Statistica Sinica **24** (2014), 871-896
doi:http://dx.doi.org/10.5705/ss.2012.126

BAYESIAN SENSITIVITY ANALYSIS OF STATISTICAL MODELS WITH MISSING DATA

Hongtu Zhu[1], Joseph G. Ibrahim[1] and Niansheng Tang[2]

[1] *University of North Carolina at Chapel Hill and* [2] *Yunnan University*

Abstract: Methods for handling missing data depend strongly on the mechanism that generated the missing values, such as missing completely at random (MCAR) or missing at random (MAR), as well as other distributional and modeling assumptions at various stages. It is well known that the resulting estimates and tests may be sensitive to these assumptions as well as to outlying observations. In this paper, we introduce various perturbations to modeling assumptions and individual observations, and then develop a formal sensitivity analysis to assess these perturbations in the Bayesian analysis of statistical models with missing data. We develop a geometric framework, called the Bayesian perturbation manifold, to characterize the intrinsic structure of these perturbations. We propose several intrinsic influence measures to perform sensitivity analysis and quantify the effect of various perturbations to statistical models. We use the proposed sensitivity analysis procedure to systematically investigate the tenability of the non-ignorable missing at random (MNAR) assumption. Simulation studies are conducted to evaluate our methods, and a dataset is analyzed to illustrate the use of our diagnostic measures.

Key words and phrases: Influence measure, missing data mechanism, perturbation manifold, sensitivity analysis.

1. Introduction

It is common to have missing data in surveys, clinical trials, and longitudinal studies. Various statistical methods have been developed to handle missing data. These methods depend on the missing data mechanism that generates the missing values and other modeling assumptions at various stages, and the resulting estimates and tests can be sensitive to these assumptions. Sensitivity analyses are commonly performed to perturb the model assumptions and/or individual observations to check the sensitivity of a specific influence measure (e.g., a parameter of interest). There is an extensive literature on sensitivity analysis for missing data problems in frequentist analysis (Copas and Eguchi (2005), Little and Rubin (2002), Zhu and Lee (2001), Copas and Li (1997), van Steen, Molenberghs, and Thijs (2001), Troxel (1998), Jansen et al. (2006), Jansen et al. (2003), Verbeke et al. (2001), Troxel, Ma, and Heitjan (2004), Shi, Zhu, and Ibrahim (2009), Hens et al. (2006), Daniels and Hogan (2008)).

The literature on influence measures include Copas and Eguchi (2005), Zhu and Lee (2001), Troxel, Ma, and Heitjan (2004), Copas and Li (1997), van Steen, Molenberghs, and Thijs (2001), Troxel (1998), Jansen et al. (2006), Jansen et al. (2003), Hens et al. (2006), Verbeke et al. (2001), Shi, Zhu, and Ibrahim (2009), and Daniels and Hogan (2008). For instance, in frequentist analysis, Copas and Eguchi (2005) developed a general formulation for assessing the bias of maximum likelihood estimates in the presence of small model perturbations for missing data problems. The local influence method in Cook (1986) was successfully applied to carry out sensitivity analyses for various statistical models with missing data (van Steen, Molenberghs, and Thijs (2001), Troxel (1998), Jansen et al. (2006), Hens et al. (2006), Jansen et al. (2003), Verbeke et al. (2001)). Shi, Zhu, and Ibrahim (2009) further systematically investigated the local influence methods proposed in Zhu et al. (2007) for GLMs with missing at random (MAR) covariates as well as missing not at random (MNAR) covariates, often referred to as nonignorable missing covariates.

In contrast, in the Bayesian literature, several analogues of Cook's local influence (Cook (1986)) were developed to carry out model assessment by using either the curvature of some influence measures (Millar and Stewart (2007)), Linde (2007), Lavine (1991)) or the Fréchet derivative of the posterior with respect to the prior (Dey, Ghosh, and Lou (1996), Gustafson (1996a,b), Berger (1994)). Daniels and Hogan (2008) examined several global and local sensitivity methods in the Bayesian analysis of pattern mixture models (Little (1994), Andridge and Little (2011)). Recently, Zhu, Ibrahim, and Tang (2011) developed a general framework of Bayesian influence analysis for assessing various perturbation schemes to the data, the prior and the sampling distribution for a class of statistical models without missing data.

The aim of this paper is to develop a formal Bayesian sensitivity analysis framework in statistical models with missing data. We introduce various perturbations to the modeling of the missing data mechanism, individual observations, and the prior. We develop a geometric framework, called the Bayesian perturbation manifold, to characterize the intrinsic structure of these perturbations. We examine several influence measures for sensitivity analysis and for quantifying the effect of various perturbations to statistical models with missing data.

We develop a Bayesian perturbation manifold for a large class of statistical models with missing data; examine three Bayesian influence measures including the ϕ-divergence, the posterior mean distance, and the Bayes factor; focus on assessing the missing data mechanism, while simultaneously perturbing other distributional assumptions, the prior, and individual observations.

To motivate our methodology, we consider data on 1,116 female sex workers in Philippine cities from a study of the relationship between Acquired Immune

Deficiency Syndrome (AIDS) and the use of condoms (Morisky et al. (1998)), which is discussed in more detail in Section 3. The data contains items about knowledge of AIDS, attitudes toward AIDS, belief, and self efficiency of condom use. Nine variables in the original data set (items 33, 32, 31, 43, 72, 74, 27h, 27e, and 27i in the questionnaire) were taken as responses. The primary interest here was to find how the threat of AIDS is associated with aggressiveness of the sex worker and the fear of contracting AIDS. The responses and covariates are missing at least once for 361 workers (32.35%). In Section 3, we carry out a Bayesian analysis of a structural equations model with both missing covariates and responses to analyze this data set, and present a formal Bayesian sensitivity analysis.

The rest of this paper is organized as follows. In Section 2, we construct a Bayesian perturbation manifold to characterize various perturbations to statistical models with missing data and derive its associated geometric quantities. We propose global and local influence measures to quantify the effects of perturbing the missing data mechanism, while simultaneously perturbing the data, the prior, and other modeling assumptions on posterior quantities of interest. In Section 3, we present simulation studies and a data analysis to illustrate the importance of the proposed method in assessing the missing data mechanism and other potential misspecifications.

2. Bayesian Sensitivity Analysis

2.1. Statistical models with missing data

Let $\mathbf{z}_{obs} = (\mathbf{z}_{1,o}, \ldots, \mathbf{z}_{n,o})$ and $\mathbf{z}_{mis} = (\mathbf{z}_{1,m}, \ldots, \mathbf{z}_{n,m})$ be the observed and missing data, respectively, and $\mathbf{z}_{com} = (\mathbf{z}_{1,c}, \ldots, \mathbf{z}_{n,c}) = (\mathbf{z}_{mis}, \mathbf{z}_{obs})$ be the complete data, where $\mathbf{z}_{i,c} = (\mathbf{z}_{i,o}, \mathbf{z}_{i,m})$ for $i = 1, \ldots, n$. In applications, the dimensions of $\mathbf{z}_{i,c}$, $\mathbf{z}_{i,o}$ and $\mathbf{z}_{i,m}$ may be different across i. For instance, the number of observations may vary across clusters for clustered data.

For missing data problems, we consider a statistical model $p(\mathbf{z}_{com} \mid \boldsymbol{\theta})$ for the complete data such that $p(\mathbf{z}_{com} \mid \boldsymbol{\theta})$ is the product of a model for the observed data $p(\mathbf{z}_{obs} \mid \boldsymbol{\theta})$ and a model for the missing data given the observed data $p(\mathbf{z}_{mis} \mid \mathbf{z}_{obs}, \boldsymbol{\theta})$. This class of statistical models for missing data includes generalized linear models with missing covariates and/or responses, generalized linear mixed models, nonlinear models, parametric survival models, and many others. To carry out Bayesian inference, we usually use Markov chain Monte Carlo (MCMC) methods to simulate samples from the posterior distribution of the observed data, given by

$$p(\boldsymbol{\theta} \mid \mathbf{z}_{obs}) \propto p(\mathbf{z}_{obs} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) \propto \int p(\mathbf{z}_{com} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\mathbf{z}_{mis}. \tag{2.1}$$

Example 1 (Missing Covariate Data)**.** Consider n independent observations $\mathbf{z}_{com} = \{\mathbf{z}_{i,c} = (\mathbf{x}_i, \mathbf{c}_i, \mathbf{r}_i, y_i),\ i = 1, \ldots, n\}$, where y_i is the response variable, $\mathbf{x}_i$ is a $p_1 \times 1$ vector of completely observed covariates, and $\mathbf{c}_i = (\mathbf{c}_{i,m}, \mathbf{c}_{i,o})$ is a $p_2 \times 1$ vector of partially observed covariates, where $\mathbf{c}_{i,m}$ and $\mathbf{c}_{i,o}$ denote the missing and observed components of $\mathbf{c}_i$, respectively. Let $\mathbf{r}_i$ be a $p_2 \times 1$ vector whose j^{th} component, r_{ij}, equals 1 if the j^{th} component of $\mathbf{c}_i$, denoted by c_{ij}, is observed, and 0 if c_{ij} is missing. We assume that $p(\mathbf{x}_i, \mathbf{c}_i, \mathbf{r}_i, y_i|\boldsymbol{\theta}) = p(y_i|\mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\theta})p(\mathbf{x}_i, \mathbf{c}_i|\boldsymbol{\theta})\ p(\mathbf{r}_i|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the vector of unknown parameters. In this case, $\mathbf{z}_{i,m} = \mathbf{c}_{i,m}$ and $\mathbf{z}_{i,o} = (\mathbf{x}_i, \mathbf{c}_{i,o}, \mathbf{r}_i, y_i)$ for all i.

We assume the generalized linear model (GLM)

$$p(y_i|\mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\beta}, \tau) = \exp[a_i^{-1}(\tau)\{y_i\eta_i(\boldsymbol{\beta}) - b_1(\eta_i(\boldsymbol{\beta}))\} + b_2(y_i, \tau)] \tag{2.2}$$

for $i = 1, \ldots, n$, where $a_i(\cdot)$, $b_1(\cdot)$, and $b_2(\cdot, \cdot)$ are known functions, $\eta_i = \eta(\mu_i)$ and $\mu_i = g((\mathbf{x}_i', \mathbf{c}_i')\boldsymbol{\beta})$, in which $g(\cdot)$ is a known link function, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and $p = p_1 + p_2$. We assume that

$$p(\mathbf{x}_i, \mathbf{c}_i|\boldsymbol{\alpha}) = p(c_{ip_2}|c_{i,p_2-1}, \ldots, c_{i1}, \mathbf{x}_i, \boldsymbol{\alpha}_{2p_2}) \times \cdots \times p(c_{i1}|\mathbf{x}_i, \boldsymbol{\alpha}_{21})p(\mathbf{x}_i|\boldsymbol{\alpha}_1). \tag{2.3}$$

Similarly, we model the missing-data mechanism as

$$p(\mathbf{r}_i|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}) = p(r_{ip_2}|r_{i,p_2-1}, \ldots, r_{i1}, y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_{p_2}) \times \cdots \times p(r_{i1}|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_1). \tag{2.4}$$

To carry out a full Bayesian analysis, we need to specify a prior for $\boldsymbol{\theta}$. We can take independent priors for the components of $\boldsymbol{\theta}$ such that $p(\boldsymbol{\theta}) = p(\tau)p(\boldsymbol{\beta})p(\boldsymbol{\xi})p(\boldsymbol{\alpha})$. For τ and $\boldsymbol{\beta}$, we can take $\tau \sim \text{gamma}(\alpha_0/2, \lambda_0/2)$ and $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, where α_0, λ_0, $\boldsymbol{\mu}_0$ ($p \times 1$), and $\boldsymbol{\Sigma}_0$ ($p \times p$ positive definite matrix) are pre-specified hyperparameters. If $\lambda_{min}(\boldsymbol{\Sigma}_0)$ converges to ∞, then $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ tends to an improper prior. In contrast, if $\lambda_{max}(\boldsymbol{\Sigma}_0)$ is very small, then $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ tends to a strongly informative prior. For $\boldsymbol{\alpha}$, we can take a prior of the form $p(\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}_1)p(\boldsymbol{\alpha}_{21}) \cdots p(\boldsymbol{\alpha}_{2p_2})$. To make valid Bayesian inferences about $\boldsymbol{\beta}$, we need an appropriate prior $p(\boldsymbol{\theta})$ and the correct specification of the sampling distributions (2.2)−(2.4), therefore, it is crucial to assess the robustness of both the prior and the sampling distribution with respect to posterior estimate of $\boldsymbol{\beta}$. Particularly, there is a growing awareness of the need for formal methods for investigating the sensitivity of inferences to the missing-data mechanism (Copas and Eguchi (2005), Little and Rubin (2002), Zhu and Lee (2001), Troxel, Ma, and Heitjan (2004), Copas and Li (1997), van Steen, Molenberghs, and Thijs (2001), Troxel (1998), Jansen et al. (2006), Jansen et al. (2003), Verbeke et al. (2001), Shi, Zhu, and Ibrahim (2009), Daniels and Hogan (2008), Ibrahim et al. (2005)).

