University of Michigan School of Public Health

An Extended General Location Model for Causal Inference from Data Subject to Noncompliance and Missing Values

Yahong Peng[*]   Rod Little[†]

Trivellore E. Raghuanthan[‡]

[*]Merck Laboratories, yahong_peng@merck.com

[†]University of Michigan, rlittle@umich.edu

[‡]University of Michigan, teraghu@umich.edu

An Extended General Location Model for Causal Inference from Data Subject to Noncompliance and Missing Values

Yahong Peng, Rod Little, and Trivellore E. Raghuanthan

Abstract

Noncompliance is a common problem in experiments involving randomized assignment of treatments, and standard analyses based on intention-to treat or treatment received have limitations. An attractive alternative is to estimate the Complier-Average Causal Effect (CACE), which is the average treatment effect for the sub-population of subjects who would comply under either treatment (Angrist, Imbens and Rubin, 1996, henceforth AIR). We propose an Extended General Location Model to estimate the CACE from data with non-compliance and missing data in the outcome and in baseline covariates. Models for both continuous and categorical outcomes and ignorable and latent ignorable (Frangakis and Rubin, 1999) missing data mechanisms are developed. Inferences for the models are based on the EM algorithm and Bayesian MCMC methods. We present results from simulations that investigate sensitivity to model assumptions and the influence of missing-data mechanism. We also apply the method to the data from a job search intervention for the unemployed workers.

An Extended General Location Model for Causal Inferences from Data Subject to Noncompliance and Missing Values

Yahong Peng[1], Roderick J. A. Little[2], Trivellore E. Raghunathan[2]

[1]Merck Research Laboratories, BL 3-2, West Point, PA 19486

E-mail: *yahong_peng@merck.com*

[2]Biostatistics Department, SPH, University of Michigan, Ann Arbor, MI 48105

SUMMARY. Noncompliance is a common problem in experiments involving randomized assignment of treatments, and standard analyses based on intention-to treat or treatment received have limitations. An attractive alternative is to estimate the Complier-Average Causal Effect (CACE), which is the average treatment effect for the subpopulation of subjects who would comply under either treatment (Angrist, Imbens and Rubin, 1996, henceforth AIR). We propose an Extended General Location Model to estimate the CACE from data with non-compliance and missing data in the outcome and in baseline covariates. Models for both continuous and categorical outcomes and ignorable and latent ignorable (Frangakis and Rubin, 1999) missing data mechanisms are developed. Inferences for the models are based on the EM algorithm and Bayesian MCMC methods. We present results from simulations that investigate sensitivity to model assumptions and the influence of missing-data mechanism. We also apply the method to the data from a job search intervention for the unemployed workers.

KEY WORDS: Causal inference; EM algorithm; General location model; Missing data; Non-compliance.

1   Introduction

Inferences for treatment effects are relatively straightforward in randomized trials with perfect compliance with the assigned treatments. However, for scientific experiments involving human participants, non-compliance and partial compliance are very common in practice. Compliance often varies according participant characteristics, and may be associated with

effects of the treatments. For example, in a clinical trial to test the efficacy of a new drug, severe side effects may result in lack of compliance. Misleading conclusions are likely to result if compliance information is ignored in such settings. Missing values in the outcome or the covariates further complicate the data analysis. This article concerns methods for analyzing data from a clinical trial with all-or-nothing compliance and missing values in the outcomes and covariates.

Intent-to-treat and as-treated methods provide simple estimates of treatment effects and are widely applied in practice, but both these methods have limitations. The intent-to-treat analysis compares the outcome distributions of the treatments as randomized, ignoring compliance information. It provides a valid estimate of the effect of the treatment assignment, but a potentially biased estimate of the effect of the treatment itself, which is often of more interest. The as-treated analysis compares the outcome distributions between the treatments actually received. It focuses more directly on the treatment effect itself, but is subject to selection bias when there is noncompliance, since the randomization is violated. An attractive alternative approach proposed by AIR is to estimate the Complier-Average Causal Effect (CACE), the average causal effect for the sub-population of subjects who would comply with either of the treatments if assigned to them. We call compliance under both treatments "principal compliance" to distinguish it from "observed compliance", since it is a form of principal stratification (Frangakis and Rubin, 2000) for correct causal inference with post-treatment variables. Unlike observed compliance, principal compliance is incompletely observed since compliance under the treatment not assigned is unknown.

Analysis methods for the CACE have received considerable attention recently (Imbens and Rubin, 1996, 1997; Baker, 1998; Frangakis and Rubin,1999, 2000; Yau and Little, 2001; Robins, 1994; Robins, 1999). Building in this work, we propose an extended general location (EGL) model for the CACE for data with non-compliance and missing values. Our model extends methods in Little and Yau (1998) to accommodate discrete as well as continuous outcomes, missing values in baseline covariates as well as in outcomes, and two alternative assumptions about the compliance and missing-data mechanisms.

We motivate and illustrate our model using data from the JOBS II (Vinokur, Price, &

Schul, 1995), a randomized trial at the University of Michigan to test a job search intervention for unemployed workers. In this trial unemployed individuals were randomly assigned to a control treatment consisting of a booklet describing job-search methods and tips, or an experimental treatment group consisting of the same booklet and five 4-hour job search seminars. Follow-up questionnaires were mailed periodically to all the participants, and measures of outcomes such as depression, financial strain and re-employment were obtained. We focus here on a binary outcome, re-employment at a fixed date after the intervention, but we also consider models for a continuous outcome. The JOBS II data have missing values, and only about one half of the individuals assigned to the experimental treatment complied, in the sense that they attended the seminars. Since "compliance" with the control treatment is not an issue here, it is assumed that individuals would comply with the control treatment, but may or not comply with the experimental treatment. Thus compliance (in the CACE sense) is known for participants assigned to the experimental group, but is not known for participants assigned to the control group, since their participation if assigned the intervention is not observed. Our methods for estimating the CACE are illustrated for this behavioral intervention, but are also directly applicable to trials involving medical interventions with all-or-nothing compliance.

Section 2 presents our EGL model for inferences about the CACE, and inferences based on the EM algorithm and Bayesian MCMC methods. An alternative approach based on instrumental variables is given in Section 2.4. Section 3 presents simulation studies that compare our likelihood-based approach with instrumental variable methods for causal inference (Goldberg, 1972; Bloom, 1984; Heckman and Hotz, 1989). In Section 4 we apply our methods to JOBS II data. Section 5 contains conclusions and discussion.

2 Extended General Location Model

2.1 The Complete-Data Model

We define the following variables for each participant i in a randomized clinical trial:

$x_i = (x_{i1}, \cdots, x_{ip})^T = p$ continuous baseline covariates that are conditioned in the analysis. While these might treated as fixed in a complete-data analysis, we assign them a normal distribution to allow cases with missing values in these variables to be included in the analysis. Transformations may be needed to improve the normality assumption.

b_i = discrete baseline covariates. For simplicity, we assume b_i is scalar and binary, although the extension to more than two levels is immediate.

r_i = randomization indicator for the treatment assigned. In the JOBS II application, $r_i = 1$ for the intervention, $r_i = 0$ for the control treatment

y_i = an outcome variable, which we initially assume is continuous and normal. Extensions to vector outcomes are outlined in Section 6.

c_i = principal compliance indicator, taking the value 1 if participant i would comply with either treatment, and 0 otherwise. As noted above, in the JOBS II setting c_i is observed for participants assigned to the experimental group, taking the value 1 for individuals who attended the seminars and 0 otherwise. There was a small amount of partial compliance in JOBS II in that a few individuals attended a subset of the seminars but not all of them; for the purposes of our analysis these participants are treated as compliers. For individuals assigned to the control group, c_i is missing since we do not know if they would have attended the seminars if they had been assigned to the experimental group.

