JOURNAL OF MULTIVARIATE ANALYSIS **61**, 102–128 (1997)
ARTICLE NO. MV971660

Efficiency Comparisons in Multivariate Multiple Regression with Missing Outcomes

Andrea Rotnitzky*

Harvard School of Public Health

Christina A. Holcroft†

University of Massachusetts Lowell

and

James M. Robins‡

Harvard School of Public Health

We consider a follow-up study in which an outcome variable is to be measured at fixed time points and covariate values are measured prior to start of follow-up. We assume that the conditional mean of the outcome given the covariates is a linear function of the covariates and is indexed by occasion-specific regression parameters. In this paper we study the asymptotic properties of several frequently used estimators of the regression parameters, namely the ordinary least squares (OLS), the generalized least squares (GLS), and the generalized estimating equation (GEE) estimators when the complete vector of outcomes is not always observed, the missing data patterns are monotone and the data are missing completely at random (MCAR) in the sense defined by Rubin [11]. We show that when the covariance of the outcome given the covariates is constant, as opposed to the nonmissing data case: (a) the GLS estimator is more efficient than the OLS estimator, (b) the GLS estimator is inefficient, and (c) the semiparametric efficient estimator in a model that imposes linear restrictions only on the conditional mean of the last occasion regression can be less efficient than the efficient estimator in a model that imposes

Received March 21, 1996; revised November 11, 1996.

AMS subject classification: primary 62J05; secondary 62B99.

Key words and phrases: generalized estimating equations, generalized least squares, missing data, repeated measures, semiparametric efficient.

* Supported in part by the National Institutes of Health under Grant 1-R29-GM48704-01A1.

† Supported by the National Institutes of Health under Grants 5T32 CA09337-12&13&14 and 1-R29-GM48704-01A1.

‡ Supported in part by the National Institutes of Health under Grants 2 P30 ES00002, R01AI32475, R01-ES03405, and K04-ES00180.

0047-259X/97 $25.00

linear restrictions on the conditional means of all the outcomes. We provide formulae and calculations of the asymptotic relative efficiencies of the considered estimators in three important cases: (1) for the estimators of the occasion-specific means, (2) for estimators of occasion-specific mean differences, and (3) for estimators of occasion-specific dose-response model parameters.

1. INTRODUCTION

Many randomized and nonrandomized follow-up studies are designed so that outcomes Y_{it}, $t=1, ..., T$, corresponding to the ith subject are to be measured at T prespecified time points and a vector of covariates X_i is to be measured at baseline. In randomized studies, X_i may record a treatment arm indicator as well as pretreatment variables such as age, sex, and race. Often the conditional mean of the outcome Y_{it} given X_i is assumed to be linear in X_i, that is $E(Y_{it} \mid X_i) = \beta_t X_i$, and the goal of the study is to make inferences about the unknown regression parameters β_t. For example, if X_i represents dose levels of a drug administered at baseline, investigators are often interested in estimating the parameter β_t indexing an occasion-specific linear dose-response model. Often a subset of the outcome vector $Y_i = (Y_{i1}, ..., Y_{iT})^{\mathrm{T}}$ is missing for some subjects. In this paper we assume that the outcomes are missing completely at random (MCAR) in the sense defined by Rubin [11] and that the nonresponse patterns are monotone, that is once a subject misses a cycle of the study he or she misses also all subsequent cycles. Monotone patterns of MCAR data arise, for example, in randomized studies with staggered entry and a fixed termination calendar time. Monotone MCAR data also arises if subjects drop out of the study for reasons unrelated to Y_i.

Extensive literature exists on the estimation of parameters $\beta = (\beta_1^T, ..., \beta_T^T)^{\mathrm{T}}$ in the absence of missing data. When the covariance of Y_i given X_i, $\Sigma(X_i)$, is known, then the generalized least squares (GLS) estimator $\hat{\beta}_G$ of β is best linear unbiased [7, p. 301]. Chamberlain [3] showed that the asymptotic variance of $\hat{\beta}_G$ attains the semiparametric variance bound for regular estimators of β in the semiparametric model defined solely by the linear model restrictions on the marginal means. When $\Sigma(X_i)$ is unknown, $\hat{\beta}_G$ is unfeasible because it depends on the unknown covariance function. Carroll and Ruppert [2] showed that when $\Sigma(X_i)$ is a smooth function of X_i, then the two-stage generalized least squares estimator $\tilde{\beta}_G$ that uses a nonparametric estimator of $\Sigma(X_i)$, has the same asymptotic distribution as $\hat{\beta}_G$. The generalized estimating equations (GEE) estimator $\tilde{\beta}_{\mathrm{GEE}}$ proposed by Liang and Zeger [5] is a generalized least squares estimator of β that uses an estimate of $\Sigma(X_i)$ from a, possibly misspecified, parametric model for the covariance function. When the parametric model for the covariance function is correctly specified then $\tilde{\beta}_{\mathrm{GEE}}$ is asymptotically equivalent to $\tilde{\beta}_G$ and $\hat{\beta}_G$.

When the true covariance function does not depend on X_i, i.e., $\Sigma(X_i) = \Sigma$ for all i, then $\hat{\beta}_G$ is exactly equal to $\hat{\beta}_{\mathrm{OLS}} = (\hat{\beta}^T_{1,\mathrm{OLS}}, ..., \hat{\beta}^T_{T,\mathrm{OLS}})^{\mathrm{T}}$, where $\hat{\beta}_{t,\mathrm{OLS}}$ is the ordinary least squares (OLS) estimator of the coefficient in the linear regression for the tth outcome Y_{it} on the covariates X_i [4, pp. 300–307; 12, pp. 395–401]. Thus, when the covariance function is constant, the ordinary least squares estimator of β_t is semiparametric efficient in a model that imposes solely linear restrictions on the conditional means of the outcomes Y_{it} given X_i, $t = 1, ..., T$. The estimator $\hat{\beta}_{t,\mathrm{OLS}}$ is also semiparametric efficient in the model defined by the linear restriction on the tth mean only, i.e., $E(Y_{it} \mid X_i) = \beta_t X_i$, but without restrictions imposed on the conditional means of the remaining outcomes, i.e., $E(Y_{ij} \mid X_i)$, $j \neq t$, is unspecified [9]. Thus, with full data, when $\Sigma(X_i)$ is not a function of X_i, knowledge that the means of the remaining outcomes are linear in X_i does not asymptotically add information about the regression parameter β_t corresponding to the tth outcome. Furthermore, since $\hat{\beta}_{t,\mathrm{OLS}}$ is also the semiparametric efficient estimator of β_t when the outcomes Y_{ij}, $j \neq t$, are not recorded, then we conclude that only the outcome Y_{it} conveys information about β_t when no Y_{it} are missing and $\Sigma(X_i)$ in constant.

With monotone MCAR outcomes the estimators $\hat{\beta}_G$, $\tilde{\beta}_G$, $\tilde{\beta}_{\mathrm{GEE}}$ and $\hat{\beta}_{\mathrm{OLS}}$ are consistent for estimating β but they may be less efficient than the semiparametric efficient estimator $\hat{\beta}_{\mathrm{EFF}}$ of β in the model defined by the linear restrictions on the conditional means of the vector Y_i given X_i and the MCAR condition [9]. The goal of this paper is to compare and explain the asymptotic relative efficiencies of the estimators $\tilde{\beta}_G$, $\tilde{\beta}_{\mathrm{GEE}}$, and $\hat{\beta}_{\mathrm{OLS}}$ relative to $\hat{\beta}_{\mathrm{EFF}}$. In Section 2 we describe the model assumptions. In Section 3 we review well-known results about the estimation of β when the complete vector Y_i is observed for all subjects. In Section 4 we review a class of estimators introduced by Robins and Rotnitzky [9] that includes estimators that are asymptotically equivalent to $\tilde{\beta}_G$, $\tilde{\beta}_{\mathrm{GEE}}$, $\hat{\beta}_{\mathrm{OLS}}$, and $\hat{\beta}_{\mathrm{EFF}}$. In Section 5 we use a representation of the asymptotic variance of the estimators in this class that helps interpreting the source of differences among the asymptotic variances of the various considered estimators. Asymptotic relative efficiencies are explicitly calculated for the various estimators of β in three important special cases, namely, (1) when $X_i = 1$, (2) when $X_i = (1, X_i^*)$ and X_i^* is binary, and (3) when $X_i = (1, X_i^*)$ and X_i^* is an arbitrary explanatory variable. Section 6 contains some final remarks.

2. MODEL

With $i = 1, ..., n$ indexing subject, let Y_{it} be the outcome of the ith subject at the tth follow-up cycle of the study, $t = 1, ..., T$. Let X_i denote a $p \times 1$

vector of explanatory variables for the ith subject measured just prior to start of follow-up. We assume that the first element of the vector X_i is the constant 1. Define $R_{it}=1$ if Y_{it} is observed and $R_{it}=0$ otherwise. We assume that the missing data patterns are monotone, that is $R_{it}=0$ implies $R_{i(t+1)}=0$. We also assume that X_i is completely observed for all subjects and that the vectors (X_i^T, Y_i^T, R_i^T), $i=1, ..., n$, are independent and identically distributed, where $Y_i=(Y_{i1}, ..., Y_{iT})^{\mathrm{T}}$ and $R_i=(R_{i1}, ..., R_{iT})^{\mathrm{T}}$. We further assume that the missing data process satisfies

$$P(R_{it}=1 \mid R_{i(t-1)}=1, Y_i, X_i)=P(R_{it}=1 \mid R_{i(t-1)}=1, X_i). \tag{1}$$

and that

$$P(R_{it}=1 \mid R_{i(t-1)}=1, X_i)>\sigma>0, \tag{2}$$

for some $\sigma>0$. Condition (1) is equivalent to the condition that the data are missing completely at random [11]. Condition (2) says that all subjects have a probability of having the full vector Y_i completely observed that is bounded away from zero. We suppose that the conditional mean of Y_{it} given X_i follows the linear regression model

$$E(Y_{it} \mid X_i)=\beta_{0t}^T X_i, \tag{3}$$

where β_{0t} is a $p\times 1$ vector of unknown parameters, $t=1, ..., T$. Throughout we refer to the semiparametric model defined by restriction (3) as the "all-linear-means" model. The goal of this article is to compare the asymptotic relative efficiencies of several commonly used estimators of β_{0t} when the outcomes Y_{it} are not always observed, the missingness patterns are monotone, and the data are missing competely at random, i.e., Eq. (1) is true.

3. ESTIMATION WITHOUT MISSING DATA

In this section we briefly review well-known results about the estimation of β_{0t} when Y_i is observed for all subjects. Let $\varepsilon_{it}(\beta_t)=Y_{it}-\beta_t^T X_i$, $\varepsilon_i(\beta)=(\varepsilon_{i1}(\beta_1), ..., \varepsilon_{iT}(\beta_T))^{\mathrm{T}}$ with $\beta=(\beta_1^T, ..., \beta_T^T)^{\mathrm{T}}$, and let $d(X_i)$ be a $p\times T$ fixed matrix of functions of X_i. When Y_i is observed for all subjects, then under mild regularity conditions, the estimating equation

$$\sum_{i=1}^{n} d(X_i)\, \varepsilon_i(\beta)=0, \tag{4}$$

has a root that is consistent and asymptotically normal for estimating β_0. Several commonly used estimators of β_0 are solutions to Eq. (4) for some specific choice of $d(X_i)$.

When $\Sigma(X_i)$, the covariance of Y_i given X_i, is known, the generalized least squares estimator $\hat{\beta}_G$ solves (4) that uses $d^*_{\mathrm{GLS}}(X_i) = (I \otimes X_i)\,\Sigma(X_i)^{-1}$, where I is the $T \times T$ identity matrix and $\otimes$ denotes the Kronecker product. The Kronecker product of an $a \times b$ matrix T and a $c \times d$ matrix S is the $ac \times bd$ matrix with block elements $\{T_{ij} S\}$ (Seber, 1984, p. 7). When $\Sigma(X_i)$ is unknown and satisfies certain smoothness conditions, Carroll and Ruppert [2] showed that the two-stage generalized least squares estimator $\tilde{\beta}_G$ that solves (4) with $d_{\mathrm{GLS}}(X_i) = (I \otimes X_i)\,\hat{\Sigma}(X_i)^{-1}$, where $\hat{\Sigma}(X_i)^{-1}$ is a preliminary consistent nonparametric estimator of $\Sigma(X_i)$, has the same asymptotic distribution as $\hat{\beta}_G$.