Example 2 (Missing Response Data)**.** We consider n independent observations $\mathbf{z}_{com} = \{\mathbf{z}_{i,c} = (\mathbf{x}_i, \mathbf{r}_i, \mathbf{y}_i),\ i = 1, \ldots, n\}$, where $\mathbf{y}_i = (\mathbf{y}_{i,m}, \mathbf{y}_{i,o})$ is a $p_y \times 1$ response vector, in which $\mathbf{y}_{i,m}$ and $\mathbf{y}_{i,o}$ denote the missing and observed components of $\mathbf{y}_i$, respectively, and $\mathbf{x}_i$ is a $p_x \times 1$ vector of completely observed covariates. Moreover, $\mathbf{r}_i$ is a $p_y \times 1$ vector, whose j^{th} component, r_{ij}, equals 1 if the j^{th} component of $\mathbf{y}_i$, denoted by y_{ij}, is observed, and 0 if y_{ij} is missing. It is common to model the joint distribution of $(\mathbf{y}_i, \mathbf{r}_i)$ given $\mathbf{x}_i$ as

$$p(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}) = p(\mathbf{y}_{i,m}, \mathbf{y}_{i,o} | \mathbf{x}_i, \boldsymbol{\theta}_I) p(\mathbf{r}_i | \mathbf{y}_{i,m}, \mathbf{y}_{i,o}, \mathbf{x}_i, \boldsymbol{\theta}_N), \tag{2.5}$$

where $\boldsymbol{\theta}_I$ is the vector of parameters of interest and $\boldsymbol{\theta}_N$ includes all parameters in the missing data mechanism $p(\mathbf{r}_i | \mathbf{y}_{i,m}, \mathbf{y}_{i,o}, \mathbf{x}_i, \boldsymbol{\theta}_N)$. In this case, $\mathbf{z}_{i,m} = \mathbf{y}_{i,m}$ and $\mathbf{z}_{i,o} = (\mathbf{x}_i, \mathbf{y}_{i,o}, \mathbf{r}_i)$ for all i.

To carry out a full Bayesian analysis, we need to specify a prior for $\boldsymbol{\theta}$ and the missing data mechanism. For instance, a well-known ignorability condition (Rubin, 1976) is commonly used to carry out posterior inference on $\boldsymbol{\theta}_I$ without specifying the missing data mechanism. Specifically, a missing data mechanism is said to be ignorable if it is MAR, (2.5) is true and $p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}_I)p(\boldsymbol{\theta}_N)$. Although it is computationally easier to assume the ignorability condition, most missing data mechanisms are nonignorable (Daniels and Hogan (2008)). An alternative method for nonignorable missing data is to use the extrapolation factorization

$$p(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}) = p(\mathbf{y}_{i,m} | \mathbf{y}_{i,o}, \mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\theta}_N) p(\mathbf{y}_{i,o}, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}_I). \tag{2.6}$$

In this case, $p(\mathbf{y}_{i,m} | \mathbf{y}_{i,o}, \mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\theta}_N)$ is an extrapolation model and cannot be identifiable by the observed data, while $p(\mathbf{y}_{i,o}, \mathbf{r}_i | \mathbf{x}_i, \boldsymbol{\theta}_I)$ is an observed data model. Here, the components in $\boldsymbol{\theta}_N$ are called sensitivity parameters (Daniels and Hogan (2008)).

2.2. Bayesian perturbation manifold

We introduce a perturbation vector $\boldsymbol{\omega} = \boldsymbol{\omega}(\mathbf{z}_{com}, \boldsymbol{\theta})$ in a set Ω to perturb the complete-data model $p(\mathbf{z}_{com}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\mathbf{z}_{com} \mid \boldsymbol{\theta})$. To ensure that the perturbation $\boldsymbol{\omega}$ is meaningful and sensible, we require the following. (1) $p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega})$ is the probability density of $(\mathbf{z}_{com}, \boldsymbol{\theta})$ for the perturbed model as $\boldsymbol{\omega}$ varies in a set Ω; (2) There is an $\boldsymbol{\omega}^0 \in \Omega$ such that $p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}^0) = p(\mathbf{z}_{com}, \boldsymbol{\theta})$ and $p(\mathbf{z}_{obs}, \boldsymbol{\theta} \mid \boldsymbol{\omega}^0) = \int p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}^0) d\mathbf{z}_{mis} = p(\mathbf{z}_{obs}, \boldsymbol{\theta})$ for all $(\mathbf{z}, \boldsymbol{\theta})$. The $\boldsymbol{\omega}^0$ can be regarded as the 'central point' of Ω representing no perturbation. See Gustafson (2006) and Daniels and Hogan (2008) for general discussions of model expansion from a Bayesian viewpoint.

Example 1 (Continued)**.** We are interested in perturbing the missing-data mechanism $p(\mathbf{r}_i|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi})$ in (2.4). For instance, when (2.4) is assumed to be MAR, we can consider a general perturbation scheme

$$\begin{aligned} &p(\mathbf{r}_i|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}, \boldsymbol{\omega}) \\ &\quad = p(r_{ip_2}|r_{i,p_2-1}, \ldots, r_{i1}, y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_{p_2}, \boldsymbol{\omega}) \cdots p(r_{i1}|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_1, \boldsymbol{\omega}), \end{aligned} \tag{2.7}$$

where $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_m)^T$ is an $m \times 1$ vector. The perturbation (2.7) is commonly used to perturb the given GLM with MAR covariates in the direction of MNAR (Shi, Zhu, and Ibrahim (2009), Verbeke et al. (2001)). We can also consider the individual-specific infinitesimal perturbation (Verbeke et al. (2001), Hens et al. (2006), Jansen et al. (2006), Jansen et al. (2003))

$$\begin{aligned} &p(\mathbf{r}_i|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}, \boldsymbol{\omega}_i) \\ &\quad = p(r_{ip_2}|r_{i,p_2-1}, \ldots, r_{i1}, y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_{p_2}, \boldsymbol{\omega}_i) \cdots p(r_{i1}|y_i, \mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\xi}_1, \boldsymbol{\omega}_i). \end{aligned} \tag{2.8}$$

Large effects of $\boldsymbol{\omega}_i$ in (2.8) can provide insight into which cases have large influence. Influence measures developed for the perturbation (2.8) are closely related to Bayesian case influence measures, such as the Conditional Predictive Ordinate (CPO) (Geisser (1993), Gelfand, Dey, and Chang (1992)).

We develop a geometric framework, called a Bayesian perturbation manifold, to delineate the effect of introducing each perturbation $\boldsymbol{\omega}$ in Ω. Under some mild conditions, $\mathcal{M} = \{p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}) : \boldsymbol{\omega} \in \Omega\}$ is a Riemannian Hilbert manifold (Lang (1995)). On $\mathcal{M}$, we consider a smooth curve $C(t)$ given by

$$\begin{aligned} C(t) = \Big\{ &p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) : [-\epsilon, \epsilon] \to \mathcal{M}, C(0) = p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}), \text{ and} \\ &\int \dot{\ell}(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t))^2 p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) d\mathbf{z}_{com} d\boldsymbol{\theta} < \infty \Big\}, \end{aligned} \tag{2.9}$$

in which $\dot{\ell}(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) = d \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t))/dt$ is called the tangent (or derivative) vector. The tangent vectors for all possible curves of the form $C(t)$ form the tangent space of $\mathcal{M}$ at $\boldsymbol{\omega}$, denoted by $T_{\boldsymbol{\omega}}\mathcal{M}$. The *inner product* of any two tangent vectors $\boldsymbol{v}_1(\boldsymbol{\omega})$ and $\boldsymbol{v}_2(\boldsymbol{\omega})$ in $T_{\boldsymbol{\omega}}\mathcal{M}$ is given by

$$< \boldsymbol{v}_1, \boldsymbol{v}_2 > (\boldsymbol{\omega}) = \int \{\boldsymbol{v}_1(\boldsymbol{\omega}) \boldsymbol{v}_2(\boldsymbol{\omega})\} p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}) d\mathbf{z}_{com} d\boldsymbol{\theta}. \tag{2.10}$$

It can be shown that the length of the curve $C(t)$ from t_1 to t_2 is

$$S_C(\boldsymbol{\omega}(t_1), \boldsymbol{\omega}(t_2)) = \int_{t_1}^{t_2} \sqrt{< \dot{\ell}(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)), \dot{\ell}(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) >}\, dt. \tag{2.11}$$

We consider the concept of a geodesic as a direct extension of the straight line in Euclidean space on $\mathcal{M}$. For a real function $f(\boldsymbol{\omega})$ defined on $\mathcal{M}$, we take

$df[\boldsymbol{v}](\boldsymbol{\omega}) = \lim_{t\to 0} t^{-1}(f[p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t))] - f[p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0))])$ as the directional derivative of f at the perturbation distribution $p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega})$ in the direction of $\boldsymbol{v}(\boldsymbol{\omega}) \in T_{\boldsymbol{\omega}}\mathcal{M}$. For any two smooth vector fields $\boldsymbol{u}(\boldsymbol{\omega})$ and $\boldsymbol{v}(\boldsymbol{\omega})$ in $T_{\boldsymbol{\omega}}\mathcal{M}$, we define the directional derivative $d\boldsymbol{u}[\boldsymbol{v}](\boldsymbol{\omega}) = \lim_{t\to 0} t^{-1}\{\boldsymbol{u}(\boldsymbol{\omega}(t)) - \boldsymbol{u}(\boldsymbol{\omega}(0))\}$ of a vector field $\boldsymbol{u}(\boldsymbol{\omega})$, called the *connection*, at the perturbation distribution in the direction of $\boldsymbol{v}(\boldsymbol{\omega})$. The popular Levi-Civita connection, denoted by $\nabla_{\boldsymbol{v}}\boldsymbol{u}(\boldsymbol{\omega})$, is

$$d\boldsymbol{u}[\boldsymbol{v}](\boldsymbol{\omega}) - 0.5\Big\{\boldsymbol{u}(\boldsymbol{\omega})\boldsymbol{v}(\boldsymbol{\omega})p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \omega) - \int \boldsymbol{u}(\omega)\boldsymbol{v}(\omega)p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \omega)d\mathbf{z}_{com}d\boldsymbol{\theta}\Big\}. \tag{2.12}$$

A *geodesic* on the manifold $\mathcal{M}$ is a smooth curve $\boldsymbol{\gamma}(t) = p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t))$ on $\mathcal{M}$ with $\dot{\ell}(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) = \boldsymbol{v}(\boldsymbol{\omega}(t))$ such that $\nabla_v \boldsymbol{v}(\boldsymbol{\omega}(t)) = 0$. The geodesic is (locally) the shortest path between points on $\mathcal{M}$. Finally, based on these geometric quantities of $\mathcal{M}$, we define $(\mathcal{M}, <\boldsymbol{u}, \boldsymbol{v}>, \nabla_{\boldsymbol{v}}\boldsymbol{u})$ as the *Bayesian perturbation manifold* (BPM) with an inner product $<\boldsymbol{u}, \boldsymbol{v}>$ and the Levi-Civita connection $\nabla_{\boldsymbol{v}}\boldsymbol{u}$.

Compared to existing sensitivity analysis methods, a key advantage of using the BPM is that it provides a framework for quantifying simultaneous perturbations to the prior, the missing data mechanism, other distributional assumptions, and individual observations. Such simultaneous perturbations can be important, since it can allow one to disentangle the uncertainty about unverifiable missing data mechanism assumptions from the misspecification of the prior and other distributional assumptions, as well as detect the presence of outliers. According to the best of our knowledge, no methods currently exist for handling such simultaneous perturbations.