We write $[A|B]$ to denote the distribution of A given B, N for a univariate normal and N_p for a p-variate normal distribution. We assume observations are independently distributed given the randomization indicator r_i, and factor $[y_i, x_i, c_i, b_i|r_i]$ as $[y_i, x_i, c_i, b_i|r_i] = [c_i, b_i|r_i] \cdot [x_i|c_i, b_i, r_i] \cdot [y_i|x_i, c_i, b_i, r_i]$. We then model the components of this factorization as follows:

$$[c_i, b_i|r_i] \quad \sim \quad Multinomial(\alpha), \tag{1}$$

$$[x_i|c_i = c, b_i = b, r_i = r] \quad \sim \quad N_p(\mu_{cb}, \Sigma), \tag{2}$$

$$[y_i|x_i, c_i, b_i, r_i] \quad \sim \quad N(g(x_i, c_i, b_i, r_i; \beta), \tau^2), \tag{3}$$

where $g(x_i, c_i, b_i, r_i; \beta) = \beta_0 + \beta_C c_i + \beta_{CR} c_i r_i + \beta_B b_i + \beta_X x_i + \beta_{BC} b_i c_i +$

$$\beta_{XC} x_i c_i + \beta_{BCR} b_i c_i r_i + \beta_{XCR} x_i c_i r_i. \tag{4}$$

4

The parameters of the model are thus: $\theta = (\alpha, \{\mu_{cb}\}, \Sigma, \beta, \tau^2)$, where μ_{cb} is a $(p \times 1)$ vector, $\Sigma = (\sigma_{ij})$ is a $(p \times p)$ matrix, and $\beta^T = (\beta_0, \beta_C, \beta_{CR}, \beta_B, \beta_X, \beta_{BC}, \beta_{XC}, \beta_{BCR}, \beta_{XCR})$. The parameterization of this model embodies the following important assumptions, described in more detail in AIR:

(1) Parameters $(\alpha, \mu_{cb}, \Sigma)$ do not depend on r_i because of randomization of treatments ensures that r_i is independent of c_i, b_i, x_i.

(2) The mean structure in (4) omits the effects of r_i, $b_i r_i$ and $x_i r_i$, because we assume that for noncompliers ($c_i = 0$) the effect of treatment is the same whether subject i is randomized to control (and has no access to the treatment) or randomized to the treatment (and fails to comply). This is the crucial exclusion restriction (ER) assumption discussed in AIR. Eq. (4) implies that the mean of y_i for the non-compliers with baseline covariates ($b_i = b, x_i = x$) is $g(x, 0, b, r; \beta) = \beta_0 + \beta_B b + \beta_X x$, and the CACE for compliers with covariates ($b_i = b, x_i = x$) is then $CACE(b, x) = g(x, 1, b, 1; \beta) - g(x, 1, b, 0; \beta) = \beta_{CR} + \beta_{BCR} b + \beta_{XCR} x$. An overall CACE can be defined by averaging over the distribution $p(b, x | c = 1)$ of B and X for compliers:

$$CACE = \int_{b,x} CACE(b, x) p(b, x | c = 1) dx db. \tag{5}$$

(3) The independence assumption of the model implictly requires the stable unit treatment value assumption (SUTVA) that potential outcomes for each unit are unrelated to the treatments assigned and the treatments actually received by the other units (AIR,). While this assumption is almost universally invoked, it is somewhat suspect in the JOBS II setting, since the experimental treatment takes place in a group setting.

We call this an extended general location (EGL) model because the distribution of the continuous variables $[x_i, y_i | c_i, b_i, r_i]$ is normal with a covariance matrix that depends on the discrete variables (c_i, b_i, r_i), rather than being a constant as in the standard general location model (Olkin and Tate, 1961, Little and Rubin 2002, Section 14.2). As in that model, restrictions can be imposed on the EGL model, either by assuming a loglinear model for $[c_i, b_i | r_i]$, or by contraining the means μ_{cd} of the normal model for $[x_i | c_i, b_i, r_i]$. Restrictions are particularly helpful in reducing the number of parameters when the sample size is small. The distinctness of the parameters of the models for $[c_i, b_i | r_i]$ and for $[x_i | c_i, b_i, r_i]$ is preserved,

5

so estimation procedures are not radically different than for unrestricted models. With appropriate restrictions on the model parameters $(\alpha, \mu_{cb}, \Sigma)$ of $[x_i, c_i, b_i | r_i]$, the EGL model yields $[c_i | x_i, b_i, r_i]$ similar to those in Little and Yau (1998) after re-parameterization (Peng, 2001).

Eq. (3) of the EGL model can also be adapted to a binary outcome, say y_i^*. We assume there is an underlying latent continuous variable y_i for y_i^*, where $y_i^* = I(y_i > 0)$. The distribution $[y_i | x_i, c_i, b_i, r_i]$ is given by equation (3) with $\tau^2 = 1$. This implies a probit model for y_i^*, namely,

$$\Pr(y_i^* = 1 | x_i, c_i, b_i, r_i) = \Phi(g(x_i, c_i, b_i, r_i; \beta)), \tag{6}$$

where $\Phi(\cdot)$ is the standard normal cumulative probability function. When the outcome of interest is binary, $\Phi(g(x, 1, b, 1; \beta)) - \Phi(g(x, 1, b, 0; \beta))$ is the CACE for subjects with covariates $(b_i = b, x_i = x)$, and the overall CACE can be calculated as in Eq. (5). The probit model (6) for $[y_i^* | x_i, c_i, b_i, r_i]$ is proposed rather than the more common logit model because it is more convenient for computation, as discussed in Section 2.3.

2.2 Missing Data Mechanisms

We now suppose there are missing values of (x_i, b_i, y_i) in the data set, in addition to the missing values of c_i in the control group. The validity of methods for handling the missing data depends on assumptions about mechanisms that create missing values. For subject i, let $w_i = (x_i, b_i, y_i)$ denote the set of values if they were fully observed, and write $w_{obs,i}$ for the observed values of w_i and $w_{mis,i}$ for the missing values. Also, let m_i be the vector of missing-data indicators for w_i, with entries 1 if the corresponding entries of w_i are missing and 0 otherwise. The missing-data mechanism for w_i is specified via a distribution $\Pr(m_i | w_i, c_i, r_i; \phi)$ for m_i given w_i and compliance and randomization indicators (c_i, r_i), indexed by unknown parameters ϕ. We consider two alternative assumptions about the missing-data mechanism:

Ignorable missing-data mechanism (IMD). The missing data mechanism of w_i is said to be ignorable if the data are missing at random (MAR, Rubin, 1976; Little and Rubin, 2002), that is: $\Pr(m_i | w_i, c_i, r_i = 1; \phi) = \Pr(m_i | w_{obs,i}, c_i, r_i = 1; \phi)$, for all $w_{mis,i}$, $\Pr(m_i | w_i, c_i, r_i = 0; \phi) = \Pr(m_i | w_{obs,i}, r_i = 0; \phi)$, for all $w_{mis,i}, c_i$, and the parameters ϕ are

distinct from the parameters θ of the complete-data model. ML and Bayes estimation under this assumption does not require a term for the missing data mechanism to be included in the likelihood.