The GEE estimator [5], $\tilde{\beta}_{\mathrm{GEE}}$, solves (4) with $d_{\mathrm{GEE}}(X_i) = (I \otimes X_i) \times \hat{C}(X_i)^{-1}$, where $\hat{C}(X_i) = C(X_i; \hat{\alpha})$ and $\hat{\alpha}$ is a consistent estimator of α_0 in the model

$$\Sigma(X_i) = C(X_i; \alpha_0), \tag{5}$$

where α_0 is a $q \times 1$ unknown parameter vector and $C(X_i; \alpha)$ is, for each α, a $T \times T$ symmetric and positive definite matrix function of X_i. Liang and Zeger [5] showed that the solution to (4) that uses $d_{\mathrm{GEE}}(X_i)$ will be a consistent and asymptotically normal estimator of β_0 even when (5) is misspecified. In fact, it is standard to show that $\tilde{\beta}_{\mathrm{GEE}}$ will have the same asymptotic distribution as $\hat{\beta}_{\mathrm{GEE}}$ solving Eq. (4) that uses $d^*_{\mathrm{GEE}}(X_i) = (I \otimes X_i)\,C(X_i; \alpha^*)^{-1}$, where α^* is the probability limit of $\hat{\alpha}$ (see, for example, [8]). Thus, when (5) is correctly specified, $d^*_{\mathrm{GEE}}(X_i) = d^*_{\mathrm{GLS}}(X_i)$, and hence $\tilde{\beta}_{\mathrm{GEE}}$ and $\hat{\beta}_G$ have the same asymptotic distribution.

The estimator $\hat{\beta}_{\mathrm{OLS}} = (\hat{\beta}^T_{1,\mathrm{OLS}}, ..., \hat{\beta}^T_{T,\mathrm{OLS}})^{\mathrm{T}}$ in which each $\hat{\beta}_{0t}$ is the ordinary least squares estimator of β_{0t} from the regression of Y_{it} on X_i is also obtained as a solution to Eq. (4). In fact, $\hat{\beta}_{\mathrm{OLS}}$ solves (4) that uses $d_{\mathrm{OLS}}(X_i) = I \otimes X_i$.

Robins and Rotnitzky [9] showed that the solutions to the estimating Eq. (4) essentially constitute all regular and asymptotically linear (RAL) estimators of β_0. That is, any RAL estimator of β_0 is asymptotically equivalent to a solution of Eq. (4) for some choice of function $d(X_i)$. Two estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ of μ_0 are said to be asymptotically equivalent if $\sqrt{n}(\hat{\mu}_1 - \hat{\mu}_2)$ converges to 0 in probability. If $\hat{\mu}_1$ and $\hat{\mu}_2$ are asymptotically equivalent then $\sqrt{n}(\hat{\mu}_1 - \mu_0)$ and $\sqrt{n}(\hat{\mu}_2 - \mu_0)$ have the same asymptotic distribution. An estimator $\hat{\beta}$ is said to be asymptotically linear if $(\hat{\beta} - \beta_0)$ is asymptotically equivalent to a sample average of n i.i.d. mean zero, finite variance random variables. For example, the solution to an estimating equation $\sum_{i=1}^n m(Y_i, X_i; \beta) = 0$ is, under smoothness conditions for $m(Y_i, X_i; \beta)$, asymptotically linear because using a standard Taylor series expansion,

$(\hat{\beta}-\beta_0)$ can be shown to be asymptotically equivalent to the sample average of $E(\partial m(Y_i, X_i; \beta)/\partial\beta|_{\beta=\beta_0}]^{-1} m(Y_i, X_i; \beta_0)$. Regularity is a technical condition that prohibits super-efficient estimators by specifying that the convergence of the estimator to its limiting distribution is locally uniform.

Chamberlain [3] showed that the asymptotic variance of $\hat{\beta}_G$ achieves the semiparametric variance bound for regular estimators of β_0 in the sense defined by Begun, Hall, Huang, and Wellner [1]. The semiparametric variance bound for β_0 in a semiparametric model is the supremum of the Cramer–Rao variance bounds for β_0 over all regular parametric submodels nested within the semiparametric model and it is therefore a lower bound for the asymptotic variance of all regular estimators of β_0.

When $\Sigma(X_i)$ is not a function of X_i, it can easily be shown that $\hat{\beta}_{\mathrm{GLS}}$ and $\hat{\beta}_{\mathrm{OLS}}$ are algebraically identical (see, for example, [4, p. 307]). Thus, $\hat{\beta}_{\mathrm{OLS}}$ coincides with the semiparametric efficient estimator $\hat{\beta}_G$ and it is therefore locally semiparametric efficient in the "all-linear-means" model at the additional restriction that $\Sigma(X_i)$ is constant. A locally semiparametric efficient estimator of a parameter β_0 in model A at an additional restriction B is an estimator that attains the semiparametric variance bound for β_0 in model A when both A and B are true and remains consistent when A is true but B is false.

Consider now the estimation of β_{0T}, the coefficient of the regression of the outcome Y_{iT} on X_i, in a model that does not impose restrictions on the conditional means $E(Y_{it} \mid X_i)$ for $t<T$. Specifically, under the new model, which throughout we call the "last-mean-linear" model, data on X_i and the vector Y_i are observed, $i=1, ..., n$, but the model imposes only a linear restriction on the last conditional mean, i.e.,

$$E(Y_{iT} \mid X_i) = \beta_{0T}^T X_i. \tag{6}$$

Robins and Rotnitzky [9] showed that $\hat{\beta}_{T,\mathrm{OLS}}$ is locally semiparametric efficient for β_0 in the "last-mean-linear" model at the additional restriction that $\mathrm{Var}(Y_{iT} \mid X_i)$ is not a function of X_i. Thus, since $\hat{\beta}_{T,\mathrm{OLS}}$ is also a locally semiparametric efficient estimator of β_{0T} in the "all-linear-means" model at the restriction that $\Sigma(X_i)$ is constant, then it follows that when Y_i is observed for all subjects and $\Sigma(X_i)$ is constant, knowledge that the conditional means $E(Y_{it} \mid X_i)$ for the preceding outcomes $\bar{Y}_{iT} = (Y_{i1}, ..., Y_{i(T-1)})^{\mathrm{T}}$ are linear in X_i does not asymptotically add information about β_{0T}. Furthermore, since $\hat{\beta}_{T,\mathrm{OLS}}$ is also a locally semiparametric estimator of β_{0T} in the model (6) at the restriction that $\mathrm{Var}(Y_{iT} \mid X_i)$ is constant when data on $\bar{Y}_{iT}$ are not recorded [3], then it follows that when $\Sigma(X_i)$ is constant and the "all-linear-means" model holds, data on $\bar{Y}_{iT}$ does not provide information about β_{0T}.

4. ESTIMATION WITH MONOTONE MCAR DATA

In this section we review results about the estimation of β_0 when Y_i is not fully observed for all subjects and the missing data process satisfies (1) and (2). Let $\lambda_{it} \equiv P(R_{it}=1 \mid R_{i(t-1)}=1, X_i)$ and $\pi_{it} \equiv \prod_{i=1}^{t} \lambda_{it}$. Suppose first that

$$\lambda_{it} \text{ are known for all } i \text{ and } t. \tag{7}$$

Robins and Rotnitzky [9] showed that the estimating equation

$$\sum_{i=1}^{n} U_i(d, \phi; \beta) = 0, \tag{8}$$

where

$$U_i(d, \phi\,; \beta) = \frac{R_{iT}}{\pi_{iT}} d(X_i)\, \varepsilon_i(\beta) - \sum_{t=1}^{T} \frac{(R_{it} - \lambda_{it} R_{i(t-1)})}{\pi_{it}} \phi_t(\bar{Y}_{it}, X_i)$$

with $\phi_t(\bar{Y}_{it}, X_i)$, $t=1, ..., T$, an arbitrary $p \times 1$ function of $\bar{Y}_{it} \equiv (Y_{i1}, ..., Y_{i(t-1)})^{\mathrm{T}}$ and X_i chosen by the investigator, has, under mild regularity conditions, a solution $\hat{\beta}(d, \phi)$ that is a consistent and asymptotically normal estimator of β_0. The asymptotic variance of $\hat{\beta}(d, \phi)$ is given by

$$\Gamma(d)^{-1}\, \Omega(d, \phi)\, \Gamma(d)^{-1, T} \tag{9}$$

where $\Gamma(d) = E\{d(X_i)[I \otimes X_i]^T\}$ and $\Omega(d, \phi) = \mathrm{Var}\{U_i(d, \phi; \beta_0)\}$. They also showed that the solutions of (8) are essentially all RAL estimators of β_0 in the "all-linear-means" model with the additional restrictions (1), (2), and (7). Furthermore the solution of (8), $\hat{\beta}(d_{\mathrm{eff}}, \phi_{\mathrm{eff}})$, that uses

$$d_{\mathrm{eff}}(X_i) = (I \otimes X_i) \left\{ \mathrm{Var}\left[\left(\frac{R_{iT}}{\pi_{iT}} \varepsilon_i - \sum_{t=1}^{T} \frac{(R_{it} - \lambda_{it} R_{i(t-1)})}{\pi_{it}} \right.\right.\right.$$
$$\left.\left.\left. \times E(\varepsilon_i \mid \bar{Y}_{it}, X_i) \right) \,\right|\, X_i \right] \right\}^{-1} \tag{10}$$

and

$$\phi_{\mathrm{eff},\, t}(\bar{Y}_{it}, X_i) = d_{\mathrm{eff}}(X_i)\, E(\varepsilon_i \mid \bar{Y}_{it}, X_i), \tag{11}$$

where $\varepsilon_i \equiv \varepsilon_i(\beta_0)$, is semiparametric efficient in this model. In addition, they showed that knowledge of the nonresponse probabilities λ_{it} does not asymptotically provide information about β_0 since the semiparametric efficiency bound for β_0 remains unchanged if the restriction (7) is dropped.

That is, the semiparametric variance bound for β_0 is the same in the models: (a) defined by (1), (2), (3), and (7); (b) defined by (1), (2), and (3). They further showed that all RAL estimators of β_0 in model (b) are asymptotically equivalent to the solution of (8) for some choice of $d(X_i)$ and $\phi_t(\bar{Y}_{it}, X_i)$.

Consider now Eq. (4) restricted to the available observations, i.e.,

$$\sum_{i=1}^{n} d^{\text{obs}}(X_i)\, \varepsilon_i^{\text{obs}}(\beta) = 0, \tag{12}$$

where $\varepsilon_i^{\text{obs}}(\beta)$ is the vector of observed residuals for the ith subject and $d^{\text{obs}}(X_i)$ is the corresponding submatrix of $d(X_i)$. Liang and Zeger [5] showed that (12) has a solution that is consistent and asymptotically normal for estimating β_0. Thus, since this solution is a RAL estimator of β_0, it must have the same asymptotic distribution as a solution of Eq. (8) for some specific $d(X_i)$ and $\phi_t(\bar{Y}_{it}, X_i)$. The estimators $\hat{\beta}_G$, $\tilde{\beta}_G$, $\tilde{\beta}_{\text{GEE}}$ and $\hat{\beta}_{\text{OLS}}$ calculated from the available observations all solve the Eq. (12) using the corresponding submatrices of their respective functions $d^*_{\text{GLS}}(X_i)$, $d_{\text{GLS}}(X_i)$, $d_{\text{GEE}}(X_i)$, and $d_{\text{OLS}}(X_i)$ defined in Section 3. They are therefore asymptotically equivalent to the solution of Eq. (8) for specific functions $d(X_i)$ and $\phi_t(\bar{Y}_{it}, X_i)$. Define

$$\begin{aligned} d^C_{\text{lin}}(X_i) = (I \otimes X_i) \Bigg\{ \text{Var}_C \Bigg[\frac{R_{iT}}{\pi_{iT}} \varepsilon_i - \sum_{t=1}^{T} \frac{(R_{it} - \lambda_{it} R_{i(t-1)})}{\pi_{it}} \\ \times E_{\text{lin},\,C}(\varepsilon_i \mid \bar{Y}_{it}, X_i) \,\Bigg|\, X_i \Bigg] \Bigg\}^{-1}, \end{aligned}$$

$$\phi^C_{\text{lin},\,t}(\bar{Y}_{it}, X_i) = d^C_{\text{lin}}(X_i)\, E_{\text{lin},\,C}(\varepsilon_i \mid \bar{Y}_{it}, X_i),$$

and

$$E_{\text{lin},\,C}(\varepsilon_i \mid \bar{Y}_{it}, X_i) = \text{Cov}_C[(Y_i, \bar{Y}_{it}) \mid X_i)]\, \text{Var}_C(\bar{Y}_{it} \mid X_i)^{-1}\, \bar{\varepsilon}_{it},$$

where $\bar{\varepsilon}_{it}$ is the $(t-1) \times 1$ vector with jth element equal to $Y_{ij} - \beta_{0j}^T X_i$, and C, when used as a subscript of Var and Cov, indicates that the conditional variances and covariances are calculated assuming $\text{Cov}(Y_i \mid X_i) = C(X_i)$, where $C(X_i)$ is a given $T \times T$ symmetric positive definite matrix function of X_i. In the Appendix we show