Example 1 (Continued)**.** Consider the simultaneous perturbation model

$$p(\boldsymbol{\theta}|\boldsymbol{\omega}_\theta)\prod_{i=1}^{n}\{p(y_i|\mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\beta}, \tau, \boldsymbol{\omega}_{iy})p(\mathbf{x}_i, \mathbf{c}_i|\boldsymbol{\alpha}, \boldsymbol{\omega}_{ic})p(\mathbf{r}_i|\mathbf{x}_i, \mathbf{c}_i, y_i, \boldsymbol{\xi}, \boldsymbol{\omega}_{ir})\}, \tag{2.13}$$

where $\boldsymbol{\omega}$ includes $\boldsymbol{\omega}_\theta$ and $\boldsymbol{\omega}_i = (\boldsymbol{\omega}_{iy}^T, \boldsymbol{\omega}_{ic}^T, \boldsymbol{\omega}_{ir}^T)$ for all i and all components of $\boldsymbol{\omega}$ are assumed to be independent of $\mathbf{z}_{com}$ and $\boldsymbol{\theta}$. The three terms on the right hand side of (2.13) are assumed to be probability densities and $\boldsymbol{\omega}_\theta$, $\boldsymbol{\omega}_{iy}$, $\boldsymbol{\omega}_{ic}$, and $\boldsymbol{\omega}_{ir}$ for all i have no components in common. In this case, the BPM is given by

$$\mathcal{M} = \Big\{p(\boldsymbol{\theta}|\boldsymbol{\omega}_\theta)\prod_{i=1}^{n}p(\mathbf{x}_i, \mathbf{c}_i, \mathbf{r}_i, y_i|\boldsymbol{\theta}, \boldsymbol{\omega}_i) : (\boldsymbol{\omega}_\theta, \boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_n) \in \Omega\Big\}, \tag{2.14}$$

where $p(\mathbf{x}_i, \mathbf{c}_i, \mathbf{r}_i, y_i|\boldsymbol{\theta}, \boldsymbol{\omega}_i)$ denotes the product of the three terms on the right hand side of (2.13). Consider $\boldsymbol{\omega}(t)$ as a vector of smooth functions of t and $\boldsymbol{v}_h = d\boldsymbol{\omega}(0)/dt$. It follows from the arguments in Zhu, Ibrahim, and Tang (2011) that

$T_{\boldsymbol{\omega}}\mathcal{M}$ is spanned by the functions $\partial_{\boldsymbol{\omega}_\theta} \log p(\boldsymbol{\theta} \mid \boldsymbol{\omega}_\theta)$, $\partial_{\boldsymbol{\omega}_{iy}} \log p(y_i|\mathbf{x}_i, \mathbf{c}_i, \boldsymbol{\beta}, \tau, \boldsymbol{\omega}_{iy})$, $\partial_{\boldsymbol{\omega}_{ic}} \log p(\mathbf{x}_i, \mathbf{c}_i|\boldsymbol{\alpha}, \boldsymbol{\omega}_{ic})$, and $\partial_{\boldsymbol{\omega}_{ir}} \log p(\mathbf{r}_i|\mathbf{x}_i, \mathbf{c}_i, y_i, \boldsymbol{\xi}, \boldsymbol{\omega}_{ir})$, where $\partial_{\boldsymbol{\omega}} = \partial/\partial_{\boldsymbol{\omega}}$. By using the chain rule, we have

$$\boldsymbol{v}(\boldsymbol{\omega}(0)) = \boldsymbol{v}_h \partial_{\boldsymbol{\omega}} \ell(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0)) \text{ and } < \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}(0)) = \boldsymbol{v}_h^T G(\boldsymbol{\omega}(0)) \boldsymbol{v}_h, \quad (2.15)$$

where

$$G(\boldsymbol{\omega}(0)) = \int [\partial_{\boldsymbol{\omega}} \ell(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0))]^{\otimes 2} p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}) d\boldsymbol{z}_{com} d\boldsymbol{\theta} \quad (2.16)$$

is the Bayesian Fisher information matrix with respect to $\boldsymbol{\omega}$ (Daniels and Hogan (2008)). Geometrically, $\boldsymbol{\omega}_\theta$, $\boldsymbol{\omega}_{iy}$, $\boldsymbol{\omega}_{ic}$, and $\boldsymbol{\omega}_{ir}$ are orthogonal to each other with respect to the inner product defined in (2.10) (Cox and Reid (1987)). Similar to Zhu et al. (2007), one can easily separate out the influence of the missing data mechanism from that of the data, the prior, and other distributional assumptions.

Example 2 (Continued)**.** The sensitivity parameters in (2.6) can be either fixed at a range of values, or assigned an appropriate distribution (Daniels and Hogan (2008)). Here, we take the first approach and treat $\boldsymbol{\theta}_N$ or its parametrization as a perturbation vector. Generally, we consider a simultaneous perturbation model

$$p(\boldsymbol{\theta}|\boldsymbol{\omega}_\theta) \prod_{i=1}^n \{p(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, \mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\omega}_N) p(\mathbf{y}_{i,o}, \mathbf{r}_i|\mathbf{x}_i, \boldsymbol{\theta}_I, \boldsymbol{\omega}_I)\}, \quad (2.17)$$

where $\boldsymbol{\omega}$ includes $\boldsymbol{\omega}_\theta$, $\boldsymbol{\omega}_N$, and $\boldsymbol{\omega}_I$, which represent the perturbation vectors to the prior, the extrapolation model, and the observed data model, respectively. For simplicity, we assume that $\boldsymbol{\omega}_\theta$, $\boldsymbol{\omega}_N$, and $\boldsymbol{\omega}_I$ do not share any common components and are independent of $\mathbf{z}_{com}$ and finite dimensional parameters. Moreover, it is assumed that $p(\boldsymbol{\theta}|\boldsymbol{\omega}_\theta)$, $p(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, \mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\omega}_N)$, and $p(\mathbf{y}_{i,o}, \mathbf{r}_i|\mathbf{x}_i, \boldsymbol{\theta}_I, \boldsymbol{\omega}_I)$ are probability densities for all i. Generally, it is possible that $\boldsymbol{\omega}_N$ and $\boldsymbol{\omega}_I$ may depend on $\mathbf{z}_{com}$ and vary across i.

Consider $\boldsymbol{\omega}(t)$ as a vector of smooth functions of t and $\boldsymbol{v}_h = d\boldsymbol{\omega}(0)/dt$. In this case, $T_{\boldsymbol{\omega}}\mathcal{M}$ is spanned by $\partial_{\boldsymbol{\omega}_\theta} \log p(\boldsymbol{\theta} \mid \boldsymbol{\omega}_\theta)$, $\sum_{i=1}^n \partial_{\boldsymbol{\omega}_N} \log p(\mathbf{y}_{i,m}|\mathbf{y}_{i,o}, \mathbf{r}_i, \mathbf{x}_i, \boldsymbol{\omega}_N)$, and $\sum_{i=1}^n \partial_{\boldsymbol{\omega}_I} \log p(\mathbf{y}_{i,o}, \mathbf{r}_i|\mathbf{x}_i, \boldsymbol{\theta}_I, \boldsymbol{\omega}_I)$. Subsequently, we can calculate the Bayesian Fisher information matrix $G(\boldsymbol{\omega}(0))$ according to (2.16). Geometrically, $\boldsymbol{\omega}_\theta$, $\boldsymbol{\omega}_N$, and $\boldsymbol{\omega}_I$ are also orthogonal to each other with respect to the inner product defined in (2.10) (Cox and Reid (1987)).

2.3. Intrinsic influence measures

As the purpose of a sensitivity analysis is to assess the uncertainty of the parameter of interest as $\boldsymbol{\omega}$ varies in Ω given the data at hand, we take an Intrinsic Influence Measure (IFM) to be a functional of $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega})$ as $\boldsymbol{\omega}$ varies in Ω, where $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega})$ is the perturbed posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{z}_{obs}$ and

$\boldsymbol{\omega}$. Generally, let $\text{IF}(\boldsymbol{\omega}) = \text{IF}(p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}))$ be the intrinsic influence measure. Three common intrinsic influence measures are the ϕ-divergence function, the posterior mean, and the Bayes factor (Kass, Tierney, and Kadane (1989), Kass and Raftery (1995)).

For the missing data mechanism, one can fix an $\boldsymbol{\omega}_0 \in \Omega$ corresponding to MAR and then develop a Relative Intrinsic Influence Measure (RIFM) as a functional of $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega})$ and $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}_0)$,

$$\text{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \text{RI}(p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}), p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}_0)). \tag{2.18}$$

For instance, $\text{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}^0)$ can be the total variation distance of $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}^0)$ and $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega})$ (Dey, Ghosh, and Lou (1996)). One can take $\text{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \text{IF}(\boldsymbol{\omega}) - \text{IF}(\boldsymbol{\omega}_0)$ as the difference between IFMs at $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_0$. See more examples in Section 2.4.

We also suggest rescaling $\text{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ by using the minimal geodesic distance between $p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega})$ and $p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}_0)$, $g(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$, on the BPM $\mathcal{M}$. Thus, we define the intrinsic influence measure for comparing $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega})$ to $p(\boldsymbol{\theta} \mid \boldsymbol{z}_{obs}, \boldsymbol{\omega}_0)$ as

$$\text{IGI}_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \frac{\text{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)^2}{g(\boldsymbol{\omega}, \boldsymbol{\omega}_0)^2}. \tag{2.19}$$

The proposed $\text{IGI}_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ can be interpreted as the ratio of the change of the objective function relative to the minimal distance $p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega})$ and $p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}_0)$ on $\mathcal{M}$. In practice, one can identify the most influential $\boldsymbol{\omega}$ in Ω, denoted by $\hat{\boldsymbol{\omega}}_I$, which maximizes $\text{IGI}_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ for all $\boldsymbol{\omega} \in \Omega$.

We consider the local behavior of $\text{RI}(\boldsymbol{\omega}(t), \boldsymbol{\omega}_0)$ as t approaches zero along all possible smooth curves $p(\boldsymbol{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t))$ passing through $\boldsymbol{\omega}(0) = \boldsymbol{\omega}_0$. Since $\text{RI}(\boldsymbol{\omega}(t), \boldsymbol{\omega}_0)$ is a function from R to R, it follows from a Taylor's series expansion that

$$\text{RI}(\boldsymbol{\omega}(t), \boldsymbol{\omega}_0) = \text{RI}(\boldsymbol{\omega}(0), \boldsymbol{\omega}_0) + \partial\text{RI}(\boldsymbol{\omega}(0))t + 0.5\partial^2\text{RI}(\boldsymbol{\omega}(0))t^2 + o(t^2),$$

where $\partial\text{RI}(\boldsymbol{\omega}(0))$ and $\partial^2\text{RI}(\boldsymbol{\omega}(0))$ denote the first- and second order derivatives of $\text{RI}(\boldsymbol{\omega}(t), \boldsymbol{\omega}^0)$ with respect to t evaluated at $t = 0$. We need to distinguish between $\partial\text{RI}(\boldsymbol{\omega}(0)) \neq 0$ for some smooth curves $\boldsymbol{\omega}(t)$ and $\partial\text{RI}(\boldsymbol{\omega}(0)) = 0$ for all smooth curves $\boldsymbol{\omega}(t)$. For the case $\partial\text{RI}(\boldsymbol{\omega}(0)) \neq 0$, $\partial\text{RI}(\boldsymbol{\omega}(0)) = d(\text{RI})[\boldsymbol{v}](\boldsymbol{\omega}(0))$ is the directional derivative of RI in the direction of $\boldsymbol{v} \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$ (Lang (1995)). The first-order local influence measure is defined as

$$\text{FI}_{RI}[\boldsymbol{v}](\boldsymbol{\omega}(0)) = \lim_{t\to 0} \text{IGI}_{RI}(\boldsymbol{\omega}(0), \boldsymbol{\omega}(t)) = \frac{\{d(\text{RI})[\boldsymbol{v}](\boldsymbol{\omega}(0))\}^2}{< \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}(0))}. \tag{2.20}$$

We use the tangent vector $\boldsymbol{v}_{FI,\max}$ in $T_{\boldsymbol{\omega}(0)}\mathcal{M}$ that maximizes $\text{FI}_{RI}[\boldsymbol{v}](\boldsymbol{\omega}(0))$, to carry out a sensitivity analysis.

For the case $\partial \mathrm{RI}(\boldsymbol{\omega}(0)) = 0$, we use $\partial^2 \mathrm{RI}(\boldsymbol{\omega}(0))$ to assess the second-order local influence of $\boldsymbol{\omega}$ to a statistical model (Zhu et al. (2007)). The second-order influence measure in the direction $\boldsymbol{v} \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$ is defined as

$$\mathrm{SI}_{RI}[\boldsymbol{v}](\boldsymbol{\omega}(0)) = \frac{\partial^2 \mathrm{RI}(\boldsymbol{\omega}(0))}{< \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}(0))}. \tag{2.21}$$

Geometrically, $\mathrm{SI}_{RI}[\boldsymbol{v}](\boldsymbol{\omega}(0))$ is invariant to scalar transformations and smooth transformations. To carry out a sensitivity analysis, we use the tangent vector $\boldsymbol{v}_{S,\max}$ in $T_{\boldsymbol{\omega}(0)}\mathcal{M}$ that maximizes $\mathrm{SI}_{RI}[\boldsymbol{v}](\boldsymbol{\omega}(0))$ for all $\boldsymbol{v} \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$.

2.4. Bayesian sensitivity analysis

Our sensitivity analysis consists of four steps.

1. Introduce a Bayesian perturbation manifold based on $p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega})$.
2. Calculate the geometric metric $< \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}_0)$ of the perturbation manifold.
3. Choose an intrinsic influence measure $\mathrm{IF}(\boldsymbol{\omega})$. If $\partial \mathrm{RI}(\boldsymbol{\omega}(0)) \neq \mathbf{0}$, then we calculate $\boldsymbol{v}_{FI,\max}$ to assess local influence of minor perturbations to the model. If $\partial \mathrm{RI}(\boldsymbol{\omega}(0)) = \mathbf{0}$, then we compute $\boldsymbol{v}_{S,\max}$. We inspect $\boldsymbol{v}_{FI,\max}$ (or $\boldsymbol{v}_{S,\max}$) in order to detect the most influential components of $\boldsymbol{\omega}$.
4. For the most influential subcomponents of $\boldsymbol{\omega}$, we calculate $\mathrm{IGI}_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ and $\hat{\boldsymbol{\omega}}_I = \mathrm{argmax}_{\omega \in \Omega} \mathrm{IGI}_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$.

