Likelihood-Based Inference for the Finite Population Mean with Post-Stratification Information Under Nonignorable Nonresponse

Sahar Z Zangeneh[1,2] and Roderick J Little[3]

[1]RTI International, Research Triangle Park, NC, U.S.A.

[2]University of Washington, Seattle WA, U.S.A.

[3]Department of Biostatistics, The University of Michigan, Ann Arbor, MI, U.S.A.

Abstract

We describe models and likelihood-based estimation of the finite population mean for a survey subject to unit nonresponse, when post-stratification information is available from external sources. A feature of the models is that they do not require the assumption that the data are missing at random (MAR). As a result, the proposed models provide estimates under weaker assumptions than those required in the absence of post-stratification information, thus allowing more robust inferences. In particular, we describe models for estimation of the finite population mean of a survey outcome with categorical covariates and externally observed categorical post-stratifiers. We compare inferences from the proposed method with existing design-based estimators via simulations. We apply our methods to school-level data from California Department of Education to estimate the mean academic performance index (API) score in years 1999 and 2000. We end with a discussion.

Key words: Maximum likelihood; missing not at random; non-ignorable models; post-stratification; raking; unit nonresponse.

1 Introduction

It is truly an honor to contribute an article to this special issue celebrating Nan Laird's award of the 2021 International Prize in Statistics. We start by connecting the topic of our article with some aspects of Nan's methodological work. A useful feature of likelihood-based methods of statistical inference – in particular, Bayesian inference or asymptotic inference based on maximum likelihood (ML) – is that the methods can be applied to non-rectangular datasets, such as arise when there are missing data. Two of Nan Laird's most cited papers, on ML estimation using the Expectation Maximization (EM) algorithm (Dempster et al., 1977) and ML estimation of mixed models for unbalanced longitudinal data (Laird and Ware, 1982), exploit this property.

Standard ML software for missing data is based on the assumption that the missingness mechanism is ignorable, which means that inference can be based on the likelihood derived from the complete-data model for the study variables, without modeling the missingness mechanism. A key sufficient condition for ignoring the missingness mechanism is that the data are missing at random (MAR), as discussed in Rubin's famous (1976) paper (Rubin, 1976). An interesting feature of our paper is that it includes an simple practical example where missingness is missing not at random (MNAR) but the mechanism is nevertheless ignorable, thus showing that the MAR condition is a sufficient but not always a necessary condition for ignorability.

Our paper concerns the analysis of nonresponse in survey sample data when there is post-stratified data, specifically marginal distributions of survey variables available for the population or a random sample of the population from sources external to the survey. Such data are increasingly important in survey sampling settings, with the rising levels of survey nonresponse and increased reliance on data that are not randomly sampled. As we show, the presence of post-stratified data allows the MAR assumption to be relaxed, and certain MNAR models to be fitted.

In finite population survey sampling, likelihoods can be defined by so-called "superpopulation" modeling, where the finite population is assumed to be sampled from an infinite-sized "superpopulation," and inference is based on a statistical models for the survey variables in this superpopuation (Chambers et al., 2012; Valliant et al., 2000). This approach leads to likelihood functions for model parameters, and inferences about finite population parameters based (in effect) on prediction of the values of survey variables for nonrespondents and non-sampled units. However, concerns over model misspecification lead many statisticians trained in probability sampling to prefer the so-called design-based or randomization approach to statistical inference. In this approach, which is predominant in classic survey sampling texts (e.g. Kish (1965), Cochran (2007)), the survey variables are treated as fixed quantities and not assigned a distribution; rather inference is based on the probability distribution that underlying probabilistic selection of the sample. This approach is not strictly applicable when there is survey nonresponse, but the "quasi-randomization" approach, which acts as if we have a probability sample after conditioning on auxiliary data available for respondents and nonrespondents, can be thought of as extending the randomization approach to handle nonresponse. In this article we adopt a superpopulation modeling perspective to surveys, but our simulations include some comparisons with common design-based approaches.

We describe likelihood-based estimation of the finite population mean of a survey variable Y, when (a) Y and a set of post-stratifiers Z are observed for r respondents but missing for n–r nonrespondents in the sample, (b) a set of covariates X is observed for all n units in the survey, and (c) the marginal distribution of each $Z_k, k = 1, \ldots, K$ is also observed for the same target population, from a larger survey or a census. A Z_k could represent a set of variables, provided their joint distribution is available from auxiliary data. For ease of presentation, we consider univariate auxiliary Z_k margins throughout this paper. The structure of the data is depicted in Figure 1. For unit $i = 1, \ldots, n$ in the survey,

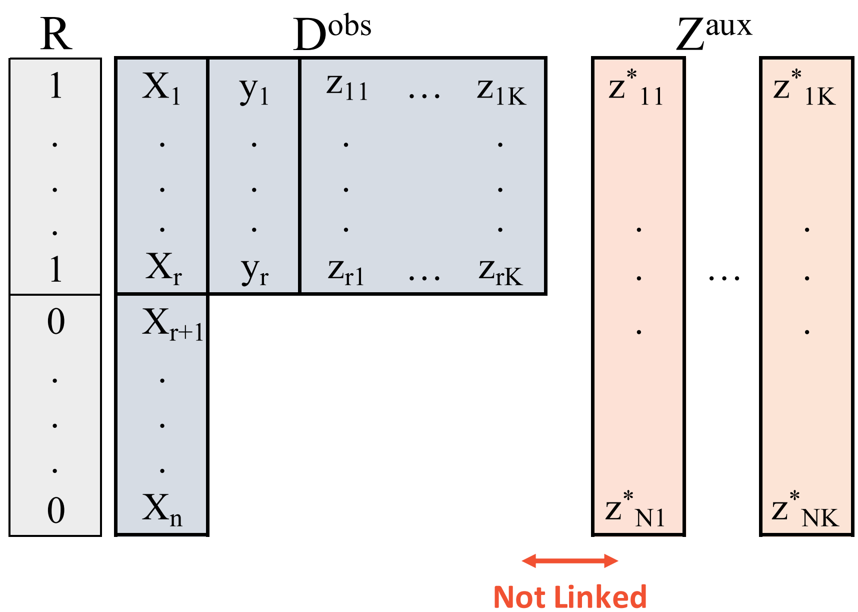


Figure 1: Missing data pattern with postsratifying information.

let $d_i = (x_i, y_i, z_i)$ denote the values of (X, Y, Z), and R_i denote the value of the response indicator R, where $R_i = 1$ if (y_i, z_i) is observed and $R_i = 0$ if (y_i, z_i) is missing. We denote by D the full data matrix for the survey, $D = (d_1, \ldots, d_n)^T$, $d_i = (x_i, y_i, z_i)$, D^{obs} the observed survey data, namely $\{d_i, i = 1, \ldots, r\}$ and $\{x_i, i = r + 1, \ldots, n\}$ and Z^{aux} the auxiliary data consisting of the marginal distributions of Z_k, $k = 1, ..., K$. Note that the units in the auxiliary data Z^{aux} are not linked with the units in the survey. This scenario occurs frequently in settings where post-stratification is used for nonresponse adjustment.

We assume throughout the paper the probability that (z_i, y_i) is observed may depend on x_i and z_i but does not depend on y_i, given x_i and z_i, that is:

$$\mathbb{P}(R_i = 1|x_i, z_i, y_i, \psi) = \mathbb{P}(R_i = 1|x_i, z_i, \psi). \tag{1}$$

where ψ represents unknown model parameters for the conditional response propensity model. The resulting mechanism is missing not at random (MNAR) (Rubin, 1976; Little and Rubin, 2019), if missingness depends on z_i, because z_i is not observed for survey units i that are missing. We describe circumstances where the auxiliary margins Z^{aux} provide us with the information needed to estimate the parameters governing the joint distribution of X and Z, allowing ML or Bayesian inference. We focus here on models for the important case where X and Z consist of categorical variables, although our general approach can also be applied to problems where some or all of X or Z are continuous.

Standard design-based approaches to this data structure include post-stratification (Holt and Smith, 1979) and extensions such as raking, where respondents are weighted to match the distribution of the discrete post-stratifiers in the population. Calibration methods, extend post-stratification to encompass known population totals of continuous auxiliary variables (Deville and Sarndal, 1992; Deville et al., 1993; Särndal et al., 2003; Lumley, 2010; Kott and Chang, 2010; Kott and Liao, 2017, 2018): These methods minimize the distance between the original sampling weights and new calibration weights subject to known sums of auxiliary variables (Deville and Sarndal, 1992). Kalton and Flores-Cervantes (2003) describe estimators obtained from alternative choices of distance functions. One advantage of our likelihood-based approach is that it does not require the choice of a distance function, which appears to us to be somewhat arbitrary.

Model-based inference, on the other hand, treats the survey outcomes as well as the inclusion and response indicators as random variables in a statistical model: The model is used to (i) infer the population parameters of interest or (ii) predict the unobserved values of Y. Two main variants of model-based inference are frequentist superpopulation modeling, where inferences are based on repeated samples from the sample and the superpopulation, and Bayesian inference, where a prior distribution is chosen for the parameters, and inferences are based on the posterior distribution of the finite population quantities of interest given the observed data (Little, 2004). Little (1993) justifies post-stratified and raking estimates for categorical post-strata as ML estimates for particular models. Gelman and Little (1997) propose multilevel regression and post-stratification, a Bayesian multilevel modeling approach to post-stratified survey data. This approach was further developed by Si et al. (2017) and Si and Zhou (2019). None of these articles consider MNAR models for missing data, which is the focus of this paper.

Section 2 outlines likelihood-based inference for surveys with the data pattern of Figure 1. Section 3 compares and contrasts repeated sampling properties of the proposed model-based estimators to commonly used design-based estimators for a variety of assumed missing data mechanisms using simulated categorical data. Section 4 applies the proposed methods to real data with continuous outcomes from the California Department of Education. Section 5 ends with a discussion and directions for future research.

2 Models for unit nonresponse with auxiliary information

2.1 Overview

Denoting density functions by $f(.)$, we consider models that are i.i.d. over the units i, where the joint distribution of X, Z, Y and R is factored as

$$f_{X,Z,Y,R}(\mathbf{x}_i, \mathbf{z}_i, y_i, r_i|\theta, \phi) = f_{Y|X,Z,R}(y_i|\mathbf{x}_i, \mathbf{z}_i, \theta, R_i = r_i) f_{X,Z,R}(x_i, z_i, r_i|\phi)$$

$$= f_{Y|X,Z,R}(y_i|\mathbf{x}_i, \mathbf{z}_i, \theta) f_{X,Z,R}(\mathbf{x}_i, \mathbf{z}_i, r_i|\phi), \tag{2}$$

and θ and ϕ are distinct parameters (Little and Rubin, 2019). Note that the distribution of Y given (X, Z, R) in the second line of Eq. (2) does not depend on R. This is justified because the assumption in Eq. (1) about the missingness mechanism implies that R is independent of Y given X and Z. This means that the parameters θ of this conditional distribution can be estimated from the component of the likelihood based on the survey respondents. The remaining parameters ϕ are then estimated by assuming a model for the joint distribution of X, Z and R for which these parameters are identified from the survey and auxiliary data. We consider here cases where X and Z are categorical, in which case the available data lead to incomplete contingency tables with supplemental margins. We can thus apply methods for this data structure discussed in Little and Rubin (2019).