LEMMA 1. *Let $\hat{\beta}_{\text{lin}}(C)$ be the solution of Eq.* (8) *that uses $d^C_{\text{lin}}(X_i)$ and $\phi^C_{\text{lin},\,t}(\bar{Y}_{it}, X_i)$ instead of $d(X_i)$ and $\phi_t(\bar{Y}_{it}, X_i)$. Then,*

(a) *$\hat{\beta}_G$ and $\tilde{\beta}_G$ are asymptotically equivalent to $\hat{\beta}_{\text{lin}}(\Sigma)$, where $\Sigma(X_i) = \text{Cov}(Y_i \mid X_i)$ is the true conditional covariance of Y_i given X_i;*

(b) $\tilde{\beta}_{\text{GEE}}$ *and* $\hat{\beta}_{\text{GEE}}$ *are asymptotically equivalent to* $\hat{\beta}_{\text{lin}}(C_{\alpha^*})$, *where* $C_{\alpha^*}(X_i) \equiv C(X_i; \alpha^*)$ *is the "working covariance" function defined in Eq.* (5) *evaluated at* α^*. *Here,* α^* *is the probability limit of* $\hat{\alpha}$ *estimated from model* (5); *and*

(c) $\hat{\beta}_{\text{OLS}}$ *is asymptotically equivalent to* $\hat{\beta}_{\text{lin}}(I)$, *where* I *is the* $T \times T$ *identity matrix.*

Part (a) of Lemma 1 was shown by Robins and Rotnitzky [9] and is included here for completeness. Robins and Rotnitzky [9] also showed that the asymptotic variance of $\hat{\beta}_{\text{lin}}(\Sigma)$ is equal to $\Omega(d^{\Sigma}_{\text{lin}}, \phi^{\Sigma}_{\text{lin}})^{-1}$. Part (b) of Lemma 1 implies that when model (5) is correctly specified, $\tilde{\beta}_{\text{GEE}}$ has the same asymptotic distribution as $\hat{\beta}_G$.

Robins and Rotnitzky [9] showed that $\hat{\beta}_G$ has the smallest asymptotic variance in the class of estimators that are solutions to Eq. (12). They also showed that $\hat{\beta}_G$ and the semiparametric efficient estimator $\hat{\beta}(d_{\text{eff}}, \phi_{\text{eff}})$ have the same asymptotic variance if and only if $E_{\text{lin}, \Sigma}(\varepsilon_i \mid \bar{Y}_{it}, X_i) = E(\varepsilon_i \mid \bar{Y}_{it}, X_i)$, i.e., when the conditional expectation of ε_i is linear in $\bar{Y}_{it}$.

In this section we have shown that the estimators $\hat{\beta}_G$, $\tilde{\beta}_G$, $\tilde{\beta}_{\text{GEE}}$, $\hat{\beta}_{\text{GEE}}$, and $\hat{\beta}_{\text{OLS}}$ calculated from all available observations are asymptotically equivalent to solutions of Eq. (8) for specific choices of functions $d(X_i)$ and $\phi_t(\bar{Y}_{it}, X_i)$ when the MCAR condition (1) holds and the missing data patterns are monotone. In Section 3 we noted that, in the absence of missing data, $\hat{\beta}_G$ and $\tilde{\beta}_G$ were semiparametric efficient. As argued previously with monotone MCAR data, $\hat{\beta}_G$ is no longer efficient if the conditional means $E(\varepsilon_i \mid \bar{Y}_{it}, X_i)$ are nonlinear functions of $\bar{Y}_{it}$. In Section 3 we further noted that when $\Sigma(X_i)$ is constant, $\hat{\beta}_G$ and $\hat{\beta}_{\text{OLS}}$ are algebraically identical. This is no longer true with monotone MCAR data. In fact, in the next section we show that large efficiency gains can be obtained by using $\hat{\beta}_G$ instead of $\hat{\beta}_{\text{OLS}}$.

Consider now the estimation of β_{0T} in the "last-mean-linear" model defined by restriction (6) when Y_i is not observed for all subjects and the data are MCAR and monotone. Rotnitzky and Robins [10] showed that all RAL estimators of β_{0T} in the model defined by (1), (2), and (6) are asymptotically equivalent to a solution $\hat{\beta}_T(d^*, \phi^*)$ of

$$\sum_{i=1}^{n} S_i(d^*, \phi^*; \beta_T) = 0,$$

for some specific $p \times 1$ functions $d^*(X_i)$ and $\phi^*(\bar{Y}_{it}, X_i)$, $t = 1, ..., T$. The estimating function S_i is defined as

$$S_i(d^*, \phi^*; \beta_T) = \frac{R_{iT}}{\pi_{iT}} d^*(X_i)\, \varepsilon_{iT}(\beta_T) - \sum_{t=1}^{T} \frac{(R_{it} - \lambda_{it} R_{i(t-1)})}{\pi_{it}} \phi^*_i(\bar{Y}_{it}, X_i).$$

Robins and Rotnitzky [9] also showed that the solution $\hat{\beta}_T(d^*_{\text{eff}}, \phi^*_{\text{eff}})$ that uses

$$d^*_{\text{eff}} = X_i \left\{ \text{Var}\left[\frac{R_{iT}}{\pi_{iT}} \varepsilon_{iT} - \sum_{t=1}^{T} \frac{(R_{it} - \lambda_{it} R_{i(t-1)})}{\pi_{it}} E(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i) \,\middle|\, X_i \right] \right\}^{-1}$$

and $\phi^*_{\text{eff}}(\bar{Y}_{it}, X_i) = d^*_{\text{eff}}(X_i)\, E(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i)$ has asymptotic variance equal to $\Omega^{-1}_{\text{last}} = \text{Var}[S_i(d^*_{\text{eff}}, \phi^*_{\text{eff}}; \beta_{0T})]^{-1}$ that attains the semiparametric variance bound for estimating β_{0T} in the model defined by restrictions (1), (2), and (6).

Since $\hat{\beta}_T(d_{\text{eff}}, \phi_{\text{eff}})$ has asymptotic variance that attains the semiparametric bound in the model that additionally assumes the linearity of the conditional means of Y_{it} given X_i, $t < T$, then if we let the inverse of the variance bound of β_0 represent the amount of information about β_0 in a given model, we have that

$$\frac{\text{AVar}\{\hat{\beta}_T(d_{\text{eff}}, \phi_{\text{eff}})\}^{-1} - \text{AVar}\{\hat{\beta}_T(d^*_{\text{eff}}, \phi^*_{\text{eff}})\}^{-1}}{\text{AVar}\{\hat{\beta}_T(d_{\text{eff}}, \phi_{\text{eff}})\}^{-1}}$$

represents the fraction of the information about β_{0T} associated with the knowledge that $E(Y_{it} \mid X_i)$ is a linear function of X_i for all $t < T$, where for any estimator $\hat{\mu}$ of a parameter μ_0, $\text{AVar}(\hat{\mu})$ denotes the variance of the asymptotic distribution of $\sqrt{n}(\hat{\mu} - \mu_0)$. In Section 5 we examine this fraction for the special case in which $X_i = (1, X^*_i)^{\text{T}}$ for an arbitrary explanatory variable X^*_i.

5. EFFICIENCY COMPARISONS

In this section we compare the asymptotic relative efficiency (ARE) of the various estimators of β_{0t} discussed in Section 4 in the model

$$E(Y_{it} \mid X_i) = \beta_{0,0,t} + \beta_{0,1,t} X^*_i,$$

where X^*_i is a scalar random variable. We start with the case $X^*_i = 0$ which corresponds to the problem of estimating the mean $\beta_{0,0,t}$ of Y_{it}, $t = 1, ..., T$. We then consider the case in which X^*_i is a binary variable and finally the case of an arbitrary covariate X^*_i. Without loss of generality, we focus on the efficiency comparisons of the estimators of the coefficients $\beta_{0,0,T}$ and $\beta_{0,1,T}$ of the model for the conditional mean of the last outcome Y_{iT} given X_i.

5.1. *Estimation of Occasion-Specific Means*

Suppose that X_i consists solely of the constant 1. In this case we are interested in estimating $\beta_{0,0,T}$, the mean of the outcome Y_{iT} measured at the last occasion. To illustrate the asymptotic behavior of the estimators of $\beta_{0,0,T}$ we consider first the simple but pedagogical case in which $T=2$ and Y_{i1} is observed for all subjects. The semiparametric efficient estimator $\hat{\beta}_2(d_{\text{eff}}, \phi_{\text{eff}})$ of $\beta_{0,0,2}$ has asymptotic variance equal to the lower rightmost element of $\Omega_{\text{eff}}^{-1} \equiv \Omega(d_{\text{eff}}, \phi_{\text{eff}})^{-1}$ which can be easily calculated to be

$$\text{Var}(\varepsilon_{i2}) + \frac{1-\lambda_2}{\lambda_2} E[\text{Var}(\varepsilon_{i2} \mid Y_{i1})]. \tag{13}$$

Since $\hat{\beta}_2(d_{\text{eff}}, \phi_{\text{eff}})$ is semiparametric efficient, $I_{\text{MIS}} = \text{AVar}\{\hat{\beta}_2(d_{\text{eff}}, \phi_{\text{eff}})\}^{-1}$ represents the information available for estimating $\beta_{0,0,2}$ when asymptotically, a fraction $1-\lambda_2$ of the outcomes Y_{i2} are missing. Since $I_{\text{FULL}} = \text{Var}(\varepsilon_{i2})^{-1}$ is the information for estimating $\beta_{0,0,2}$ when all Y_{i2}'s are observed then with $\Phi_{\text{eff}} = \lambda_2^{-1}(1-\lambda_2)\, E[\text{Var}(\varepsilon_{i2} \mid Y_{i1})]$,

$$\frac{I_{\text{FULL}} - I_{\text{MIS}}}{I_{\text{FULL}}} = \frac{\Phi_{\text{eff}}}{\text{Var}(\varepsilon_{i2}) + \Phi_{\text{eff}}}$$

represents the fraction of information lost due to missing Y_{i2}'s. This fraction is equal to 0 when $\Phi_{\text{eff}} = 0$, which occurs when $\lambda_2 = 1$, i.e., when Y_{i2} is observed for all subjects, or when $\text{Var}(\varepsilon_{i2} \mid Y_{i1}) = 0$, i.e., when Y_{i1} is a perfect predictor of Y_{i2}.

The asymptotic variance of $\hat{\beta}_{\text{lin}}(C)$ is given by the lower rightmost element of $\Gamma(d_{\text{lin}}^C)^{-1}\, \Omega(d_{\text{lin}}^C, \phi_{\text{lin}}^C)\, \Gamma(d_{\text{lin}}^C)^{-1, T}$. It is easy to show that this element is equal to

$$\text{Var}(\varepsilon_{i2}) + \frac{1-\lambda_2}{\lambda_2} E\{[\varepsilon_{i2} - E_{\text{lin},\, C}(\varepsilon_{i2} \mid \varepsilon_{i1})]^2\}. \tag{14}$$

Formula (14) with $C(X_i) = C(X_i; \alpha^*)$ is, in view of Lemma 1, the asymptotic variance of $\tilde{\beta}_{\text{GEE},2}$, the GEE estimator of $\beta_{0,0,2}$, that uses the "working covariance" model (5). In particular, taking $C(X_i) = I$, the asymptotic variance of $\hat{\beta}_{\text{OLS},2}$ is given by

$$\text{Var}(\varepsilon_{i2}) + \frac{1-\lambda_2}{\lambda_2} [\text{Var}(\varepsilon_{i2})]. \tag{15}$$

Notice that (15) coincides with $\text{Var}(\varepsilon_{i2})/\lambda_2$, which is equal to $\text{Var}(\varepsilon_{i2})$, the asymptotic variance of the normalized estimator of the sample mean of Y_{i2} had no Y_{i2} been missing, divided by λ_2, the fraction of subjects with Y_{i2} observed for large *n*.

