ depends on the specific density function for the error terms in (3); however, to avoid making assumptions about the density, we use bootstrap resampling, which provides reliable finite-sample estimators of $\text{Var}\,(\widehat{\boldsymbol{\beta}}_\theta)$ in quantile regression models (Buchinksy (1995)).

The practical implication for estimation of MSQ models is the same as for estimation of marginal structural mean models: under specific assumptions, weighted estimating equations applied to observed data, namely (6), can be used to estimate the structural (causal) model parameter $\boldsymbol{\beta}_\theta$. The fundamental assumptions include these: (i) the cdf's of potential outcomes distributions indexed by $\overline{a}_t$ can differ only by a location shift parameterized through a known function of $\overline{a}_t$; (ii) treatment receipt is sequentially ignorable, given confounders (no unmeasured confounders); (iii) the probability of receiving treatment as a function of confounders is known up to a finite dimensional parameter and can be consistently estimated; (iv) each individual has a positive probability of receiving treatment at each time point. Assumption (iii) can be critiqued using standard lack of fit diagnostics. Neither (i) nor (ii) can be empirically verified. Sensitivity analyses are available for checking violations of (ii) while maintaining (i), (iii) and

(iv); see Ko et al. (2003), Robins (1999b) and Scharfstein et al. (1999, Section 7). Our ability to understand sensitivity to violations of (i) (model mis-specification) is far more limited. We can never know the true potential outcomes models, but we can examine sensitivity to inferences over a wide range of models. We have also made, for convenience, the assumption that treatment history can be summarized as a scalar $g(\overline{a}_t)$. The foregoing results on consistent estimation of $\boldsymbol{\beta}_\theta$ are not dependent on this; see Ko et al. (2003) for illustration of other functionals.

4. Estimating Treatment Effects from a Natural History Study

In this section we illustrate the models and estimation procedures using two examples from the HIV Epidemiology Research Study (HERS), an observational cohort study of the natural history of HIV in women (Smith et al. (1997) and Mayer et al. (2003)). The objective is to estimate the causal effect of highly-active antiviral therapy (HAART) on changes in two important markers, viral load (copies of HIV-1 RNA per ml) and CD4 cell count ($\times$1000 cells/ml). See Ko et al. (2003) for the specific definition of HAART used in HERS. Both viral load and CD4 count are measured at enrollment and then every six months thereafter, for up to six years (twelve visits total). For these analyses, we use data from visits 7 onward (treating visit 7 as a baseline) because HERS enrollment began in 1993 but HAART was not in widespread use until 1996, meaning that HAART was generally unavailable to participants prior to their seventh visit.

4.1. Selection of confounders and estimation of weights

Because both viral load and CD4 count are markers of HIV progression, we use the same set of candidate confounders for both analyses; they are listed in Table 1. These variables were selected because they are believed to affect, either directly or indirectly, both the probability of receiving HAART and the clinical potential outcomes. For cases where a time-varying covariate is not available at the specific measurement time, the last value is carried forward and used to represent the true value of the missing outcome. Viral loads that are used as covariates are in the $\log_{10}$ scale. In constructing the weight model, viral load *covariate* values that fall below the lower detection limit of 50 copies/ml are replaced by a randomly-generated value from the uniform distribution on the interval $[0, \log_{10} 50)$. The weights were estimated by fitting (5), assuming $f(\overline{A}_{i,t-1}, \mathbf{c}_{it}, \mathbf{x}_{it}; \boldsymbol{\gamma})$ is a linear function of visit indicators (i.e., time-specific intercepts) and the covariates listed in Table 1. Because data are available prior to our visit 7 baseline, it is possible to make use of confounding variables prior to baseline.