Note that the mechanism is allowed to depend on c_i for experimental subjects ($r_i = 1$) since c_i is observed for these subjects, but is not allowed to depend on c_i for control subjects ($r_i = 0$) since c_i is missing for these subjects. The latter may not be realistic, since the probability of non-response for noncompliers may be systematically higher than that for compliers, regardless of treatment. To handle such cases, Frangakis and Rubin (1999) propose an instrumental variable estimator for an alternative missing data mechanism, which they call latent ignorability:

Latent ignorable missing-data mechanism (LIMD). The missing data mechanism of w_i is said to be latent ignorable if the data are latent MAR (LMAR) within each level of the latent compliance status, that is: $\Pr(m_i|w_i, c_i, r_i = 1; \phi) = \Pr(m_i|w_{obs,i}, c_i, r_i = 1; \phi)$, for all $w_{mis,i}$, $\Pr(m_i|w_i, c_i, r_i = 0; \phi) = \Pr(m_i|w_{obs,i}, c_i, r_i = 0; \phi)$, for all $w_{mis,i}$, and the parameters ϕ are distinct from the parameters of the complete-data model. See also Barnard et al. (2002). Note that LMAR is weaker than MAR since the mechanism is allowed to depend on c_i for control as well as experimental subjects. This missing data mechanism is in general nonignorable, because c_i is unknown in the control group.

As LIMD is weaker than IMD, a stronger exclusion restriction assumption is needed to identify the parameters when the complete-data model is saturated, and inference may be vulnerable to missspecification when the complete-data model is unsaturated. Hence, in our LIMD models we replace the ER assumption above by the following stronger assumption:

Compound exclusion restriction (CER). The joint distribution (y_i, m_i) for a non-complier randomized to the control group is the same as the distribution (y_i, m_i) for a non-complier randomized to the treatment group.

Our LIMD model involves a particular form of LIMD where missingness of y_i is allowed to depend on missing values of c_i, but missingness of (x_i, b_i) is not. Accordingly, we write $m_i = (m_i^Y, m_i^{XB})$, where m_i^Y is the missing-data indicator for y_i and m_i^{XB} is the vector of missing-data indicators for x_i and b_i, and assume for all $w_{mis,i}, c_i$, $\Pr(m_i^Y, m_i^{XB}|w_i, c_i, r_i =$

7

$0; \phi) = \Pr(m_i^Y | w_{obs,i}, c_i, r_i = 0; \phi_1) \Pr(m_i^{XB} | m_i^Y, w_{obs,i}, r_i = 0, \phi_2)$, where $\phi = (\phi_1, \phi_2)$ are distinct from θ. By a straightforward extension of Rubin's (1976) theory, only the component m_i^Y of m_i distribution needs to be modelled in this case. We then factor $[y_i, x_i, c_i, b_i, m_i^Y | r_i] = [c_i, b_i | r_i] \cdot [x_i | c_i, b_i, r_i] \cdot [m_i^Y | x_i, c_i, b_i, r_i] \cdot [y_i | m_i^Y, x_i, c_i, b_i, r_i]$, where (i) $[y_i | m_i^Y, x_i, c_i, b_i, r_i] = [y_i | x_i, c_i, b_i, r_i]$ because of the latent ignorability assumption; (ii) $[c_i, b_i | r_i]$, $[x_i | c_i, b_i, r_i]$ and $[y_i | x_i, c_i, b_i, r_i]$ can be modelled as in Eqs. (1) to (3); and (iii)$[m_i^Y | x_i = x, c_i = c, b_i = b, r_i = r] \sim Bernoulli(\phi_{xcbr})$. To model the latter distribution, we assume there is an underlying continuous variable x_i^M for m_i^Y, such that $m_i^Y = I(x_i^M > 0)$. We then assume that, given (x_i, c_i, b_i, r_i), x_i^M follows a normal distribution with mean $g(x_i, c_i, b_i, r_i; \nu)$ and variance 1, where $g(\cdot)$ is defined in (4) with regression coefficients β replaced by ν. This implies a probit model for $[m_i^Y | x_i, c_i, b_i, r_i]$. The CER assumption implies that the mean structure of x_i^M does not include the effects of $r_i, x_i r_i, b_i r_i$, that is: $E(x_i^M | x_i, c_i = 0, b_i, r_i = 0) = E(x_i^M | x_i, c_i = 0, b_i, r_i = 1) = \nu_0 + \nu_D b_i + \nu_X x_i$. This is also an extended general location model for $[y_i, x_i, x_i^M, c_i, b_i | r_i]$. The extension for binary y_i^* is carried out in the same way as before.

In next section, we consider likelihood-based estimates for the CACE in these settings.

2.3 Maximum Likelihood and Bayes Estimation

Under IMD and ER, maximum likelihood (ML) estimates for the parameters of our EGL model can be obtained via the EM algorithm (Dempster, Laird and Rubin, 1977), and is similar to EM for the general location model (Little and Rubin, 2002). We outline the EM algorithm for the unrestricted extended general location model of Section 3.1. (Details of the algorithm and codes in Splus can be found at Little's web page). Let $d_i = (d_{i1}, d_{i2}, d_{i3}, d_{i4})$ be the vector of indicators for the discrete variables (c_i, b_i), with $d_{i1} = I(c_i = b_i = 0)$, $d_{i2} = I(c_i = 1, b_i = 0)$, $d_{i3} = I(c_i = 0, b_i = 1)$, $d_{i4} = I(c_i = b_i = 1)$, with corresponding probabilities $\alpha = (\alpha_1, \alpha_2, \alpha_3, a_4)$. The complete-data likelihood can be written as

$$L_{com}(\theta) \propto \prod_i [\{\prod_{j=1}^{4} (\alpha_j^{d_{ij}})\} \cdot |\Sigma|^{-\frac{p}{2}} exp\{-\frac{1}{2}(x_i - \mu_{c_i b_i})^T \Sigma^{-1} (x_i - \mu_{c_i b_i})\} \cdot$$
$$\frac{1}{\tau} exp\{-\frac{1}{2\tau^2}(y_i - g(x_i, c_i, b_i, r_i; \beta))^2\}],$$

8

The logarithm of this likelihood is linear in complete-data sufficient statistics $\sum_i r_i$, $\sum_i d_{ij}$, $\sum_i d_{ij}x_i$, $\sum_i d_{ij}y_i$, $\sum_i d_{ij}x_i x_i^T$, $\sum_i d_{ij}x_i y_i$, $\sum_i d_{ij}y_i^2$, $\sum_i r_i d_{ij}$, $\sum_i r_i d_{ij}x_i$, $\sum_i r_i d_{ij}y_i$, $\sum_i r_i d_{ij}x_i x_i^T$, $\sum_i r_i d_{ij}x_i y_i$, and $\sum_i r_i d_{ij}y_i^2$, for $j = 1, ...4$. The maximization step (M-step) of EM calculates the parameter estimates that maximize the complete-data likelihood given the sufficient statistics calculated from the current E-step, and is straightforward. Specifically, α_j is estimated by the expected count in the corresponding cell formed by the discrete variables, μ_{cb} by the sample mean of X with subjects with $(c_i = c, b_i = b)$, Σ by the pooled within-cell covariance matrix of X, β and τ^2 by the regression y_i on the components of (x_i, c_i, b_i, r_i) included in the model.