Formula (14) says that the asymptotic variance of $\tilde{\beta}_{\mathrm{GEE},2}$ depends on the probability limit of the estimated working covariance only through $E\{[\varepsilon_{i2}-E_{\mathrm{lin},\,C}(\varepsilon_{i2}\mid\varepsilon_{i1})]^2\}$. In particular,

$$\mathrm{AVar}(\tilde{\beta}_{\mathrm{GEE}})-\mathrm{AVar}(\hat{\beta}_{\mathrm{OLS}})=\frac{1-\lambda_2}{\lambda_2}\{E\{[\varepsilon_{i2}-E_{\mathrm{lin},\,C}(\varepsilon_{i2}\mid\varepsilon_{i1})]^2\}-E(\varepsilon_{i2}^2)\}. \tag{16}$$

It follows from (16) that $\hat{\beta}_{\mathrm{OLS},2}$ is not necessarily less efficient than $\hat{\beta}_{\mathrm{GEE},2}$ since for certain choices of working covariance C, the right-hand side of (16) will be positive. For example, if the working covariance model specifies that the covariance of Y_i is constant and equal to

$$\mathrm{Cov}(Y_i)=\begin{pmatrix}1 & -1/2\\ -1/2 & 1\end{pmatrix} \tag{17}$$

but the true covariance of Y_i is $\left(\begin{smallmatrix}1 & \rho_0\\ \rho_0 & 1\end{smallmatrix}\right)$ for some $\rho_0\neq -1/2$, then (16) is equal to $(1-\lambda_2)\,\lambda_2^{-1}$ which is positive if $\rho_0>-(0.5)^2$. This result says that $\tilde{\beta}_{\mathrm{GEE}}$ is not necessarily a more efficient estimator than the ("working-independence") OLS estimator $\hat{\beta}_{\mathrm{OLS}}$. Of course in our example, since $X_i=1$, the GEE estimator that uses an unrestricted model for $\mathrm{Cov}(Y_i\mid X_i)$, that is,

$$\mathrm{Cov}(Y_i\mid X_i)=\begin{pmatrix}\alpha_{01} & \alpha_{02}\\ \alpha_{02} & \alpha_{03}\end{pmatrix},$$

for some unknown parameters α_{01}, α_{02}, and α_{03} is semiparametric efficient and feasible, and it will be preferred to GEE estimators using, possibly incorrect, constant-valued working covariances, such as (17). The point of our example was to show that $\tilde{\beta}_{\mathrm{GEE}}$ can be less efficient that $\hat{\beta}_{\mathrm{OLS}}$ and the working covariance models should be chosen carefully if efficiency improvements over $\hat{\beta}_{\mathrm{OLS}}$ are desired.

The asymptotic variance of the generalized least squares estimator $\hat{\beta}_{G,2}$ is by Lemma 1, equal to part (a)(14) with $C(X_i)=\mathrm{Cov}(Y_i\mid X_i)$ and thus can be written as

$$\mathrm{Var}(\varepsilon_{i2})+\frac{1-\lambda_2}{\lambda_2}E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{i2}\mid Y_{i1})], \tag{18}$$

where $\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{i2}\mid Y_{i2})=\mathrm{Var}(Y_{i2})-\{\mathrm{Cov}(Y_{i1},Y_{i2})^2/\mathrm{Var}(Y_{i1})\}$ is the residual variance from the population linear regression of Y_{i2} on Y_{i1}. In the Appendix we show that $E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{i2}\mid Y_{i1})]$ is equal to $E[\mathrm{Var}(\varepsilon_{i2}\mid Y_{i1}]$ if and only if $E(\varepsilon_{i2} \mid Y_{i1})$ is a linear function of Y_{i1}. Thus $\hat{\beta}_G$ is semiparametric efficient only when $E(\varepsilon_{i2} \mid Y_{i1})$ is linear in Y_{i1}. As noted in Section 4, this has been previously observed by Robins and Rotnitzky [9]. The following argument helps to understand why $\hat{\beta}_{G,2}$ may fail to be semiparametric efficient. For the ith subject with observed outcome Y_{i1}, let $\hat{Y}_{i2}$ be the predicted value of Y_{i2} from the linear regression of Y_{i2} on Y_{i1} based on subjects with observed outcomes at both occasions. That is, letting $\hat{\delta}_1$ and $\hat{\delta}_2$ be the solution of

$$\sum_{i=1}^{n} R_{i2} \begin{pmatrix} 1 \\ Y_{i1} \end{pmatrix} (Y_{i2} - \delta_1 - \delta_2 Y_{i1}) = 0,$$

we define $\hat{Y}_{i2} = \hat{\delta}_1 + \hat{\delta}_2 Y_{i1}$, $i = 1, ..., n$. In the Appendix we show that $\hat{\beta}_{G,2}$ has the same asymptotic distribution as the solution $\hat{\beta}_{\text{IMP},2}$ of

$$\sum_{i=1}^{n} R_{i2}(Y_{i2} - \beta_2) + (1 - R_{i2})(\hat{Y}_{i2} - \beta_2) = 0.$$

The solution $\hat{\beta}_{\text{IMP},2}$ coincides with the regression imputation estimator of $\beta_{0,0,2}$ described by Little and Rubin [6, pp. 45–47]. This estimator is calculated by first imputing the missing Y_{i2}'s with their predicted values from the linear regression of Y_{i2} on Y_{i1} based on the complete data, and then averaging the observed and imputed values of Y_{i2}. The loss of efficiency of $\hat{\beta}_{\text{IMP},2}$ and therefore of $\hat{\beta}_{G,2}$, arises because the missing Y_{i2} are imputed from a model that assumes that $E(Y_{i2} \mid Y_{i1})$ is linear in Y_{i1}. Rotnitzky and Robins [10] showed that, when Y_1 is discrete, $\hat{\beta}_{\text{IMP}}$ can be made semiparametric efficient by replacing $\hat{Y}_{i2}$ by $\hat{E}(Y_{i2} \mid Y_{i1})$, the nonparametric maximum likelihood estimator of $E(Y_{i2} \mid Y_{i1})$.

A comparison of formulas (13), (15), and (18) helps to understand the efficiency differences among $\hat{\beta}_{\text{EFF},2}$, $\hat{\beta}_{G,2}$, and $\hat{\beta}_{\text{OLS},2}$. Since $E[\text{Var}(\varepsilon_{i2} \mid Y_{i1})] \leqslant E[\text{Var}_{\text{lin}}(\varepsilon_{i2} \mid Y_{i1})] \leqslant E[\text{Var}(\varepsilon_{i2})]$, $\hat{\beta}_{\text{OLS},2}$ can never be more efficient that $\hat{\beta}_{G,2}$, which, in turn, can never be more efficient than $\hat{\beta}_{\text{EFF},2}$. In the Appendix we show that $E[\text{Var}(\varepsilon_{i2})] = E[\text{Var}_{\text{lin}}(\varepsilon_{i2} \mid Y_{i1})]$ only when $\text{Cov}(Y_{i1}, Y_{i2}) = 0$ and therefore $\hat{\beta}_{G,2}$ and $\hat{\beta}_{\text{OLS},2}$ will have the same asymptotic variance only when Y_{i1} and Y_{i2} are uncorrelated. The greater efficiency of $\hat{\beta}_{G,2}$ relative to $\hat{\beta}_{\text{OLS},2}$ is therefore explained because $\hat{\beta}_{G,2}$, as opposed to $\hat{\beta}_{\text{OLS},2}$, exploits the correlation between Y_{i1} and Y_{i2} for estimation of β_{02} via the linear regression imputation of the missing Y_{i2}'s. However, as noted earlier, the linear regression imputation of Y_{i2} will only lead to efficient estimators of β_{02} when $E(Y_{i2} \mid Y_{i1})$ is linear in Y_{i1}, and except for this case, $\hat{\beta}_{G,2}$ will fail to extract all the information available in Y_{i1} and Y_{i2} about $\beta_{0,0,2}$.

Given two estimators $\tilde{\mu}$ and $\hat{\mu}$ of a scalar parameter μ, the asymptotic relative efficiency of $\tilde{\mu}$ compared to $\hat{\mu}$ is denoted by $\mathrm{ARE}(\tilde{\mu}, \hat{\mu})$ and is defined by $\mathrm{ARE}(\tilde{\mu}, \hat{\mu}) = \mathrm{AVar}(\hat{\mu})/\mathrm{AVar}(\tilde{\mu})$. With $\hat{\beta}_{\mathrm{EFF}} \equiv \hat{\beta}(d_{\mathrm{eff}}, \phi_{\mathrm{eff}})$, (13) and (18) imply that

$$\mathrm{ARE}(\hat{\beta}_{\mathrm{OLS},\,2}, \hat{\beta}_{\mathrm{EFF},\,2}) = 1 - (1-\lambda_2)\left\{\frac{\mathrm{Var}(\varepsilon_{i2}) - E[\mathrm{Var}(\varepsilon_{i2} \mid Y_{i1})]}{\mathrm{Var}(\varepsilon_{i2})}\right\}$$

and

$$\begin{aligned}&\mathrm{ARE}(\hat{\beta}_{G,\,2}, \hat{\beta}_{\mathrm{EFF},\,2})\\ &\quad = 1 - \frac{(1-\lambda_2)}{1-(1-\lambda_2)\,\rho^2}\left\{\frac{\mathrm{Var}(\varepsilon_{i2}) - E[\mathrm{Var}(\varepsilon_{i2} \mid Y_{i1})]}{\mathrm{Var}(\varepsilon_{i2})} - \rho^2\right\},\end{aligned}$$

where $\rho = \mathrm{Corr}(Y_{i1}, Y_{i2})$. The efficiency loss of $\hat{\beta}_{G,\,2}$ relative to $\hat{\beta}_{\mathrm{EFF},\,2}$ is summarized in the term

$$\frac{(1-\lambda_2)}{1-(1-\lambda_2)\,\rho^2}\left\{\frac{\mathrm{Var}(\varepsilon_{i2}) - E[\mathrm{Var}(\varepsilon_{i2} \mid Y_{i1})]}{\mathrm{Var}(\varepsilon_{i2})} - \rho^2\right\}.$$

The factor $\{\mathrm{Var}(\varepsilon_{i2}) - E[\mathrm{Var}(\varepsilon_{i2} \mid Y_{i1}]\}\ \mathrm{Var}(\varepsilon_{i2})^{-1} - \rho^2$ can be interpreted as a measure of the degree of non-linearity in $E(Y_{i2} \mid Y_{i1})$. This factor is equal to 0 when $E(Y_{i2} \mid Y_{i1})$ is linear in Y_{i1}, and it can be as large as 1. The upper bound 1 is achieved when Y_{i1} and Y_{i2} are uncorrelated but Y_{i1} is a perfect predictor of Y_{i2}, for example if Y_{i1} is normally distributed with zero mean and $Y_{i2} = Y_{i1}^2$. The factor $(1-\lambda_2)/\{1-(1-\lambda_2)\,\rho^2\}$ quantifies the efficiency loss as a function of the fraction of missing Y_{i2}.