Following Hernán et al. (2002), we also make use of a stabilizing factor to reduce variability in the weights. The stabilizing factor can be a function of model covariates $\mathbf{x}_{it}$ and of observed treatment history $\overline{A}_{i,t-1}$. Its main purpose is to reduce variability introduced by very large weights (i.e., from individuals with very small $\widehat{\pi}_{it}$) and therefore reduce variability in $\widehat{\boldsymbol{\beta}}_\theta$. The stabilizing factors form the numerators of the weights, and are computed similarly to the $\widehat{\pi}_{it}$. We fit a model

$$\text{logit}\{\lambda^*_{it}(\boldsymbol{\gamma}^*)\} = \mathbf{v}_{it}\boldsymbol{\gamma}^*_0 + A_{i,t-1}\gamma^*_1 + A_{i,t-2}\gamma^*_2 + A_{i,t-1} * A_{i,t-2}\gamma^*_3,$$

where $\lambda^*_{it}(\boldsymbol{\gamma}^*) = \text{pr}(A_{it} = 1 \mid \overline{A}_{i,t-1}, \mathbf{x}_{it}; \boldsymbol{\gamma}^*)$, and $\mathbf{v}_{it} \subset \mathbf{x}_{it}$ is a $1 \times T$ vector of visit indicators with elements $v_{itj} = I(t = j)$. The stabilizing factor, denoted π^*_{it}, is then computed similarly to π_{it}, as shown in (7). Our estimation of the structural quantile models uses estimated stabilized weights $\widehat{w}_{it} = \widehat{\pi}^*_{it}/\widehat{\pi}_{it}$.

Table 1. Confounding variables used for analysis of HERS data. Note ART = antiviral therapy, and denotes a treatment regimen that is not HAART, and typically consists of single antiviral agent.

Variable	Unit or category	Timing
HAART status	yes/no	$t-1$, $t-2$
ART status	yes/no	enrollment, $t-1$
AIDS status	yes/no	$t-1$
HIV symptom scale	0–10	$t-1$
ln CD4	ln cells/mm^3	enrollment, $t-1$
$\log_{10}$ HIV RNA	$\log_{10}$ copies/mm^3	enrollment, $t-1$
ln CD4 $\times$ $\log_{10}$ HIV RNA		$t-1$
HAART $\times$ ln CD4		$t-1$
HAART $\times$ $\log_{10}$ HIV RNA		$t-1$
recent IV drug use	yes/no	enrollment only
lifetime IV drug use	yes/no	enrollment only
Race	Black, White, Other	enrollment only
Years aware of HIV status	0, 1–5, 6+ years	enrollment only

4.2. Fitting structural models to CD4 and viral load data

4.2.1. Descriptive statistics and exploratory analyses

Treating visit 7 as baseline, we use outcome data on 478 women from visits 8 through 12. A total of 2105 CD4 counts and 1990 viral loads are available; 443 (or 22 percent) of the viral loads are left-censored at 50 copies/ml, the lower limit of detection for assays being used.

Because the effect of HAART can depend on underlying HIV progression status (Carpenter et al. (2000), Hernán et al. (2000) and Ko et al. (2003)), we assume that treatment effect depends on baseline CD4 count, categorized as < 200, $200-500$, and > 500. Prior to fitting our model, we estimated treatment-specific cdf's for both markers at each visit and within each category of baseline CD4. We assume for these empirical summaries, and for the structural models in the next subsection, that $g(\overline{a}_t) = a_t$. For the empirical summaries within a baseline CD4 count stratum, this implies that the family of marginal c.d.f.'s $\{F_{Y_t(\overline{a}_t)}(y) : \overline{a}_t \in \mathcal{A}_t\}$ has only two members, $F_{Y_t(1)}$ and $F_{Y_t(0)}$. Their crude estimators are

$$\widehat{F}_{Y_t(0)}(y) = \frac{\sum_{i=1}^{n_t}(1-\widehat{w}_{it})(1-A_{it})I(Y_{it} \le y)}{\sum_{i=1}^{n_t}(1-\widehat{w}_{it})(1-A_{it})},$$
$$\widehat{F}_{Y_t(1)}(y) = \frac{\sum_{i=1}^{n_t}\widehat{w}_{it}A_{it}I(Y_{it} \le y)}{\sum_{i=1}^{n_t}\widehat{w}_{it}A_{it}},$$