In practice, we iteratively perform the four-step influence analysis as described above. We start with a simultaneous perturbation to $\mathbf{z}_{com}$, $p(\boldsymbol{\theta})$ and $p(\mathbf{z}_{com}|\boldsymbol{\theta})$. We decide on a set of parametric perturbations characterized by a finite dimensional $\boldsymbol{\omega}$ such that the perturbed model is large enough to cover a large class of candidate models for the data set. With parametric perturbations, it is computationally simple to carry out the Bayesian sensitivity analysis, and a perturbation model with a large number of perturbations can approximate most of the interesting perturbation models. We start with a local influence analysis to examine the sensitivity of all components, and then focus on a few influential components using an intrinsic influence analysis. For instance, if a few influential hyperparameters to the prior are identified, one further perturbs their associated prior distributions using the additive ϵ-contamination class and then carries out an intrinsic influence analysis. After combining the information learned from our influence analysis, we might choose a new sampling distribution and/or a new prior. This procedure can be run iteratively until a certain degree of satisfaction is reached.

2.5. Examples of Bayesian influence measures

We focus on assessing the influence of a perturbation scheme $\boldsymbol{\omega}$ to the posterior distribution based on $\phi-$divergence, the posterior mean distance, and the

Bayes factor. The Bayes factor, the ϕ-divergence, and the posterior mean quantify the effects of introducing $\boldsymbol{\omega}$ on the overall assumed model, on the overall posterior distribution, and on the posterior mean of $\boldsymbol{\theta}$, respectively. Since the Bayes factor measures the overall difference between $p(\mathbf{z}_{obs}|\boldsymbol{\omega})$ and $p(\mathbf{z}_{obs}|\boldsymbol{\omega}_0)$, it can be more sensitive to some discrepancies between the assumed model and the observed data. As the ϕ-divergence measures the overall difference between $p(\mathbf{z}_{mis}, \boldsymbol{\theta}|\mathbf{z}_{obs}, \boldsymbol{\omega})$ and $p(\mathbf{z}_{mis}, \boldsymbol{\theta}|\mathbf{z}_{obs}, \boldsymbol{\omega}_0)$, and such a difference may include the mean, median, etc. It can be more sensitive to some changes of the posterior distributions, but the posterior mean distance is more sensitive to a subtle change in the posterior mean.

Example 3 (Bayes factor)**.** The logarithm of the Bayes factor for comparing $\boldsymbol{\omega}$ with $\boldsymbol{\omega}_0$ is

$$\begin{aligned}\mathrm{BF}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) &= \log(p(\mathbf{z}_{obs} \mid \boldsymbol{\omega})) - \log(p(\mathbf{z}_{obs} \mid \boldsymbol{\omega}_0)) \\ &= \log\Big(\int p(\mathbf{z}_{com} \mid \boldsymbol{\theta}, \boldsymbol{\omega}) p(\boldsymbol{\theta} \mid \boldsymbol{\omega}) d\mathbf{z}_{mis} d\boldsymbol{\theta}\Big) - \log\Big(\int p(\mathbf{z}_{com} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\mathbf{z}_{mis} d\boldsymbol{\theta}\Big).\end{aligned}$$

The value of $\mathrm{BF}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ can be regarded as a statistic for testing hypotheses of $\boldsymbol{\omega}$ against $\boldsymbol{\omega}_0$ (Kass and Raftery (1995)). Under some smoothness conditions, $\mathrm{BF}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$ is a continuous map from $\mathcal{M}$ to R.

We set $\mathrm{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \mathrm{BF}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$, where $\boldsymbol{\omega}(t)$ is a smooth curve on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}_0$ and $d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) \mid_{t=0} = \boldsymbol{v}(\boldsymbol{\omega}_0) \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$, where $d_t = d/dt$. It can be shown that

$$\partial \mathrm{RI}(\boldsymbol{\omega}(0)) = E\{d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) \mid \mathbf{z}_{obs}, \boldsymbol{\omega}(t)\} \mid_{t=0},$$

where the conditional expectation is taken with respect to $p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}(t))$. We can use MCMC methods to draw samples $\{(\boldsymbol{\theta}^{(s)}, \mathbf{z}_{mis}^{(s)}) : s = 1, \ldots, S_0\}$ from $p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs})$ and then approximate $\partial \mathrm{RI}(\boldsymbol{\omega}(0))$ by using $S_0^{-1} \sum_{s=1}^{S_0} d_t \log p(\mathbf{z}_{obs}, \mathbf{z}_{mis}^{(s)}, \boldsymbol{\theta}^{(s)} \mid \boldsymbol{\omega}_0)$.

We consider a simultaneous perturbation to both the prior and the sampling distribution. We have

$$\mathrm{FI}_{BF}[\boldsymbol{v}](\boldsymbol{\omega}(0)) = \frac{E\{d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0)) \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0\}^2}{< \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}_0)}.$$

For instance, for the perturbation to the prior given by $p(\boldsymbol{\theta}; t) = p(\boldsymbol{\theta}) + t\{g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})\}$, it can be shown that

$$\mathrm{FI}_{BF}[\boldsymbol{v}](\boldsymbol{\omega}(0)) = \frac{\mathrm{E}\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta}) \mid \mathbf{z}_{obs}\}^2}{\mathrm{var}_P\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta})\}} = \frac{\{p_g(\mathbf{z}_{obs})/p(\mathbf{z}_{obs})\}^2}{\mathrm{var}_P\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta})\}},$$

where $p(\mathbf{z}_{obs}) = \int p(\mathbf{z}_{com}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\mathbf{z}_{mis}d\boldsymbol{\theta}$ and $p_g(\mathbf{z}_{obs}) = \int p(\mathbf{z}_{com}|\boldsymbol{\theta})g(\boldsymbol{\theta})d\mathbf{z}_{mis}d\boldsymbol{\theta}$. Since the ratio of $p_g(\mathbf{z}_{obs})$ to $p(\mathbf{z}_{obs})$ is the Bayes factor in favor of $g(\boldsymbol{\theta})$ against $p(\boldsymbol{\theta})$, the first-order local influence measure is the square of the normalized Bayes factor of $g(\boldsymbol{\theta})$ against $p(\boldsymbol{\theta})$.

Example 4 ($\phi-$divergence)**.** The $\phi-$divergence between two posterior distributions for $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}$ is

$$\Phi_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \int \phi(R(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \boldsymbol{\omega}, \boldsymbol{\omega}_0))p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)d\mathbf{z}_{mis}d\boldsymbol{\theta},$$

where $R(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \boldsymbol{\omega}, \boldsymbol{\omega}_0) = p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega})/p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)$ and $\phi(\cdot)$ is a convex function with $\phi(1) = 0$, such as the Kullback-Leibler divergence or the χ^2-divergence (Kass, Tierney, and Kadane (1989)).

We set $\mathrm{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \Phi_{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$, where $\boldsymbol{\omega}(t)$ is a smooth curve on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}_0$ and $d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) \mid_{t=0} = \boldsymbol{v}(\boldsymbol{\omega}_0) \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$. It can be shown that $\partial\mathrm{RI}(\boldsymbol{\omega}(0)) = 0$ and

$$\partial^2\mathrm{RI}(\boldsymbol{\omega}(0)) = \ddot{\phi}(1)\int [d_t \log p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}(t))]^2 p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)d\mathbf{z}_{mis}d\boldsymbol{\theta} \mid_{t=0},$$

where $\ddot{\phi}(t) = d^2\phi(t)/dt^2$. Here, we need a computational formula. Note that

$$\begin{aligned} & d_t \log p(\mathbf{z}_{mis}, \boldsymbol{\theta}|\mathbf{z}_{obs}, \boldsymbol{\omega}(t)) \\ & = d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0)) - \int [d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(0))] p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)d\mathbf{z}_{mis}. \end{aligned}$$

In practice, we use MCMC methods to draw samples $\{(\boldsymbol{\theta}^{(s)}, \mathbf{z}_{mis}^{(s)}) : s = 1, \ldots, S_0\}$ from $p(\boldsymbol{\theta}, \mathbf{z}_{mis} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)$ and then approximate $\partial^2\mathrm{RI}(\boldsymbol{\omega}(0))$ using

$$\ddot{\phi}(1)S_0^{-1}\sum_{s=1}^{S_0}\Big[d_t \log p(\mathbf{z}_{mis}^{(s)}, \mathbf{z}_{obs}, \boldsymbol{\theta}^{(s)} \mid \boldsymbol{\omega}(0)) - S_0^{-1}\sum_{s'=1}^{S_0} d_t \log p(\mathbf{z}_{mis}^{(s')}, \mathbf{z}_{obs}, \boldsymbol{\theta}^{(s')} \mid \boldsymbol{\omega}(0))\Big]^2.$$

For perturbation schemes to the prior distribution, it can be shown that

$$< \boldsymbol{v}, \boldsymbol{v} > (\boldsymbol{\omega}(0)) = \int [d_t \log p(\boldsymbol{\theta} \mid \boldsymbol{\omega}(0))]^2 p(\boldsymbol{\theta} \mid \boldsymbol{\omega}(0))d\boldsymbol{\theta}$$

and $\partial^2\mathrm{RI}(\boldsymbol{\omega}(0)) = \ddot{\phi}(1)\mathrm{var}[d_t \log p(\boldsymbol{\theta} \mid \boldsymbol{\omega}(0)) \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0]$, which are, respectively, the Fisher information matrices of $\boldsymbol{\omega}(t)$ based on the prior and posterior distributions, where $\mathrm{var}(\cdot \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)$ denotes the posterior variance. For instance, for $p(\boldsymbol{\theta} \mid \boldsymbol{\omega}(\boldsymbol{\theta})) = p(\boldsymbol{\theta}) + t\{g(\boldsymbol{\theta}) - p(\boldsymbol{\theta})\}$, we can show that

$$\mathrm{SI}_{\Phi_{RI}}[\boldsymbol{v}]\{\boldsymbol{\omega}(0)\} = \frac{\ddot{\phi}(1)\mathrm{var}\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta}) \mid \mathbf{z}_{obs}\}}{\mathrm{var}_P\{g(\boldsymbol{\theta})/p(\boldsymbol{\theta})\}},$$

where $\mathrm{var}_P(\cdot)$ denotes the prior variance.

Example 5 (Posterior mean distance)**.** We measure the distance between the posterior means of $h(\boldsymbol{\theta})$ for $\boldsymbol{\omega}_0$ and $\boldsymbol{\omega}$ (Kass, Tierney, and Kadane (1989), Gustafson (1996b)). The posterior mean of $h(\boldsymbol{\theta})$ after introducing $\boldsymbol{\omega}$ is

$$M_h(\boldsymbol{\omega}) = \int h(\boldsymbol{\theta}) p(\mathbf{z}_{mis}, \boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}) d\mathbf{z}_{mis} d\boldsymbol{\theta}.$$

Cook's posterior mean distance for characterizing the influence of $\boldsymbol{\omega}$ is then

$$\mathrm{CM}_h(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \{M_h(\boldsymbol{\omega}) - M_h(\boldsymbol{\omega}_0)\}^T G_h \{M_h(\boldsymbol{\omega}) - M_h(\boldsymbol{\omega}_0)\}, \tag{2.22}$$

where G_h is a positive definite matrix. Henceforth, G_h is the inverse of the posterior covariance matrix of $h(\boldsymbol{\theta})$ for $p(\boldsymbol{\theta} \mid \mathbf{z}_{obs}, \boldsymbol{\omega}_0)$.

We set $\mathrm{RI}(\boldsymbol{\omega}, \boldsymbol{\omega}_0) = \mathrm{CM}_h(\boldsymbol{\omega}, \boldsymbol{\omega}_0)$, where $\boldsymbol{\omega}(t)$ is a smooth curve on $\mathcal{M}$ with $\boldsymbol{\omega}(0) = \boldsymbol{\omega}_0$ and $d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) \mid_{t=0} = \boldsymbol{v}(\boldsymbol{\omega}_0) \in T_{\boldsymbol{\omega}(0)}\mathcal{M}$. It can be shown that $\partial \mathrm{RI}(\boldsymbol{\omega}(0)) = 0$ and $\partial^2 \mathrm{RI}(\boldsymbol{\omega}(0)) = \dot{M}_h(\boldsymbol{v})^T G_h \dot{M}_h(\boldsymbol{v})$, where

$$\dot{M}_h(\boldsymbol{v}) = d_t M_h(\boldsymbol{\omega}(0)) = \mathrm{Cov}\{h(\boldsymbol{\theta}), d_t \log p(\mathbf{z}_{com}, \boldsymbol{\theta} \mid \boldsymbol{\omega}(t)) \mid \mathbf{z}_{obs}, \boldsymbol{\omega}^0\} \mid_{t=0} .$$

We can use MCMC methods to approximate $\dot{M}_h(\boldsymbol{v})$ and G_h.

2.6. A simple theoretical example

We consider a simple example involving missing responses (Daniels and Hogan (2008)). Consider a data set $\mathbf{z}_{com} = ((y_1, r_1), \ldots, (y_n, r_n))^{\mathrm{T}}$, where $r_i = 1$ if y_i is observed and 0 if y_i is missing. We focus on perturbing the missing-data mechanism.

First, we fit a pattern mixture model for (y_i, r_i) such that

$$y_i | r_i = 1 \sim N(\mu_1, \sigma^2), \quad y_i | r_i = 0 \sim N(\mu_0, \sigma^2), \quad r_i \sim Ber(\phi). \tag{2.23}$$

Model (2.23) assumes that the observed and missing responses differ in their mean but share the same variance. Since the observed data do not contain any information on μ_0, we assume $\mu_0 = \mu_1 + \omega_\mu$.