EXAMPLE. To illustrate the relative efficiencies of $\hat{\beta}_{\mathrm{OLS},\,2}$ and $\hat{\beta}_{G,\,2}$ compared to the semiparametric efficient estimator $\hat{\beta}_{\mathrm{EFF},\,2}$, consider $Y_{i1} = Z_{i1}^{7/3}$ with

$$\begin{pmatrix} Z_{i1} \\ \varepsilon_{i2} \end{pmatrix} \overset{\text{iid}}{\sim} \mathrm{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2\begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}\right). \tag{19}$$

Since Y_{i1} is a one-to-one transformation of Z_{i1}, $E(\varepsilon_{i2} \mid Y_{i1}) = E(\varepsilon_{i2} \mid Z_{i1})$ and since, by normality, $E(\varepsilon_{i2} \mid Z_{i1})$ is linear in Z_{i1}, then $E(\varepsilon_{i2} \mid Y_{i1}) = a + bY_{i1}^{3/7}$ for some constants a and b. Thus, the conditional mean of Y_{i2} given Y_{i1} is a nonlinear function of Y_{i1}. In the Appendix we show that

$$\mathrm{Corr}(Y_{i1}, Y_{i2}) = \theta\,\mathrm{Corr}(Z_{i1}, Z_{i2}), \tag{20}$$

where $\theta = E(Z_i^{10/3})\, E(Z_i^{14/3})^{-1/2}$. Using the average of 10,000 simulated values of $Z_{i1}^{14/3}$ and $Z_{i1}^{10/3}$ we calculated $\theta \approx 0.88$. Furthermore, $\text{Var}(Y_{i2} \mid Y_{i1})$ is, by Y_{i1} a one-to-one transformation of Z_{i1}, equal to $\text{Var}(Y_{i2} \mid Z_{i1}) = \sigma^2(1-\eta^2)$ and in view of (20), $\text{Var}(Y_{i2} \mid Y_{i1}) = \sigma^2[1-(0.88)^2\,\rho^2]$. Setting $\lambda_2 = 0.5$, the ARE's of $\hat{\beta}_{\text{OLS},\,2}$ and $\hat{\beta}_{G,\,2}$ compared to $\hat{\beta}_{\text{EFF},\,2}$ reduce to

$$\text{ARE}(\hat{\beta}_{\text{OLS},\,2}, \hat{\beta}_{\text{EFF},\,2}) = 1 - (0.5)(0.88)^2\,\rho^2$$

and

$$\text{ARE}(\hat{\beta}_{G,\,2}, \hat{\beta}_{\text{EFF},\,2}) = 1 - \frac{0.5\{1-(0.88)^2\}\,\rho^2}{\{1-0.5\rho^2\}},$$

where $\rho = \text{Corr}(Y_{i1}, Y_{i2})$. Figure 1 plots $\text{ARE}(\hat{\beta}_{\text{OLS},\,2}, \hat{\beta}_{\text{EFF},\,2})$ and $\text{ARE}(\hat{\beta}_{G,\,2}, \hat{\beta}_{\text{EFF},\,2})$ as a function of ρ for $\lambda_2 = 0.5$. The plots indicate that the efficiency of both $\hat{\beta}_{\text{OLS},\,2}$ and $\hat{\beta}_{G,\,2}$ decreases as a function of $|\rho|$. The relatively small efficiency loss of $\hat{\beta}_{G,\,2}$ is due to the relatively small fraction of missing data, i.e., $1-\lambda_2 = 0.5$, and the fact that $E(Y_{i2} \mid Y_{i1})$ is well-approximated by a linear function of Y_{i1}, for values of Y_{i1} lying in a region of high probability. We have also calculated $\text{ARE}(\hat{\beta}_{G,\,2}, \hat{\beta}_{\text{EFF},\,2})$ for $\lambda_2 = 0.2$ (results not presented) and obtained that the ARE reached a minimum of 0.52.

Consider now the estimation of β_{0T} for $T \geqslant 2$. The asymptotic variances of $\hat{\beta}_{\text{EFF},\,T}$ and $\hat{\beta}_{\text{lin}}(C)$ are the lower rightmost elements of $\Omega(d_{\text{eff}}, \phi_{\text{eff}})^{-1}$ and $\Gamma(d^C_{\text{lin}})^{-1}\,\Omega(d^C_{\text{lin}}, \phi^C_{\text{lin}})\,\Gamma(d^C_{\text{lin}})^{-1,\,T}$, respectively. A straightforward calculation gives

$$\text{AVar}(\hat{\beta}_{\text{eff},\,T}) = \text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} E[\text{Var}(\varepsilon_{iT} \mid \bar{Y}_{it})] \tag{21}$$

and

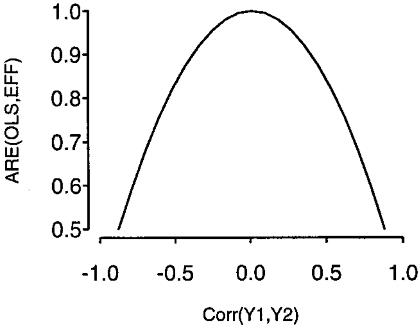
$$\text{AVar}\{\hat{\beta}_{\text{lin},\,T}(C)\} = \text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} E\{[\varepsilon_{iT} - E_{\text{lin},\,C}(\varepsilon_{iT} \mid \bar{Y}_{it})]^2\}. \tag{22}$$

Thus, by Lemma 1, the asymptotic variances of $\hat{\beta}_{G,\,T}$ and $\hat{\beta}_{\text{OLS},\,T}$ are

$$\text{AVar}(\hat{\beta}_{G,\,T}) = \text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} E[\text{Var}_{\text{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it})] \tag{23}$$

and

$$\text{AVar}(\hat{\beta}_{\text{OLS},\,T}) = \text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} E[\text{Var}(\varepsilon_{iT})], \tag{24}$$



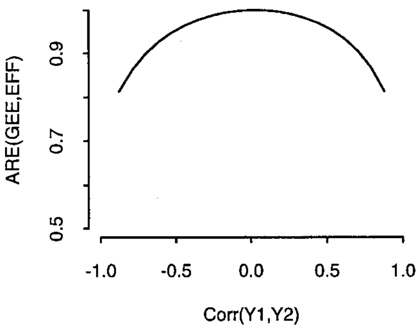


FIG. 1. ARE's for estimating the mean of Y_{i2} when Y_{i1} is always observed and $P(Y_{i2}$ missing$) = 0.5$.

where $\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it}) = \mathrm{Cov}(Y_{iT}, \bar{Y}_{it}) \, \mathrm{Var}(\bar{Y}_{it})^{-1} \, \mathrm{Cov}(\bar{Y}_{it}, Y_{iT})$. Thus, differences in the asymptotic variances of $\hat{\beta}_{\mathrm{EFF},\,T}$, $\hat{\beta}_{G,\,T}$, and $\hat{\beta}_{\mathrm{OLS},\,T}$ are driven by differences among $E[\mathrm{Var}(\varepsilon_{iT} \mid \bar{Y}_{it})]$, $E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it})]$ and $E[\mathrm{Var}(\varepsilon_{iT})]$. Analogously to the case $T=2$, $E[\mathrm{Var}(\varepsilon_{iT} \mid \bar{Y}_{it})] = E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it})]$ if and only if $E[\varepsilon_{iT} \mid \bar{Y}_{it}]$ is a linear function of $\bar{Y}_{it}$, $t = 1, ..., T$, which is then the necessary and sufficient condition for $\hat{\beta}_{G,\,T}$ to be fully efficient. When $\bar{Y}_{iT}$ and ε_{iT} are independent, then $\mathrm{Var}(\varepsilon_{iT} \mid \bar{Y}_{it}) = \mathrm{Var}(\varepsilon_{iT})$ and $\hat{\beta}_{\mathrm{OLS},\,T}$ is efficient. Analogously to the case $T=2$, it can be shown that $\hat{\beta}_{G,\,T}$ is asymptotically equivalent to a regression imputation estimator of the Tth mean in which a missing Y_{iT} from a subject with data observed up to time $t-1$, is imputed with its predicted value from the linear regression of Y_{iT} on $\bar{Y}_{it}$ based on subjects with complete data. Thus, the efficiency loss of $\hat{\beta}_{G,\,T}$ relative to $\hat{\beta}_{\mathrm{EFF},\,T}$ is due to the imputation of the missing Y_{iT} from, possibly misspecified, linear models for $E(Y_{iT} \mid \bar{Y}_{it})$. Since $E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it})] = \mathrm{Var}(\varepsilon_{iT})$ holds for all $t = 1, ..., T$ if and only if $\mathrm{Cov}(\varepsilon_{iT}, \bar{Y}_{iT}) = 0$, it follows that $\hat{\beta}_{\mathrm{OLS}, T}$ and $\hat{\beta}_{G, T}$ will have the same asymptotic distribution only when ε_{iT} and $\bar{Y}_{iT}$ are uncorrelated. Also, $\hat{\beta}_{\mathrm{OLS}, T}$ will be fully efficient only when $\mathrm{Var}(\varepsilon_{iT} \mid \bar{Y}_{iT}) = \mathrm{Var}(\varepsilon_{iT})$.

Finally, as in the case $T = 2$, it can be shown from formula (22) and Lemma 1 that the asymptotic variance of $\tilde{\beta}_{\mathrm{GEE}, T}$ can be larger than the asymptotic variance of $\hat{\beta}_{\mathrm{OLS}, T}$ for some misspecified working covariance models (5).

5.2. *Estimation of Occasion-Specific Mean Differences*

Suppose that X_i^* is a binary indicator variable and consider the model

$$E(Y_{it} \mid X_i) = \beta_{0, 0, t} + \beta_{0, 1, t} X_i^*.$$

In a randomized placebo-controlled follow-up trial for comparing treatment A versus placebo, for example, $X_i^* = 0$ if subject i is assigned to the placebo arm and $X_i^* = 1$ if subject i is assigned to the treatment A arm. Thus, $\beta_{0, 0, t} = E(Y_{it} \mid X_i^* = 0)$ is the occasion-specific mean in the placebo arm and $\beta_{0, 1, t} = E(Y_{it} \mid X_i^* = 1) - E(Y_{it} \mid X_i^* = 0)$ is the occasion-specific difference between the treatment A and placebo means.

Consider now the estimation of $\beta_0 = (\beta_{0, 0, 1}, \beta_{0, 1, 1}, ..., \beta_{0, 0, T}, \beta_{0, 1, T})^{\mathrm{T}}$. Let $\hat{\beta}_{0, G}$ be the generalized least squares estimator of the vector of occasion-specific means in the placebo arm, $\beta_{0, 0} = (\beta_{0, 0, 1}, ..., \beta_{0, 0, T})^{\mathrm{T}}$ computed from placebo-arm data only. Similarly, let $\tilde{\beta}_{0, \mathrm{GEE}}$, $\hat{\beta}_{0, \mathrm{OLS}}$, and $\hat{\beta}_{0, \mathrm{EFF}}$ be the GEE, OLS, and semiparametric efficient estimators of $\beta_{0, 0}$ computed from placebo-arm data only. Define analogously the estimators $\hat{\beta}_{1, G}$, $\tilde{\beta}_{1, \mathrm{GEE}}$, $\hat{\beta}_{1, \mathrm{OLS}}$, and $\hat{\beta}_{1, \mathrm{EFF}}$ of $\beta_{0, 1} = (\beta_{0, 1, 1}, ..., \beta_{0, 1, T})^{\mathrm{T}}$ computed from treatment A-arm data only. In the Appendix we show that the estimators $\hat{\beta}_G$, $\tilde{\beta}_{\mathrm{GEE}}$, $\hat{\beta}_{\mathrm{OLS}}$, and $\hat{\beta}_{\mathrm{EFF}}$ of β_0 can be expressed respectively in terms of $\hat{\beta}_{j, G}$, $\tilde{\beta}_{j, \mathrm{GEE}}$, $\hat{\beta}_{j, \mathrm{OLS}}$, and $\hat{\beta}_{j, \mathrm{EFF}}$, $j = 0, 1$. Specifically, $\hat{\beta}_{G, 0, t}$, the generalized least squares estimator of the intercept of the tth-regression, $t = 1, ..., T$, based on data on both treatment arms coincides with the generalized least squares of the tth mean in the placebo arm, i.e.,

$$\hat{\beta}_{G, 0, t} = \hat{\beta}_{0, G, t}. \tag{25}$$

The generalized least squares estimator $\hat{\beta}_{G, 1, t}$ of the slope in the tth regression, $t = 1, ..., T$, based on data from both treatment arms is equal to the difference between the arm-specific generalized least squares estimators of the tth occasion means, i.e.,

$$\hat{\beta}_{G, 1, t} = \hat{\beta}_{1, G, t} - \hat{\beta}_{0, G, t}. \tag{26}$$

Relationships (25) and (26) hold also for the GEE, OLS, and semiparametric efficient estimators of β_0.

Equation (25) implies that the ARE of the GLS and OLS estimators of the occasion-specific intercepts $\beta_{0,0,t}$ compared to the semiparametric efficient estimator of $\beta_{0,0,t}$ are equal to the ratios of the asymptotic variances given in (23) and (24) to the asymptotic variance given in (22).