where n_t is the number of available measures at t and $\widehat{w}_{it}$ is the estimated stabilized weight described above. Under (A0), $\widehat{w}_{it} = 1$ and under (A1) the estimated weights are used. Figure 1 shows, for visit 9 ($t = 2$), $\widehat{F}_{Y_t(0)}(y)$ and $\widehat{F}_{Y_t(1)}(y)$ for both CD4 and $\log_{10}$ viral load, estimated using data from those with baseline CD4< 200. Each is estimated under both (A0) and (A1), and the pattern is typical to other visits and baseline CD4 groups. For both markers, the location shift due to treatment is wider under (A1), suggesting a that confounders may be accounting for nonrandom treatment receipt such that sicker patients, and/or those more likely to benefit, are preferentially receiving HAART. The plots of viral load highlight the problem with censoring, namely that an appreciable proportion of women have viral loads below 50. For quantile regression, this implies that without making modeling assumptions to extrapolate in the left tail, we can only consider modeling quantiles above a certain proportion. Koenker and Basset (1978, Remark to Theorem 3.4) establish the minimum number of observations needed to identify a linear model of a fixed quantile θ.

4.2.2. Specification and estimation of structural quantile models

For purposes of illustrating the models, we take $g(\overline{a}_t) = a_t$, i.e., receipt of HAART during the previous six months. We model quantiles $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ of CD4 counts and $\theta \in \{0.50, 0.75, 0.90\}$ of $\log_{10}$ viral loads. The log scale is chosen for viral load because treatment effects typically are multiplicative with respect to viral load, and consequently most clinicians gauge treatment effects on the $\log_{10}$ scale. Standard errors are computed using the bootstrap (Efron and Tibshirani (1986)), where resampling is done at the individual level.

Normal quantile plots of bootstrapped model coefficients suggest that their sampling distributions are normally distributed (not shown). This was not true for models fit to absolute viral load, which suggests that for finite sample sizes similar to ours, direct construction of bootstrap confidence intervals is recommended.