Our inferences are based on the likelihood shown below in Eq. (3),

$$L(\theta, \phi | D_{obs}, Z^{\text{aux}}) = \prod_{i=1}^{r} f_{Y|X,Z}(y_i | x_i, z_i, \theta) \phi_0^r (1 - \phi_0^r)^{(n-r)} \times \prod_{i=1}^{r} f_{X,Z}(x_i, z_i | r_i = 1, \phi^{(1)})$$
$$\times \prod_{i=r+1}^{n} f_X(x_i | r_i = 0, \phi) \times \prod_{k=1}^{K} \prod_{j=1}^{N} f_{Z_k}(z^*_{jk} | \phi), \tag{3}$$

where the first r sample units are respondents, ϕ_0 is the marginal probability of response to the survey, $(D^{\text{obs}}, Z^{\text{aux}})$ represents the observed data, and the last component of the likelihood comes from the auxiliary data. We use $\phi^{(r_i)}$ to distinguish between the parameters in the observed ($r_i = 1$) and missing ($r_i = 0$) units respectively. According to Eq. (2), the parameter θ, describing the conditional distribution of Y given X and Z, is the same for the observed and missing data, but the parameter ϕ can differ between the observed and missing data. A slight simplification in Eq. (3) is that the data from each of the auxiliary margins is assumed independent of the information from the survey data. This is not quite true if the auxiliary margins and survey have units in common: however we believe that this information is negligible, and it is not easily recoverable given that the auxiliary and survey units are not linked. ML estimation of the population mean of Y is achieved by first predicting the values of Z for nonrespondents in the sample given X and the ML estimate of ϕ, and then predicting the values of Y for nonrespondents from the distribution of Y given (X, Z) and the ML estimate of θ. The Bayesian approach replaces ML estimates of the parameters with draws from their posterior distribution.

We now consider some special cases of ML inference based on Eq. (3).

2.2 Single Post-Stratifier

We first consider the simple case of a single post-stratifier Z and no covariates X. The missingness assumption in Eq. (1) then reduces to

$$\mathbb{P}(R_i = 1|Z_i, Y_i, \psi) = \mathbb{P}(R_i = 1|Z_i, \psi). \tag{4}$$

The likelihood in Eq. (3) reduces to

$$\begin{aligned} L(\theta, \phi|D_{obs}, Z^{\text{aux}}) = A(\theta) \times B(\phi) \times C(\phi) \text{ where} \\ A(\theta) = \prod_{i=1}^{r} f_{Y|Z}(y_i|z_i, \theta), \\ B(\phi) = \prod_{j=1}^{N} f_Z(z_j^*|\phi), \text{ and} \\ C(\phi) = \phi_0^r(1-\phi_0^r)^{(n-r)} \prod_{i=1}^{r} f_Z(z_i|r_i = 1, \phi^{(1)}) \end{aligned} \tag{5}$$

The parameters θ of the conditional distribution of Y given Z can be estimated from $A(\theta)$, and the parameters of the marginal distribution of Z across respondents and nonrespondents can be estimated from the auxiliary data $B(\phi)$.

For univariate categorical Z with J categories, a natural model is to assume that Z is multinomial with

$$\mathbb{P}(z_i = j) = \phi_j, \quad j = 1, \ldots, J, \quad \sum_{j=1}^{J} \phi_j = 1. \tag{6}$$

The ML estimate of ϕ_j, is simply the proportion of the auxiliary data in post-stratum j. The resulting direct estimate of the population mean of Y is

$$\bar{Y}_{\text{mod}} = \sum_{j=1}^{J} \hat{\phi}_j \bar{y}_{\text{mod}_j}, \tag{7}$$

where $\bar{Y}_{\text{mod}_j}$ is the average of observed and predicted values of Y in post-stratum j, based on the assumed model for Y given Z and $\hat{\phi}_j = N_j/N$. For example, if the model assumed that Y

was normal with mean μ_j and variance σ_j^2, then $\bar{Y}_{\text{mod}} = \bar{Y}_{\text{PS}} = \sum_{j=1}^{J} \hat{\phi}_j \bar{y}_{jR}$, where $\bar{y}_{jR}$ is the respondent sample mean in post-stratum j. This estimator is the well-known post-stratified mean, and weights respondents by the inverse of the response rate in post-stratum j.

Alternatively, we can use the models in (3) and (6) to predict or impute the unobserved values of Z for individual nonrespondents, and use them as predictors in a model for Y. The resulting predictive estimator of the population mean of Y is

$$\bar{Y}_{\text{pred}} = \frac{1}{n}\left(\sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} \hat{y}_i\right), \tag{8}$$

where $\hat{y}_i$ is the predicted value of y_i given the predicted value of $\hat{z}_i$ for nonrespondent $i = r+1, \ldots, n$. The parameters θ for the regression of Y on Z, and $\phi^{(0)}$, for the distribution of Z among nonrespondents are both estimated by ML. The ML estimate of $\phi_j^{(0)}$, the estimated proportion of nonrespondents in category j is

$$\hat{\phi}_j^{(0)} = \frac{n\hat{\phi}_j - r\hat{\phi}_j^{(1)}}{n-r}$$

where $\hat{\phi}_j^{(1)}$ is the observed proportion of respondents in category j, which can be estimated from $C(\phi)$.

Estimators based on Eqs. (7) and (8) require at least one respondent in each post-stratum, and may be unstable if the respondent sample sizes in any post-strata is small. This is particularly likely if Z is a vector of two or more variables, with their joint distribution available from auxiliary data. Instability can be addressed by assuming an unsaturated model for Y. For example if Z is bivariate, say $Z = (Z_1, Z_2)$, then we can assume an additive model for Y given (Z_1, Z_2), or a mixed model with fixed main effects of Z_1 and Z_2 and random interactions. This modeling approach to stabilizing $\bar{Y}_{\text{PS}}$ and $\bar{Y}_{\text{pred}}$ differs from the typical design-based approach, which is to modify the nonresponse weight. This example is also discussed in Little et al. (2017), who point out that the post-stratified mean is actually ML for a MNAR model.

2.3 Two or More Post-Stratifiers

Suppose now we have two categorical post-stratifiers Z_1 and Z_2, with respectively J_1 and J_2 levels, and we have auxiliary data on the marginal distributions of Z_1 and Z_2 but not their joint distribution. The model (2) becomes

$$f_{Z,Y,R}(z_{i1}, z_{i2}, y_i, r_i|\theta, \phi) = f_{Y|Z}(y_i|z_{i1}, z_{i2}, \theta) \times f_{Z,R}(z_{i1}, z_{i2}, r_i|\phi).$$

Denoting the marginal probability of response by ϕ_0, the likelihood (3) becomes

$$\begin{aligned} L(\theta, \phi|D_{obs}, Z^{\text{aux}}) &= A(\theta) \times B(\phi), \text{ where} \\ A(\theta) &= \prod_{i=1}^{r} f_{(Y|Z_1,Z_2)}(y_i|z_{i1}, z_{i2}, \theta) \text{ and} \\ B(\phi) &= \phi_0^r(1-\phi_0^r)^{(n-r)} \prod_{i=1}^{r} f_{(Z_1,Z_2)}(z_{i1}, z_{i2}|r_i = 1, \phi^{(1)}) \\ &\quad \times \prod_{j=1}^{N} f_{Z_1}(z_{j1}^*|\phi) \times \prod_{j=1}^{N} f_{Z_2}(z_{j2}^*|\phi). \end{aligned} \quad (9)$$

The ML estimates of θ are estimated from $A(\theta)$, and ML estimates of ϕ are estimated from $B(\phi)$. We focus on the latter here.

An unconstrained (or saturated) multinomial joint distribution for (Z_1, Z_2, R) has $2J_1J_2 - 1$ distinct probabilities. The data described in Figure 1 yields estimates of $J_1J_2 + J_1 + J_2 - 2$ probabilities, namely the joint distribution of (Z_1, Z_2) for respondents ($J_1J_2 - 1$ probabilities) and the marginal distributions of Z_1 ($J_1 - 1$ probabilities), Z_2 ($J_2 - 1$ probabilities) and R (1 probability). This implies that there are

$$2J_1J_2 - 1 - (J_1J_2 - J_1 - J_2 - 2) = (J_1 - 1)(J_2 - 1)$$

more parameters that are not estimable. That is, the saturated MNAR model is under-identified.

We consider the constrained MNAR "RAKE" model that assumes the marginal distributions of Z_1 and Z_2 are different for respondents and nonrespondents, but the $(J_1-1)(J_2-1)$ odds ratios of Z_1 and Z_2 are the same for respondents and nonrespondents. This yields the same number of constraints as there are under-identified parameters, that is, a just-identified model. Little and Wu (1991) showed that raking the $J_1 \times J_2$ table of respondent counts (say $\{r_{j_1j_2}\}$) to the auxiliary margins of Z_1 and Z_2 gives ML estimates $\hat{\phi}$ of ϕ under this RAKE model.

The post-stratified estimator (7) extends to

$$\bar{Y}_{\text{rake}} = \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \hat{\phi}_{j_1j_2} \bar{y}_{\text{mod}_{j_1j_2}}$$

where $\hat{\phi}_{j_1j_2}$ is the estimated proportion of the population with $Z_1 = j_1, Z_2 = j_2$ from raking, and $\bar{y}_{\text{mod}_{j_1j_2}}$ is the average of observed and predicted values of Y given $Z_1 = j_1, Z_2 = j_2$, based on the model for Y given Z_1, Z_2 with θ estimated by ML. The predictive estimator (8) uses predicted values of Z_1 and Z_2 for nonrespondents, where $\hat{\phi}^{(0)}_{j_1j_2}$ is the estimated proportion of nonrespondents with $Z_1 = j_1, Z_2 = j_2$ from raking.

With $K > 2$ auxiliary margins, raking yields ML estimates of ϕ for the model that assumes the marginal distributions of $Z_1, \ldots, Z_K$ differ for respondents and nonrespondents, but the j-way associations between $Z_1, \ldots, Z_K$ are the same for respondents and nonrespondents, for $j = 2, \ldots, K$. A more parsimonious unsaturated log-linear model for $Z_1, \ldots, Z_K$ that sets higher-order associations to zero may be needed here if the number of respondents in the cells formed by $Z_1, \ldots, Z_K$ is small. For discussion of unsaturated models for

$Z_1, \ldots, Z_K$ and R (see Little and Rubin, 2019, Chapter 13 and Section 15.4.2).

2.4 One Post-Stratifier and One Covariate

With one covariate X observed for all units in the sample, and one post-stratifier Z observed for survey respondents, the model in Eq. (2) yields the likelihood

$$\begin{aligned} L(\theta, \phi | D^{\text{obs}}, Z^{\text{aux}}) &= A(\theta) \times B(\phi), \text{ where} \\ A(\theta) &= \prod_{i=1}^{r} f_{Y|X,Z}(y_i | x_i, z_i, \theta) \text{ and} \\ B(\phi) &= \phi_0^r (1-\phi_0)^{(n-r)} \prod_{i=1}^{r} f_X(x_i | r_i = 1, \phi^{(1)}) \prod_{i=r+1}^{n} f_X(x_i | r_i = 0, \phi^{(0)}) \times \prod_{j=1}^{N} f_Z(z_j^* | \phi), \end{aligned} \tag{10}$$

where ϕ_0 is the marginal probability of response. This structure is similar to the case of two post-stratifiers, with X playing the role of one of the post-stratifiers. Here we have data on the distributions of X for respondents and nonrespondents from the sample, whereas for a post-stratifier, we have data on the marginal distribution from auxiliary data and the distribution for respondents from the sample. In particular for categorical X and Z, we can apply the RAKE model of Section 2.3 with X playing the role of Z_2. For that model, raking the joint distribution of X and Z for respondents to the auxiliary margin of Z and the margin of X from the sample yields ML estimates of ϕ and $\phi^{(0)}$.

3 Simulation study

3.1 Simulation Design and Methods Compared

The goal of this simulation study is to explore repeated sampling properties of the proposed estimators for different missingness mechanisms and different outcome regression models.

