A SEMIPARAMETRIC APPROACH FOR ANALYZING NONIGNORABLE MISSING DATA

Hui Xie, Yi Qian and Leming Qu

University of Illinois, Chicago, Northwestern University and Boise State University

Abstract: In missing data analysis, there is often a need to assess the sensitivity of key inferences to departures from untestable assumptions regarding the missing data process. Such sensitivity analysis often requires specifying a missing data model that commonly assumes parametric functional forms for the predictors of missingness. In this paper, we relax the parametric assumption and investigate the use of a generalized additive missing data model. We also consider the possibility of a nonlinear relationship between missingness and the potentially missing outcome, whereas the existing literature commonly assumes a more restricted linear relationship. To avoid computational complexity, we adopt an index approach for local sensitivity. We derive explicit formulas for the resulting semiparametric sensitivity index. The computation of the index is simple, and completely avoids the need to repeatedly fit the semiparametric nonignorable model. Only estimates from the standard software analysis are required, with a moderate amount of additional computation. Thus, the semiparametric index provides a fast and robust method to adjust the standard estimates for nonignorable missingness. An extensive simulation study is conducted to evaluate the effects of misspecifying the missing data model and to compare the performance of the proposed approach with the commonly used parametric approaches. The simulation study suggests that the proposed method helps reduce bias that might arise from the misspecification of the functional forms of predictors in the missing data model. We illustrate the method in a Wage Offer dataset.

Key words and phrases: Generalized additive model, MNAR, nonignorability, semiparametric joint selection model, sensitivity analysis.

1. Introduction

Missing data arise frequently in studies across different disciplines, including public health, medicine, economics, business, and the social sciences. Missingness can be due to nonresponse in household surveys, attrition in longitudinal studies, or patient noncompliance in experimental studies and clinical trials. In missing data analysis, *ignorability* has been a standard assumption regarding the missing data mechanism (Rubin (1976)). Under ignorability, a valid likelihood/Bayesian inference can ignore modeling the missing data mechanism. In the likelihood/Bayesian inference, ignorability holds under the assumptions of

missing-at-random (MAR) and parameter distinctness (Rubin (1976), Heitjan and Rubin (1991)).

Although ignorability is a convenient and useful assumption, it is usually an approximation to reality when missingness is not by design. There are important situations in which this assumption is questionable. In the analysis of potentially nonignorable missing data, the selection model is a popular class of models in which one augments the model for the complete data with a missing data model. In practice, a parametric binary regression model has been commonly employed for modeling the missing data process. When there are continuous missingness predictors, misspecifying the functional forms of these predictors by the parametric model can lead to severe bias in the inference of the primary parameters of interest. In this paper, we propose a data-driven procedure to adaptively choose the functional forms of the continuous predictors. Specifically, we propose using the generalized additive model (GAM) to relax the linearity assumptions in the missing data model.

One can perform a direct estimation of such a semiparametric joint selection model (Chen and Ibrahim (2006)). The direct estimation can yield valid inferences when the model is correctly specified, although its computation is heavy and requires specialized programming. More importantly, such a joint selection model is often weakly or non-identified (Little (1995), Troxel (1998), Troxel, Harrington and Lipsitz (1998), Chen and Ibrahim (2006), and the results can be highly sensitive to untestable model assumptions (Kenward (1998)). To tackle this problem, the use of a nonignorable selection model has been proposed as a tool for assessing sensitivity of inference to nonignorable missingness (Little (1995), Vach and Blettner (1995), Copas and Li (1997), Scharfstein, Rotnizky and Robins (1999), Copas and Eguchi (2001)). A global sensitivity analysis usually involves repeatedly fitting the nonignorable model for a range of magnitudes of nonignorability, which can be computationally burdensome. To avoid the computational burden, we utilize an index approach to local sensitivity (Troxel, Ma and Heitjan (2004)) that uses a Taylor series expansion to approximate the estimates in the neighborhood of the MAR model. The index method has been applied in various settings (Xie and Heitjan (2004, 2009), Ma, Troxel and Heitjan (2005), Xie (2008, 2009) Xie and Qian (2010), Zhang and Heitjan (2006, 2007)).