It follows from (26) that

$$\mathrm{AVar}(\hat{\beta}_{G,1,t}) = \mathrm{AVar}(\hat{\beta}_{1,G,t}) + \mathrm{AVar}(\hat{\beta}_{0,G,t}),$$

and the same relationship holds for the GEE, OLS, and semiparametric efficient estimator. Furthermore, it follows from (21), (22), (23), and (24) that for $j=0, 1$,

$$\mathrm{AVar}(\hat{\beta}_{j,\mathrm{EFF},t}) = P(X_i=j)^{-1}\Bigg\{\mathrm{Var}(\varepsilon_{it} \mid X_i=j) + \sum_{l=1}^{t} \frac{1-\lambda_{lj}}{\pi_{lj}} \times E[\mathrm{Var}(\varepsilon_{it} \mid \bar{Y}_{il}, X_i=j)]\Bigg\},$$

$$\mathrm{AVar}\{\hat{\beta}_{j,\mathrm{lin},t}(C)\} = P(X_i=j)^{-1}\Bigg\{\mathrm{Var}(\varepsilon_{it} \mid X_i=j) + \sum_{l=1}^{t} \frac{1-\lambda_{lj}}{\pi_{lj}} \times E\{[\varepsilon_{it} - E_{\mathrm{lin},C}(\varepsilon_{it} \mid \bar{Y}_{il}, X_i=j)]^2\}\Bigg\},$$

$$\mathrm{AVar}(\hat{\beta}_{j,G,t}) = P(X_i=j)^{-1}\Bigg\{\mathrm{Var}(\varepsilon_{it} \mid X_i=j) + \sum_{l=1}^{t} \frac{1-\lambda_{lj}}{\pi_{lj}} \times E[\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{it} \mid \bar{Y}_{il}, X_i=j)]\Bigg\}, \tag{27}$$

and

$$\mathrm{AVar}(\hat{\beta}_{j,\mathrm{OLS},t}) = P(X_i=j)^{-1}\Bigg\{\mathrm{Var}(\varepsilon_{it} \mid X_i=j) + \sum_{l=1}^{t} \frac{1-\lambda_{lj}}{\pi_{lj}} E[\mathrm{Var}(\varepsilon_{it} \mid X_i=j)]\Bigg\},$$

where $\lambda_{lj} = P(R_{il}=1 \mid R_{i(l-1)}=1, X_i=j)$ and $\mathrm{Var}_{\mathrm{lin}}(\varepsilon_{it} \mid \bar{Y}_{il}, X_i=j) = \mathrm{Cov}(Y_{it}, \bar{Y}_{il} \mid X_i=j)\,\mathrm{Var}(\bar{Y}_{il} \mid X_i=j)^{-1}\,\bar{\varepsilon}_{il}$. Thus when (a) the nonresponse probabilities λ_{lj} do not depend on the treatment arm, i.e., $\lambda_{lj}=\lambda_l$; (b) the covariance of Y_i is the same for both treatment arms, i.e., $\mathrm{Cov}(Y_i \mid X_i^*) = \mathrm{Cov}(Y_i)$; and (c) $\mathrm{Var}(\varepsilon_{it} \mid \bar{Y}_{il}, X_i^*)$ is not a function of X_i^*, $l=1, \ldots, t$,

$t=1, \ldots, T$, then the ARE of the GLS and OLS estimators of the occasion-specific slopes compared to the semiparametric efficient estimator remain the same as the ARE's of the respective estimators of the occasion-specific intercepts discussed earlier. Finally, as in Section 5.1, it can be shown that the GEE estimator of $\beta_{0,\,1,\,T}$ can be less efficient than the OLS estimator for some misspecified working covariance models.

EXAMPLE. To illustrate the dependence of the ARE's on the difference between the correlation matrices in the two treatment groups, we consider a randomized placebo-controlled study with data measured at baseline and at one follow-up point. We assume that data at baseline are always observed, i.e., $\lambda_{1j}=1$, $j=0, 1$, and that the probability that Y_{i2} is missing is the same in both treatment arms. We assume that $Y_{i1}=Z_{i1}^{7/3}$ and that given X_i^*,

$$\begin{pmatrix} Z_{i1} \\ \varepsilon_{i2} \end{pmatrix} \overset{\text{indep}}{\sim} \text{Normal}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \eta(X_i^*) \\ \eta(X_i^*) & 1 \end{pmatrix}\right). \tag{28}$$

Under (28), $E(Y_{i1} \mid X_i^*)=0$ so in this example we assume that there are no differences in the treatment means at baseline. Thus, within each treatment arm, the data follows the model (19) of Example 1. However, since $\eta(X_i^*)$ is a function of X_i^*, the covariance between Y_{i1} and Y_{i2} changes with treatment arm. A straightforward calculation shows that

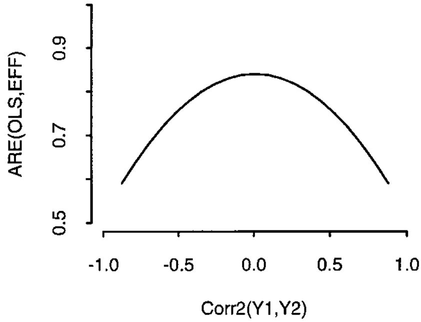
$$\text{AVar}(\hat{\beta}_{\text{EFF},\,1,\,2})=\sigma^2\left\{\frac{1}{\lambda_2 P_0 P_1}-\frac{1-\lambda_2}{\lambda_2}(0.88)^{-2}\frac{(\rho_0^2 P_1+\rho_1^2 P_0)}{P_0 P_1}\right\}, \tag{29}$$

$$\text{AVar}(\hat{\beta}_{G,\,1,\,2})=\sigma^2\left\{\frac{1}{\lambda_2 P_0 P_1}-\frac{1-\lambda_2}{\lambda_2}\frac{(\rho_0^2 P_1+\rho_1^2 P_0)}{P_0 P_1}\right\}, \tag{30}$$

and

$$\text{AVar}(\hat{\beta}_{\text{OLS},\,1,\,2})=\sigma^2\left\{\frac{1}{\lambda_2 P_0 P_1}\right\}, \tag{31}$$

where $\rho_j=\text{Corr}(Y_{i1}, Y_{i2} \mid X_i^*=j)$, $P_j=P(X_i^*=j)$, $j=0, 1$. Figure 2 plots the ARE of the OLS and GLS estimator of $\beta_{0,\,1,\,2}$, the slope in the regression model for the second occasion, compared to the semiparametric efficient estimator of $\beta_{0,\,1,\,2}$ against ρ_1 for $\lambda_2=0.5$, $\rho_0=\sqrt{0.5}$ and $P_0=P_1=0.5$. Both ARE's attain their maximum at $\rho_1=0$, but these maximums are not equal to 1. The OLS estimator is substantially less efficient than the semiparametric efficient estimator when $|\rho_1|$ is large. The GLS estimator



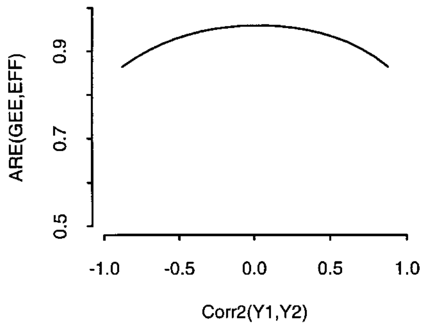


FIG. 2. ARE's for estimating the mean difference of the outcomes Y_{i2} in 2 groups. Here, Y_{i1} is always observed, $P(Y_{i2}\,\text{missing})=0.5$, $\text{Corr}(Y_{i1}, Y_{i2})=\sqrt{0.5}$ in the first group and $\text{Corr2}(Y1, Y2)=\text{Corr}(Y_{i1}, Y_{i2})$ in the second group.

performs relatively well over the whole range of ρ_1 as indicated by the theory since $E(\varepsilon_{i2} \mid Y_{i1})$ is well approximated by a linear function of Y_{i1} over the range of high probability values of Y_{i1}.

5.3. *Estimation of Occasion-Specific Slopes*

We now consider the efficiency of different estimators of β_0 in the model

$$E(Y_{it} \mid X_i) = \beta_{0,0,t} + \beta_{0,1,t} X_i^*, \tag{32}$$

for an arbitrary random variable X_i^*. In what follows it will be convenient to define $\beta_0^* = (\beta_{0,0,1}, \beta_{0,0,2}, ..., \beta_{0,0,T}, \beta_{0,1,1}, ..., \beta_{0,1,T})^{\mathrm{T}}$. The vector β_0^* is obtained by permuting the elements of β_0 so that the first T elements of β_0^*

are the time-ordered intercepts and the last T elements of β_0^* are the time-ordered slopes. The semi-parametric variance bound for estimating β_0^* in model (32) is

$$\Omega_{\text{eff}}^{*\,-1} = E\left\{\begin{pmatrix} I \\ X_i^* I \end{pmatrix} K_{\text{eff}}^{-1}(X_i)(I, X_i^* I)\right\}^{-1}, \tag{33}$$

where I is the $T \times T$ identity matrix and

$$K_{\text{eff}}(X_i) = \text{Var}(\varepsilon_i \mid X_i^*) + \sum_{t=1}^{T} \frac{1-\lambda_{it}}{\pi_{it}} E\{\text{Var}(\varepsilon_i \mid \bar{Y}_{it}, X_i^*) \mid X_i\}.$$

If, for $t = 1, ..., T$,

$$\text{Var}(\varepsilon_i \mid X_i^*),\ E\{\text{Var}(\varepsilon_i \mid \bar{Y}_{it}, X_i^*) \mid X_i^*\},\text{ and } \lambda_{it} \text{ do not depend on } X_i^*, \tag{34}$$

then $K_{\text{eff}}(X_i)$ is a constant matrix and

$$\Omega_{\text{eff}}^* = \begin{pmatrix} K_{\text{eff}}^{-1} & \mu_1 K_{\text{eff}}^{-1} \\ \mu_1 K_{\text{eff}}^{-1} & \mu_2 K_{\text{eff}}^{-1} \end{pmatrix},$$

where $\mu_1 = E(X_i^*)$ and $\mu_2 = E(X_i^{*2})$. The semiparametric variance bound for estimating the vector of occasion-specific slopes $\beta_{0,1} = (\beta_{0,1,1}, ..., \beta_{0,1,T})^{\mathrm{T}}$ is the $T \times T$ lower rightmost block matrix of $\Omega_{\text{eff}}^{*\,-1}$, which, when (34) holds is, by the formula of the inverse of a partitioned matrix, equal to

$$\begin{aligned}\Omega_{1,\text{eff}}^{-1} &= [\mu_2 K_{\text{eff}}^{-1} - \mu_1 K_{\text{eff}}^{-1} K_{\text{eff}} \mu_1 K_{\text{eff}}^{-1}]^{-1} \\ &= K_{\text{eff}}/\text{Var}(X_i^*).\end{aligned}$$

Thus, when (34) holds the semiparametric variance bound for estimating the slope at the last occasion is given by the lower rightmost element of $\Omega_{1,\text{eff}}^{-1}$ and it is equal to

$$\text{AVar}(\hat{\beta}_{\text{EFF},1,T}) = \left[\text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} E\{\text{Var}(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i^*)\}\right] \Big/ \text{Var}(X_i^*). \tag{35}$$

Consider now $\hat{\beta}_G^*$, the generalized least squares estimator of β_0^*. Its asymptotic variance is given by

$$\Omega_{\text{lin}}^{*\,-1} = \left\{\begin{pmatrix} I \\ X_i^* I \end{pmatrix} K_{\text{lin}}^{-1}(X_i)(I, X_i^* I)\right\}^{-1}, \tag{36}$$

where

$$K_{\text{lin}}(X_i) = \text{Var}(\varepsilon_i \mid X_i^*) + \sum_{t=1}^{T} \frac{1-\lambda_{it}}{\pi_{it}} E\{\text{Var}_{\text{lin}}(\varepsilon_i \mid \bar{Y}_{it}, X_i^*) \mid X_i^*\},$$

and $\text{Var}_{\text{lin}}(\varepsilon_i \mid \bar{Y}_{it}, X_i^*) = \text{Cov}(\varepsilon_i, \bar{Y}_{it} \mid X_i)\,\text{Var}(\bar{Y}_{it} \mid X_i)^{-1}\,\text{Cov}(\bar{Y}_{it}, \varepsilon_i \mid X_i)$. When λ_{it} and $\text{Var}(Y_i \mid X_i^*)$ do not depend on X_i^*, an identical argument used to derive (35) now gives

$$\text{AVar}(\hat{\beta}_{G,1,T}) = \left[\text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{1-\lambda_t}{\pi_t} \text{Var}_{\text{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it})\right] \Big/ \text{Var}(X_i^*). \tag{37}$$

Thus, when (34) holds $\hat{\beta}_{G,1,T}$ is semiparametric efficient if and only if $E\{\text{Var}_{\text{lin}}(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i)\} = E\{\text{Var}(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i)\}$ or equivalently when $E(\varepsilon_{iT} \mid \bar{Y}_{it}, X_i)$ is linear in $\bar{Y}_{it}$, as noted also by Robins and Rotnitzky [9].