To focus on the missingness mechanisms for unit nonresponse, we consider simple random samples from a finite population. To avoid distributional assumptions, we consider here the situation where all variables of interest are univariate and binary.

Let Y be a binary survey variable of interest and Z a binary post-stratifier, observed only for sample respondents. Let X denote a binary covariate, observed for all units in the sample and R the binary response indicator which is observed for all units in the sample. The marginal distribution of Z in the population is also available from an external source.

We generate data for (X, Y, Z, R) using a selection model factorization (Little and Rubin, 2019):

$$f_{X,Z,Y,R}(x_i, z_i, y_i, r_i|\theta, \phi) \\ = f_{Y|X,Z}(y_i|x_i, z_i, \theta) f_{X,Z}(x_i, z_i|\phi) f_{R|X,Z,Y}(r_i|x_i, z_i, y_i, \psi), \tag{11}$$

where

1. (X, Z) are multinomial with $\mathbb{P}(X = Z = 0) = .2$, $\mathbb{P}(Z = 0, X = 1) = .35$, $\mathbb{P}(X = 0, Z = 1) = .3$ and $\mathbb{P}(X = Z = 1) = .15$

2. Y given (Z, X) is Bernoulli with

$$\operatorname{logit} \mathbb{P}(Y = 1|X, Z) = \theta_0 + \theta_X(X - \bar{X}) + \theta_Z(Z - \bar{Z}) + \theta_{XZ}(X - \bar{X})(Z - \bar{Z})$$

   for $\theta_0 = 0.5$ and six choices of $(\theta_X, \theta_Z, \theta_{XZ})$ shown in Table 1.

3. R given (Z, X, Y) is Bernoulli with

$$\operatorname{logit} \mathbb{P}(r_i = 1|z_i, x_i, y_i, \psi) = \psi_0 + \psi_X(X - \bar{X}) + \psi_Z(Z - \bar{Z}) + \psi_{XZ}(X - \bar{X})(Z - \bar{Z}) + \psi_Y(Y - \bar{Y})$$

Table 1: Parameters for the outcome regression model: distribution of Y given X and Z.

θ_X	θ_Z	θ_{XZ}
2	2	2
2	2	0
2	0	0
0	2	0
0	0	0

for seven choices of $\psi = (\psi_X, \psi_Z, \psi_{XZ}, \psi_Y)$ shown in Table 2, chosen to reflect different relationships between R and Y, X and Z. The coefficients are chosen to give an approximate response rate of 70% for all simulated datasets.

A total of $5 \times 7 = 35$ combinations of population structures and non-response mechanisms are considered in our simulation study. All populations are generated such to avoid the presence of structural zeros. At each iteration, we generate a population of size $N = 100,000$ and draw a simple random sample with fixed sample size of $n = 1000$. We use the six estimators described below to estimate the finite population mean $\bar{Y}$.

The following methods for estimating the population mean of Y were compared in the simulation study:

1. The respondent mean, ignoring the supplemental information about X and Z. This method is labelled CC, for complete-case analysis.

2. The respondent weighted mean, with weights the inverse of the response rate within categories of X, ignoring the information about Z. We label this method NR, for nonresponse weighted analysis.

Table 2: Parameters for the response propensity model: distribution of R given X,Z and Y.

MD Scenario	ψ_X	ψ_Z	ψ_{XZ}	ψ_Y
Scenario 1	2	2	2	2
Scenario 2	2	2	2	0
Scenario 3	2	2	0	2
Scenario 4	2	2	0	0
Scenario 5	2	0	0	0
Scenario 6	0	2	0	0
Scenario 7	0	0	0	0

3. The post-stratified weighted mean, with weights obtained by matching to the Z auxiliary margin, ignoring the information about X. We label this method PSZ, for post-stratification based on Z.

4. NRPS: the weighted mean, with weights from one iteration of raking to the X sample margin and then the Z auxiliary margin. This is a standard design-based approach.

5. RAKEXZ: Similar to NRPS, but iteratively raking on the X and Z margins until convergence. This yields ML estimates of the joint distribution of X and Z under the RAKE model of Section 2.3 and takes the form of the estimator in Eq. (7) based on a logistic regression with X and Z interactions for Y.

6. PRED1: Predictive model-based estimator in the form of (8), where X and Z are jointly imputed for the nonrespondents using the RAKE model of Section 2.3, assuming the odds ratios of X and Z are the same for respondents and nonrespondents. Nonrespondent values of Y are imputed assuming a saturated logistic model for Y

given X and Z.

7. PRED2: Same as PRED1, except the interactions of X and Z are not included in the logistic model for Y given X and Z.

Inferences for CC, NR, PSZ, RAKEXZ and NRPS are performed using the the `survey` package in R (Lumley, 2009). We use the R package `nlme` (Bates, 2005) to fit the regression models in the two predictive estimators PRED1 and PRED2, and use bootstrap replicates for standard errors.

3.2 Simulation Results

Tables 4 and 5 compare the absolute root mean square error and the absolute empirical bias of the six different estimators described in Section 3.1 in repeated random samples. Tables 6 and 7 compare the non-coverage and the average relative width of 95% confidence intervals from the six different estimators in repeated random samples. When the response depends on the outcome Y (MD Scenarios 1 and 3), none of the methods perform well, with high relative bias and relative RMSE, and confidence coverage far below the nominal 95% level. On the other hand, when the data is MCAR (MD Scenario 7), all methods perform well.

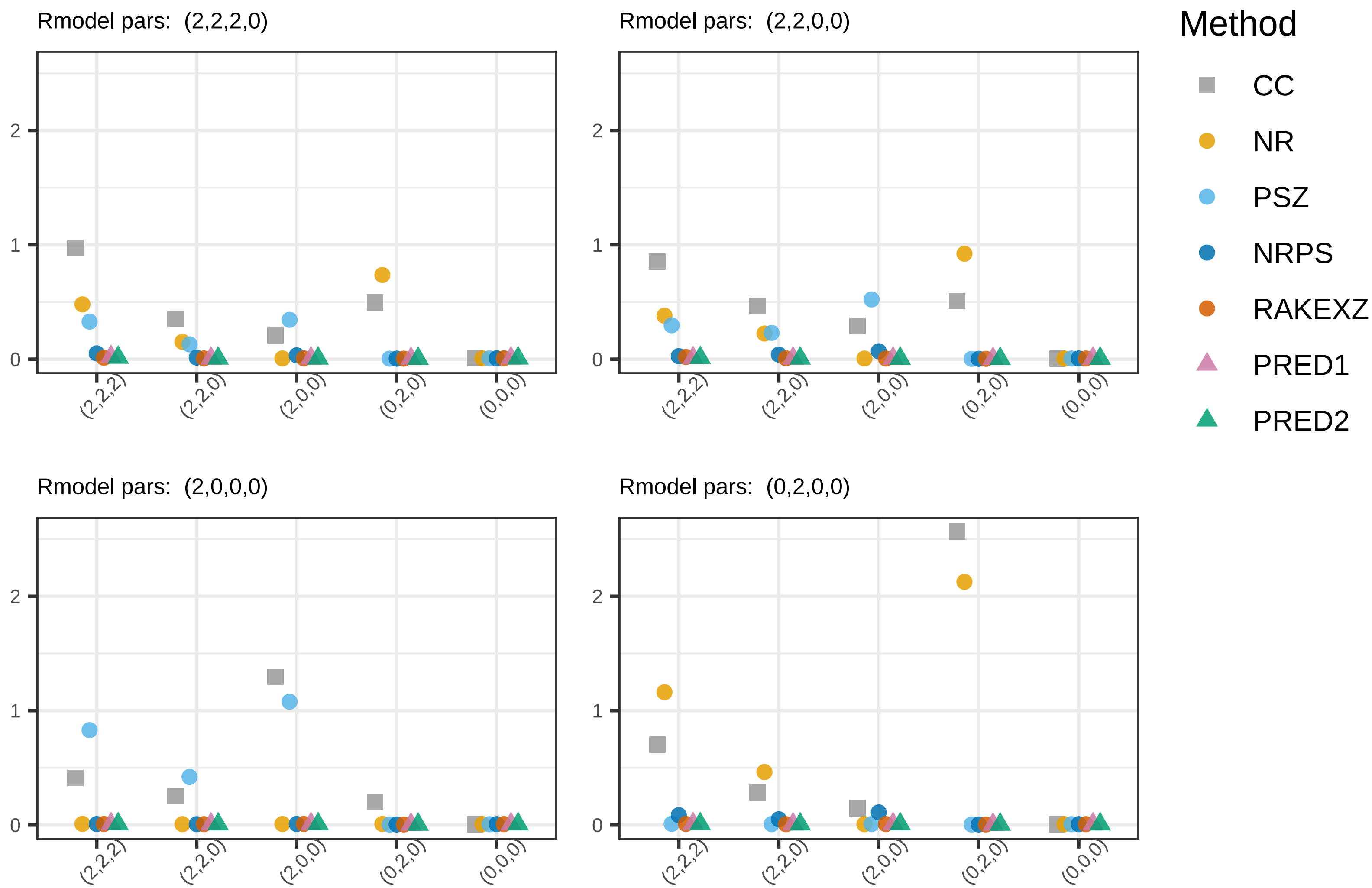


Figure 2: Relative root mean square error of the six different estimators for $\bar{Y}$ displayed as a percentage of the true value of $\bar{Y}$.

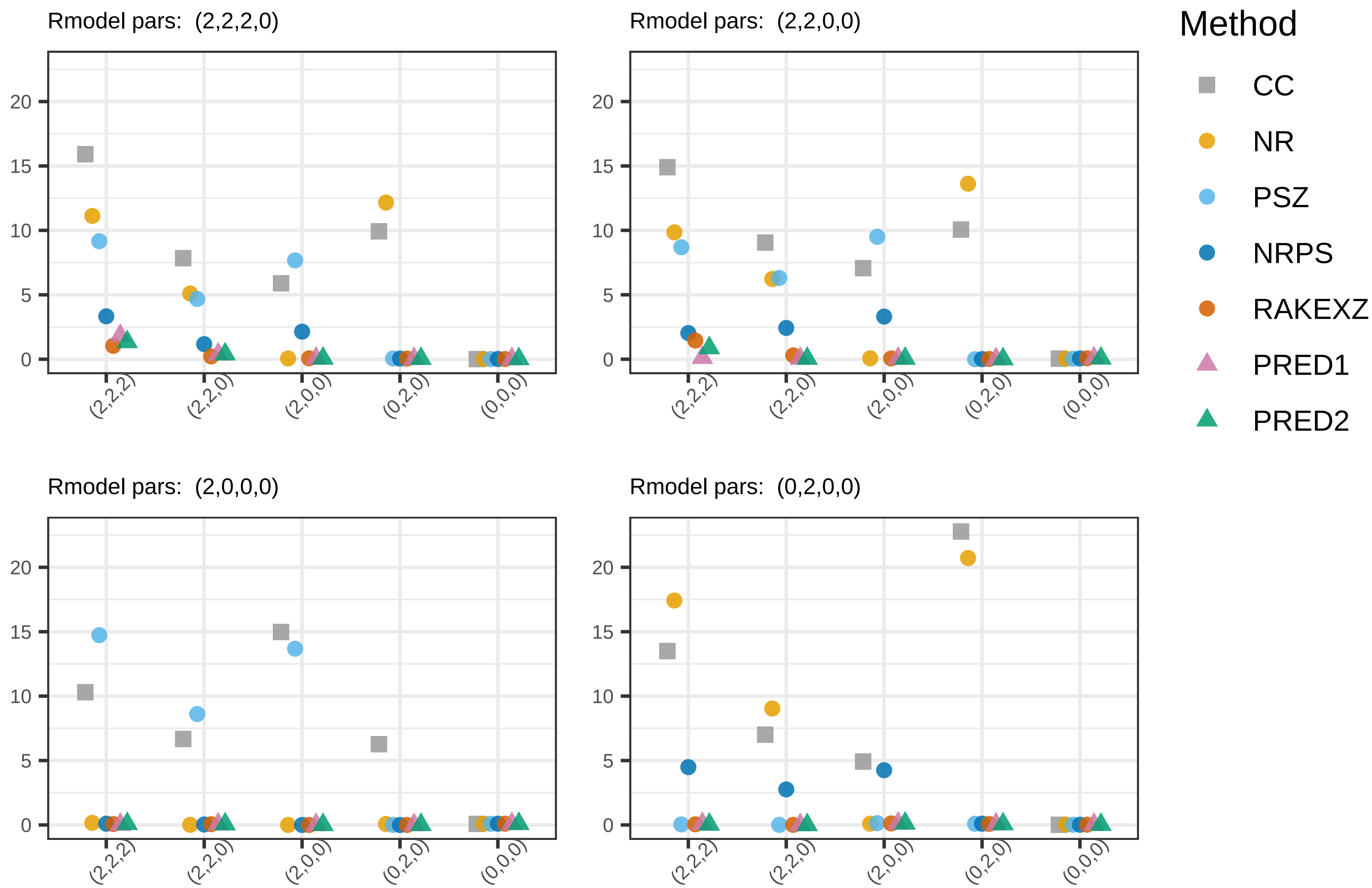


Figure 3: Relative absolute empirical bias of the six different estimators for $\bar{Y}$ displayed as a percentage of the true value of $\bar{Y}$.