The local sensitivity method proposed in the literature utilizes the linear logistic regression for modeling the missing data process. In this article we relax the linearity assumption and investigate the usage of a generalized additive model for a more robust and flexible modeling of the missing data mechanism. Further, the proposed method is computationally less complex than the alternative global sensitivity method. Specifically, our approach avoids fitting any complicated semiparametric joint selection model. Only estimates from a MAR

analysis of the outcome model and a MAR GAM for the missing data process are required to evaluate sensitivity. Both estimates can be obtained using standard software packages such as SAS or S-Plus/R. For instance, the MAR GAM for the missing data process can be fitted using PROC GAM in SAS or the S-Plus/R function *gam*. In summary, the proposed approach renders sensitivity analysis for nonignorable missingness simple to perform by avoiding excessive additional computation, and robust to model misspecification by automatically adjusting for potentially complex missing data mechanisms.

The rest of the paper is organized as follows. In Section 2 we describe the semiparametric joint selection model. In Section 3 we review the ISNI (*index of local sensitivity to nonignorability*) methodology. In Section 4 we investigate the use of GAM to model the missing data process; we also consider the possibility of a nonlinear relationship between missingness and the potentially missing outcome, and present specific formulas for the sensitivity index in the Appendix when the relationship follows a quadratic form. In Section 5 we report on simulation studies to compare the performance of the parametric and semiparametric approaches for modeling the missing data process with respect to their ability to reduce the bias of the MAR estimates. In Section 6 we apply the methodology to an application on estimating a wage offer function. We conclude with a discussion in Section 7.

2. Selection Model for Nonignorable Missingness

Consider data (Y_i, X_i, Z_i, G_i) from the unit i, $i = 1, \ldots, n$. The underlying ideal outcome, Y_i, arises independently from a distribution with density $f_\theta(Y_i|X_i)$, where X_i contains a set of fully observed covariates. We are interested in drawing inferences on θ or a subset of it. For our purpose here, we restrict attention to the case in which Y_i is univariate. For various reasons, Y_i is subject to missingness. Let G_i be 1 if y_i is observed and 0 if y_i is missing. We assume the missing data model $f_\gamma(g_i|y_i, z_i)$, where $G_i|Y_i, Z_i \sim Bernoulli(P_\gamma(G_i = 1|Y_i, Z_i))$ and

$$P_\gamma(G_i = 1|Y_i = y_i, Z_i = z_i) = h(\eta_{\gamma_0}(z_i) + \eta_{\gamma_1}(y_i)), \tag{2.1}$$

h is the inverse of a monotonic link function, Z_i is a set of fully observed predictors for missingness, $\eta_{\gamma_0}(\cdot)$ and $\eta_{\gamma_1}(\cdot)$ are smooth functions, and their functional forms are as yet unspecified. Let $\gamma = (\gamma_0, \gamma_1)$ in which γ_0 is a vector of parameters that associates the probability of missingness with observed data, and γ_1 associates the probability of missingness with potentially unobserved data. In the model, $\eta_{\gamma_1}(y)$ represents the form of nonignorable missingness: when $\eta_{\gamma_1}(y)$ is constant in y, the missing data mechanism is MAR; when $\eta_{\gamma_1}(y)$ depends on y, it becomes missing not at random (MNAR).

Let (Y, X, Z, G) be the data stacked over all the units. We rewrite Y as $(Y_{\mathrm{obs}}, Y_{\mathrm{mis}})$, where Y_{obs} refers to the observed components of Y and Y_{mis} refers to the missing components of Y. The covariates X and Z are considered fixed, and the conditioning on them in $f_\theta(y_i|x_i)$ and $f_\gamma(g_i|y_i, z_i)$ is suppressed for notational simplicity. The data to be modeled are thus (Y_{obs}, G). The correct log-likelihood of the model parameters is

$$
\begin{aligned}
L_C(\theta,\ \gamma;\ y_{\mathrm{obs}}, g) &\propto \ln f_{\theta,\gamma}(y_{\mathrm{obs}}, g) \\
&= \ln \int f_\gamma(g|y_{\mathrm{obs}}, y_{\mathrm{mis}})\, f_\theta(y_{\mathrm{obs}}, y_{\mathrm{mis}})\, dy_{\mathrm{mis}}.
\end{aligned}
\tag{2.2}
$$

Under the MAR condition, $f_\gamma(g|y_{\mathrm{obs}}, y_{\mathrm{mis}}) = f_\gamma(g|y_{\mathrm{obs}})$, and can be moved out of the integral. With the parameter distinctness, this results in a simpler log-likelihood for θ,

$$
L_I(\theta;\ y_{\mathrm{obs}}) \propto \ln f_\theta(y_{\mathrm{obs}}) = \ln \int f_\theta(y_{\mathrm{obs}}, y_{\mathrm{mis}})\, dy_{\mathrm{mis}}.
\tag{2.3}
$$

In practice, L_I is often used because it avoids modeling the missing data mechanism. However, in the general case of nonignorable missingness, L_I is not proportional to L_C and the inference based on L_I is potentially biased.