The expectation step (E-step) of EM calculates the conditional expectations of the complete-data sufficient statistics, given the observed data and the current estimates of the parameters from the M-step. It is easier to calculate the conditional expectations for the sufficient statistics $\sum_i d_i$ first, and then calculate the conditional expectations of the sufficient statistics that involve continuous variables (x_i, y_i) and the interaction terms between the discrete and continuous variables. Calculations of the conditional expectations of the sufficient statistics that contain continuous variables (x_i, y_i) involve the parameters of conditional distributions of the multivariate normal. They are easily computed by the SWEEP operator (e.g. Little and Rubin, 2002). The E-step can be carried out by missing-data pattern to avoid unnecessary SWEEP computations.

The initial values for EM cannot be computed using data from the set of complete cases, because the values of c_i are completely unknown in the control group. To avoid this, we first estimate α by using only the data from the treatment group subjects, and then impute values of c_i for the control group subjects by the estimated α. We then calculate the sufficient statistics, ignoring those cases with missing values in (x_i, b_i, y_i), and do an M-step based on the imputed sufficient statistics to obtain initial estimates for β's. The method of moments can also be used to provide initial estimates after imputing the compliance status for control group subjects.

The observed data likelihood is readily computed from by-products of the E-step with each iteration, providing a check of the computations. In addition, we can use the observed

data likelihood to perform likelihood ratio tests for choosing between models.

For Bayesian inferences, we assume α, (μ, Σ) and (β, τ^2) are a priori independent with conventional non-informative priors, $p(\alpha) \propto \prod_{j=1}^{4} \alpha_j^{-0.5}$, $p(\mu, \Sigma) \propto |\Sigma|^{-\frac{p+1}{2}}$ and $p(\beta, \tau^2) \propto \frac{1}{\tau^2}$. These distributions yield a complete-data posterior distribution that is Dirichlet for α, inverse-Wishart for Σ and multivariate normal for μ given Σ, scaled inv-χ^2 for τ^2 and multivariate normal distribution for β given τ^2: $[\alpha|Data] \propto \prod_{j=1}^{4} \alpha_j^{\sum_i d_{ij}-0.5}$, $(\Sigma|Data, \alpha) \sim W^{-1}(N-4, N\widehat{\Sigma}^{-1})$, $(\mu_{cb}|Data, \alpha, \Sigma) \sim N(\widehat{\mu_{cb}}, (\Sigma/\sum_i I(c_i = c, b_i = b)))$, $(\tau^2|Data, \alpha, \mu, \Sigma) \sim \widehat{\tau}^2/\chi_{N-k}^2$, $(\beta|Data, \alpha, \Sigma, \mu, \tau^2) \sim N(\widehat{\beta}, \tau^2(Z^T Z)^{-1})$, where Z is the design matrix of the regression of y_i on x_i, c_i, b_i, r_i, $(\widehat{\mu}, \widehat{\Sigma}, \widehat{\beta}, \widehat{\tau}^2)$ are the ML estimates for model parameters $(\mu, \Sigma, \beta, \tau^2)$ based upon the complete data likelihood. The Gibbs sampler algorithm iterates between the P-step and the I-step. The tth P-step of the Gibbs sampler algorithm simulates random draws, $(\alpha^{(t)}, \mu^{(t)}, \Sigma^{(t)}, \beta^{(t)}, \tau^{2(t)})$, for the model parameters based on the complete-data posterior distribution with missing data filled in from the previous I-Step. The I-step is very similar to the E-step of EM. Instead of calculating the conditional expectations of missing values in terms of the sufficient statistics, random draws for the missing values in the data set are generated from their predictive distribution, using current draws of the parameters from the previous P-step. When the observable outcome of interest is discrete, $\tau^2 = 1$, and drawing values of y_i involves simulating random numbers from truncated normal distributions, since we only observe the sign of y_i.

We used the method of Gelman and Rubin (1992) to assess convergence of the Gibbs sequences. To obtain reasonably over-dispersed random numbers for the starting points for the Gibbs draws, m bootstrap samples were generated from the original data set with a sample size of half the size of the original data set. For each of the bootstrap samples, ML estimates are obtained via the EM algorithm and used as the starting values for the Gibbs algorithm.

Quite straightforward modifications of these algorithms yield ML and Bayes' estimates for the LIMD/CER model of the previous section. For the M-step of EM, everything remains the same except that we need an additional step for estimating ν. Because the regression models involving ν and β have the same structure, the M-step for ν is the same as

10

that for β, with x_i^M replacing y_i in expressions involving these terms. For the E-step, conditional expectations also need to be computed for complete-data sufficient statistics associated with the model for x_i^M, namely: $\{\sum_i x_i^M, \sum_i d_{ij} x_i^M, \sum_i x_i^M x_i, \sum_i d_{ij} x_i^M x_i, \sum_i r_i x_i^M, \sum_i r_i d_{ij} x_i^M, \sum_i r_i x_i^M x_i, \sum_i r_i d_{ij} x_i^M x_i\}$, $j = 1, \cdots 4$.

For Bayesian inference we assume a flat prior for ν, $p(\nu) \propto constant$, and the priors for the other parameters in the IMD model. The P-step is as before except that we need additional draws for ν from its posterior distribution $N(\hat{\nu}, (Z'Z)^{-1})$, where $\hat{\nu}$ is the ML estimate of ν obtained via the EM algorithm based on the imputed data set. For the I-step, we draw the missing values in the discrete variables (c_i, b_i) first, then draw the missing continuous variables in (x_i, x_i^M, y_i) based on the corresponding predictive distribution given the current draws for the missing discrete variable and the observed data.

2.4 Bloom's method and Extensions

We conducted simulation studies to compare the model-based methods in Section 3 with Bloom's (1984) instrumental variable (IV) approach for continuous outcomes with non-compliance, and extensions discussed below.