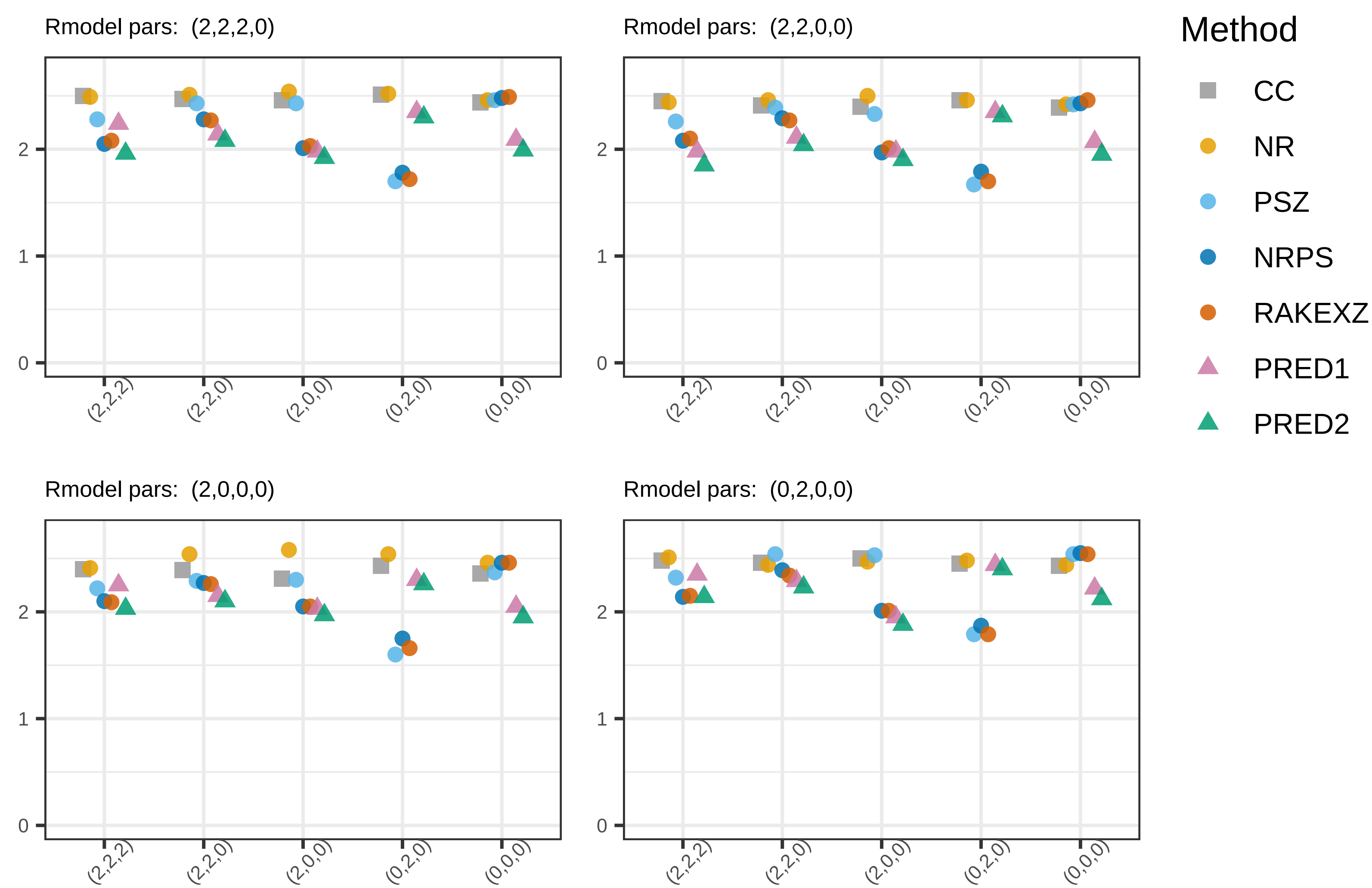


Figure 4: Relative average width of the 95% confidence intervals for the six different estimators of $\bar{Y}$ displayed as a percentage of the true value of $\bar{Y}$.

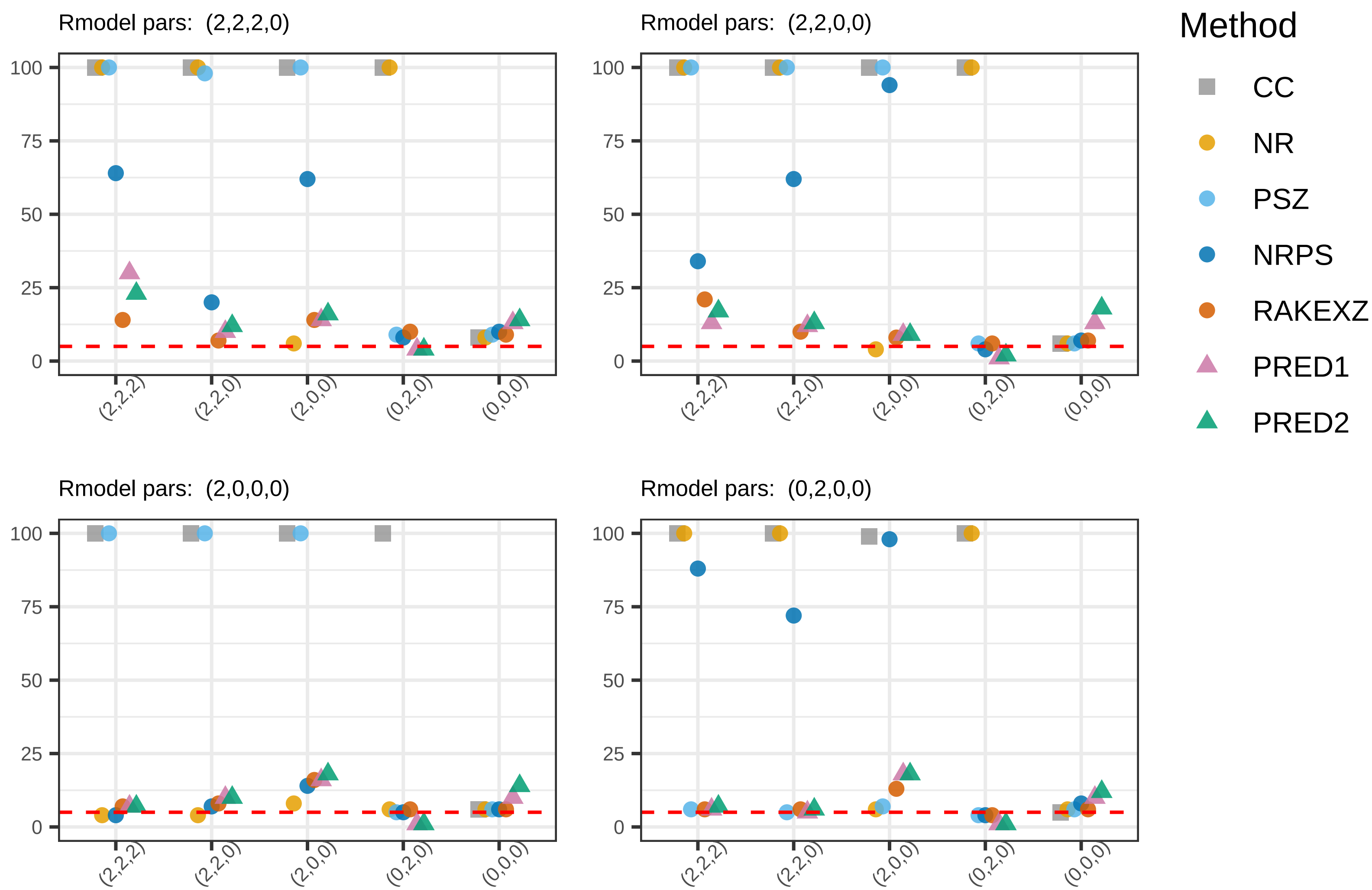


Figure 5: Non coverage of the 95% confidence intervals six different estimators for $\bar{Y}$. The red horizontal dashed line represents the nominal non-coverage of 5%.

Figures 2 – 4 display the intermediate missing data mechanism in MD scenarios 2, 4, 5 and 6. In these scenarios, the three model-based estimators give more efficient point estimates as indicated by lower root mean square errors (Figure 2), mainly due to the reduced bias (Figure 3). These estimators also yield tighter confidence intervals (Figure 4) while achieving nominal coverage (Figure 5). These estimators adapt to the optimal design-based estimators which give point estimates with low RMSE and empirical bias with good inferences when the weights appropriately adjust for missingness: In missing data scenarios 2 and 4, where missingness depends on X and Z, the NRPS and RAKEXZ estimators achieve this. In missing data scenario 5, missingness depends on X, and weighting methods that adjust for X give efficient results while in missing data scenario 6 where missingness depends on Z, weighting methods that adjust for X give efficient results. It summary, as long as the response mechanism does not depend on the X and Z interaction, the model-based estimators remove much of the bias.

4 Application

We apply the six estimators in Section 3.1 to data from the Academic Performance Index (API), a standardized test of students which sought to measure academic performance and progress of public schools in the state of California (Kim and Sunderman, 2005). The API was administered by California Department of Education and used to guide statewide policy through 2017, when it was replaced by a new accountability system. The `apipop` dataset in the R package `survey` contains information on 37 variables for all 6194 schools with at least 100 students.