3. ISNI Methodology

As indicated above, for a dataset with missingness the correct log-likelihood is $L_C(\theta,\ \gamma;\ y_{\mathrm{obs}}, g)$. We define

$$
(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1)) = \underset{\theta \in \Omega_\theta, \gamma_0 \in \Omega_{\gamma_0}}{\mathrm{argmax}}\ L_C(\theta, \gamma_0, \gamma_1\,;\, y_{\mathrm{obs}}, g)\quad \text{for a fixed } \gamma_1.
$$

One can then vary γ_1 in a plausible range and investigate how the other parameter estimates in the model are affected. In the existing literature, such sensitivity analysis commonly assumes a linear binary regression model for the missing data process, that is, a restriction that $\eta_{\gamma_0}(z_i) = \gamma_0^T z_i$ and $\eta_{\gamma_1}(y_i) = \gamma_1 y_i$. Here the likelihood $L_C(\theta, \gamma_0;\ \gamma_1)$ is proportional in θ to the likelihood $L_I(\theta)$ for all $\theta \in \Omega_\theta$ when $\gamma_1 = 0$; $\hat{\theta}(0)$ is then the MLE of θ in the ignorable model. The difference between $\hat{\theta}(0)$ and $\hat{\theta}(\gamma_1)$ is a measure of the sensitivity of the MLE when γ_1 is perturbed around the ignorable model. The idea of a local sensitivity analysis is to approximate $\hat{\theta}(\gamma_1)$ by a Taylor series expansion as

$$
\hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \left.\frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1}\right|_{\gamma_1=0} \times \gamma_1,
\tag{3.1}
$$

where $\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\right|_{\gamma_1=0}$ measures the changing rate of $\hat{\theta}$ as a function of γ_1, and is referred to as the *index of local sensitivity to nonignorability* (ISNI) (Troxel, Ma and Heitjan (2004)) In this approximation, $\hat{\theta}(0)$ is obtained by maximizing the simpler log-likelihood L_I. As shown in Troxel, Ma and Heitjan (2004), a simple formula for $\left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\right|_{\gamma_1=0}$ is

$$\text{ISNI} \;=\; \left.\frac{\partial\hat{\theta}(\gamma_1)}{\partial\gamma_1}\right|_{\gamma_1=0} = -\nabla^2 L_{\theta,\theta}^{-1}\nabla^2 L_{\theta,\gamma_1}, \tag{3.2}$$

with

$$\nabla^2 L_{\theta,\theta} = \left.\frac{\partial^2 L_C}{\partial\theta\partial\theta^T}\right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0} \;;\; \nabla^2 L_{\theta,\gamma_1} = \left.\frac{\partial^2 L_C}{\partial\theta\partial\gamma_1}\right|_{\hat{\theta}(0),\hat{\gamma}_0(0),\gamma_1=0}.$$

The first term $\nabla^2 L_{\theta,\theta}$ is the observed Hessian matrix of the ignorable model that is usually readily available; the second term evaluates the orthogonality of θ and γ_1. One limitation of ISNI, as a local sensitivity method, is that the local approximation in (3.1) might not be sufficiently accurate for extreme nonignorability (i.e., $|\gamma_1|$ is large). Thus, ISNI is most useful for moderate nonignorability (e.g., a rich set of observed predictors for missingness has been conditioned on so that the remaining nonignorability is not extreme).

4. Extending ISNI Using a Generalized Additive Model

When some components of Z are continuous, a linear predictor, as considered above, may not be adequate, and the misspecification of the functional forms for Z may lead to severe bias in the estimation of θ, the parameter of primary interest. It is thus desirable to use extended models that describe a wider range of selection mechanisms. In this section we investigate the use of GAM for a robust and flexible modeling of the missingness probability. Rather than prespecifying $\eta_{\gamma_0}(Z_i)$ as a linear form, we let $\eta_{\gamma_0}(Z_i)$ follow a GAM (Hastie and Tibshirani (1990)):

$$\eta_{\gamma_0}(Z_i) = \gamma_{00} + \eta_{01}(Z_{i1}) + \eta_{02}(Z_{i2}) + \ldots + \eta_{0m}(Z_{im}),$$

where Z_i is composed of m missingness predictors, $(Z_1,\ldots,Z_m)$, and η_{0j} is an arbitrary smooth and mean zero function for the jth covariate Z_j, $j = 1,\ldots,m$. Because of the additivity in the nonparametric component, the model is termed a generalized additive model.