Here ω_μ can be regarded as a perturbation and $\boldsymbol{\theta} = (\mu_1, \sigma^2, \phi)$. The complete -data likelihood function is

$$p(\mathbf{z}_{com}, \boldsymbol{\theta} | \omega_\mu) = \phi^{\sum_i r_i} (1-\phi)^{n - \sum_i r_i} \Big\{ \prod_{i, r_i = 0} p(y_i | \mu_1 + \omega_\mu, \sigma^2) \Big\} \Big\{ \prod_{i, r_i = 1} p(y_i | \mu_1, \sigma^2) \Big\},$$

where $p(y|\mu, \sigma^2)$ denotes the normal density function with mean μ and variance σ^2. Regardless of the prior for $\boldsymbol{\theta}$, it can be shown that $G(\omega_\mu) = \sum_{i=1}^n (1 - r_i)/\sigma^2$, which is independent of ω_μ, and thus $\mathcal{M}$ is flat and $g(\omega_{\mu,1}, \omega_{\mu,2}) = c|\omega_{\mu,1} -$

$\omega_{\mu,2}|$, where c is a scalar (Zhu et al. (2007)). Moreover, since the observed-data likelihood function $\int p(\mathbf{z}_{com}, \boldsymbol{\theta}|\omega_\mu)\ d\mathbf{z}_{mis}$ does not depend on ω_μ, all IFs and IFMs based on $p(\boldsymbol{\theta}|\mathbf{z}_{obs},\ \omega_\mu)$ are zero. This indicates that varying ω_μ does not influence the posterior inferences on $\boldsymbol{\theta}$ given $\mathbf{z}_{obs}$. Instead, if we consider the posterior mean $\mu_1 + (1-\phi)\omega_\mu$, the marginal mean of y_i, as the influence measure, then we have

$$\mathrm{IF}(\omega_\mu) = E[\mu_1 + (1-\phi)\omega_\mu|\mathbf{z}_{obs}] = E[\mu_1|\mathbf{z}_{obs}] + \{1 - E[\phi|\mathbf{z}_{obs}]\}\omega_\mu,$$
$$\mathrm{IGI}_{RI}(\omega_{\mu,1}, \omega_{\mu,2}) = \frac{[\mathrm{IF}(\omega_{\mu,1}) - \mathrm{IF}(\omega_{\mu,2})]^2}{g(\omega_{\mu,1}, \omega_{\mu,2})^2} = \frac{\{1 - E[\phi|\mathbf{z}_{obs}]\}^2\sigma^2}{\sum_{i=1}^n (1 - r_i)},$$

where $E[\cdot|\mathbf{z}_{obs}]$ denote the expectation taken with respect to $p(\boldsymbol{\theta}|\mathbf{z}_{obs})$. In this case, $\mathrm{IF}(\omega_\mu)$ does not belong to any of the three Bayesian influence measures considered in Section 2.4, but our invariant influence measure is applicable. Moreover, the constant $\mathrm{IGI}_{RI}(\omega_{\mu,1}, \omega_{\mu,2})$ indicates that any inferences about the measure of y_i are completely driven by the assumptions regarding the size of ω_μ.

Second, we fit a selection model for (y_i, r_i) such that

$$y_i \sim N(\mu_1, \sigma^2), \quad r_i \sim Ber(\phi_i) \ \text{ with } \ \mathrm{logit}(\phi_i) = \xi_1 + \omega_\xi y_i, \tag{2.24}$$

where logit$(\cdot)$ denotes the logistic function. In (2.24), $\omega_\xi = 0$ corresponds to MAR, whereas $\omega_\xi \neq 0$ corresponds to MNAR. In this case, ω_ξ can be regarded as a perturbation and $\boldsymbol{\theta} = (\mu_1, \sigma^2, \xi_1)$. The complete-data likelihood function is

$$p(\mathbf{z}_{com}, \boldsymbol{\theta}|\omega_\xi) = \phi_i^{\sum_i r_i}(1-\phi_i)^{n-\sum_i r_i}\prod_{i=1}^n p(y_i; \mu_1, \sigma^2).$$

If $p(\boldsymbol{\theta})$ is the prior for $\boldsymbol{\theta}$, it can be shown that

$$G(\omega_\xi) = n\int y^2 \frac{\exp(\xi_1 + \omega_\xi y)}{[1 + \exp(\xi_1 + \omega_\xi y)]^2} p(y; \mu_1, \sigma^2)p(\boldsymbol{\theta})dyd\boldsymbol{\theta},$$

which does not have a simple form. Moreover, since the observed-data likelihood function $\int p(\mathbf{z}_{com}, \boldsymbol{\theta}|\omega_\xi)d\mathbf{z}_{mis}$ does depend on ω_ξ, all IFs and IFMs based on $p(\boldsymbol{\theta}|\mathbf{z}_{obs}, \omega_\xi)$ can be numerically calculated according to the formulas given in Sections 2.3−2.4. Generally, in the selection model, varying ω_ξ does not influence the posterior inferences about $\boldsymbol{\theta}$ given $\mathbf{z}_{obs}$.

3. Simulation Study

We consider a two-level model. We assume that the data are obtained from N individuals nested within J groups, with group j containing n_j individuals, where $N = \sum_{j=1}^J n_j$. The level-1 units are the individuals and the level-2 units are the groups. At level-1, for each group j ($j = 1, \ldots, J$), the within-group model is given by

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta}_j + \varepsilon_{ij}, \quad i = 1, \ldots, n_j, \tag{3.1}$$

where y_{ij} is the outcome variable, $\mathbf{x}_{ij}$ is a q-vector with explanatory variables (including a constant), $\boldsymbol{\beta}_j$ is a q-vector of regression coefficients, and ε_{ij} is the residual. At level-2, we further assume $\boldsymbol{\beta}_j$ to be a vector of random regression coefficients,

$$\boldsymbol{\beta}_j = Z_j \boldsymbol{\gamma} + \mathbf{u}_j, \tag{3.2}$$

where Z_j is a $q \times r$ matrix with explanatory variables (including a constant) obtained at the group level, $\boldsymbol{\gamma}$ is an r-vector containing fixed coefficients, and $\mathbf{u}_j$ is a q-vector of residuals. Assume that $\mathbf{u}_j$ is independent of ε_{ij}, $\mathbf{u}_j \sim N_q(0, \Sigma)$, and $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. We assume that the covariates $\mathbf{x}_{ij}$ and Z_j are completely observed for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, but the responses y_{ij} may be missing.

We simulated a data set according to (3.1)−(3.2). We set $J = 100$, $q = 2$, and $r = 3$, and then we chose varying values of n_j in order to create a scenario with different cluster sizes. Specifically, we set $n_1 = \ldots = n_{10} = 3$, $n_{91} = \ldots = n_{100} = 20$, and $n_i \in \{5, 7, 8, 10, 12, 13, 15, 17\}$ for $i = 11, \ldots, 90$. We independently generated all components (except the intercept) of $\mathbf{x}_{ij}$ and Z_j as $U(0, 1)$. We assumed that the y_{ij}'s were MAR with missing data mechanism

$$\Pr(r_{ij} = 1 \mid \mathbf{x}_{ij}, \varphi) = \frac{\exp(\varphi_0 + \varphi_x^T \mathbf{x}_{ij})}{1 + \exp(\varphi_0 + \varphi_x^T \mathbf{x}_{ij})}, \tag{3.3}$$

where $\varphi = (\varphi_0, \varphi_x)$, $r_{ij} = 1$ if y_{ij} is missing and $r_{ij} = 0$ if y_{ij} is observed. We set $\varphi_0 = -2.0$, $\varphi_x = (0.5, 0.5)^T$, $\boldsymbol{\gamma} = (0.8, 0.8, 0.8)^T$, $\Sigma = 0.5 1_2 1_2^T + 0.5 I_2$, and $\sigma_\varepsilon^2 = 1.0$. The missing data fraction of the responses is about 18.4%. To add some outliers, we modified the simulated data set by generating new $\{y_{ij} : j = 1, 99, 100; i = 1, \ldots, n_j\}$ from a $N(\mathbf{x}_{ij}^T Z_j \boldsymbol{\gamma} + \mathbf{x}_{ij}^T \mathbf{u}_j, \sigma_\varepsilon^2)$ distribution with $\mathbf{u}_j \sim N(5.6 1_2, 1.96 I_2 + 0.3\Sigma)$ ($j = 1, 99, 100$).

We fit (3.1)−(3.3) to the simulated data set and used MCMC sampling to carry out the Bayesian influence analysis (Chen, Shao, and Ibrahim (2000)). We took

$$p(\boldsymbol{\gamma}) \overset{D}{=} N(\boldsymbol{\gamma}^0, H_{0\varepsilon}), \quad p(\sigma_\varepsilon^{-2}) \overset{D}{=} \Gamma(\alpha_{0\varepsilon}, \beta_{0\varepsilon}), \quad p(\Sigma) \overset{D}{=} IW_q(\rho_0, R^0),$$

where $\boldsymbol{\gamma}^0, H_{0\varepsilon}, \alpha_{0\varepsilon}, \beta_{0\varepsilon}, R^0$, and ρ_0 are hyperparameters whose values are pre-specified. We assumed that $p(\varphi) \overset{D}{=} N(\varphi^0, H_{0\varphi})$, where φ^0 and $H_{0\varphi}$ are the given hyperparameters. Furthermore, we set $\boldsymbol{\gamma}^0 = (0.8, 0.8, 0.8)^T$, $R^0 = 2I_2 + 2 1_2 1_2^T$, $\varphi^0 = (-2.0, 0.5, 0.5)^T$, $H_{0\varphi} = I_3$, $\alpha_{\varepsilon 0} = 10.0$, $\beta_{\varepsilon 0} = 8.0$, $\rho_0 = 10$, and $H_{0\varepsilon} = \mathrm{diag}(0.2, 0.2, 0.2)$.

We simultaneously perturbed the distributions of $\mathbf{u}_j$ and the prior distributions of $\boldsymbol{\gamma}$, Σ, and σ_ε^2, whose perturbed complete-data joint (unnormalized) log-posterior density is given by

$$\begin{aligned}\ell(\mathbf{z}_{com},\boldsymbol{\theta}\mid\boldsymbol{\omega})=&\sum_{j=1}^{J}\frac{1}{2}\left\{-q\log\left(\frac{2\pi}{\boldsymbol{\omega}_j}\right)-\log\mid\Sigma\mid-\boldsymbol{\omega}_j\mathbf{u}_j^T\Sigma^{-1}\mathbf{u}_j\right\}\\&+\frac{1}{2}\left\{-r\log\left(\frac{2\pi}{\boldsymbol{\omega}_{\boldsymbol{\gamma}}}\right)-\log\mid H_{0\varepsilon}\mid-\boldsymbol{\omega}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}-\boldsymbol{\gamma}^0)^TH_{0\varepsilon}^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}^0)\right\}\\&-(\rho_0-q-1)\log\frac{\mid\Sigma\mid}{2}-\frac{1}{2}\boldsymbol{\omega}_\Sigma\mathrm{tr}(R^0\Sigma^{-1})-\frac{1}{2}q\rho_0\log(2)\\&+\rho_0\log\frac{|\boldsymbol{\omega}_\Sigma R_0|}{2}-\log\Gamma_q\left(\frac{\rho_0}{2}\right)+\log[p(\sigma_\varepsilon^{-2})+\boldsymbol{\omega}_\sigma\{g(\sigma_\varepsilon^{-2})-p(\sigma_\varepsilon^{-2})\}],\end{aligned}$$

where $g(\sigma_\varepsilon^{-2})$ is the density of a Gamma $(\alpha_{0\varepsilon}+3,\beta_{0\varepsilon}+1)$ distribution and $\boldsymbol{\omega}=(\boldsymbol{\omega}_1,\ldots,\boldsymbol{\omega}_J,\boldsymbol{\omega}_{\boldsymbol{\gamma}},\boldsymbol{\omega}_\Sigma,\boldsymbol{\omega}_\sigma)^T$. In this case, $\boldsymbol{\omega}^0=(1,1,\ldots,1,0)'$ represents no perturbation. By differentiating $\ell(\mathbf{z}_{com},\boldsymbol{\theta}\mid\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$, after some calculations, we have

$$G(\boldsymbol{\omega}^0)=\mathrm{diag}\Big[q\frac{I_J}{2},\frac{r}{2},\mathrm{var}_\Sigma\frac{\{\mathrm{tr}(R^0\Sigma^{-1})\}}{4},\mathrm{var}_{\sigma_\varepsilon^2}\Big\{\frac{g(\sigma_\varepsilon^{-2})}{p(\sigma_\varepsilon^{-2})}\Big\}\Big],$$

where var_Σ and $\mathrm{var}_{\sigma_\varepsilon^2}$ denote the variance with respect to the priors of Σ and σ_ε^2, respectively. Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}}=\boldsymbol{\omega}^0+G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega}-\boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\boldsymbol{v}_{F,\max}=\mathrm{argmax}\mathrm{FI}_{BF}[\boldsymbol{v}]\{\tilde{\boldsymbol{\omega}}(0)\}$, $\mathrm{SI}_{\Phi_{IR}}[\mathbf{e}_j]$, and $\mathrm{SI}_{CM_h}[\mathbf{e}_j]$, in which $\phi(\cdot)$ was chosen to be the Kullback-Leibler divergence divergence and $h(\boldsymbol{\theta})=\boldsymbol{\theta}$. Note that the numbers of observations in groups 1, 99, and 100 were, respectively, 3, 20, and 20. Groups 1, 99 and 100 were detected to be influential by all our local influence measures. Selected results for $\mathrm{SI}_{\Phi_{IR}}[\mathbf{e}_j]$ are presented in Figure 1(a).