We consider two numeric outcomes, the mean api scores in year 1999 and 2000 which we denote by Y_1 and Y_2 respectively. It is plausible to assume that missingness of school-level data would depend on whether or not a school had all its pupils tested. However this variable will be measured once the survey is taken, and not necessarily available through official statistics or past surveys— we thus consider as our covariate X, the binary variable which is equal to one if 100% of students in a school are tested and zero otherwise. Missingness of information on a school can also depend on the school's overall performance. One such measure is whether a school is eligible for awards. The proportion of schools eligible for awards can be assumed to be obtainable from official statistics, and we thus consider `awards` as our binary post-stratifier Z. We use R_1 and R_2 to denote the binary response indicator variables for Y_1 and Y_2 respectively. We consider the following models for the missing data mechanisms

$$\text{logit}[\mathbb{P}(R_1 = 1|X = x, Z = z)] = 1 + \psi_x(x - p_x) + \psi_z(z - p_z) + \psi_{xz}(x - p_x)(z - p_z)$$

and

$$\text{logit}[\mathbb{P}(R_2 = 1|X = x, Z = z)] = 1 + \psi_x(x - p_x) + \psi_z(z - p_z) + \psi_{xz}(x - p_x)(z - p_z)$$

using the same values of ψ_x, ψ_z and ψ_{xz} shown in Table 3. Here, we only consider the scenarios where missingness depends on X and Z. Similar to our simulation study, the coefficients in Table 3 are also chosen to give a response rate of approximately 70%. The five different missingness mechanisms are similar to those considered in Section 3, reflecting different dependency structures of R. We draw repeated samples from the `apipop` dataset and apply the proposed estimators to each observed dataset.

Table 3: Models for R given X and Z.

MD Scenario	ψ_X	ψ_Z	ψ_{XZ}
Scenario 1	2	2	2
Scenario 2	2	2	0
Scenario 3	2	0	0
Scenario 4	0	2	0
Scenario 5	0	0	0

We use the same six estimators considered in Section 3. After verifying normality assumptions of Y_1 and Y_2 for respondents, we use linear regression with binary predictors to model the distribution of the two different outcomes given X and Z. We use 50 bootstrap samples for the RAKE model at the first step, and 50 predictive draws using the residual standard errors of the linear regression of Y on X and Z. The design-based methods were all derived using the `survey` package in `R`, and residual standard errors of the linear model were extracted using the `R` software package `arm` (Gelman et al., 2018).

Figures 6 compares the point estimates for the mean API score in the years 1999 and Figure 7 compares its interval estimates. Similar qualitative results were observed for the mean API score in the years 2000 (see Figures 1 and 2 in the Online Supplement). The qualitative patterns are in general similar for both survey outcomes. Our results suggest that that all methods perform well when the data is MCAR. The three model-based estimators, namely RAKEXZ, PRED1 and PRED2 all perform well and show robustness to the missing data mechanisms, as evident by the relatively flat RMSEs and EBs for all other missing data mechanisms. In these simulations, we also see that the methods involving PS, namely PS and NRPS perform relatively well. However, we see methods CC and NR give very high RMSEs, especially for $\bar{Y}_2$ and empirical bias for the first two missing data mechanisms, and CC and NR still performing poorly for the fourth missing data mechanism.

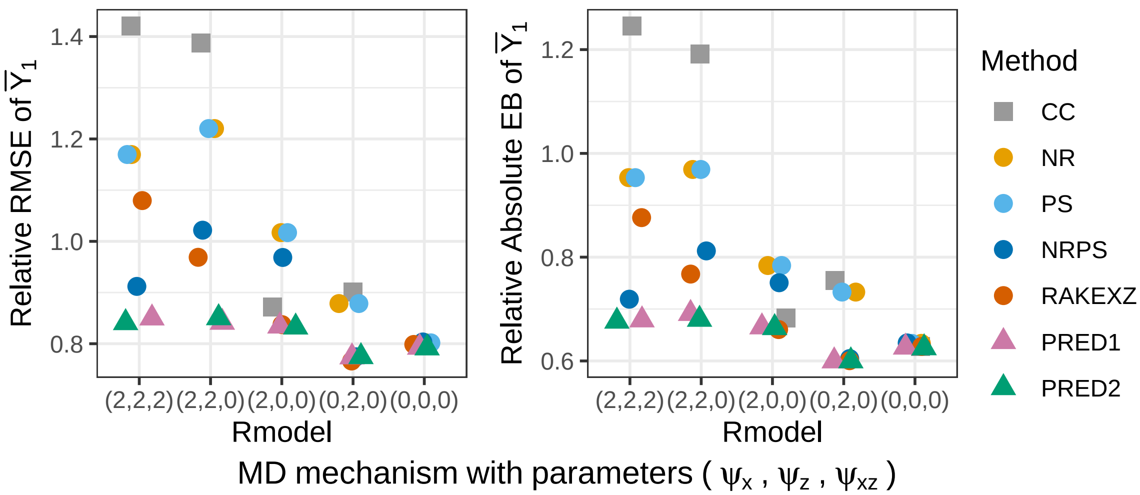


Figure 6: Comparison of Point Estimates of six different estimators for the population mean API score in 1999 (Y1).

In terms of the interval estimates displayed in Figure 6, the three model-based methods perform well in the sense of yielding tight confidence intervals that achieve nominal coverage when the missing data structure conforms with the method of choice. While these findings agree in general, with our simulation results based on a binary outcome, they are more pronounced here. The two model based methods give conservative intervals and achieve nominal coverage throughout. However, this comes at the cost of wide confidence intervals. We observe similar qualitative patterns for both outcomes.

The qualitative patterns are in general similar for both survey outcomes. Our results suggest that that all methods perform well when the data is MCAR. The three model-based estimators all perform well and show robustness to the missing data mechanisms, as evident by the relatively flat RMSEs and EBs for all other missing data mechanisms. In these simulations, we also see that the methods involving PS, namely PS and NRPS perform relatively well. However, we see methods CC and NR give very high RMSEs, especially for Y_2 and empirical bias for the first two missing data mechanisms, and CC and NR still performing poorly for the fourth missing data mechanism.

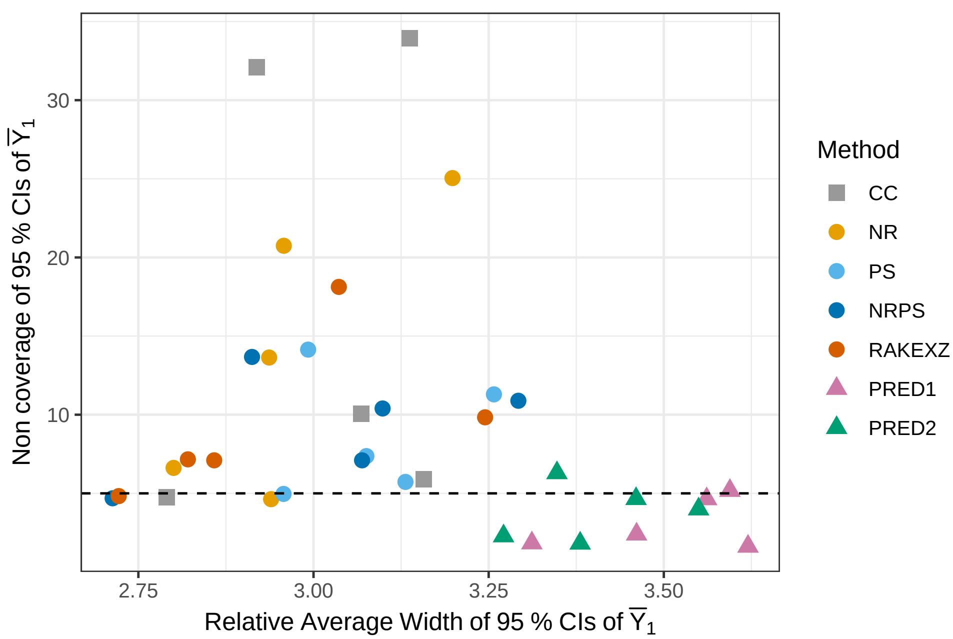


Figure 7: Non-coverage vs relative average width of resulting 95% CI of the population mean API score in 1999 (Y1).

5 Discussion

We describe likelihood-based inference for survey nonresponse when post-stratification variables are observed for survey nonrespondents but not nonrespondents, and marginal distributions of these variables are available from auxiliary data. Models assume that missingness does not depend on the survey variable subject to nonresponse, but are MNAR when missingness depends on the post-stratification variables. By formally modeling the joint distribution of X and Z, the auxiliary information provides us with the data to identify MNAR models, weakening assumptions about the mechanism. A novel feature of the paper is to describe how post-stratification information from external sources can be formally incorporated into the likelihood function. Thus, we are not aware of the basic missingness assumption of Eq. (2) and the likelihood function of Eq. (3) having been described in previous literature. The model-based estimates considered here are maximum likelihood, with standard errors estimated using bootstrap replicates. For small samples where the asymptotic properties of ML do not apply, an attractive alternative approach is to add prior distributions for the parameters and base inferences on Bayesian posterior distributions.

Advantages of this modeling approach are that (a) the model assumptions clarify conditions under which particular estimates are asymptotically optimal; (b) unsaturated models allow for situations where the data do not support saturated models for the joint distribution of (Z, X and R) or Y given Z and X; and (c) the approach avoids arbitrary choices of distance functions required for methods that modify the survey weights. There has been recent interest in likelihood-based with auxiliary information. Chatterjee et al. (2016) and Chen et al. (2015) developed methodology for regression models. Chatterjee et al. (2016) also relaxed the simple random sampling assumption by considering more general sampling designs such as two-phase sampling. These and other work discussed in the introduction do not consider non-ignorable nonresponse models.

We focused here on simple random sampling designs and categorical covariates and post-stratifiers. Stratified random sampling can be accommodated by including stratum indicators as X variables in the model, and cluster and multistage sampling by hierarchical models that include random effects to model clustering. These extensions, and models that include continuous variables within X and Z, are topics for future research.

References

Bates, D. (2005). Fitting linear mixed models in R. *R news*, 5(1):27–30.

Chambers, R. L., Steel, D. G., Wang, S., and Welsh, A. (2012). *Maximum likelihood estimation for sample surveys*. CRC Press.

Chatterjee, N., Chen, Y.-H., Maas, P., and Carroll, R. J. (2016). Constrained maximum likelihood estimation for model calibration using summary-level information from external big data sources. *Journal of the American Statistical Association*, 111(513):107–117.

Chen, A., Owen, A. B., and Shi, M. (2015). Data enriched linear regression. *Electronic Journal of Statistics*, 9(1):1078–1112.

Cochran, W. (2007). *Sampling techniques*. Wiley-India.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Deville, J. and Sarndal, C. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, pages 376–382.

Deville, J., Sarndal, C., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, pages 1013–1020.

Gelman, A. and Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*.

Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., Zheng, T., Dorie, V., and Su, M. Y.-S. (2018). Package 'arm'.

Holt, D. and Smith, T. (1979). Post stratification. *Journal of the Royal Statistical Society. Series A (General)*, pages 33–46.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting methods. *Journal of Official Statistics -Stockkolm-*, 19(2):81–98.

Kim, J. S. and Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational researcher*, 34(8):3–13.

Kish, L. (1965). Survey sampling.

Kott, P. and Chang, T. (2010). Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse. *Journal of the American Statistical Association*, 105(491):1265–1275.

Kott, P. S. and Liao, D. (2017). Calibration weighting for nonresponse that is not missing at random: Allowing more calibration than response-model variables. *Journal of Survey Statistics and Methodology*, 5(2):159–174.

Kott, P. S. and Liao, D. (2018). Calibration weighting for nonresponse with proxy frame variables (so that unit nonresponse can be not missing at random). *Journal of Official Statistics (JOS)*, 34(1).

Laird, N. and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974.

Little, R. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association*, pages 1001–1012.

Little, R. (2004). To model or not to model? Competing modes of inference for finite population sampling. *Journal of the American Statistical Association*, 99(466):546–556.

Little, R. and Wu, M. (1991). Models for contingency tables with known margins when target and sampled populations differ. *Journal of the American Statistical Association*, pages 87–95.

Little, R. J. and Rubin, D. B. (2019). *Statistical analysis with missing data*, volume 793. John Wiley & Sons.