With complete data and no covariates, the standard IV estimator of the CACE is $\widehat{\delta_{IV}} = (\overline{y}_1 - \overline{y}_0)/\widehat{\pi}_c$, where $\overline{y}_j$ is the mean outcomes of all subjects with $r_i = j$, and $\widehat{\pi}_c$ is the proportion of compliers in the experimental group, $r_i = 1$. When baseline covariates x_i, b_i are available, Bloom's (1984) adjusted estimate is $\widehat{\delta_{IVn}} = \widehat{\beta_P} + \frac{1-\widehat{\pi}_c}{\widehat{\pi}_c} \cdot \widehat{\beta_{NS}}$, where $\widehat{\beta_P}$ and $\widehat{\beta_{NS}}$ are the least square estimates from the regression model $y_i = \beta_0 + \beta_P(r_i c_i) + \beta_{NS} r_i(1 - c_i) + \beta_B b_i + \beta_X x_i + \epsilon$. The estimator reduces to the standard IV estimator in the absence of baseline covariates. We call $\widehat{\delta_{IVn}}$ the naive instrumental variable (IVn) estimator, as it implicitly makes the strong assumption that the compliance rate π_c is constant for different covariates x, i.e., the compliance is independent of x. A better estimator can be obtained by first estimating the CACE $\delta(x_i, b_i)$ for compliers with covariates (x_i, b_i) as, $\widehat{\delta(x_i, b_i)} = \widehat{\beta_P} + \frac{1-\widehat{\pi}_{ci}}{\widehat{\pi}_{ci}} \cdot \widehat{\beta_{NS}}$, where $\widehat{\pi}_{ci}$ is the estimated compliance rate for subjects with covariates (x_i, b_i), and then estimating the overall causal effect δ as follows, $\widehat{\delta_{IV}} = \int_{X,B} \widehat{\delta(x, b)} p(x, b) dx db$. The method IV in the simulation is this estimator applied to the complete cases. We also considered two

11

IV estimators for the situation where the missing data mechanism for y_i is latent ignorable. The naive latent IV estimate is $\widehat{\delta_{IVLn}} = \widehat{\beta_P} + \frac{(1-\widehat{\pi_c})(1-\widehat{\phi_{01}})}{(1-\widehat{\phi_0})-(1-\widehat{\pi_c})(1-\widehat{\phi_{01}})} \cdot \widehat{\beta_{NS}}$, where $\widehat{\phi_{01}}$ and $\widehat{\phi_0}$ are the nonresponse rates in the treatment noncomplier and control groups, respectively. With a homogenious compliance rate, this is a consistent estimate of the CACE given by Frangakis and Rubin (1999) for LIMD. The IVL estimator with heterogenous compliance rate $\widehat{\delta_{IVL}}$ is obtained by replacing $\widehat{\pi_c}$ in this expression by a predicted compliance rate $\widehat{\pi_{ci}}$ that depends on the covariates (x_i, b_i), and then integrating the resulting CACE estimate over the covariates as before.

3   Simulation Study

3.1   Description of the Study

In the simulation, values of r_i, c_i, b_i and y_i were generated as follows, $[r_i] \sim Bernoulli(p_R)$, $[c_i, b_i | r_i] \sim Multinomial(\alpha_{CB})$, $[y_i | b_i, c_i, r_i] \sim F(\cdot) + 0.5 - 0.5c_i - c_i r_i - 0.5b_i + c_i b_i$. To resemble the JOBS II data, we choose $(\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11}) = (0.3, 0.35, 0.15, 0.20)$ and $p_R = 0.65$. To test the sensitivity of the normality assumption of y_i, three distributions of $F(\cdot)$ are chosen: $N(0,1)$, $T(df = 4)$ and $\Gamma(shape = 5, scale = 0.2)$. The random numbers from $F(\cdot)$ are re-centered and re-scaled so that they have mean 0 and variance 1.

Let m_i^B or m_i^Y be the missing data indicators for b_i or y_i respectively. Four missing data mechanisms are considered: two of them (MAR1 and MAR2) are ignorable, while the other two (LMAR and NMAR) are non-ignorable. Specifically,

(1) Missing at random (MAR1), where the probability of missingness depends on observed values in the data set, and the missing data mechanism satisfies the latent ignorability conditions and the compound exclusion restriction in Frangakis and Rubin (1999): $\Pr(m_i^Y = 1 | C, R, X, Y) = \Pr(m_i^B = 1 | C, R, X, Y) = \Phi(\alpha_0 + \alpha_{CR}CR)$, where $\Phi(\cdot)$ is the cumulative standard normal probability function.

(2) Missing at random (MAR2), where the probability of missingness only depends on the observed values in the data set, and the missing data mechanism satisfies the latent

ignorability conditions but not the compound exclusion restriction in Frangakis and Rubin (1999): $\Pr(m_i^Y = 1|c_i, r_i, b_i, y_i) = \Pr(m_i^B = 1|c_i, r_i, b_i, y_i) = \Phi(\alpha_0 + \alpha_R r_i + \alpha_{CR} c_i r_i)$. Hence $\Pr(m_i^Y = 1|c_i = 0, r_i = 0, b_i, y_i) = \Phi(\alpha_0) \neq \Pr(m_i^Y = 1|c_i = 0, r_i = 1, b_i, y_i) = \Phi(\alpha_0 + \alpha_R r_i)$.

(3) Latent missing at random (LMAR) (Frangakis and Rubin, 1999), where the probability that y_i is missing depends upon the compliance status c_i which is unobservable for the control group subjects. In addition, the conditional distribution $(y_i, m_i|c_i, r_i)$ for never-takers satisfies the CER assumption. The probability that b_i is missing depends only on r_i: $[y_i, m_i^Y|c_i = 0, r_i = 0, b_i] = [y_i, m_i^Y|c_i = 0, r_i = 1, b_i], \Pr(m_i^Y = 1|c_i, r_i, b_i, y_i) = \Phi(\alpha_0 + \alpha_C c_i + \alpha_{CR} c_i r_i)$, and $\Pr(m_i^B = 1|c_i, r_i, b_i, y_i) = \Phi(\alpha_0 + \alpha_R r_i)$.

(4) Not missing at random (NMAR), where the probability that y_i is missing depends upon values of baseline covariate b_i, which contain missing values: $\Pr(m_i^Y = 1|c_i, r_i, b_i, y_i) = \Pr(m_i^B = 1|c_i, r_i, b_i, y_i) = \Phi(\alpha_0 + \alpha_X b_i)$.

For each generated data set, the following methods are applied for causal inferences: IVn and IV are the naive IV and the IV estimators defined section 4.1, and $IVnL$ and IVL are the corresponding extended IV estimators assuming latent ignorable mechanisms. For likelihood-based methods, estimates of the CACE as well as other model parameters are obtained based on extended general location models. Specifically, MLBD (for maximum likelihood before deletion) is the ML estimate of the CACE of the data before generating missing values in b_i and y_i; MLI and BYSI are the ML and Bayesian estimates of the CACE assuming ignorable mechanisms; MLLI and BYSLI are the ML and Bayesian estimates of the CACE assuming latent ignorable mechanisms; and MLCC and BYSCC are the ML and Bayesian (posterior mean) estimates of the CACE of the data after deleting observations with missing values in b_i or y_i.

3.2   Results

Simulation results of the CACE parameter are summarized in Table 1 and Table 2. For each simulation condition, biases and root mean square errors (RMSE) of the estimates of the CACE from each method are reported, after multiplied by 1000. In addition, confidence

intervals for each parameter are constructed from normal approximations based on estimates for that parameter, specifically for the ith parameter, the $100P\%$ confidence interval is constructed as $EST_i \pm Z_P \cdot SE_i$, for $P = 0.5, 0.6, 0.8, 0.9$, EST_i and SE_i are the sample mean and standard error of estimates for the ith parameter, and Z_P is the $(100(1-P)/2)$th quantile of the standard normal distribution. Standard errors are calculated as standard deviations from the Gibbs draws for likelihood-based methods and as standard deviations from 200 bootstrap sample estimates for instrumental variable methods. Coverage probabilities of the confidence intervals of parameters are then obtained and compared.