We used the same setup, except that we employed a perturbed prior distribution for $\boldsymbol{\gamma}$: $p(\boldsymbol{\gamma})\stackrel{D}{=}N(4\boldsymbol{\gamma}^0,H_{0\varepsilon})$, and then applied the same MCMC method, perturbation scheme, and local influence measures. Groups 1, 99, and 100 and the perturbed prior distribution of $\boldsymbol{\gamma}$ were identified to be influential by all our local influence measures. Selected results for $\mathrm{SI}_{\Phi_D}[e_j]$ are presented in Figure 1(b).

Next, we explored the potential deviations of the MAR mechanism in the direction of MNAR. We simulated a data set using the same setup except that the missing data mechanism for y_{ij} was

$$\Pr(r_{ij}=1\mid\mathbf{x}_{ij},y_{ij},\varphi,\varphi_y)=\frac{\exp(\varphi_0+\varphi_x^T\mathbf{x}_{ij}+\varphi_y y_{ij})}{1+\exp(\varphi_0+\varphi_x^Ts\mathbf{x}_{ij}+\varphi_y y_{ij})},\qquad(3.4)$$

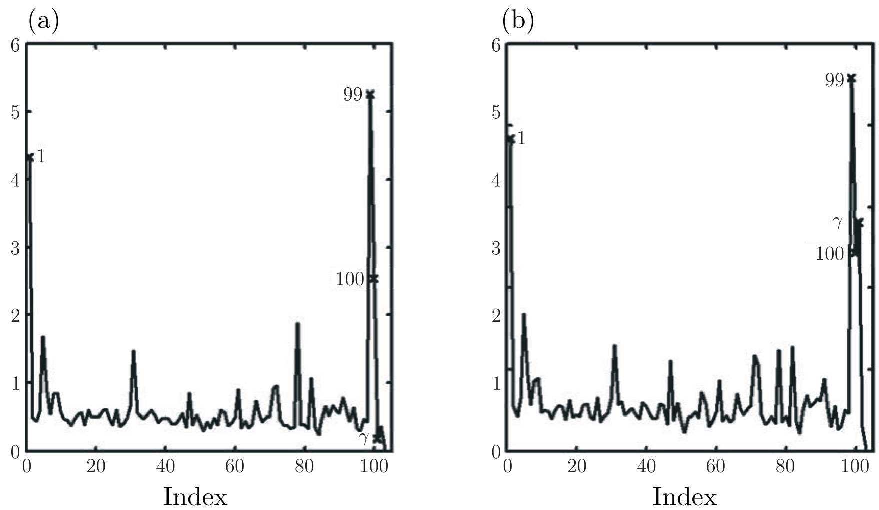

Figure 1. Simulation Study: group index plots of local influence measures for simultaneous perturbation: (a) $\mathrm{SI}_{\Phi_{IR}}[\mathbf{e}_j]$ can detect the three influential groups (1, 99, and 100); (b) $\mathrm{SI}_{\Phi_{IR}}[\mathbf{e}_j]$ can simultaneously detect the three influential groups (1, 99, and 100) and the perturbed prior distribution $p(\boldsymbol{\gamma})$.

with $\varphi_y = 0.5$ to make the missing data fraction approximately equal to 25%.

Similar to sensitivity analysis methods in missing data problems (Molenberghs and Kenward (2007), Little and Rubin (2002)), we fit the model (3.1)−(3.2) and (3.4), with φ_y fixed at a value $\boldsymbol{\omega}_y$, to the simulated data set. When $\boldsymbol{\omega}_y = 0$, the missing data is MAR and hence the missing data mechanism in (3.4) is ignorable. Thus, by varying $\boldsymbol{\omega}_y$ in an interval Ω_1, we can treat $\boldsymbol{\omega}_y$ as a perturbation scheme to the sampling distribution and then calculate the associated local influence measures. Specifically, we chose $\boldsymbol{\omega} = (\boldsymbol{\omega}_y)$ and obtained a curve $C(t)$ on $\mathcal{M}$ at $t = \boldsymbol{\omega}$.

We used the same prior distributions for $\boldsymbol{\gamma}$, φ, σ_ε^2, and Σ as before and used MCMC sampling to carry out the Bayesian influence analysis. We calculated the intrinsic influence measures $\mathrm{IGI}_f(\boldsymbol{\omega}^0, \Omega_1)$ for $\Phi_D(\boldsymbol{\omega})$ and $M_h(\boldsymbol{\theta})$, in which we chose $\phi(\cdot)$ as the Kullback-Leibler divergence, set $h(\boldsymbol{\theta}) = \boldsymbol{\gamma}$ and treated $\boldsymbol{\omega}^0 = 0$ as no perturbation. We set $\Omega_1 = [-2.0, 2.0]$ and approximated Ω_1 via $K_0 = 41$ grid points $\boldsymbol{\omega}_{g,(k)} = -2.0 + 0.1k$ for $k = 0, \ldots, 40$. For a given $\boldsymbol{\omega} \in \Omega_1$, $d(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ was calculated via a composite trapezoidal rule.

Figures 2 (a) and 2 (b) present plots of $\mathrm{IGI}_{IR}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ against $\boldsymbol{\omega} \in \Omega_1$ for $\Phi_{IR}(\boldsymbol{\omega})$ and $M_h(\boldsymbol{\omega})$, respectively. The intrinsic influence measures reach maxima near the true value of $\varphi_y = 0.5$. This indicates that the nonignorable missing data mechanism is tenable for the simulated data. We also followed a standard sensitivity analysis to compute the posterior means and standard deviations of $\boldsymbol{\gamma}$

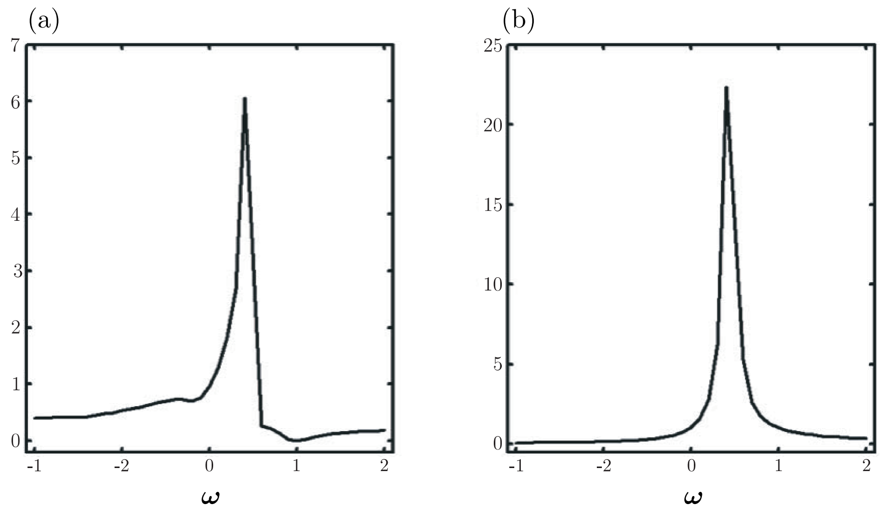

Figure 2. Simulation Study: plots of $\text{IGI}_{IR}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ against $\boldsymbol{\omega} \in \Omega_1$ for (a) $\Phi_{IR}(\boldsymbol{\omega})$ and (b) $M_h(\boldsymbol{\omega})$, in which $h(\boldsymbol{\theta}) = \boldsymbol{\gamma}$.

Table 1. Table 1. Posterior means (PMs) and standard errors (SDs) of $\boldsymbol{\gamma}$ at different values of φ_y

True $\boldsymbol{\gamma}^0 = (0.8, 0.8, 0.8)^T$	γ_1		γ_2		γ_3	
	PM	SD	PM	SD	PM	SD
$\varphi_y = 0.5$	0.831	0.174	0.721	0.251	0.809	0.255
$\varphi_y = 0.3$	0.777	0.170	0.697	0.249	0.786	0.247
$\varphi_y = 0.15$	0.738	0.167	0.661	0.243	0.776	0.249
$\varphi_y = 0.0$	0.697	0.177	0.622	0.247	0.749	0.250

for different φ_y in Table 1. Although we observed that the posterior distribution of $\boldsymbol{\gamma}$ varies with φ_y, it is hard to tell why $\varphi_y = 0.5$ is more meaningful. We also carried out a local influence analysis under this MNAR setting (not presented here) and observed that the proposed local influence method can pick up anomalous features of the data that are not necessarily associated with the missing data mechanism (Jansen et al. (2006)).

4. Real Data Example

We consider a small portion of a data set from a study of the relationship between acquired immune deficiency syndrome (AIDS) and the use of condoms (Morisky et al. (1998)). This subset contains 11 items on such topics as knowledge about AIDS and beliefs, behaviors and attitudes towards condom use collected from 1,116 female sex workers. Nine items, denoted by $\mathbf{y} = (y_1, \ldots, y_9)^T$, were taken as responses. Items (y_1, y_2, y_3) are related to a latent variable, η, which can be roughly interpreted as the threat of AIDS, while items (y_4, y_5, y_6) and (y_7, y_8, y_9) are, respectively, related to latent variables ξ_1 and ξ_2, that can be interpreted as aggressiveness of the sex worker and the worry of contracting AIDS (Lee and Tang (2006)). All response variables were treated as continuous. A continuous item x_1, on the duration as a sex worker, and an ordered categorical item x_2, on the knowledge about AIDS, were taken as covariates. The response variables and covariates are missing at least once for 361 subjects (32.35%) (see Table 4 of Lee and Tang (2006)). The covariate x_2 is completely observed.

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{i9})^T$ and $\boldsymbol{\varpi}_i = (\eta_i, \xi_{i1}, \xi_{i2})^T$. We considered the measurement and structural equations model given by

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\mu} + \Lambda \boldsymbol{\varpi}_i + \boldsymbol{\varepsilon}_i, \\ \eta_i &= b_1 x_{i1} + b_2 x_{i2} + \boldsymbol{\gamma}_1 \xi_{i1} + \boldsymbol{\gamma}_2 \xi_{i2} + \delta_i, \quad \text{for} \quad i = 1, \ldots, 1116, \end{aligned}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_9)^T$ and

$$\Lambda^T = \begin{pmatrix} 1.0^* & \lambda_{21} & \lambda_{31} & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{52} & \lambda_{62} & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 0.0^* & 1.0^* & \lambda_{83} & \lambda_{93} \end{pmatrix},$$

in which 0.0^* and 1.0^* are regarded as fixed values to identify the scale of the latent factor. We took ε_i distributed as $N(0, \Psi)$, where $\Psi = \text{diag}(\psi_1, \ldots, \psi_9)$, and $\boldsymbol{\varpi}_i$ and $\boldsymbol{\varepsilon}_i$ are independent. In the structural equation, $\Gamma = (b_1, b_2, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ is a vector of unknown parameters, $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2})^T$ is distributed as $N(0, \Phi)$, δ_i is distributed as $N(0, \psi_\delta)$, and $\boldsymbol{\xi}_i$ and δ_i are independent.

We assumed the missing data are MNAR, and hence the missingness mechanism of the response variables is non-ignorable (Ibrahim and Molenberghs (2009)). Let $r_{yij} = 1$ if y_{ij} is missing and $r_{yij} = 0$ if y_{ij} is observed. For the missing data mechanism of the response variables, we took $\text{logit}\{\text{pr}(r_{yij} = 1 \mid y_i)\} = \varphi_0 + \varphi_1 y_{i1} + \ldots + \varphi_9 y_{i9}$, where $\varphi = (\varphi_0, \varphi_1, \ldots, \varphi_9)^T$. We also assumed that the covariate x_{i1} is MNAR. Let $r_{xi1} = 1$ if x_{i1} is missing and $r_{xi1} = 0$ if x_{i1} is observed. It was assumed that x_{i1} has a $N(0, \tau_x^2)$ distribution and $\text{logit}\{\text{pr}(r_{xi1} = 1 \mid \varphi_x)\} = \varphi_{x0} + \boldsymbol{\omega} x_{i1}$. When $\boldsymbol{\omega} = 0$, the missingness mechanism reduces to MAR.

We fitted the proposed structural equation models to the AIDS data set and used MCMC sampling to carry out the Bayesian influence analysis. We specified the prior distributions for $\boldsymbol{\mu}$, Λ, Ψ, Γ, $\boldsymbol{\omega}$, Φ, ψ_δ, $\boldsymbol{\varphi}$, φ_{x0}, and τ_x as those in Lee and Tang (2006). A total of $40,000$ MCMC samples was used to compute the intrinsic and local influence measures.