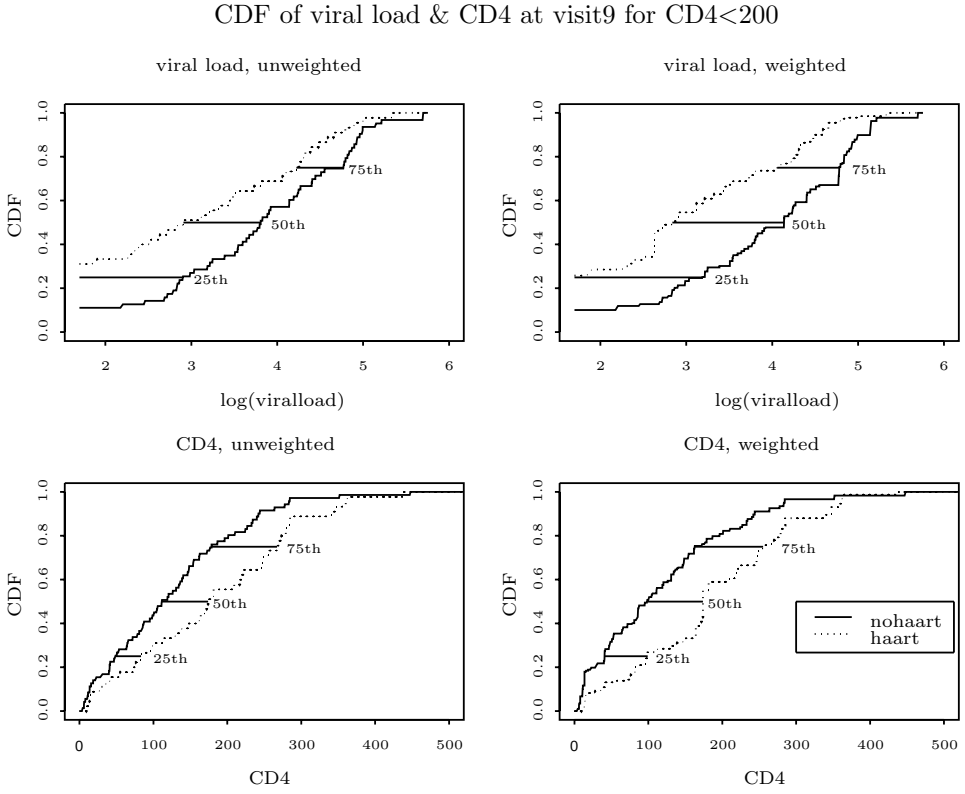

Figure 1. For women with baseline CD4 count less than 200, unweighted and weighted empirical c.d.f.'s for viral load (top row) and CD4 count (bottom row) at visit 9.

As described above, we treat visit 7 as the baseline because prior to this, very few women received HAART. The treatment effects are stratified by baseline CD4 (< 200, $200-500$, > 500), and underlying temporal variations are accounted for using visit indicators as covariates. Let Y_{i0} denote baseline CD4, and define $\mathbf{x}_{it} = (\mathbf{u}_i, \mathbf{v}_{it})$ as follows: $\mathbf{u}_i = [I(Y_{i0} < 200), I(200 \le Y_{i0} \le 500), I(Y_{i0} > 500)]$ indicates CD4 category, and $\mathbf{v}_{it} = (v_{i1}, \ldots, v_{iT})$ is the $1 \times T$ vector of visit indicators (actual visit numbers run from 8 to 12, so $T = 5$), with elements $\{v_{ij} = I(j = t) : j = 1, \ldots, T\}$. For a given θ, the structural quantile model is

$$Q_\theta\{Y_{it}(\overline{a}_t) \mid \mathbf{v}_{it}, \mathbf{u}_i\} = \mathbf{v}_{it}\boldsymbol{\beta}_{\theta 0} + \mathbf{u}_i\boldsymbol{\beta}_{\theta 1} + a_t\mathbf{u}_i\boldsymbol{\beta}_{\theta 2},$$

where $\boldsymbol{\beta}_{\theta 0}$ is a 5×1 vector of visit-specific intercepts, $\boldsymbol{\beta}_{\theta 1}$ (3×1) captures shifts attributable to baseline CD4 count, and $\boldsymbol{\beta}_{\theta 2}$ (3×1) captures (causal) shifts due to treatment within each baseline CD4 category. An important assumption underlying this model is that location shifts of quantile θ due to treatment are constant across visits, within baseline CD4 category. This assumption is made based on visit- and group-specific plots of weighted and unweighted cdf's (similar to Figure 1, not shown).

The model is fitted using the SAS/IML program REGQUANT, available from the SAS Sample Library, which implements the estimation routine given in Koenker and Bassett (1978) and Bassett and Koenker (1982). Weighted estimation under (A1) is implemented by replacing observations $\{Y_{it}, \mathbf{v}_{it}, \mathbf{u}_i, A_{it}\mathbf{u}_i\}$ with with their weighted counterparts $\{\widehat{w}_{it}Y_{it}, \widehat{w}_{it}\mathbf{v}_{it}, \widehat{w}_{it}\mathbf{u}_i, \widehat{w}_{it}A_{it}\mathbf{u}_i\}$ (Lipsitz et al. (1997)). Estimation under (A1) cannot be implemented with the current version of the 'qreg' function in Stata because it forces inclusion of an unweighted intercept term. Standard errors were calculated via bootstrap (Efron and Tibshirani (1986)), using 100 bootstrap samples of individual records. Weights were calculated in advance and were not recalculated within each bootstrap sample.