Little, R. J., Rubin, D. B., and Zangeneh, S. Z. (2017). Conditions for ignoring the missing-data mechanism in likelihood inferences for parameter Subsets. *Journal of the American Statistical Association*, 112(517):314–320.

Lumley, T. (2009). SURVEY: analysis of complex survey samples. R package version 3.11-2.

Lumley, T. (2010). *Complex surveys: a guide to analysis using R*, volume 565. Wiley.

Rubin, D. (1976). Inference and missing data. *Biometrika*, 63(3):581.

Särndal, C., Swensson, B., and Wretman, J. (2003). *Model assisted survey sampling*. Springer Verlag.

Si, Y., Trangucci, R., Gabry, J. S., and Gelman, A. (2017). Bayesian hierarchical weighting adjustment and survey inference. *arXiv preprint arXiv:1707.08220*.

Si, Y. and Zhou, P. (2019). Bayes-raking: Bayesian finite population inference with known margins. *arXiv preprint arXiv:1901.02117*.

Valliant, R., Dorfman, A., and Royall, R. (2000). *Finite population sampling and inference: a prediction approach*. Wiley, New York.

Table 4: Comparison of $10,000 \times$ relative RMSE of estimators in simulations (n=10000).

$(\psi_X, \psi_Z, \psi_{XZ}, \psi_Y)$	$(\beta_X, \beta_Z, \beta_{XZ})$	CC	NR	PS.Z	NRPS	RAKEXZ	PRED1	PRED2
(2,2,2,2)	(2,2,2)	26	42	47	75	87	83	84
(2,2,2,0)	(2,2,2)	64	87	96	131	145	144	145
(2,2,0,2)	(2,2,2)	82	73	58	51	79	79	79
(2,2,0,0)	(2,2,2)	56	32	39	42	42	42	42
(2,0,0,0)	(2,2,2)	174	183	185	195	199	198	200
(0,2,0,0)	(2,2,2)	97	48	33	5	1	2	2
(0,0,0,0)	(2,2,2)	35	15	13	1	1	1	1
(2,2,2,2)	(2,2,0)	21	1	34	3	1	1	1
(2,2,2,0)	(2,2,0)	50	74	0	1	1	1	1
(2,2,0,2)	(2,2,0)	1	1	1	1	1	1	1
(2,2,0,0)	(2,2,0)	22	41	41	76	96	89	91
(2,0,0,0)	(2,2,0)	41	64	65	103	123	123	123
(0,2,0,0)	(2,2,0)	58	67	29	36	74	74	74
(0,0,0,0)	(2,2,0)	44	15	35	38	38	38	38
(2,2,2,2)	(2,0,0)	158	166	166	177	184	184	183
(2,2,2,0)	(2,0,0)	85	38	30	3	2	1	1
(2,2,0,2)	(2,0,0)	47	22	23	4	1	1	1
(2,2,0,0)	(2,0,0)	29	1	52	7	1	1	1
(2,0,0,0)	(2,0,0)	51	92	0	0	0	0	0
(0,2,0,0)	(2,0,0)	1	1	1	1	1	1	1
(0,0,0,0)	(2,0,0)	41	1	83	1	1	1	1
(2,2,2,2)	(0,2,0)	26	1	42	1	1	1	1
(2,2,2,0)	(0,2,0)	129	1	108	1	1	1	1
(2,2,0,2)	(0,2,0)	20	1	0	0	0	0	0
(2,2,0,0)	(0,2,0)	1	1	1	1	1	1	1
(2,0,0,0)	(0,2,0)	70	116	1	9	1	1	1
(0,2,0,0)	(0,2,0)	28	46	1	5	1	1	1
(0,0,0,0)	(0,2,0)	15	1	1	11	1	1	1
(2,2,2,2)	(0,0,0)	257	213	0	0	0	0	0
(2,2,2,0)	(0,0,0)	1	1	1	1	1	1	1
(2,2,0,2)	(0,0,0)	1	1	1	1	1	1	1
(2,2,0,0)	(0,0,0)	1	1	1	1	1	1	1
(2,0,0,0)	(0,0,0)	1	1	1	1	1	1	1
(0,2,0,0)	(0,0,0)	1	1	0	0	0	0	0
(0,0,0,0)	(0,0,0)	1	1	1	1	1	1	1

Table 5: Comparison of 100 × relative absolute empirical bias of estimators in simulations (n=10000).

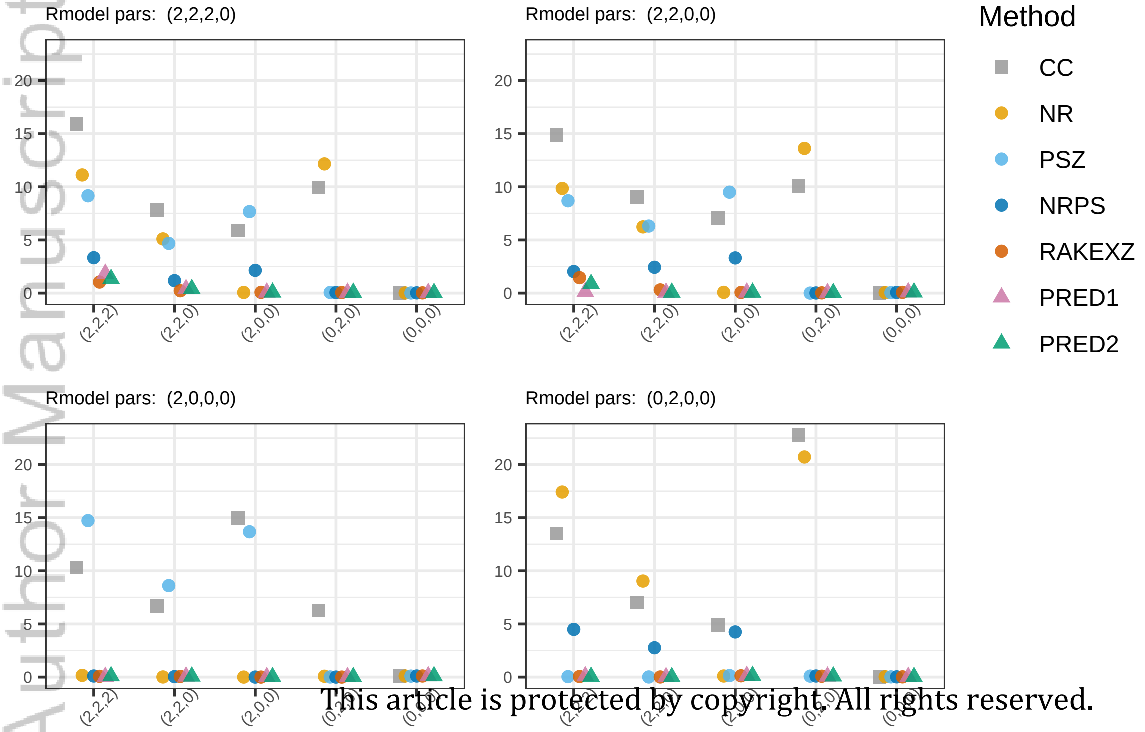
$(\psi_X, \psi_Z, \psi_{XZ}, \psi_Y)$	$(\beta_X, \beta_Z, \beta_{XZ})$	CC	NR	PS.Z	NRPS	RAKEXZ	PRED1	PRED2
(2,2,2,2)	(2,2,2)	814	1043	1109	1399	1506	1477	1483
(2,2,2,0)	(2,2,2)	1061	1243	1309	1529	1605	1603	1605
(2,2,0,2)	(2,2,2)	1194	1124	1002	935	1169	1168	1169
(2,2,0,0)	(2,2,2)	1055	799	884	914	915	914	915
(2,0,0,0)	(2,2,2)	1669	1707	1720	1765	1783	1777	1789
(0,2,0,0)	(2,2,2)	1592	1112	916	333	104	183	134
(0,0,0,0)	(2,2,2)	784	510	468	117	22	38	39
(2,2,2,2)	(2,2,0)	590	6	767	214	7	6	6
(2,2,2,0)	(2,2,0)	994	1216	6	5	5	4	4
(2,2,0,2)	(2,2,0)	1	1	1	2	2	2	2
(2,2,0,0)	(2,2,0)	743	1026	1030	1414	1583	1528	1541
(2,0,0,0)	(2,2,0)	844	1060	1068	1350	1478	1477	1477
(0,2,0,0)	(2,2,0)	996	1080	705	782	1132	1132	1131
(0,0,0,0)	(2,2,0)	939	532	836	871	874	874	874
(2,2,2,2)	(2,0,0)	1587	1627	1628	1682	1712	1713	1710
(2,2,2,0)	(2,0,0)	1489	985	869	203	145	12	87
(2,2,0,2)	(2,0,0)	906	623	631	243	30	5	5
(2,2,0,0)	(2,0,0)	706	7	950	331	6	6	5
(2,0,0,0)	(2,0,0)	1006	1362	0	1	2	2	1
(0,2,0,0)	(2,0,0)	3	4	4	6	7	8	7
(0,0,0,0)	(2,0,0)	1031	16	1473	10	7	3	9
(2,2,2,2)	(0,2,0)	667	1	861	4	6	6	6
(2,2,2,0)	(0,2,0)	1498	0	1368	0	0	1	1
(2,2,0,2)	(0,2,0)	627	8	1	0	0	1	1
(2,2,0,0)	(0,2,0)	8	10	9	10	10	9	9
(2,0,0,0)	(0,2,0)	1350	1742	4	449	5	9	4
(0,2,0,0)	(0,2,0)	700	904	1	276	1	0	0
(0,0,0,0)	(0,2,0)	491	10	14	425	12	12	12
(2,2,2,2)	(0,0,0)	2278	2072	9	10	8	6	7
(2,2,2,0)	(0,0,0)	3	3	2	2	2	3	2
(2,2,0,2)	(0,0,0)	13	12	10	9	8	18	7
(2,2,0,0)	(0,0,0)	13	10	12	9	8	9	9
(2,0,0,0)	(0,0,0)	10	7	9	6	7	6	7
(0,2,0,0)	(0,0,0)	4	3	6	6	6	4	4
(0,0,0,0)	(0,0,0)	3	3	3	3	3	3	3

Table 6: Comparison of non-coverage of 95% interval estimates in simulations (n=10000).