When λ_{it} is not a function of X_i^*, $\hat{\beta}_{\text{OLS},1,T}$ is computed from a fraction of the outcomes Y_{iT} that, as $n \to \infty$, is equal to π_T. Thus, the asymptotic variance of the OLS estimator of $\beta_{0,1,T}$ is equal to $\text{Var}(\varepsilon_{iT})/\{\pi_T \text{Var}(X_i^*)\}$. A straightforward calculation shows that this variance can be rewritten as

$$\text{AVar}(\hat{\beta}_{\text{OLS},1,T}) = \left[\text{Var}(\varepsilon_{iT}) + \sum_{t=1}^{T} \frac{(1-\lambda_t)}{\pi_t} \text{Var}(\varepsilon_{iT})\right] \Big/ \text{Var}(X_i^*). \tag{38}$$

Comparing Eqs. (37) and (38) to Eqs. (23) and (24), it follows that when (34) holds the asymptotic variances of the estimators $\hat{\beta}_{\text{OLS},1,T}$ and $\hat{\beta}_{G,1,T}$ of the occasion-specific slopes are equal to the asymptotic variances of the corresponding estimators of the occasion-specific means divided by the variance of X_i^*. We conclude that when (34) holds, the asymptotic relative efficiencies of $\hat{\beta}_{\text{OLS},1,T}$ and $\hat{\beta}_{G,1,T}$ compared to $\hat{\beta}_{\text{EFF},1,T}$ are less than or equal to those discussed in Section 5.1 for estimation of the mean of Y_{iT}.

Consider now the estimation of $\beta_{0,1,T}$ in the “last-mean-linear” model (6) with the additional restriction (1), where $X_i = (1, X_i^*)$. The semiparametric variance bound for estimating $\beta_{0,T}$ in this model is given by $\Omega_{\text{last}}^{-1} = \text{Var}\{S_i(d_{\text{eff}}^*, \phi_{\text{eff}}^*; \beta_{0T})\}^{-1}$. It is straightforward to show that

$$\Omega_{\text{last}}^{-1} = E\left\{\begin{pmatrix} K_{\text{eff},T}(X_i)^{-1} & X_i^* K_{\text{eff},T}(X_i)^{-1} \\ X_i^* K_{\text{eff},T}(X_i)^{-1} & X_i^{*2} K_{\text{eff},T}(X_i)^{-1} \end{pmatrix}\right\}^{-1},$$

where $K_{\text{eff},T}(X_i)$ is the lower rightmost element of the $T \times T$ matrix $K_{\text{eff}}(X_i)$. Thus, when (34) holds,

$$\Omega_{\text{last}}^{-1} = K_{\text{eff},T} \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}^{-1},$$

and the semiparametric variance bound for estimating $\beta_{0,1,T}$ is equal to $K_{\mathrm{eff},T}/\mathrm{Var}(X_i^*)$ which, by (35), coincides with $\mathrm{AVar}(\hat{\beta}_{\mathrm{EFF},1,T})$. This result says that when (34) holds, knowledge that the conditional means of Y_{it} given X_i^* are linear functions of X_i^* for $t=1, ..., T-1$ does not asymptotically add information about the parameter $\beta_{0,1,T}$. It is interesting to note that since, (a) $\hat{\beta}_{\mathrm{OLS},1,T}$ is semiparametric efficient when (34) holds and data on $\bar{Y}_{iT}$ are not available and (b) the asymptotic variance of $\hat{\beta}_{\mathrm{OLS},1,T}$ is larger than the asymptotic variance of $\hat{\beta}_{\mathrm{EFF},1,T}$ when, given X_i, $\bar{Y}_{iT}$ and ε_{iT} are statistically dependent; then, as opposed to the full-data case, data on $\bar{Y}_{iT}$ provide information about $\beta_{0,1,T}$ when, given X_i, $\bar{Y}_{iT}$ is a predictor of Y_{iT}. When (34) is not true, the lower rightmost elements of $\Omega_{\mathrm{eff}}^{*-1}$ and $\Omega_{\mathrm{last}}^{*-1}$ may not be equal. In such cases, knowledge of the linearity of the conditional means of Y_{it} given X_i, does provide additional information about $\beta_{0,1,T}$.

Finally, the asymptotic variance of $\tilde{\beta}_{\mathrm{GEE},T}$ is given by (9) with d_{lin} and ϕ_{lin} defined in Lemma 1(b). The results of Section 5.1 suggest that $\tilde{\beta}_{\mathrm{GEE},T}$ can be even less efficient than $\hat{\beta}_{\mathrm{OLS},T}$ for some misspecified working covariance models (5). A detailed study of which estimated covariances $\hat{C}(X_i)$ lead to $\tilde{\beta}_{\mathrm{GEE},T}$ being less efficient than $\hat{\beta}_{\mathrm{OLS},T}$ is beyond the scope of this paper.

6. FINAL REMARKS

In this paper we have examined the relative efficiencies of various estimators of the parameter β_t indexing the occasion-specific linear models for the conditional means of Y_{it} given X_i, $t=1, ..., T$, when the outcomes Y_{it} are MCAR and the missing data patterns are monotone. We have shown that, as opposed to the case in which the full-data vector Y_i is observed for all subjects, the GLS and OLS estimators can be less efficient than the semiparametric efficient estimator of β_t. We have noted that the efficiency loss of the GLS estimator of β_t is related to the degree of nonlinearity of the conditional means $E(Y_{it} \mid \bar{Y}_{it}, X_i)$ as functions of $\bar{Y}_{it}$. We also observed that, as opposed to the full-data case, the OLS estimator of β_t is inefficient since it only uses X_i and the outcomes Y_{it} recorded at the tth occasion, and with monotone missing data, the outcomes $\bar{Y}_{it}$ recorded prior to time t carry information about β_t.

Finally, the results of Lemma 1 are valid also when model (3) is replaced by $E(Y_{it} \mid X_i) = g_t(X_i, \beta_0)$, where $g_t(X_i; \beta_0)$ is a, possibly nonlinear, function of X_i and β_0. When $g_t(X_i; \beta_0)$ depends on β_0 only through the occasion-specific parameters β_{0t}, but $g_t(X_i; \beta_0)$ is not a linear function of β_{0t}, then $\hat{\beta}_{\mathrm{OLS}}$ and $\hat{\beta}_G$ are no longer equal, even when no Y_{it}'s are missing. Thus, with full-data and nonlinear conditional mean models, data on Y_{ij}, $j \neq t$, provide information about the occasion-specific parameters indexing the conditional mean of Y_{it} given X_i.

APPENDIX

Proof of Lemma 1. Part (a) is exactly Lemma 1 of Robins and Rotnitzky [9]. To prove part (b) we will show that $\hat{\beta}_{\mathrm{GEE}} = \hat{\beta}_{\mathrm{lin}}(C_{\alpha*})$ and then argue that since $\tilde{\beta}_{\mathrm{GEE}}$ is asymptotically equivalent to $\hat{\beta}_{\mathrm{GEE}}$ then $\tilde{\beta}_{\mathrm{GEE}}$ and $\hat{\beta}_{\mathrm{lin}}(C_{\alpha*})$ must have the same asymptotic distribution. The estimator $\hat{\beta}_{\mathrm{GEE}}$ solves

$$\sum_{i=1}^{n} (I \otimes X_i)\, C_{\alpha*}(X_i)^{-1}\, \Delta_i \varepsilon_i(\beta) = 0, \tag{39}$$

where $\Delta_i = \mathrm{diag}(R_{ij})$ is the $T \times T$ diagonal matrix with diagonal elements R_{ij}, $j = 1, ..., T$. Robins and Rotnitzky [9] showed that when $\mathrm{Cov}(Y_i \mid X_i) = C_{\alpha*}(X_i)$,

$$(I \otimes X_i)\, C_{\alpha*}(X_i)^{-1}\, \Delta_i \varepsilon_i = U_i(d^{C}_{\mathrm{lin}}, \phi^{C}_{\mathrm{lin}}; \beta_0), \tag{40}$$

where d^{C}_{lin} and ϕ^{C}_{lin} are defined in Section 4. By definition, $U_i(d^{C}_{\mathrm{lin}}, \phi^{C}_{\mathrm{lin}}; \beta_0)$ is a linear function of ε_i. Thus, $U_i(d^{C}_{\mathrm{lin}}, \phi^{C}_{\mathrm{lin}}; \beta_0) = a(X_i, R_i)\, \varepsilon_i$ for some $a(X_i, R_i)$. Let $b(X_i, R_i) \equiv (I \otimes X_i)\, C_{\alpha*}(X_i)^{-1}\, \Delta_i$ and $h(X_i, R_i) \equiv a(X_i, R_i) - b(X_i, R_i)$. By (40), $h(X_i, R_i)\, \varepsilon_i = 0$ when $\mathrm{Cov}(Y_i \mid X_i) = C_{\alpha*}(X_i)$. Thus, by the MCAR assumption (1), $\mathrm{Cov}[h(X_i, R_i)\, \varepsilon_i \mid X_i, R_i] = h(X_i, R_i) \times C_{\alpha*}(X_i)\, h(X_i, R_i)^{\mathrm{T}} = 0$ which, by $C(X_i)$ a positive definite matrix, implies that $h(X_i, R_i) = 0$ almost everywhere. Hence, $a(X_i, R_i) = b(X_i, R_i)$ a.e. and Eq. (40) is true even when $\mathrm{Cov}(Y_i \mid X_i) \neq C(X_i)$ which ends the proof of part (b). Part (c) follows immediately from part (b) by noting that $\hat{\beta}_{\mathrm{OLS}}$ solves (39) with $C_{\alpha*}(X_i) = I$.

Proof that $E[\mathrm{Var}_{\mathrm{lin}}(Y_2 \mid Y_1)] = E[\mathrm{Var}(Y_2 \mid Y_1)]$ *is equivalent to* $E(Y_2 \mid Y_1)$ *is linear in* Y_1. Suppose first that $E(Y_2 \mid Y_1)$ is linear in Y_1, then $E(Y_2 \mid Y_1) = E(Y_2) + \mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1$ and $\mathrm{Var}[E(Y_2 \mid Y_1)] = \mathrm{Cov}(Y_1, Y_2)^2\, \mathrm{Var}(Y_1)^{-1}$. Thus $E[\mathrm{Var}(Y_2 \mid Y_1)] = \mathrm{Var}(Y_2) - \mathrm{Var}[E(Y_2 \mid Y_1)]$ implies $E[\mathrm{Var}(Y_2 \mid Y_1)] = \mathrm{Var}(Y_2) - \mathrm{Cov}(Y_1, Y_2)^2 \times \mathrm{Var}(Y_1)^{-1}$ which proves that $E[\mathrm{Var}(Y_2 \mid Y_1)] = E[\mathrm{Var}_{\mathrm{lin}}(Y_2 \mid Y_1)]$. Suppose now that $E[\mathrm{Var}(Y_2 \mid Y_1)] = \mathrm{Var}(Y_2) - \mathrm{Cov}(Y_1, Y_2)^2\, \mathrm{Var}(Y_1)^{-1}$, then $\mathrm{Var}[E(Y_2 \mid Y_1)] = \mathrm{Cov}(Y_1, Y_2)^2\, \mathrm{Var}(Y_1)^{-1}$. Thus, $\mathrm{Var}[E(Y_2 \mid Y_1)] = \mathrm{Var}[\mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1]$. Now, $\mathrm{Var}[E(Y_2 \mid Y_1) - \mathrm{Cov}(Y_1, Y_2) \times \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1] = \mathrm{Var}[E(Y_2 \mid Y_1) + \mathrm{Var}[\mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1] - 2\,\mathrm{Cov}[E(Y_2 \mid Y_1)\, \mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1]$. But $\mathrm{Cov}[E(Y_2 \mid Y_1) \times \mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1] = E[Y_2 \varepsilon_1]\, \mathrm{Cov}(Y_1, Y_2) \times \mathrm{Var}(Y_1)^{-1} = \mathrm{Cov}(Y_1, Y_2)^2\, \mathrm{Var}(Y_1)^{-1}$. Thus $\mathrm{Var}[E(Y_2 \mid Y_1) - \mathrm{Cov}(Y_1, Y_2)\, \mathrm{Var}(Y_1)^{-1}\, \varepsilon_1] = 0$ which proves the assertion.