By varying $\boldsymbol{\omega}$ in an interval $[-2, 2]$, we can treat $\boldsymbol{\omega}$ as a perturbation parameter to the sampling distribution. In this case, $\boldsymbol{\omega}^0 = 0$ represents no perturbation. We calculated two intrinsic influence measures for the Kullback-Leibler

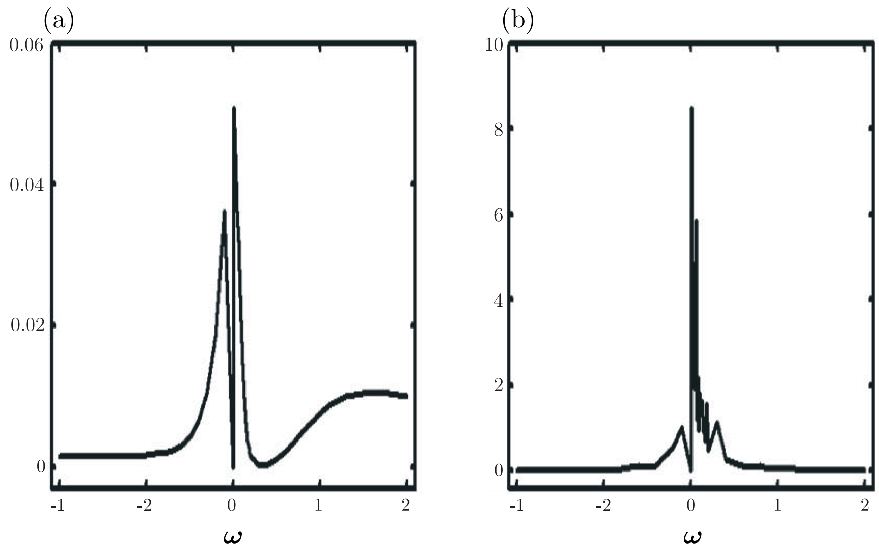

Figure 3. AIDS data analysis results: plots of $\mathrm{IGI}_{RI}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ against $\boldsymbol{\omega} \in [-2, 2]$ for (a) $\Phi_{RI}(\boldsymbol{\omega})$ and (b) $M_h(\boldsymbol{\omega})$, in which $h(\boldsymbol{\theta}) = \Gamma$.

divergence and the posterior mean distance, denoted by $\mathrm{CM}_h(\boldsymbol{\omega})$. Specifically, $\mathrm{CM}_h(\boldsymbol{\omega}, \boldsymbol{\omega}^0) = \{M_h(\boldsymbol{\omega}) - M_h(\boldsymbol{\omega}^0)\}^T C_h \{M_h(\boldsymbol{\omega}) - M_h(\boldsymbol{\omega}^0)\}$, where $M_h(\boldsymbol{\omega}) = \int h(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid z, \boldsymbol{\omega}) d\boldsymbol{\theta}$, in which $h(\boldsymbol{\theta}) = \Gamma$, and C_h is the posterior covariance matrix of Γ based on $p(\Gamma \mid z, \boldsymbol{\omega}^0)$. We calculated $\mathrm{IGI}_{RI}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ at 41 evenly spaced grid points in $[-2, 2]$ (Figure 3). An inspection of Figure 3 shows that the largest $\mathrm{IGI}_{RI}(\boldsymbol{\omega}^0, \boldsymbol{\omega})$ values are close to 0.1 for both the Kullback-Leibler divergence and $M_h(\boldsymbol{\omega})$. This indicates that the nonignorable missing data mechanism may be tenable for the AIDS data. We also carried out a standard sensitivity analysis and computed posterior means and standard deviations of Γ at different values of $\boldsymbol{\omega}$, as shown in Figure 4. Although we observe that the posterior means and standard deviations of Γ vary with $\boldsymbol{\omega}$, it is difficult to make any meaningful inference here.

We also calculated the local influence measures of the Kullback-Leibler divergence under a simultaneous perturbation scheme. The simultaneous perturbation scheme $\boldsymbol{\omega}$ includes variance perturbations $\boldsymbol{\omega}_c$ for individual observations, perturbations $\boldsymbol{\omega}_s$ to coefficients in the structural equations model, perturbations $\boldsymbol{\omega}_\xi$ to the sampling distribution of ξ_i, perturbations $\boldsymbol{\omega}_\mu$ to the prior distribution of μ, perturbations $\boldsymbol{\omega}_\Gamma$ to the prior distribution of Γ, perturbations $\boldsymbol{\omega}_\varphi$ to the prior distribution of φ, and perturbations $\boldsymbol{\omega}_x$ to the missing data mechanism. The corresponding kernel of the joint log-posterior density of $(z, \boldsymbol{\theta})$ based on the complete data is given by

$$\log p(z,\boldsymbol{\theta} \mid \boldsymbol{\omega}) \equiv \ell(z,\boldsymbol{\theta} \mid \boldsymbol{\omega}) = l_c(\boldsymbol{\omega}_c) + l_s(\boldsymbol{\omega}_s) + l_\xi(\boldsymbol{\omega}_\xi) + l_\mu(\boldsymbol{\omega}_\mu) + l_\Gamma(\boldsymbol{\omega}_\Gamma) + l_\varphi(\boldsymbol{\omega}_\varphi) + l_x(\boldsymbol{\omega}_x), \quad (4.1)$$

where

$$l_c(\boldsymbol{\omega}_c) = \frac{1}{2}\Big\{ -p\sum_{i=1}^n \log\Big(\frac{2\pi}{\boldsymbol{\omega}_i}\Big) - \sum_{i=1}^n \boldsymbol{\omega}_i (y_i-\mu-\Lambda\varpi_i)^T\Psi^{-1}(y_i-\mu-\Lambda\varpi_i)\Big\},$$

$$l_s(\boldsymbol{\omega}_s) = \frac{1}{2}\Big\{-\sum_{i=1}^n \frac{1}{\psi_\delta}(\eta_i - b_1x_{i1} - b_2x_{i2} - \boldsymbol{\gamma}_1\xi_{i1} - \boldsymbol{\gamma}_2\xi_{i2} - \boldsymbol{\omega}\boldsymbol{\gamma}_1\xi_{i1}^2 - \boldsymbol{\omega}\boldsymbol{\gamma}_2\xi_{i2}^2 - \boldsymbol{\omega}\boldsymbol{\gamma}_3\xi_{i1}\xi_{i2})^2\Big\},$$

$$l_\xi(\boldsymbol{\omega}_\xi) = \frac{1}{2}\Big\{ -nq_2\log\Big(\frac{2\pi}{\boldsymbol{\omega}_\xi}\Big) - n\log\mid\Phi\mid -\boldsymbol{\omega}_\xi\sum_{i=1}^n \xi_i^T\Phi^{-1}\xi_i\Big\},$$

$$l_\mu(\boldsymbol{\omega}_\mu) = \frac{1}{2}\{-p\log\Big(\frac{2\pi}{\boldsymbol{\omega}_\mu}\Big) - \log\mid\Sigma_0\mid -\boldsymbol{\omega}_\mu(\mu-\mu_0)^T\Sigma_0^{-1}(\mu-\mu_0)\},$$

$$l_\Gamma(\boldsymbol{\omega}_\Gamma) = \frac{1}{2}\Big\{ -(s+t)\log\Big(\frac{2\pi}{\boldsymbol{\omega}_\Gamma}\Big) - \log\mid H_\Gamma\mid -\boldsymbol{\omega}_\Gamma(\Gamma-\Gamma^0)^TH_\Gamma^{-1}(\Gamma-\Gamma^0)\Big\},$$

$$l_\varphi(\boldsymbol{\omega}_\varphi) = \frac{1}{2}\Big\{ -(p+1)\log\Big(\frac{2\pi}{\boldsymbol{\omega}_\varphi}\Big) - \log\mid H_\varphi\mid -\boldsymbol{\omega}_\varphi(\varphi-\varphi^0)^TH_\varphi^{-1}(\varphi-\varphi^0)\Big\},$$

$$l_x(\boldsymbol{\omega}_x) = \sum_{i=1}^n [r_{xi1}(\varphi_{x0}+\boldsymbol{\omega}_x x_{i1}) - \log\{1.0+\exp(\varphi_{x0}+\boldsymbol{\omega}_x x_{i1})\}].$$

In this case, $\boldsymbol{\omega}^0 = (\boldsymbol{\omega}_c^{0T}, \boldsymbol{\omega}_s^{0T}, \boldsymbol{\omega}_\xi^0, \boldsymbol{\omega}_\mu^0, \boldsymbol{\omega}_\Gamma^0, \boldsymbol{\omega}_\varphi^0, \boldsymbol{\omega}_x^0)^T$ represents no perturbation, in which $\boldsymbol{\omega}_c^0 = (1,\ldots,1)^T$, $\boldsymbol{\omega}_s^0 = (0,0,0)^T$, $\boldsymbol{\omega}_\xi^0 = \boldsymbol{\omega}_\mu^0 = \boldsymbol{\omega}_\Gamma^0 = \boldsymbol{\omega}_\varphi^0 = 1$ and $\boldsymbol{\omega}_x^0 = 0.1$.

We calculated $\partial_{\boldsymbol{\omega}}\ell(z,\boldsymbol{\theta}\mid\boldsymbol{\omega})$ and then obtained its metric tensor as

$$G(\boldsymbol{\omega}^0) = \mathrm{diag}\{G_c(\boldsymbol{\omega}_c^0), G_s(\boldsymbol{\omega}_s^0), G_\xi(\boldsymbol{\omega}_\xi^0), G_\mu(\boldsymbol{\omega}_\mu^0), G_\Gamma(\boldsymbol{\omega}_\Gamma^0), G_\varphi(\boldsymbol{\omega}_\varphi^0), G_x(\boldsymbol{\omega}_x^0)\},$$

where $G_c(\boldsymbol{\omega}_c^0) = \mathrm{diag}(p/2,\ldots,p/2)$, $G_\xi(\boldsymbol{\omega}_\xi^0) = nq_2/2$, $G_\mu(\boldsymbol{\omega}_\mu^0) = p/2$, $G_\Gamma(\boldsymbol{\omega}_\Gamma^0) = (s+t)/2$, $G_\varphi(\boldsymbol{\omega}_\varphi^0) = (p+1)/2$, $G_s(\boldsymbol{\omega}_s^0) = \mathrm{diag}[3nE_{\phi,\psi_\delta}(\phi_{11}^2/\psi_\delta), 3nE_{\phi,\psi_\delta}(\phi_{22}^2/\psi_\delta), nE_{\phi,\psi_\delta}\{(\phi_{11}\phi_{22} + 2\phi_{12}^2)/\psi_\delta\}]$, and $G_x(\boldsymbol{\omega}_x^0) = E_{\varphi_{x0},r_x,x_1}[\sum_{i=1}^n\{r_{xi1} - \exp(\varphi_{x0} + \boldsymbol{\omega}_x^0 x_{i1})/(1.0+\exp(\varphi_{x0}+\boldsymbol{\omega}_x^0x_{i1}))\}]^2$. The diagonal elements of the metric tensor $G(\boldsymbol{\omega}^0)$ reveal that $\boldsymbol{\omega}\boldsymbol{\gamma}_1$, $\boldsymbol{\omega}\boldsymbol{\gamma}_2$, $\boldsymbol{\omega}\boldsymbol{\gamma}_3$, $\boldsymbol{\omega}_\xi$, and $\boldsymbol{\omega}_x$ have larger effects compared to other perturbations (see Figure 5(a)). Then, we chose a new perturbation scheme $\tilde{\boldsymbol{\omega}} = \boldsymbol{\omega}^0 + G(\boldsymbol{\omega}^0)^{1/2}(\boldsymbol{\omega}-\boldsymbol{\omega}^0)$ and calculated the associated local influence measures $\mathrm{SI}_{\Phi_{IR}}[e_j]$ for the Kullback-Leibler divergence. The local influence measures based on the ϕ-divergence are able to detect cases {14, 25, 28, 137, 175, 408, 985} as influential observations (see Figure 5(b)), while $\boldsymbol{\omega}\boldsymbol{\gamma}_1$ and $\boldsymbol{\omega}\boldsymbol{\gamma}_3$ indicate that it