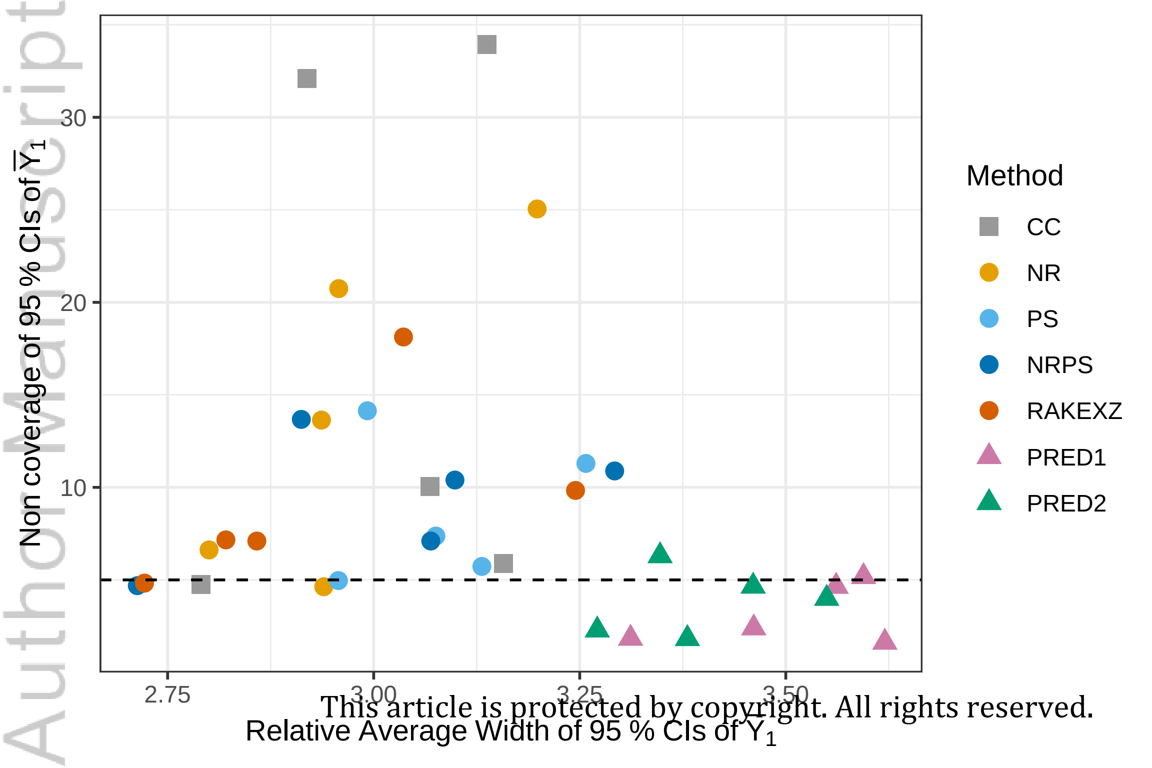
$(\psi_X, \psi_Z, \psi_{XZ}, \psi_Y)$	$(\beta_X, \beta_Z, \beta_{XZ})$	CC	NR	PS.Z	NRPS	RAKEXZ	PRED1	PRED2
(2,2,2,2)	(2,2,2)	100	100	100	100	100	100	100
(2,2,2,0)	(2,2,2)	100	100	100	100	100	100	100
(2,2,0,2)	(2,2,2)	100	100	100	100	100	100	100
(2,2,0,0)	(2,2,2)	100	100	100	100	100	100	100
(2,0,0,0)	(2,2,2)	100	100	100	100	100	100	100
(0,2,0,0)	(2,2,2)	100	100	100	64	14	30	23
(0,0,0,0)	(2,2,2)	100	100	98	20	7	10	12
(2,2,2,2)	(2,2,0)	100	6	100	62	14	14	16
(2,2,2,0)	(2,2,0)	100	100	9	8	10	4	4
(2,2,0,2)	(2,2,0)	8	8	9	10	9	13	14
(2,2,0,0)	(2,2,0)	100	100	100	100	100	100	100
(2,0,0,0)	(2,2,0)	100	100	100	100	100	100	100
(0,2,0,0)	(2,2,0)	100	100	100	100	100	100	100
(0,0,0,0)	(2,2,0)	100	99	100	100	100	100	100
(2,2,2,2)	(2,0,0)	100	100	100	100	100	100	100
(2,2,2,0)	(2,0,0)	100	100	100	34	21	13	17
(2,2,0,2)	(2,0,0)	100	100	100	62	10	12	13
(2,2,0,0)	(2,0,0)	100	4	100	94	8	9	9
(2,0,0,0)	(2,0,0)	100	100	6	4	6	1	2
(0,2,0,0)	(2,0,0)	6	6	6	7	7	13	18
(0,0,0,0)	(2,0,0)	100	4	100	4	7	7	7
(2,2,2,2)	(0,2,0)	100	4	100	7	8	10	10
(2,2,2,0)	(0,2,0)	100	8	100	14	16	16	18
(2,2,0,2)	(0,2,0)	100	6	5	5	6	1	1
(2,2,0,0)	(0,2,0)	6	6	6	6	6	10	14
(2,0,0,0)	(0,2,0)	100	100	6	88	6	6	7
(0,2,0,0)	(0,2,0)	100	100	5	72	6	5	6
(0,0,0,0)	(0,2,0)	99	6	7	98	13	18	18
(2,2,2,2)	(0,0,0)	100	100	4	4	4	1	1
(2,2,2,0)	(0,0,0)	5	6	6	8	6	10	12
(2,2,0,2)	(0,0,0)	8	6	8	11	12	10	11
(2,2,0,0)	(0,0,0)	4	4	4	6	6	9	9
(2,0,0,0)	(0,0,0)	8	4	6	12	16	20	20
(0,2,0,0)	(0,0,0)	7	7	5	5	5	1	1
(0,0,0,0)	(0,0,0)	8	8	7	7	7	19	20

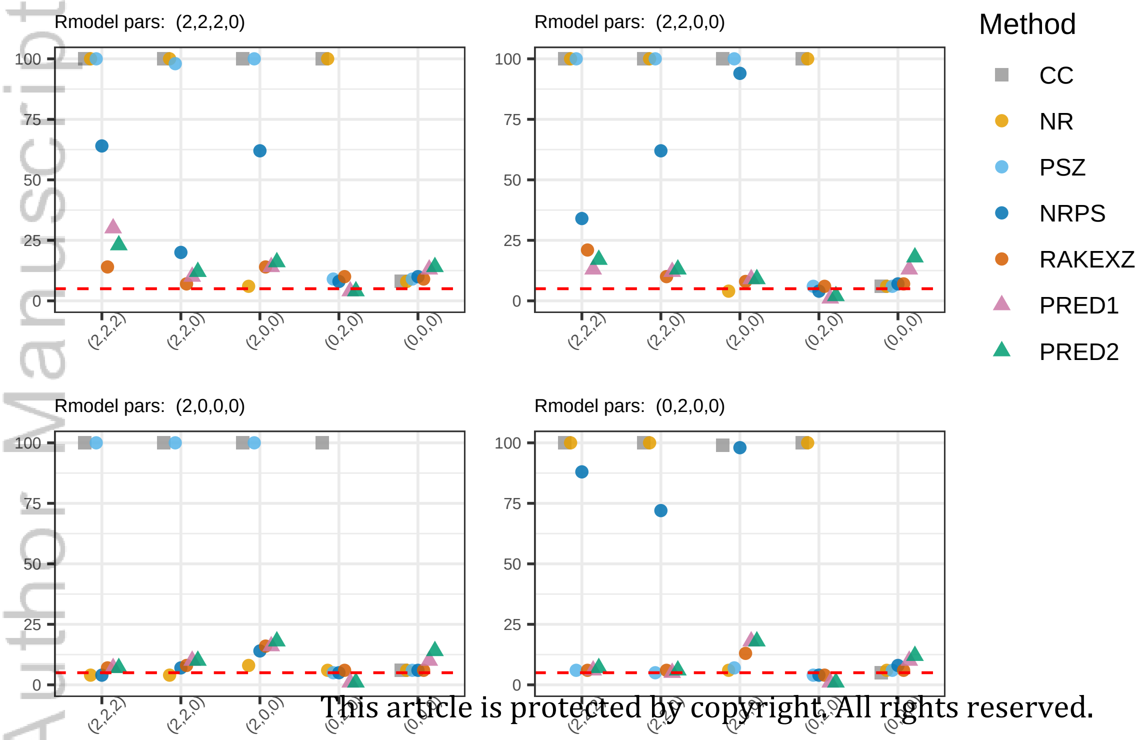
Table 7: Comparison of 100 × relative average width of 95% interval estimates in simulations (n=10000).

$(\psi_X, \psi_Z, \psi_{XZ}, \psi_Y)$	$(\beta_X, \beta_Z, \beta_{XZ})$	CC	NR	PS.Z	NRPS	RAKEXZ	PRED1	PRED2
(2,2,2,2)	(2,2,2)	227	225	204	180	181	182	172
(2,2,2,0)	(2,2,2)	258	259	246	225	224	214	210
(2,2,0,2)	(2,2,2)	260	260	263	202	201	198	191
(2,2,0,0)	(2,2,2)	249	254	157	165	159	230	227
(2,0,0,0)	(2,2,2)	266	269	269	271	273	249	236
(0,2,0,0)	(2,2,2)	250	249	228	205	208	224	196
(0,0,0,0)	(2,2,2)	247	251	243	228	227	214	208
(2,2,2,2)	(2,2,0)	246	254	243	201	203	198	192
(2,2,2,0)	(2,2,0)	251	252	170	178	172	235	230
(2,2,0,2)	(2,2,0)	244	246	246	248	249	209	199
(2,2,0,0)	(2,2,0)	225	222	206	184	184	176	165
(2,0,0,0)	(2,2,0)	253	254	245	226	224	209	205
(0,2,0,0)	(2,2,0)	255	255	254	198	196	193	187
(0,0,0,0)	(2,2,0)	245	251	155	167	157	229	226
(2,2,2,2)	(2,0,0)	261	264	263	266	269	243	228
(2,2,2,0)	(2,0,0)	245	244	226	208	210	198	185
(2,2,0,2)	(2,0,0)	241	246	239	229	227	211	204
(2,2,0,0)	(2,0,0)	240	250	233	197	201	198	190
(2,0,0,0)	(2,0,0)	246	246	167	179	170	235	231
(0,2,0,0)	(2,0,0)	239	242	242	243	246	207	195
(0,0,0,0)	(2,0,0)	240	241	222	210	209	225	203
(2,2,2,2)	(0,2,0)	239	254	229	227	226	215	210
(2,2,2,0)	(0,2,0)	231	258	230	205	205	203	197
(2,2,0,2)	(0,2,0)	243	254	160	175	166	230	226
(2,2,0,0)	(0,2,0)	236	246	237	246	246	205	195
(2,0,0,0)	(0,2,0)	248	251	232	214	215	235	214
(0,2,0,0)	(0,2,0)	246	244	254	239	234	229	223
(0,0,0,0)	(0,2,0)	250	247	253	201	201	195	188
(2,2,2,2)	(0,0,0)	245	248	179	187	179	244	240
(2,2,2,0)	(0,0,0)	243	244	254	255	254	222	212
(2,2,0,2)	(0,0,0)	233	233	218	199	196	267	189
(2,2,0,0)	(0,0,0)	239	239	230	215	213	188	185
(2,0,0,0)	(0,0,0)	238	238	234	187	187	168	165
(0,2,0,0)	(0,0,0)	241	241	159	167	159	220	218
(0,0,0,0)	(0,0,0)	233	233	233	233	233	173	168

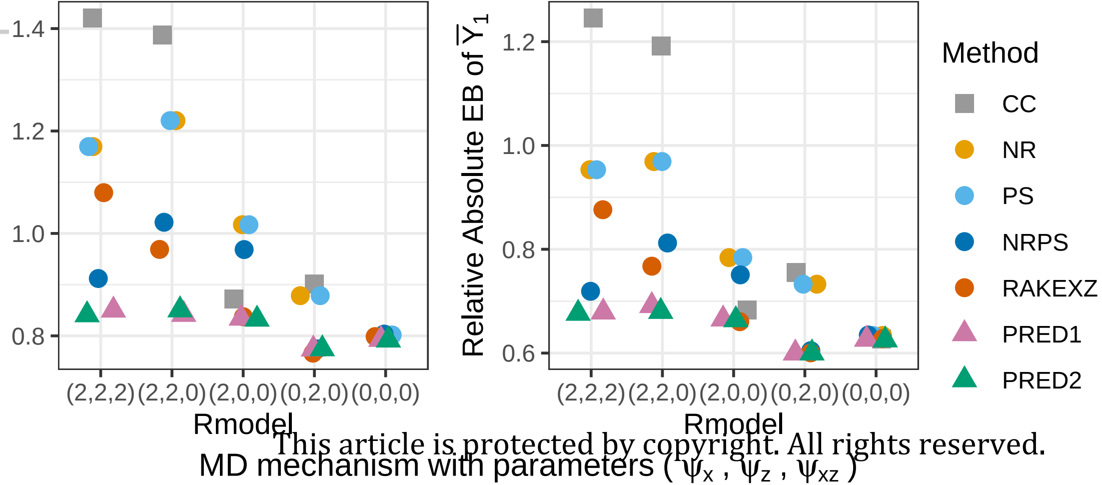


Rmodel pars: (2,2,2,0)
Rmodel pars: (2,2,0,0)
Rmodel pars: (2,0,0,0)
Rmodel pars: (0,2,0,0)
0
5
10
15
20
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Method
CC
NR
PSZ
NRPS
RAKEXZ
PRED1
PRED2