Another important modeling decision is to specify $\eta_{\gamma_1}(y_i)$. Note that y_i is missing whenever $G_i = 0$, and thus an attempt to estimate $\eta_{\gamma_1}(y_i)$ would

inevitably require imposing untestable assumptions or using external data. Because of the lack of information from the data at hand for identification, a feasible approach is to perform sensitivity analysis with respect to $\eta_{\gamma_1}(y_i)$. In the existing literature, it is common to assume a linear form for $\eta_{\gamma_1}(y_i)$, i.e., $\eta_{\gamma_1}(y_i) = \gamma_1 y_i$. One benefit of this parametrization is the ease of interpreting the sensitivity analysis result. Although the linearity assumption is reasonable in many practical applications, it is by no means universally applicable. Thus, we base our development on a more general functional form of $\eta_{\gamma_1}(y_i)$ as $\sum_{q=1}^{Q} \gamma_{1q} y_i^q$, where Q is a user-specified order.

One can consider using the following penalized log-likelihood of the resulting semi-parametric nonignorable selection model for sensitivity analysis:

$$
\begin{aligned}
L_C^P&(\theta, \gamma_{00}, \eta_{01}, \ldots, \eta_{0m}, \gamma_1; y_{\text{obs}}, g) \\
&= \ln \left[\int f_\gamma(g | \gamma_{00}, \eta_{01}, \ldots, \eta_{0m}, \gamma_1, y_{\text{obs}}, y_{\text{mis}}) f_\theta(y_{\text{obs}}, y_{\text{mis}}) dy_{\text{mis}} \right] \\
&\quad - \frac{1}{2} \sum_{j=1}^{m} \lambda_j \int \eta_{0j}''(u)^2 du,
\end{aligned}
$$

where $\lambda_j \geq 0$ is a smoothing parameter whose value can be adjusted to avoid overfitting. The sensitivity of inference with respect to nonignorable missingness can then be assessed by calculating $(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1))$ for a plausible range of values for γ_1. For any given value of γ_1, we must obtain $(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1))$ by maximizing L_C^P over (θ, γ_0). An algorithm such as the EM algorithm can be used, but the optimization can be heavy (e.g., take a long time to converge) and requires specialized programming. Moreover, it needs to be repeatedly performed for a range of γ_1 values, which further compounds the computational burden.

In contrast, the ISNI method substantially reduces the computational workload. As the parameters $\gamma_0 = (\gamma_{00}, \eta_{01}, \ldots, \eta_{0m})$ in the missing data model are orthogonal to the parameter θ in the complete data model when $\gamma_1 = 0$ (i.e., missingness is MAR), one can show that for a vector $\gamma_1 = (\gamma_{11}, \ldots, \gamma_{1Q})$, we have

$$
\left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1 = 0} = -\nabla^2 L_{\theta,\theta}^{-1} \nabla^2 L_{\theta,\gamma_1},
$$

with

$$
\begin{aligned}
\nabla^2 L_{\theta,\theta} &= \left. \frac{\partial^2 L_C^P}{\partial \theta \partial \theta^T} \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1 = 0} = \left. \frac{\partial \ln f_\theta(Y_{obs})}{\partial \theta \partial \theta^T} \right|_{\hat{\theta}(0)}, \\
\nabla^2 L_{\theta,\gamma_1} &= \left. \frac{\partial^2 L_C^P}{\partial \theta \partial \gamma_1^T} \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1 = 0} = -h_i \left. \left(\frac{\partial E_\theta(Y_i)}{\partial \theta}, \ldots, \frac{\partial E_\theta(Y_i^Q)}{\partial \theta} \right) \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1 = 0}, \quad (4.1)
\end{aligned}
$$

where $h_i = h(\hat{\gamma}_{00} + \sum_{j=1}^{m} \hat{\eta}_{0j}(Z_{ij}))$ is the predicted probability of being observed under the MAR model, and $h(\cdot)$ is the inverse of the logit link. The above calculation requires only the MAR estimates, which are obtained by optimizing the log-likelihood

$$
L_I^P(\theta, \gamma_{00}, \eta_{01}, \ldots, \eta_{0m}, \gamma_1 = 0; y_{\text{obs}}, g)
$$
$$
= \ln f_\theta(y_{\text{obs}}) + \ln f_{\gamma_0, \gamma_1 = 0}(g|y_{\text{obs}}) - \frac{1}{2} \sum_{j=1}^{m} \lambda_j \int \eta_{0j}''(u)^2 du.
$$