Estimates of the causal contrasts $\boldsymbol{\beta}_{\theta 2}$ are presented in Table 2. Positive shifts in CD4 (increased immune function) and negative shifts in viral load (decreased viral burden) indicate beneficial treatment effect. For all categories of baseline CD4, HAART exhibits a beneficial (or at least non harmful) effect on both markers. Turning first to those with baseline CD4$<$ 200, the adjustment for confounding increases the treatment effect on the entire distribution of CD4, with the most pronounced effect in the upper tail (effect on the 90th percentile is 32 under (A0) and 51 under (A1)). This pattern also is evident for viral load. Overall, with the exception of the lower tail (10th percentile), HAART therapy shifts nearly the entire CD4 distribution 50 units higher. Among women with baseline CD4 200$-$500, the effect of weighting is similar but more pronounced at the median (where treatment effect is 43 under (A0) and 64 under (A1), a difference of 1.5 standard errors). By contrast, estimated treatment effects for viral load are similar under the two different assumptions, but indicate a clear benefit to HAART therapy. The stratum of women with CD4$>$ 500 has the lowest proportion of women receiving HAART and therefore treatment effects have the greatest standard errors. Changing modeling assumptions has the most pronounced effect for this stratum, where the shift in CD4 due to treatment actually changes direction when moving from (A0) to (A1). There is relatively little change for viral load. It is worth noting that, for CD4 at least, comparing estimates to their standard errors under (A1) does not lead to the conclusion of significant treatment effect. For viral load, there is some evidence of a shift in median. Estimated treatment effects are greater when adjusting for confounders,

which suggests those who are less healthy, and/or those who are less likely to respond well, are more likely to receive HAART (note that both characteristics may be present in the same individual).

Table 2. Treatment effects and bootstrapped standard errors estimated from structural quantile model under assumptions A0 (unweighted) and A1 (weighted). Confounders used under assumption A1 are listed in Table 1.

		CD4 Count		$\log_{10}$ Viral Load	
Baseline CD4	Quantile (θ)	Unweighted	Weighted	Unweighted	Weighted
< 200	0.10	14 (16)	16 (19)		
	0.25	58 (14)	62 (13)		
	0.50	58 (21)	66 (22)	-0.82 (0.35)	-0.97 (0.31)
	0.75	56 (25)	63 (29)	-0.30 (0.15)	-0.43 (0.19)
	0.90	32 (30)	51 (34)	-0.34 (0.10)	-0.50 (0.19)
200 to 500	0.10	5 (16)	-10 (20)		
	0.25	20 (19)	27 (23)		
	0.50	43 (16)	64 (15)	-0.96 (0.23)	-0.82 (0.23)
	0.75	43 (23)	54 (28)	-0.55 (0.14)	-0.48 (0.17)
	0.90	91 (39)	98 (52)	-0.39 (0.12)	-0.35 (0.12)
> 500	0.10	6 (32)	41 (45)		
	0.25	-7 (31)	32 (35)		
	0.50	-55 (52)	25 (47)	-1.03 (0.32)	-0.82 (0.42)
	0.75	-57 (70)	22 (71)	-0.01 (0.25)	0.14 (0.22)
	0.90	-73 (110)	87 (92)	0.06 (0.18)	0.03 (0.21)

To summarize, receipt of HAART during the previous six months has a pronounced therapeutic effect on both CD4 and viral load among women whose baseline CD4 is less than or equal to 500 (the first two strata in Table 2. Treatment effects may or may not be as pronounced for those with CD4 greater than 500: for CD4, the point estimates for distributional shift are on the same order as those in the other two strata, but standard errors are high and the estimates cannot be distinguished from zero; for viral load, except for the median, point estimates indicate very little shift due to treatment. This could be a result of the relatively small degree of variability in viral load for those with baseline CD4$>$ 500, who are among the healthiest HIV-positive women and will tend to have low viral burden.