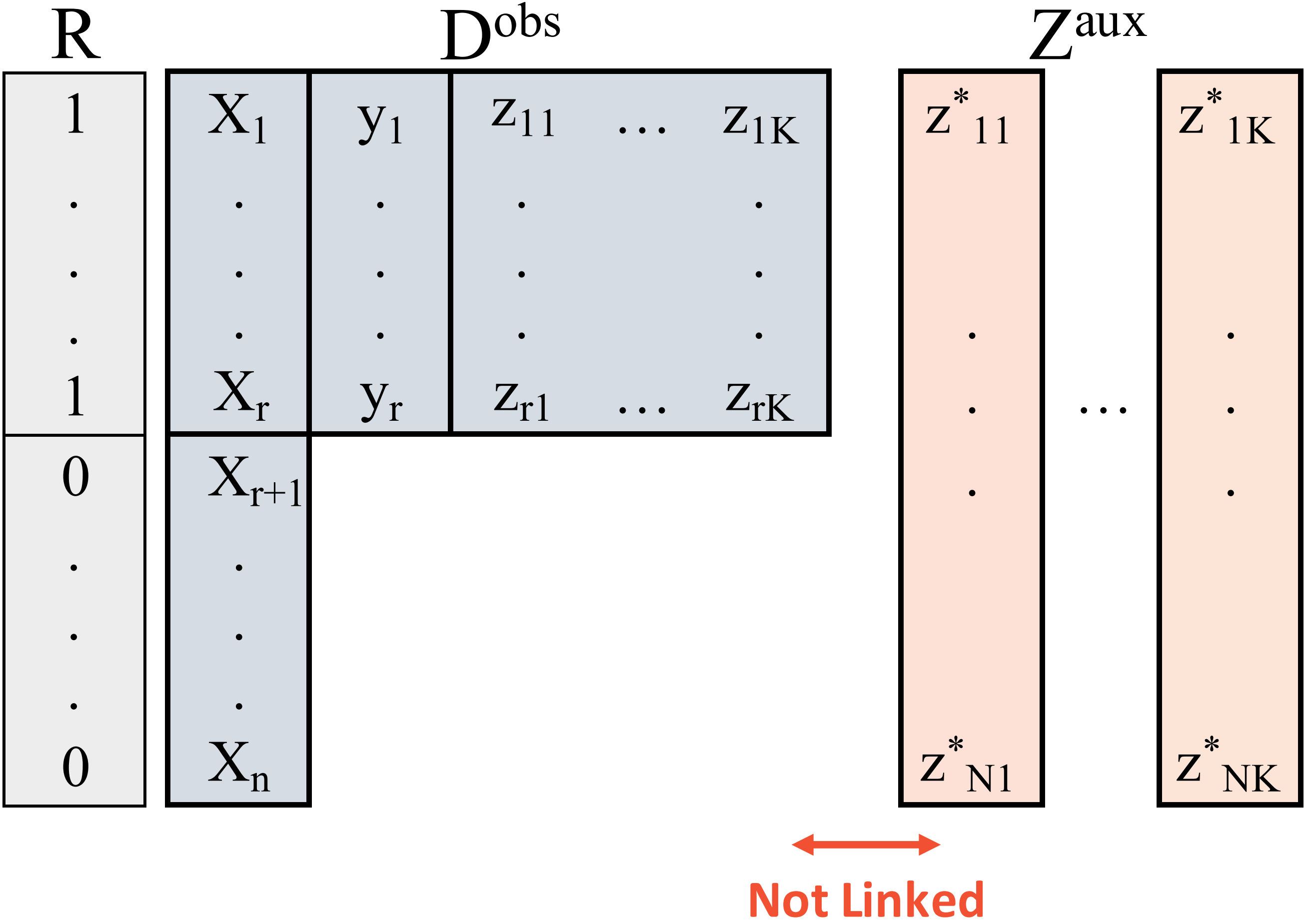

Method
CC
NR
PS
NRPS
RAKEXZ
PRED1
PRED2
30
20
10
2.75
3.00
3.25
3.50
Non coverage of 95 % CIs of $\overline{Y}_1$
Relative Average Width of 95 % CIs of $\overline{Y}_1$

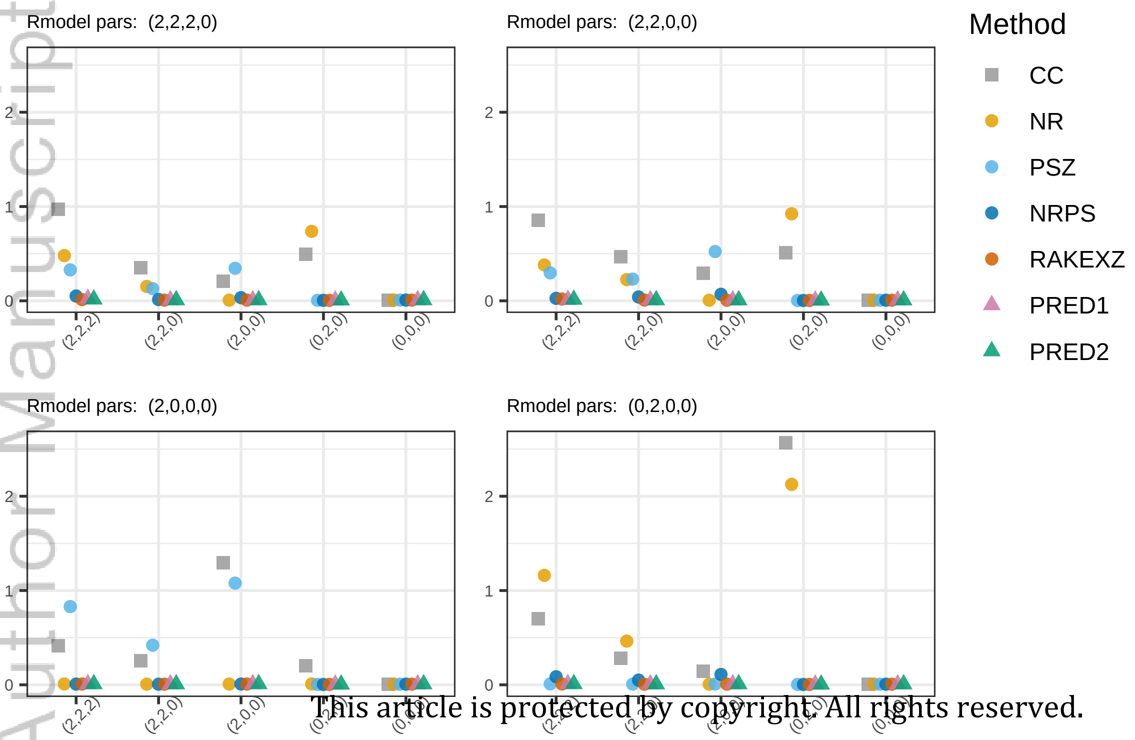


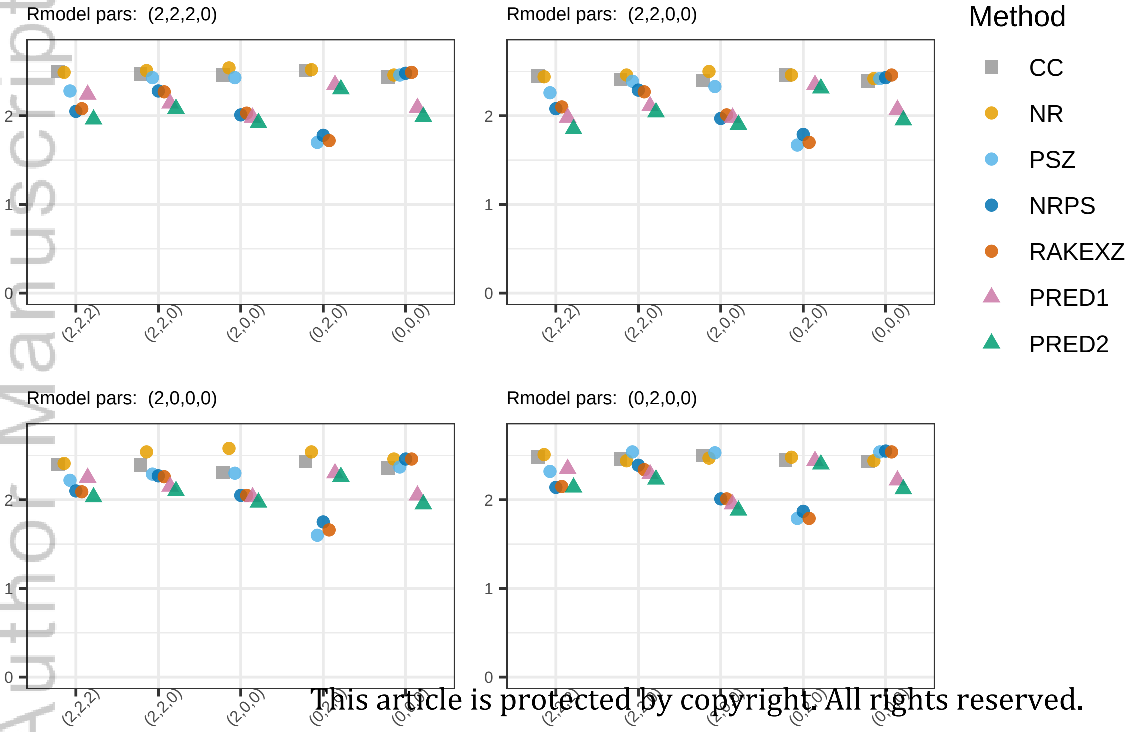
Rmodel pars: (2,2,2,0)
Rmodel pars: (2,2,0,0)
Rmodel pars: (2,0,0,0)
Rmodel pars: (0,2,0,0)
100
75
50
25
0
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Method
CC
NR
PSZ
NRPS
RAKEXZ
PRED1
PRED2



Relative RMSE of $\overline{Y}_1$
1.4
1.2
1.0
0.8
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Rmodel
Relative Absolute EB of $\overline{Y}_1$
1.2
1.0
0.8
0.6
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Rmodel
Method
CC
NR
PS
NRPS
RAKEXZ
PRED1
PRED2
MD mechanism with parameters (ψ_x , ψ_z , ψ_{xz})


R
D^{obs}
Z^{aux}
1
1
0
0
X_1
X_r
X_{r+1}
X_n
y_1
y_r
z_{11} … z_{1K}
z_{r1} … z_{rK}
z^*_{11}
z^*_{N1}
z^*_{1K}
z^*_{NK}
Not Linked


Rmodel pars: (2,2,2,0)
Rmodel pars: (2,2,0,0)
Rmodel pars: (2,0,0,0)
Rmodel pars: (0,2,0,0)
0
1
2
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Method
CC
NR
PSZ
NRPS
RAKEXZ
PRED1
PRED2



Rmodel pars: (2,2,2,0)
Rmodel pars: (2,2,0,0)
Rmodel pars: (2,0,0,0)
Rmodel pars: (0,2,0,0)
0
1
2
(2,2,2)
(2,2,0)
(2,0,0)
(0,2,0)
(0,0,0)
Method
CC
NR
PSZ
NRPS
RAKEXZ
PRED1
PRED2