In particular, the calculation of the extended ISNI requires the estimation of $\eta_0(Z)$ under the MAR model, i.e., $\gamma_1 = 0$. The fit maximizes the penalized log-likelihood

$$
PELL(\gamma_{00}, \eta_{01}, \ldots, \eta_{0m}; g) = \ln f_{\gamma_0, \gamma_1 = 0}(g|y_{\text{obs}}) - \frac{1}{2} \sum_{j=1}^{m} \lambda_j \int \eta_{0j}''(u)^2 du.
$$

The conventional algorithm for the estimation of a GAM is the local scoring procedure (Hastie and Tibshirani (1990)), which maximizes the PELL. Equation (4.1) shows that a missing observation is given more weight in the calculation of the ISNI if its predicted probability of nonmissingness, h_i, is large (i.e., unexpected missingness). The quantity h_i is related to the missing data mechanism and plays an important role in assessing the sensitivity. Using a generalized additive model to describe the missing data mechanism is useful here because we need accurate and robust estimates of these probabilities of being observed.

For ease of interpretation, one can reparameterize $\eta_{\gamma_1}(y) = \gamma_{11} \sum_{q=1}^{Q} r_q y^q$, where $r_q = \gamma_{1q}/\gamma_{11}$, and then define

$$
\text{ISNI}_r = \left. \frac{\partial \hat{\theta}(\gamma_{11}, r)}{\partial \gamma_{11}} \right|_{\gamma_{11}=0} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} \left. \frac{\partial \gamma_1}{\partial \gamma_{11}} \right|_{\gamma_1 = \gamma_{11} \cdot r} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} r, \quad (4.2)
$$

where $r = (r_1, \ldots, r_Q)^T$. Given a user-specified r, one can approximate the potential change of $\hat{\theta}(\gamma_1)$ when γ_{11} is perturbed from 0 to a given value as

$$
\hat{\theta}(\gamma_1) - \hat{\theta}(0) \approx \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} \gamma_1 = \text{ISNI}_r * \gamma_{11}. \quad (4.3)
$$

Using the extended ISNI method, it is convenient for a data analyst to entertain plausible choices of Q and r to explore the sensitivity with respect to the functional forms of $\eta_{\gamma_1}(y)$. In the Appendix we derive explicit ISNI_r formulas for $Q = 2$ (i.e., a quadratic function) when the outcome is modeled by a generalized linear model.

5. A Comparison Using Simulated Data

In this section we report on simulation studies to compare the performance of the parametric and semiparametric approaches for modeling the missing data process. Specifically, we simulated data from both linear and nonlinear missing data models, and then investigated whether the ISNI based on a GAM missing data model provided a more faithful and robust adjustment of an MAR estimate than those based on various linear logistic missing data models. We followed the steps below to perform the simulation studies.

Step 1: Generate the hypothetical complete data, (Y_i, X_i), independently from the bivariate normal distribution

$$\begin{bmatrix} Y_i \\ X_i \end{bmatrix} \sim BVN \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right],$$

where $i = 1, \ldots, n$, and the sample size $n = 500$. The parameter ρ takes the value -0.5, 0, or 0.5. We are interested in the conditional distribution of $Y_i|X_i$, which is $N(\beta_0 + \beta_1 x_i, \sigma^2)$, where $\beta_1 = \rho$ is the parameter of interest.

Step 2: Generate the missingness pattern. Y_i is subject to missingness with the probability of nonmissingness given by the missing data model

$$\text{logit}(P_\gamma(G_i = 1|Y_i = y_i, X_i = x_i)) = \eta_{\gamma_0}(x_i) + \gamma_1 y_i, \tag{5.1}$$

According to the exact form of $\eta_{\gamma_0}(x)$, we have the following configurations

- Case 1.

  - 1. Linear Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 + x + \gamma_1 y$.

  - 2. Quadratic Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 + x + x^2 + \gamma_1 y$.

  - 3. Cubic Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 + x + x^2 + x^3 + \gamma_1 y$.

  - 4. Sine Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 2\sin(2x) + \gamma_1 y$.

- Case 2.

  - 1. Linear Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 + x + \gamma_1 y$.

  - 2. Quadratic Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 - (x - 1)^2 + \gamma_1 y$.