5. Summary and Discussion

The current work invites a number of possible extensions and topics for further study. Foremost among these is a strategy for sensitivity analysis to gauge the effects of departures from the critical assumptions about confounding

(namely A1). Robins (1999b) gives details for marginal structural models of the mean (see Ko et al. (2003) for an illustration), but extension to quantiles is less obvious. Another possible extension involves second moments, for example, understanding the causal effect on covariation between two markers would expand understanding of how treatment works in two dimensions (see Liang, Wu and Carroll (2003)). Missing data and dropout is an important issue that we have not addressed directly; however it is rather easily handled by introducing a second set of weights along the lines of Lipsitz et al. (1997). See Hernán et al. (2000) for application in a survival analysis context. In a previous analysis of mean CD4 data from HERS (Ko et al. (2003)), the use of additional weights for missing data yielded only small changes in point estimates of model parameters.

Acknowledgement

The authors thank an associate editor, two referees and Jamie Robins for very helpful comments on an earlier version of the manuscript, Lytt Gardner for permission to use the data, and the organizers of the 2002 AMA/SIAM/IMS Summer Research Conference for providing the opportunity to contribute to this special issue. The material in the Appendix is adapted from a set of lecture notes prepared by Marie Davidian (http://www.stat.ncsu.edu/biostat/msm.ps); we thank her for sharing these notes and assume full responsibility for any errors. Support for this project was provided by grants R01-AI50505 and P30-AI42853 from the U.S. National Institutes of Health. Data from the HER Study were collected under grant U64-CCU10675 from the U.S. Centers for Disease Control and Prevention.

Appendix

The following argument demonstrates that $\mathbf{U}(\boldsymbol{\beta}_\theta) = \mathbf{0}$ is an unbiased estimating equation. A fully general study of the properties of weighted estimating equations for estimating marginal structural models is found in Robins (1999a).

Consider the unweighted observed-data estimating equation $\mathbf{U}_0(\boldsymbol{\beta}_\theta) = \mathbf{0}$, where

$$\mathbf{U}_0(\boldsymbol{\beta}_\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \psi(\mathbf{x}_{it}, Y_{it}(\overline{A}_{it}), \boldsymbol{\beta}_\theta),$$

which yields a consistent estimator of $\boldsymbol{\beta}_\theta$ under assumption (A0). Let $R_t = |\mathcal{A}_t| = 2^t$ denote the number of unique treatment sequences $\overline{a}_t$, and let $a_j^{(r)}$ denote treatment received at time $j \leq t$ under the treatment sequence r. The function $\mathbf{U}_0(\boldsymbol{\beta}_\theta)$ can be re-written as

$$\mathbf{U}_0(\boldsymbol{\beta}_\theta) = \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{r=1}^{R_t} \Big\{ \prod_{j=1}^{t} I(A_{ij} = a_j^{(r)}) \Big\} \psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t^{(r)}), \boldsymbol{\beta}_\theta).$$

For a correctly specified MSQ model, $E\{\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t^{(r)}), \boldsymbol{\beta}_\theta)\} = \mathbf{0}$. However, in general the solution to $\mathbf{U}_0(\boldsymbol{\beta}_\theta) = \mathbf{0}$ will be biased for $\boldsymbol{\beta}_\theta$ unless A_{ij} is independent of $\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t^{(r)}), \boldsymbol{\beta}_\theta)$ for all r and for all $j \le t$ (i.e., unless A_{ij} is exogenous).

Now consider forming $\mathbf{U}(\boldsymbol{\beta}_\theta)$ by weighting each term in $\mathbf{U}_0(\boldsymbol{\beta}_\theta)$ inversely by

$$\pi_{it}(\overline{a}_t^{(r)} \mid \mathbf{x}_{it}, \overline{\mathbf{c}}_{it}) = \prod_{j=1}^{t} \text{pr}\left\{A_{ij} = a_j^{(r)} \;\middle|\; \overline{A}_{i,j-1} = \overline{a}_{j-1}^{(r)}, \mathbf{x}_{ij}, \overline{\mathbf{c}}_{ij}\right\}.$$