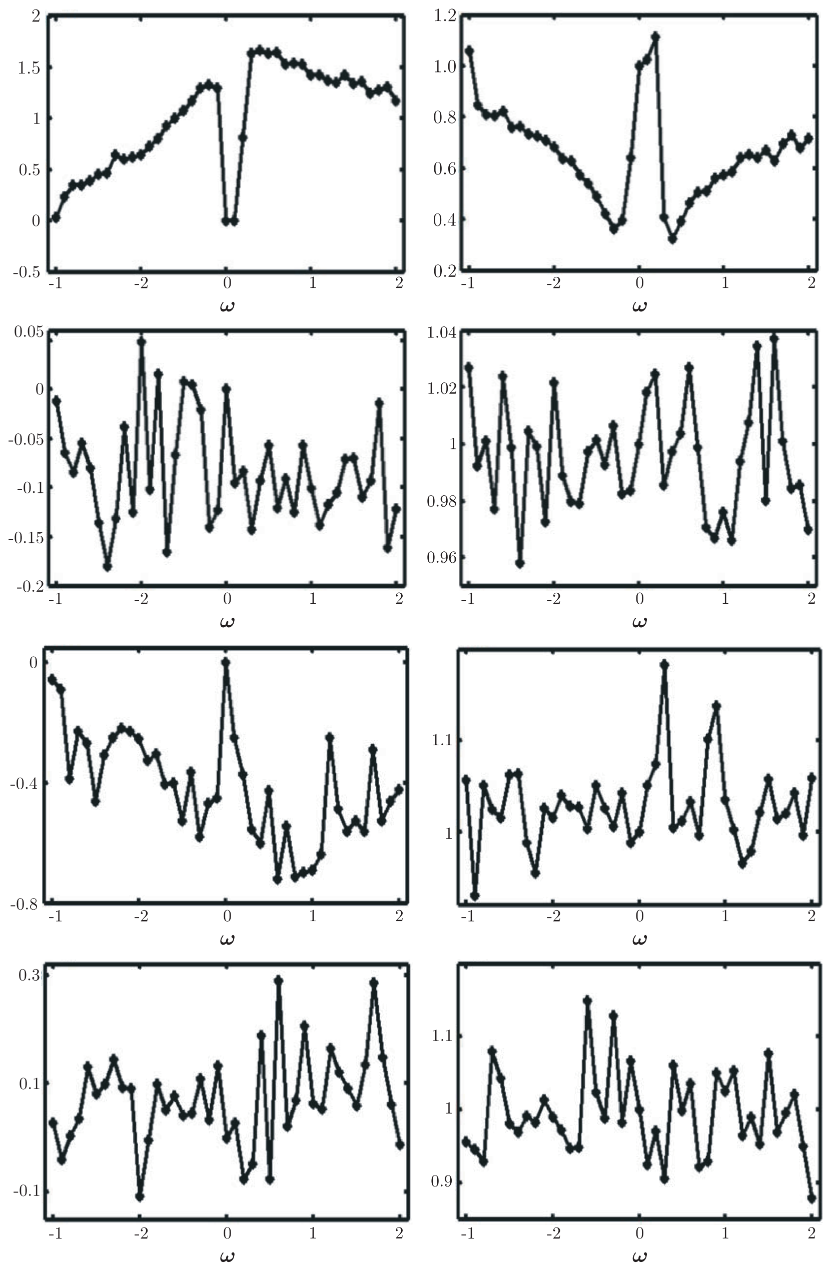

Figure 4. AIDS data analysis results: plots of (posterior means-posterior mean at $\omega = 0$)/(posterior standard deviation at $\omega = 0$) ((a),(c),(e),(g)) and the ratio of posterior standard deviations divided by the posterior standard deviation at $\omega = 0$ ((b),(d),(f),(h)) of b_1, b_2, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ as a function of $\boldsymbol{\omega} \in [-2, 2]$.

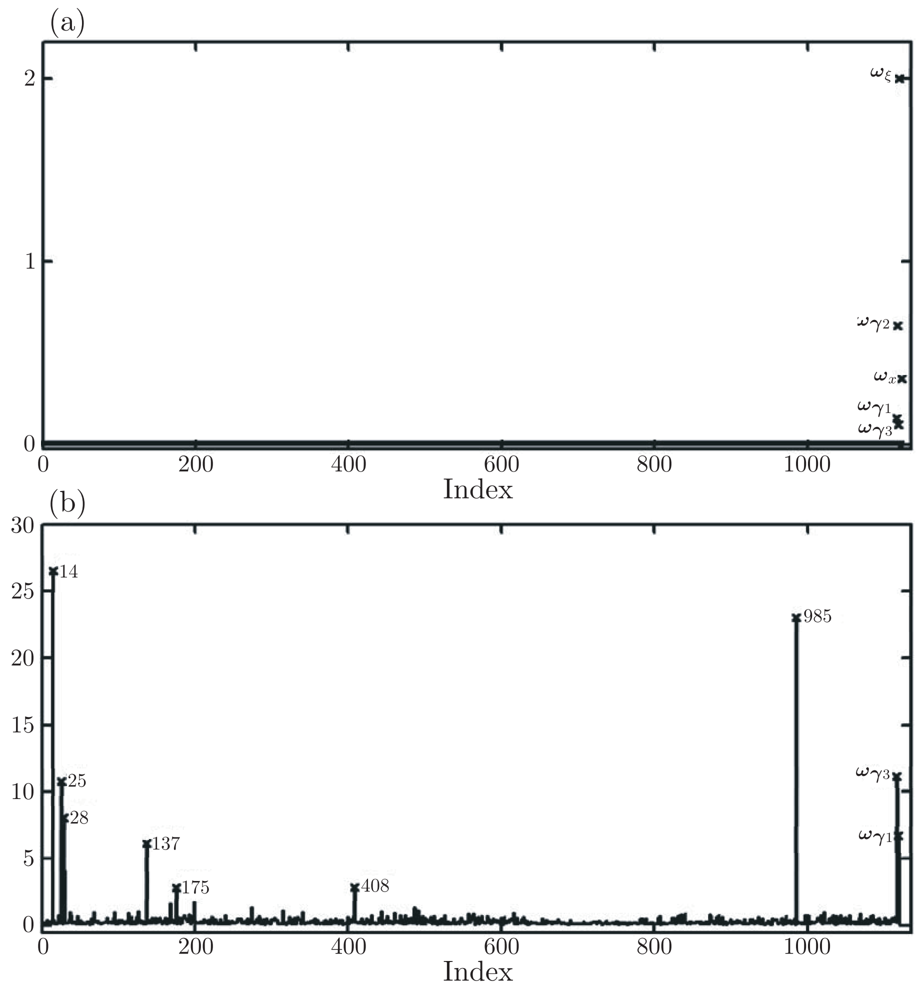

Figure 5. AIDS data analysis results: index plots of (a) metric tensor $g_{jj}(\boldsymbol{\omega}^0)$ and (b) local influence measures $\mathrm{SI}_{\Phi_{IR}}[e_j]$ for simultaneous perturbation.

may be important to include ξ_{i1}^2 and $\xi_{i1}\xi_{i2}$ in the structural model (see Figure 5(b)).

5. Discussion

We have developed Bayesian sensitivity analysis methods for assessing various perturbations to statistical methods with missing data. We have developed a Bayesian perturbation manifold to characterize the intrinsic structure of the perturbation model and to quantify the degree of each perturbation in the perturbation model. We have developed global and local influence measures for selecting the most influential perturbation based on various objective functions and their statistical properties. Finally, we have also examined a number of examples to highlight the broad spectrum of applications of this Bayesian influence analysis method in missing data problems.

Many issues merit further research. Our Bayesian sensitivity analysis method can be extended to more complex data structures (e.g., survival data) and other parametric and semiparametric models with nonparametric priors. In further research, we will generalize our methodology to the setting of estimating equations and empirical likelihood of generalized estimating equations for missing data problems. We will develop Bayesian sensitivity analysis methods to deal with the well-known masking and swamping effects in the diagnostic literature.

References

Andridge, R. R. and Little, R. J. A. (2011). Proxy pattern-mixture analysis for survey nonresponse. *Journal of Official Statistics* **27**, 153-180.

Berger, J. O. (1994). An overview of robust bayesian analysis. *Test* **3**, 5-58.

Chen, M. H., Shao, Q. M. and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.

Cook, R. D. (1986). Assessment of local influence (with Discussion). *J. Roy. Statist. Soc. Ser. B* **48**, 133-169.

Copas, J. and Eguchi, S. (2005). Local model uncertainty and incomplete data bias (with discussion). *J. Roy. Statist. Soc. Ser. B* **67**, 459- 512.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **59**, 55-96.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **49**, 1-39.

Daniels, M. J. and Hogan, J. W. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Chapman and Hall, London.

Dey, D. K., Ghosh, S. K. and Lou, K. R. (1996). On local sensitivity measures in bayesian (with discussion). In *Bayesian Robustness* (Edited by J. O. Berger, B. Betrò, e. Moreno, l. R. Pericchi, F. ruggeri, G. Salinetti and L. Wasserman). IMS Lecture Notes-Monograph Series Volume 29, 21-39.

Geisser, S. (1993). *Predictive Inference: An Introduction*. Chapman and Hall, London.

Gelfand, A. E., Dey, D. K. and Chang, H. (1992). Model determination using predictive distributions, with implementation via sampling-based methods (disc: P160-167). In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, editors, *Bayesian Statistics* **4**, 147-159. Oxford University Press, Oxford.

Gustafson, P. (1996a). Local sensitivity of inferences to prior marginals. *J. Amer. Statist. Assoc.* **91**, 774-781.

Gustafson, P. (1996b). Local sensitivity of posterior expectations. *Ann. Statist.* **24**, 174-195.

Gustafson, P. (2006). On model expansion, model contraction, identifability, and prior information: two illustrative scenarios involving mismeasured variables (with discussion). *Statist. Sci.* **20**, 111-140.

Hens, N., Aerts, M., Molenberghs, G., Thijs, H. and Verbeke, G. (2006). Kernel weighted in influence measures. *Comput. Statist. Data Anal.* **48**, 467-487.

Ibrahim, J. G., Chen, M. H., Lipsitz, S. R. and Herring, A. (2005). Missing-data methods for generalized linear models: a comparative review. *J. Amer. Statist. Assoc.* **100**, 332-346.

Ibrahim, J. G. and Molenberghs, G. (2009). Missing data methods in longitudinal studies: a review. *Test* **18**, 1-43.

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G. and Kenward, M. G. (2006). The nature of sensitivity in monotone missing not at random models. *Comput. Statist. Data Anal.* **50**, 830-858.

Jansen, I., Molenberghs, G., Aerts, M., Thijs, H. and Van Steen, K. (2003). A local influence approach to binary data from a psychiatric study. *Biometrics* **59**, 410-419.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.* **90**, 773-795.

Kass, R. E., Tierney, L. and Kadane, J. B. (1989). Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika* **76**, 663-674.

Lang, S. (1995). *Differential and Riemannian Manifolds.* 3ed edition. Springer-Verlag, New York.

Lavine, M. (1991). Sensitivity in bayesian statistics: the prior and the likelihood. *J. Amer. Statist. Assoc.* **86**, 396-399.

Lee, S. Y. and Tang, N. S. (2006). Analysis of nonlinear structural equation models with nonignorable missing covariates and ordered categorical data. *Statist. Sinica* **16**, 1117-1141.

Little, R. J. A. (1994). A class of patternmixture models for normal incomplete data. *Biometrika* **81**, 471-483.

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

Millar, R. B. and Stewart, W. S. (2007). Assessment of locally influential observations in bayesian models. *Bayesian Anal.* **2**, 365-384.

Molenberghs, G. and Kenward, G. (2007). *Missing Data in Clinical Studies.* Wiley, New York.

Morisky, D. E., Tiglao, T. V., Sneed, C. D., Tempongko, S. B., Baltazar, J. C., Detels, R. and Stein, J. A. (1998). The effects of establishment practices, knowledge and attitudes on condom use among filipina sex workers. *AIDS Care* **10**, 213-320.

Shi, X. Y., Zhu, H.T. and Ibrahim, J. G. (2009). Local influence for generalized linear models with missing covariates. *Biometrics* **65**, 1164-1174.

Troxel, A. B. (1998). A comparative analysis of quality of life data from a southwest oncology group randomized trial of advanced colorectal cancer. *Statisti. Medicine* **17**, 767-779.

Troxel, A. B., Ma, G. and Heitjan, D. F. (2004). An index of local sensitivity to nonignorability. *Statist. Sinica* **14**, 1221-1237.

van der Lindem, A. (2007). Local influence on posterior distributions under multiplicative modes of perturbation. *Bayesian Anal.* **2**, 319-332.

van Steen, K., Molenberghs, G. and Thijs, H. (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling: An International J.* **1**, 125-142.

Verbeke, G., Molenberghs, G., Thijs, H., Lasaffre, E. and Kenward, M. G. (2001). Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics* **57**, 43-50.

Zhu, H. T., Ibrahim, J. G., Lee, S. Y. and Zhang, H.P. (2007). Perturbation selection and influence measures in local influence analysis. *Ann. Statist.* **35**, 2565-2588.

Zhu, H. T., Ibrahim, J. G. and Tang, N. S. (2011). Bayesian local influence analysis: a geometric approach. *Biometrika* **98**, 307-323.

Zhu, H. T. and Lee, S. Y. (2001). Local influence for incomplete-data models. *J. Roy. Statist. Soc. Ser. B* **63(1)**, 111-126.

Department of Biostatistics, University of North Carolina at Chapel Hill, 3109 McGavran-Greenberg Hall, Campus Box 7420, Chapel Hill, North Carolina 27516, U.S.A.

E-mail: hzhu@bios.unc.edu

Department of Biostatistics, University of North Carolina at Chapel Hill, 3109 McGavran-Greenberg Hall, Campus Box 7420, Chapel Hill, North Carolina 27516, U.S.A.

E-mail: ibrahim@bios.unc.edu

Department of Statistics, Yunnan University, Kunming 650091, P. R. China

E-mail: nstang@ynu.edu.cn

(Received April 2012; accepted April 2013)