As expected, estimates from the MLBD are the best, with the smallest RMSE, and estimates from IV_n and IV_nL are the worst with the largest biases and RMSEs. Naive IV methods assume a constant compliance rate for different values X and this is not consistent with the data generating process. Both the IV method and the likelihood-based methods are robust for the CACE estimates based on this simulation results. We observe no evidence that the IV methods are more robust to model misspecificaiton than the likelihood-based methods. The RMSEs from IV are slightly larger than those from the likelihood-based methods. Probability coverages are good and consistent with the corresponding point estimates in general.

For the likelihood-based methods, ML estimates agree closely with the Bayesian estimates. RMSRs from the MLCC and BYSCC estimates are larger than the RMSEs from MLI and BYSI estimates. Also, estimates from models with the correct mechanisms perform better than those with incorrect mechanisms. Specifically, they are estimates from MLI, BYSI, MLLI and BYSLI under MAR1, from MLI and BYSI under MAR2, and from MLLI and BYSLI under LMAR. When the missing data mechanism is NMAR, i.e., missingness depends upon X, estimates from MLCC and BYSCC show no significant biases for $N(0,1)$ and $\Gamma(5,0.2)$ cases.

For the $N(0,1)$ case, estimates from MLLI, BYSLI, BYSLI and IVL, which assume latent ignorable mechanisms, have larger biases than those from MLI, BYSI and IVI when the latent ignorability conditions are not satisfied. Estimates from MLLI, BYSLI and IVL are significantly biased when the true underlying missing data mechanism is NMAR.

4 Application

As described in the introduction, the JOBS II (Vinokur, Price, & Schul, 1995) intervention project developed at the University of Michigan was to test a preventive intervention for unemployed workers. In the JOBS (JOB Search) II intervention setting, respondents were recruited from four offices of the Michigan Employment Security Commission in the south-eastern Michigan. Based on the screening questionnaire (T_0) data, only people who met the screening criteria were invited, randomized to the field study and mailed pretest (T_1) questionnaires.

A total of 1801 people returned their T_1 questionnaires and were enrolled in the study. Each subject in the control group was mailed a booklet describing job-search methods and tips, and each subject in the experimental treatment group was mailed the same booklet and an invitation to participate in five 4-hour job search seminars. Follow-up questionnaires were mailed to all the respondents two months (T_2) and six months (T_3) after the week of the job-search seminars. Measures of depression, financial strain, assertiveness, risk score, distress symptom, role and emotional functioning, job search self-efficacy, self-esteem, internal control orientation, mastery, intervention process, re-employment as well as demographics variables were obtained through the questionnaires. The re-employment outcome at T_3 is the outcome of interest here, and it is binary.

Previous analyses (Vinokur and Price, 1995; Little and Yau, 1998) demonstrate that the intervention primarily benefited the re-employment and the mental health outcomes of respondents who are at high risk of experiencing depression in the future. The analysis focuses on data from the high risk group, which includes 715 subjects. Because people who were randomized to the control group are not allowed to attend the seminars, there are only compliers and never-takers in the study.

Table 3 presents the summary statistics of the covariates measured at baseline for the high risk group subjects, by the randomly assigned treatment R and the compliance C (for treatment group only, as C is unknown for control group subjects). Comparison of the first two columns shows that the randomization did a fair job in balancing the covariate

distributions across the two treatment groups.

About 2/3 of the high risk group subjects (481 out of 715) were randomized to the treatment group. Among the 481 subjects in the treatment group, 58% (280 out of 481) of them attended at least one job-search seminar and they are the compliers. The others didn't attend any seminar and they are the never-takers. Table 3 shows that older subjects with more education, stronger motivation to attend the intervention, higher income and lower assertiveness are more likely to be compliers.

Two models with exactly the same form are fitted, with different subjects included. While the "Partial" models include only those observations with no missing values in the baseline covariates, 600 subjects in the high risk group, and the "Full" models include all 715 subjects in the analysis. The covariates involved in the model fitting are: indicator of not married, indicator for non-white, age, school grade completed, motivation to attend, assertiveness, economic hardship and attitude toward job search. Because of the normality assumption for the continuous variables in the extended general location models, logarithm transformations are applied to school grade completed and income.

Table 4 summarizes the estimation results of the probit models assuming ignorable missing data mechanism, where the standard errors of the ML estimates are calculated from 250 Bootstrap estimates. The 95% confidence intervals are estimated from the posterior distribution from the Gibbs' draws, by the 2.5th and the 97.5th percentiles. The CACEs are the differences between the probability of being reemployed at T_3 for the compliers who participate in the job-search seminars and the same probability for compliers who do not participate in the job-search seminars, evaluated at the mean values of the continuous baseline covariates, and then taken the weighted average over the distribution of the discrete covariates.

Table 4 shows that the ML estimates and the corresponding Bayesian estimates are very similar for each model. Estimates from the the "Full" models are similar to those from the "Partial" model. The standard errors from the "Full" model are smaller than those from the "Partial" model for almost all the parameters. This is to be expected, because the "Full" model uses more of the information contained in the data set and hence increases the

efficiency over the "Partial" model.

From the "Full" model that uses all the observations, the ML estimate of the probability of being reemployed at T_3 is increased by 16.1% by the intervention for a complier who has average education (grade 13) and the average score of attitude toward the job search. It is statistically significant with a 95% confidence interval of (0.004, 0.296). However, the ML estimate of the same CACE parameter from the "Partial" model is only 10% with a 95% confidence interval of (-0.051, 0.258) and it is not statistically significant. This difference is mainly due to the differences of the parameter estimates of the CR term, which is closely related to the CACE parameter. The ML estimate of the coefficient for CR from the "Full" model is 0.41 with a 95% confidence interval of (0.010, 0.764), while the estimate of the coefficient for CR from the "Partial" model is reduced to 0.268 with a 95% confidence interval of (-0.139, 0.668). Two sources may contribute to the difference between the two estimates from the "Full" and the "Partial" models. First, more information is utilized in the "Full" model than in the "Partial" model and leads to smaller standard error of the estimate. This is demonstrated by both the ML inference and the Bayesian inference. Second, the differences are also determined by the true underlying missing data mechanism of the baseline covariates. Our model assumes that the mechanism is ignorable.

In addition, white people with more education and a higher score of attitude toward job search have better chances to be reemployed in general. The effects of education and attitude toward the job search are statistically significant at a 95% confidence level.

5   Conclusions and Discussion

We have proposed an extended general location model to make inference for the CACE for data subject to non-compliance. The EGL model allows observations with missing values in the outcome or the baseline covariates to be included in the analysis, and it yields the same conditional distribution for the outcome and compliance conditioning on randomization and covariates as in Little and Yau (1998), thus generalizing that analysis. Models were presented for both ignorable and latent ignorable missing data mechanisms, together with EM

and Gibbs algorithms for computation. Simulations comparing the proposed methods with instrumental variable methods suggest that the IV methods and likelihood-based methods are reasonably robust to certain departures from the model assumptions, and the likelihood-based methods are more efficient than the IV methods. An application to the JOBS II data with a binary re-employment outcome was presented.