Then each term of the (weighted) sum in $\mathbf{U}(\boldsymbol{\beta}_\theta)$ has expectation

$$E\left[\frac{\{\prod_{j=1}^{t} I(A_{ij} = a_j)\}\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t), \boldsymbol{\beta}_\theta)}{\prod_{j=1}^{t} \text{pr}(A_{ij} = a_j \mid \overline{A}_{i,j-1} = \overline{a}_{j-1}, \mathbf{x}_{ij}, \overline{\mathbf{c}}_{ij})}\right] \tag{8}$$

(where superscript r has been dropped for clarity). It remains to show that this expectation equals zero. Recall that under Assumption (A1), for any t and for any $\overline{a}_t \in \mathcal{A}_t$,

$$E\{I(A_{it} = a_t) \mid Y_{it}(\overline{a}_t), \overline{A}_{i,t-1} = \overline{a}_{t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it}\} = \text{pr}(A_{it} = a_t \mid \overline{A}_{i,t-1} = \overline{a}_{t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it}).$$

Taking expectation of (8) conditional on $Y_{it}(\overline{a}_t), \overline{A}_{i,t-1}, \mathbf{x}_{it}$ and $\overline{\mathbf{c}}_{it}$, we have

$$\begin{aligned}
&E\left[E\left\{\frac{\{\prod_{j=1}^{t} I(A_{ij} = a_j)\}\psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t), \boldsymbol{\beta}_\theta)}{\prod_{j=1}^{t} \text{pr}(A_{ij} = a_j \mid \overline{A}_{i,j-1} = \overline{a}_{j-1}, \mathbf{x}_{ij}, \overline{\mathbf{c}}_{ij})}\right\} \middle| Y_{it}(\overline{a}_t), \overline{A}_{i,t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it}\right] \\
&= E\left[\frac{\{\prod_{j=1}^{t-1} I(A_{ij} = a_j)\}\ \psi(\mathbf{x}_{it}, Y_{it}(\overline{a}_t), \boldsymbol{\beta}_\theta)}{\prod_{j=1}^{t-1} \text{pr}(A_{ij} = a_j \mid \overline{A}_{i,j-1} = \overline{a}_{j-1}, \mathbf{x}_{ij}, \overline{\mathbf{c}}_{ij})}\right. \\
&\qquad \left. \cdot \frac{E\{I(A_{it} = a_t) \mid Y_{it}(\overline{a}_t), \overline{A}_{i,t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it}\}}{\text{pr}(A_{it} = a_t \mid \overline{A}_{i,t-1} = \overline{a}_{t-1}, \mathbf{x}_{it}, \overline{\mathbf{c}}_{it})}\right],
\end{aligned} \tag{9}$$

and (A1) implies that the second fraction inside the expectation (9) equals one. Next, we take the expectation of (9) conditionally on $Y_{i,t-1}(\overline{a}_{t-1}), \overline{A}_{i,t-2}, \mathbf{x}_{i,t-1}$ and $\overline{\mathbf{c}}_{i,t-1}$ and obtain a similar simplification. Repeating down to $t = 1$ gives the result.

References

Abadie, A., Angrist J. D. and Imbens, G. W. (1998). Instrumental variables estimation of quantile treatment effects. Technical working paper 229, National Bureau of Economic Research, Cambridge, Mass.

Basset, G. and Koenker, R. (1982). An empirical quantile function for linear models with iid errors. *J. Amer. Statist. Assoc.* **77**, 407-415.

Buchinsky, M. (1998). Recent advances in quantile regression methods: a practical guideline for empirical research. *J. Human Resources* **33**, 88-126.

Carpenter, C. C. J., Cooper, D. A., Fischl, J. M., Gatell, J. M., Gazzard, B. G., Hammer, S. M., Hirsch, M. S., Jacobsen, D. M., Katzenstein, D. A., Montaner, J. S. G., Richman, D. D., Saag, M. S., Schechter, M., Schooley, M., Thompson, M. A., Vella, S., Yeni, P. G. and Volberding, P. A. (2000). Antiretroviral therapy in adults:updated recommendations of the International AIDS Society-USA Panel. *J. Amer. Medical Assoc.* **283**, 381-391.

Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1**, 54-77.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21-29.

Geng, Z., Guo, J. and Fung, W. (2002). Criteria for confoundes in epidemiological studies. *J. Roy. Statist. Soc. Ser. B* **64**, 3-15.

Greenland, S., Robins, J. M. and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statist. Sci.* **14**, 29-46.

Hernán, M., Brumback, B. and Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* **11**, 561-570.

Hernán, M. A., Brumback, B. and Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *J. Amer. Statist. Assoc.* **96**, 440-448.

Hernán, M. A., Brumback, B. and Robins, J. M. (2002). Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Statist. Medicine* **21**, 1689-1709.

Holland, P. W. (1986). Statistics and causal inference (with discussion). *J. Amer. Statist. Assoc.* **81**, 945-970.

Hughes, M. D. (2000). Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. *Statist. Medicine* **19**, 3171-3191.

Ko, H., Hogan, J. W. and Mayer, K. H. (2003). Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models. *Biometrics* **59**, 152-162.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Liang, H., Wu, H. and Carroll, R. J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics* **4**, 297-312.

Lipsitz, S. R., Fitzmaurice, G. M., Molenberghs, G. and Zhao, L. P. (1997). Quantile regression methods for longitudinal data with drop-outs: Application to CD4 cell counts of patients infected with the Human Immunodeficiency Virus. *Appl. Statist.* **46**, 463-476.

Mayer, K. H., Hogan, J. W., Smith, D., Klein, R. S., Schuman, P., Margolick, J. B., Korkontzelou, C., Farzedegan, H., Vlahov, D. and Carpenter, C. C. J. (2003). Clinical and immunologic progression in HIV-infected US women before and after the introduction of highly active antiretroviral therapy. *J. Acquired Immune Deficiency Syndrome* **33**, 614-624.

Pratt, J. W. and Schlaifer, R. (1984). On the nature and discovery of structure (with discussion). *J. Amer. Statist. Asssoc.* **79**, 9-33.

Robins, J. M. (1999a). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials* (Edited by M. E. Halloran and D. Berry), 95-134. Springer-Verlag, New York.

Robins, J. M. (1999b). Association, causation and marginal structural models. *Synthese* **121**, 151-179.

Robins, J. M., Greenland, S. and Hu, F. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome (with discussion). *J. Amer. Statist. Assoc.* **94**, 687-712.

Robins, J. M., Hernan, M. A. and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550-560.

Robins, J. M., Rotnitzky, A. and Zhao, L.-P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90**, 106-121.

Rosenbaum, P. and Rubin, D. B. (1983). The central role of the propensity score in observational studies of causal effects. *Biometrika* **70**, 41-55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educational Psychology* **66**, 688-701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Ann. Statist.* **6**, 34-58.

Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. (1999). Adjusting for nonignorable dropout using semiparametric nonresponse models (with discussion). *J. Amer. Statist. Assoc.* **94**, 1096-1146.

Smith, D. K., Warren, D., Vlahov, P., Schuman, P., Stein, M. D., Greenberg, B. L. and Holmberg, S. D. (1997). Design and baseline participant characteristics of the Human Immunodeficiency Virus Epidemiology Research (HER) Study: a prosepective cohort study of human immunodeficiency virus infection in US women. *Amer. J. Epidemiology* **146**, 459-469.

Wasserman, L. (1999). Comment on 'Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary resonse'. *J. Amer. Statist. Assoc.* **94**, 704-706.

Center for Statistical Sciences, Department of Community Health, Brown University, Providence, RI 02912, U.S.A.

E-mail: jwh@brown.edu

Center for Statistical Sciences, Department of Community Health, Brown University, Providence, RI 02912, U.S.A.

E-mail: jooyeon@alexander.stat.brown.edu

(Received March 2003; accepted March 2004)