Proof that $\mathrm{Var}_{\mathrm{lin}}(Y_2 \mid Y_1) = \mathrm{Var}(Y_2)$ *is equivalent to* $\mathrm{Cov}(Y_1, Y_2) = 0$. By definition $\mathrm{Var}_{\mathrm{lin}}(Y_2 \mid Y_1) = \mathrm{Var}(Y_2) - \mathrm{Cov}(Y_1, Y_2)^2\, \mathrm{Var}(Y_1)^{-1}$; thus

$\mathrm{Var}_{\mathrm{lin}}(Y_2 \mid Y_1) = \mathrm{Var}(Y_2) \Leftrightarrow \mathrm{Cov}(Y_1, Y_2)^2 \, \mathrm{Var}(Y_1)^{-1} = 0$ which is equivalent to $\mathrm{Cov}(Y_1, Y_2) = 0$.

Proof that $\hat{\beta}_{G,2}$ and $\hat{\beta}_{\mathrm{IMP},2}$ are asymptotically equivalent. Since $\hat{\beta}_G = \hat{\beta}_{\mathrm{lin}}(\Sigma)$, then by definition of $\hat{\beta}_{\mathrm{lin},2}(\Sigma)$,

$$\sqrt{n}(\hat{\beta}_{G,2} - \beta_0) = n^{-1/2} \times \sum_{i=1}^{n} \left\{ \frac{R_{i2}}{\lambda_{i2}} \varepsilon_{i2} - \frac{R_{i2} - \lambda_{i2}}{\lambda_{i2}} \mathrm{Cov}(Y_{i1}, Y_{i2}) \, \mathrm{Var}(Y_{i1})^{-1} \, \varepsilon_{i1} \right\}. \tag{41}$$

Also, by definition of $\hat{\beta}_{\mathrm{IMP},2}$,

$$\sqrt{n}(\hat{\beta}_{\mathrm{IMP},2} - \beta_0) = n^{-1/2} \sum_{i=1}^{n} \left\{ R_{i2}\varepsilon_{i2} + (1 - R_{i2}) \left[\bar{Y}_{2,\mathrm{obs}} + \frac{\widehat{\mathrm{Cov}}(Y_1, Y_2)}{\widehat{\mathrm{Var}}(Y_1)} \hat{\varepsilon}_{i1} - \beta_2 \right] \right\},$$

where $\widehat{\mathrm{Cov}}(Y_1, Y_2)$ and $\widehat{\mathrm{Var}}(Y_1)$ are the sample covariance of Y_{i1} and Y_{i2} and the sample variance of Y_{i1} among subjects with $R_{i2} = 1$, $\hat{\varepsilon}_{i1} = Y_{i1} - \bar{Y}_{1,\mathrm{obs}}$ and $\bar{Y}_{j,\mathrm{obs}}$, $j = 1, 2$, is the sample average of Y_{ij} from subjects with $R_{i2} = 1$. Now,

$$\sum_{i=1}^{n} \{ R_{i2}\varepsilon_{i2} + (1 - R_{i2})(\bar{Y}_{2,\mathrm{obs}} - \beta_2) \} = \frac{n \sum_{i=1}^{n} R_{i2}\varepsilon_{i2}}{\sum_{i=1}^{n} R_{i2}}$$

and

$$\sum_{i=1}^{n} (1 - R_{i2}) \frac{\widehat{\mathrm{Cov}}(Y_1, Y_2)}{\widehat{\mathrm{Var}}(Y_1)} \hat{\varepsilon}_{i1} = n \left\{ \frac{\sum_{i=1}^{n} \varepsilon_{i1}}{n} - \frac{\sum_{i=1}^{n} R_{i2}\varepsilon_{i1}}{\sum_{i=1}^{n} R_{i2}} \right\} \frac{\widehat{\mathrm{Cov}}(Y_1, Y_2)}{\widehat{\mathrm{Var}}(Y_1)}.$$

Thus,

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_{\mathrm{IMP},2} - \beta_0)
&= \sqrt{n} \left\{ \frac{\sum_{i=1}^{n} R_{i2}\varepsilon_{i2}}{\sum_{i=1}^{n} R_{i2}} + \frac{\widehat{\mathrm{Cov}}(Y_1, Y_2)}{\widehat{\mathrm{Var}}(Y_1)} \left[\frac{\sum_{i=1}^{n} \varepsilon_{i1}}{n} - \frac{\sum_{i=1}^{n} R_{i2}\varepsilon_{i1}}{\sum_{i=1}^{n} R_{i2}} \right] \right\} \\
&= \sqrt{n} \left\{ \frac{\sum_{i=1}^{n} R_{i2}\varepsilon_{i2}}{n\lambda_2} + \frac{\mathrm{Cov}(Y_1, Y_2)}{\mathrm{Var}(Y_1)} \left[\frac{\sum_{i=1}^{n} \varepsilon_{i1}}{n} - \frac{\sum_{i=1}^{n} R_{i2}\varepsilon_{i1}}{n\lambda_2} \right] \right\} + o_p(1) \\
&= n^{-1/2} \left\{ \sum_{i=1}^{n} \frac{R_{i2}\varepsilon_{i2}}{\lambda_2} - \frac{\mathrm{Cov}(Y_1, Y_2)}{\mathrm{Var}(Y_1)} \sum_{i=1}^{n} \frac{R_{i2} - \lambda_2}{\lambda_2} \varepsilon_{i1} \right\} + o_p(1),
\end{aligned} \tag{42}$$

where the second equality follows by Slutsky's theorem. Thus, by (41) and (42) and the central limit theorem, $\hat{\beta}_{\mathrm{IMP},2}$ and $\hat{\beta}_{G,2}$ are asymptotically equivalent.

Proof of Eqs. 25 *and* 26. Let

$$M_C(\varepsilon_i) = \frac{R_{iT}}{\pi_{iT}}\varepsilon_i - \sum_{t=1}^{T} \frac{[R_{it} - \lambda_{it} R_{i(t-1)}]}{\pi_{it}} \operatorname{Cov}_C(\varepsilon_i, \bar{\varepsilon}_{it}) \operatorname{Var}_C(\bar{\varepsilon}_{it})^{-1} \bar{\varepsilon}_{it},$$

where Cov_C and Var_C are calculated under the assumption that $\operatorname{Cov}(Y_i \mid X_i^*) = C(X_i^*)$. The generalized least squares estimator $\hat{\beta}_G$ is asymptotically equivalent to $\hat{\beta}_{\mathrm{lin}}(C)$ that solves

$$\sum_{i=1}^{n} (I \otimes X_i)\, K^{-1}(X_i)\, M_C[\varepsilon_i(\beta)] = 0, \tag{43}$$

where $X_i = (1, X_i^*)^T$, $K(X_i) = \operatorname{Var}[M_C(\varepsilon_i) \mid X_i^*]$ and $C(X_i) = \operatorname{Var}(Y_i \mid X_i^*)$. When X_i^* is a binary variable (43) is equivalent to

$$\sum_{X_i^*=0} \left(I \otimes \begin{pmatrix}1\\0\end{pmatrix}\right) K_0^{-1} M_C[\varepsilon_i^{(0)}(\beta)] + \sum_{X_i^*=1} \left(I \otimes \begin{pmatrix}1\\1\end{pmatrix}\right) K_1^{-1} M_C[\varepsilon_i^{(1)}(\beta)] = 0, \tag{44}$$

where $K_j^{-1} = K^{-1}(X_i^* = j)$, $\varepsilon_i^{(0)}(\beta)$ is the $T \times 1$ vector with the jth element equal to $(Y_{ij} - B_{0j})$ and $\varepsilon_i^{(1)}(\beta)$ is the $T \times 1$ vector with the jth element equal to $(Y_{ij} - B_{0j} - B_{1j})$. The system (44) consists of $2T$ equations. Rearranging these equations so that the equations occupying odd numbered places in (44) come first, we have

$$\sum_{X_i^*=0} IK_0^{-1} M_C[\varepsilon_i^{(0)}(\beta)] + \sum_{X_i^*=1} IK_1^{-1} M_C[\varepsilon_i^{(1)}(\beta)] = 0 \tag{45}$$

$$\sum_{X_i^*=1} IK_1^{-1} M_C[\varepsilon_i^{(1)}(\beta)] = 0. \tag{46}$$

Thus, $\hat{\beta}_{0j}$, $j = 1, ..., T$ solves

$$\sum_{X_i^*=0} M_C[\varepsilon_i^{(0)}(\beta)] = 0, \tag{47}$$

and it is therefore equal to the generalized least squares estimator of $\beta_{0,0}$ based on subjects with $X_i^* = 0$. Similarly, $\widehat{\beta_{0j} + \beta_{0j}}$, $j = 1, ..., T$, solves

$$\sum_{X_i^*=1} M_C[\varepsilon_i^{(1)}(\beta)] = 0, \tag{48}$$

which is the generalized least squares estimator of the mean vector among subjects with $X_i^* = 1$. Thus it follows that $\hat{\beta}_{1j} = \widehat{\beta_{0j} + \beta_{1j}} - \beta_{0j}$ is the difference between the generalized least squares estimator of the mean vector among subjects with $X_i^* = 1$ and the GLS estimator of the mean vector among subjects with $X_i^* = 0$. That relationships (47) and (48) hold also for the GEE, OLS, and semiparametric efficient estimators follows by an analogous argument by considering the appropriate functions $M_C[\varepsilon_i(\beta)]$ in each case.

Proof of Eq. (20). $E(Y_{i1}) = 0$ since: (1) $Y_{i1} = Z_{i1}^{7/3}$, (2) the function $h(Z) = Z^{7/3}$ is odd, and (3) Z has a symmetric distribution with zero mean. Thus, $\mathrm{Var}(Y_{i1}) = E(Y_{i1}^2) = E(Z_{i1}^{14/3})$. Also, $\mathrm{Cov}(Y_{i1}, Y_{i2}) = E(Y_{i1}, \varepsilon_{i2})$ and $E(Y_{i1}\varepsilon_{i2}) = E[Y_{i1} E(\varepsilon_{i2} \mid Y_{i1})]$. But $E(\varepsilon_{i2} \mid Y_{i1}) = E(\varepsilon_{i2} \mid Z_{i1})$ because $h(Z) = Z^{7/3}$ is a one-to-one function. Thus, $E[Y_{i1} E(\varepsilon_{i2} \mid Y_{i1})] = E(Y_{i_1} \rho Z_{i1}) = \rho E(Z_i^{10/3})$. Finally, $\mathrm{Corr}(Y_{i1}, Y_{i2}) \equiv \mathrm{Cov}(Y_{i1}, Y_{i2}) / \{\mathrm{Var}(Y_{i1})\ \mathrm{Var}(Y_{i2})\}^{1/2} = \rho E(Z_i^{10/3}) / \sqrt{E(Z_{i1}^{14/3})}$ because $\mathrm{Var}(Y_{i2}) = 1$.

ACKNOWLEDGMENT

This work was conducted as part of Christina Holcroft's doctoral dissertation.

REFERENCES

1. Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Ann. Statist.* **11** 432–452.
2. Carroll, R. J., and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Ann. Statist.* **10** 429–441.
3. Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34** 305–324.
4. Johnson, R. A., and Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*, 2nd ed. Prentice–Hall, Englewood Cliffs, NJ.
5. Liang, K-Y, and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73** 13–22.
6. Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.
7. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.
8. Robins, J. M., Mark, S. D., and Newey, W. K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics* **48** 479–495.
9. Robins, J. M., and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Am. Statist. Assoc.* **90** 122–129.
10. Rotnitzky, A., and Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82** 805–820.
11. Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.
12. Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York.