Bloom's naive method (1984), which assumes constant compliance rate across different values of the covariates, may provide very biased results when compliance varies across covariates. Also, estimates from our LIMD model with the compound exclusion restriction did well when its assumptions were satisfied, but yielded estimates that were more biased than estimates from the IMD model for the nonignorable missing data mechanism, where missingness of the outcome depended on a partially observed covariate. However, we emphasize that this finding is limited to our particular choice of LIMD model and simulation conditions. We think the LIMD assumption is likely to be more realistic than IMD in many settings, so other models based on this assumption are worthy of further study.

In this article we focused on a single outcome. The EGL model is fairly readily generalized to repeated measures normal data, by replacing the normal distribution of Eq. (3) by a multivariate normal with covariance structure suited to the repeated measures. Repeated binary outcomes can be accommodated by treating the outcomes in a normal model as latent thresholds, as we have done in the univariate case. Compliance is assumed all-or-none here. However, partial compliance is possible in practice. For example, in the JOBS II intervention trial, the treatment includes five job-search seminars and some participants only receive part of them. In a clinical trial, patients may take only a fraction of the intended dosage of medications because of side effects. Thus models that allow for partial compliance are also worthy of future study.

5.1 Acknowledgments

REFERENCES

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental variables. *Journal of American Statistical Association,* **91,** 444-472.

Baker, Stuart G. (1998). Analysis of Survival Data From a Randomized Trial With All-or-none Compliance: Estimating the Cost-effectiveness of a Cancer Screening Program. *Journal of the American Statistical Association,* **93,** 929–934.

Barnard, J., Frangakis, C.E., Hill, J. and Rubin, D.B. (2002). School Choices in NY City: a Bayesian Analysis of an Imprecfect Randomized Experiment (with discussion). pps. 3-97 in *Case Studies in Bayesian Statistics*, **5**, R.Kass et al. (eds.). New York: Springer-Verlag.

Bloom, H. S. (1984). Accounting for No-shows in Experimental Evaluation Designs. *Evaluation Review*, **8**, 225-246.

Dempster, A., Laird, N. and Rubin, D. B. (1977). Maximum Likelihood Estimation from Incomplete Data Using the EM Algorithm. *Journal of Royal Statistical Society, Ser. B*, **39**, 1-38.

Frangakis, C. E. and Rubin, D. B. (1999). Addressing Complications of Intent-to-treat Analysis in the Combined Presence of All-or-none Treatment-noncompliance and Subsequent Missing Outcomes. *Biometrika*, 86, 365-379.

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

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences (with discussion). *Statistical Science*, 7, 457-472.

Goldberger, A. S. (1972). Structural Equation Methods in The Social Sciences. *Econometrica*, 40, 979-1001.

Heckman, J. and Hotz, V. J. (1989). Choosing Among Alternative Non-experimental Method for Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of American Statistical Association*, 84, 862-881.

Imbens, G.W. and Rubin, D.B. (1996). Bayesian Inference For Causal Effects in Randomized Experiments with Noncompliance. T*he Annals of Statistics*, 25, 305-327.

Imbens, G.W. and Rubin, D.B. (1997). Estimating Outcome Distributions for Compliers in Instrumental Variables Models. *Review of Economic Studies*, 64. 555-574.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis With Missing Data*, 2nd. edition. New York, Wiley.

Little, R. J. A. and Yau, L. H. Y. (1998). Statistical Techniques for Analyzing Data from Prevention Trials: Treatment of No-shows Using Rubin's Causal Model. *Psychological Methods*, 3, 147-159.

Olkin, I. and Tate, R. F. (1961). Multivariate Correlation Models with Mixed Discrete and Continuous Variables. *Annals of Mathematical Statistics*, 32, 448-465.

Peng, Y. (2001) Causal Inference for Data with Non-compliance and Missing Values. Ph.D. Dissertation. University of Michigan.

Robins, J.M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics*, 23: 2379-2412.

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

Rubin, D. R. (1976). Inference and Missing data. *Biometrika*, 63, 581-592.

Vinokur, A. D., Price, R. H., and Schul, Y. (1995). Impact of the JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, 23, 39-74.

Yau, L. and Little, R.J.A. (2001). Inference for the Complier-Average Causal Effect from Longitudinal Data Subject to Noncompliance and Missing Data, with Application to a Job Training Assessment for the Unemployed. *Journal of the American Statistical Association*, 96, 1232-1244.

Table 1: 10^3* Bias and 10^3* RMSE of the Estimate of the CACE under Different Missing Data Mechanisms of Y, Different Distributions for Y and Different Methods, with Sample Size 1000

Mechanism to Generate Missing Data

Method	$Y \sim N(0,1)$ MAR1 B	R	MAR2 B	R	LMAR B	R	NMAR B	R	$Y \sim \mathrm{T}(4)$ MAR1 B	R	MAR2 B	R	LMAR B	R	NMAR B	R	$Y \sim \Gamma(5,0.2)$ MAR1 B	R	MAR2 B	R	LMAR B	R	NMAR B	R
MLBD	-6	114	10	117	13	108	5	112	-10	114	-5	116	-8	116	-12	107	8	106	-2	112	7	116	7	119
MLI	-2	128	5	123	13	122	-1	134	-11	125	-1	123	-8	131	-17*	135	6	139	-7	125	10	133	2	141
BYSI	-3	128	4	123	13	121	-3	134	-10	124	-3	123	-7	133	-19*	136	4	139	-4	127	9	133	-1	139
MLLI	-1	128	10	124	3	114	70*	220	-11	125	-1	123	-1	125	-34*	211	4	139	-4	127	9	133	-1	139
BYSLI	-2	127	8	124	2	112	68*	218	-11	125	-14	122	-3	123	-30*	214	3	139	-1	126	3	128	-9	150
MLCC	-12	133	4	125	14	126	2	168	-20*	130	-13	127	-12	143	-19	161	12	141	-8	130	7	136	-4	158
BYSCC	-13	133	4	127	15	124	1	166	-18*	129	-14	127	-9	143	-23*	163	11	141	-12	130	5	136	-9	158
IV	-22*	139	17*	129	7	131	-10	171	-12	146	-8	133	-6	142	27*	168	11	147	-12	138	13	140	-25*	171
IVL	-21*	141	28*	126	15	124	77*	344	94*	282	110*	277	111*	295	97*	319	11	147	-23*	135	21*	138	-93*	391
IVn	84*	262	92*	259	110*	270	137*	345	94*	282	110*	277	111*	295	97*	319	102*	282	95*	270	104*	273	101*	317
IVnL	87*	268	41*	193	68*	214	101*	780	94*	289	124*	233	70*	238	138*	733	100*	281	44*	199	60*	214	90*	620

B: Bias

R: Root Mean Square Error

Bias with an asterisk when $|Bias| > 2SE$.

MAR1: ignorable and also satisfying the latent ignorable conditions.

MAR2: ignorable but not satisfying the latent ignorable conditions.

21

Table 2: Probability Coverage of the Estimate of the CACE under Different Missing Data Mechanisms of Y, Different Distributions for Y and Different Methods, with Sample Size 1000

| | | Mechanism to Generate Missing Data | | | | | | | | | | | |
| | | $Y \sim N(0,1)$ | | | | $Y \sim T(4)$ | | | | $Y \sim \gamma(5, 0.2)$ | | | |
	Method	MAR1	MAR2	LMAR	NMAR	MAR1	MAR2	LMAR	NMAR	MAR1	MAR2	LMAR	NMAR
50 %	BYSI	51.6	48.0	53.2	54.8	57.2	48.0	46.0	44.4	46.0	50.0	49.2	54.0
	BYSLI	50.4	48.4	50.0	40.4	54.8	45.6	48.0	42.4	44.4	49.6	48.4	54.8
	BYSCC	52.4	49.6	48.8	56.8	57.2	44.8	44.4	43.6	55.2	49.6	49.2	57.2
	IV	50.0	49.2	55.2	53.6	55.2	46.8	43.2	52.0	48.8	48.8	45.2	52.8
	IVL	50.4	49.2	54.8	45.6	54.0	45.6	42.0	44.0	49.2	48.0	46.8	44.8
60 %	BYSI	58.8	58.0	59.6	64.8	62.8	57.6	56.0	56.8	55.5	60.8	62.4	63.2
	BYSLI	57.6	57.2	63.6	52.8	63.6	54.4	56.4	54.0	57.2	60.4	59.6	64.0
	BYSCC	62.0	57.6	59.6	65.2	63.6	55.6	55.6	55.6	62.4	59.6	58.4	66.4
	IV	61.6	58.4	64.4	61.2	62.6	54.0	54.8	61.2	59.6	55.6	56.4	60.8
	IVL	62.4	56.8	63.6	56.8	63.6	55.6	56.4	55.2	59.6	58.0	58.8	56.0
80 %	BYSI	79.6	80.0	82.8	80.8	79.2	78.0	76.0	76.0	76.8	78.4	77.2	81.6
	BYSLI	79.6	79.2	84.8	73.2	79.2	76.4	76.0	75.6	76.8	79.6	76.4	81.6
	BYSCC	81.6	76.8	79.6	84.0	78.0	80.8	78.8	76.4	82.0	81.6	80.8	84.4
	IV	79.2	77.6	80.0	80.4	76.4	74.4	77.6	80.4	74.8	76.0	77.6	78.4
	IVL	78.8	77.6	80.4	76.8	77.2	73.6	76.4	76.0	76.0	75.2	77.6	77.6
90 %	BYSI	90.4	92.8	94.8	92.0	89.6	88.0	86.4	87.6	88.8	88.0	88.0	91.6
	BYSLI	92.0	91.6	90.8	84.8	89.6	89.2	88.0	86.8	88.4	89.2	89.2	91.2
	BYSCC	90.4	88.4	90.0	90.8	87.2	88.8	89.2	87.2	91.2	90.0	91.6	92.4
	IV	90.8	89.2	90.4	90.0	84.8	88.0	87.6	89.6	88.4	86.4	86.8	86.4
	IVL	90.4	89.2	89.6	87.2	84.0	88.0	87.2	86.8	88.0	84.8	86.0	86.0

MAR1: ignorable and also satisfying the latent ignorable conditions.

MAR2: ignorable but not satisfying the latent ignorable conditions.

22

Table 3: Summary Statistics of the Covariates by Treatment R and Compliance C, HIGH risk group

Covariate	$R = 0$ Mean (SD)	$R = 1$ Mean (SD)	$R = 1$ & $C = 1$ Mean (SD)	$R = 1$ & $C = 0$ Mean (SD)
Age	36.45 (10.08)	36.60 (10.33)	39.08 (10.18)	33.65 (9.73)
School	13.28 (2.01)	13.35 (2.02)	13.81 (2.04)	12.82 (1.86)
Motivation	5.32 (0.80)	5.31 (0.82)	5.43 (0.82)	5.16 (0.79)
Income	6.59 (3.72)	6.17 (3.84)	6.67 (4.01)	5.57 (3.54)
Assertiveness	3.05 (0.87)	3.11 (0.92)	3.01 (0.92)	3.23 (0.92)
Not-Married	0.58 (0.49)	0.65 (0.48)	0.66 (0.48)	0.64 (0.48)
Economic Hardship	3.53 (0.93)	3.62 (0.85)	3.55 (0.84)	3.71 (0.84)
Non-White	0.21 (0.41)	0.22 (0.41)	0.18 (0.39)	0.25 (0.44)
Attitude	5.77 (1.12)	6.01 (1.06)	6.05 (1.06)	5.96 (1.06)
Social Undermine	1.83 (0.78)	1.83 (0.80)	1.79 (0.78)	1.88 (0.82)
Risk(T_0)	1.68 (0.19)	1.68 (0.23)	1.67 (0.22)	1.69 (0.24)

Table 4: Model fitting for the HIGH risk group with the outcome of re-employment at T_3, where "Full" indicates the results from data with all 715 observations, and "Partial" indicates the results from the data excluding those observations with missing values in covariates, 600 observations

Parameter		MLE $\pm$ SE	Gibbs $\pm$ SE	95% CI**
Intercept	(Full)	-1.549 $\pm$ 0.508	-1.551 $\pm$ 0.459	(-2.491, -0.694)
	(Partial)	-1.566 $\pm$ 0.485	-1.564 $\pm$ 0.498	(-2.530, -0.600)
C	(Full)	-0.446 $\pm$ 0.245	-0.406 $\pm$ 0.224	(-0.852, 0.030)
	(Partial)	-0.233 $\pm$ 0.267	-0.230 $\pm$ 0.235	(-0.692, 0.216)
CR	(Full)	0.410 $\pm$ 0.194	0.379 $\pm$ 0.194	(0.010, 0.764)
	(Partial)	0.268 $\pm$ 0.226	0.260 $\pm$ 0.206	(-0.139, 0.668)
School	(Full)	0.079 $\pm$ 0.031	0.077 $\pm$ 0.027	(0.024, 0.132)
	(Partial)	0.086 $\pm$ 0.030	0.086 $\pm$ 0.030	(0.025, 0.146)
Attitude	(Full)	0.149 $\pm$ 0.053	0.151 $\pm$ 0.051	(0.052, 0.251)
	(Partial)	0.135 $\pm$ 0.054	0.136 $\pm$ 0.054	(0.033, 0.243)
Non-White	(Full)	-0.220 $\pm$ 0.134	-0.218 $\pm$ 0.133	(-0.479, 0.038)
	(Partial)	-0.249 $\pm$ 0.167	-0.253 $\pm$ 0.148	(-0.546, 0.045)
CACE*	(Full)	0.161 $\pm$ 0.075	0.148 $\pm$ 0.075	(0.004, 0.296)
	(Partial)	0.104 $\pm$ 0.086	0.100 $\pm$ 0.079	(-0.051, 0.258)

* CACE is calculated at the mean values of the school grade completed and the score of attitude toward the job search, then averaged over the distribution of non-white.

** 95% CIs are calculated based on the 2.5th and 97.5th percentiles of the Gibbs draws.