  - 3. Cubic Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 1 - (x - 1)^2 + (x - 1)^3 + \gamma_1 y$.

  - 4. Sine Function. $\text{logit}(P_\gamma(G = 1|X = x, Y = y)) = 2\sin(2(x + 1.5)) + \gamma_1 y$.

We vary γ_1 in $(-1, -0.5, -0.1, 0.1, 0.5, 1)$. The proportion of missingness in Y ranges from 15% to 65% in the generated datasets.

Step 3: With each generated dataset, compute the MAR estimate, $\hat{\beta}_1(0)$, by applying the least-square fitting of the regression model, using only the cases with observed y_i. Based on (3.1), calculate three ISNI-adjusted estimates as

$$\hat{\beta}_{1L}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNI}_L \times \gamma_1, \ \hat{\beta}_{1P}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNI}_P \times \gamma_1,$$
$$\hat{\beta}_{1G}(\gamma_1) = \hat{\beta}_1(0) + \text{ISNI}_G \times \gamma_1. \tag{5.2}$$

We use, ISNI_L, ISNI_P, and ISNI_G, in order of increasing generality, to model the missing data process. The most constrained is ISNI_L, whose calculation assumes *a priori* that $\eta_{\gamma_0}(x_i)$ at (5.1) is linear in x_i, and $\eta_{\gamma_0}(x_i) = \gamma_{00} + \gamma_{01}x_i$. The more flexible ISNI_P adds higher-order polynomial terms for x_i (i.e., quadratic term, cubic term, $\cdots$). This process stops when adding the next higher-order term of x_i to the missing data model does not significantly improve the model fit at the 0.05 level, where the improvement in model fitness is measured by the difference in model deviance. This analysis strategy represents a common parametric approach to more acceptable models for the missing data mechanism. The most general one, ISNI_G, uses the GAM method to estimate the missing data model. It uses a nonparametric scatterplot smoother, such as a smoothing spline method, for the estimation of $\eta_{\gamma_0}(x_i)$ and lets data speak to the functional form of x_i. As compared with ISNI_P, ISNI_G enjoys two advantages: GAM is clearly more general; it is an automatic procedure that avoids the manual increase of model complexity.

For our simulation model, the formula for ISNI is

$$-\hat{\sigma}^2 \Big(\sum_{i:g_i=1} \widetilde{x}_i \widetilde{x}_i^T \Big)^{-1} \sum_{i:g_i=0} h_i \widetilde{x}_i.$$

Here $\hat{\sigma}^2 = \sum_{i:g_i=1}(y_i - \hat{\beta}_0(0) - \hat{\beta}_1(0)x_i)^2 / \sum_i 1(g_i = 1)$ is the MAR estimate of the residual variance, and $\hat{\beta}_0(0)$ and $\hat{\beta}_1(0)$ are the MAR estimates of the β_0 and β_1, respectively; $\widetilde{x}_i = [1, x_i]^T$ is the vector of predictors for the unit i; h_i is the predicted probability of $G_i = 1$ under the MAR model. For ISNI_L, $h_i = h(\hat{\gamma}_{00}(0) + \hat{\gamma}_{01}(0)x_i)$. For ISNI_P, $h_i = h(\hat{\gamma}_{00}(0) + \sum_{j=1}^J \hat{\gamma}_{0j}(0)x_i^J)$, where J is the selected order of the polynomial function of x_i. For ISNI_G, $h_i = h(\hat{\gamma}_{00}(0) + \hat{\eta}_{\gamma_{01}}(x_i))$, where $\hat{\eta}_{\gamma_{01}}(x_i)$ is adaptively estimated by a smoothing spline under the MAR assumption. The *gam* function in S-Plus with a default degree of freedom of 4 is used for smoothing.

In practical applications, one can calculate the ISNI-adjusted estimates at (5.2) for a plausible range of γ_1 values, and investigate the sensitivity of the MAR estimates to nonignorable missingness. In the simulation studies, we plug in the

true value of γ_1. The performance of these ISNI-adjusted estimates can then be evaluated in terms of their ability to reduce bias of the MAR estimates for various scenarios of missing data mechanisms.

Step 4: Repeat Step 1 to Step 3 300 times for the same values of ρ and γ_1. Using the resulting sample of estimates, compute the mean squared error (MSE), bias, and standard deviation (SD) for each of the four estimators of β_1: $\hat{\beta}_1(0), \hat{\beta}_{1L}(\gamma_1), \hat{\beta}_{1P}(\gamma_1)$, and $\hat{\beta}_{1G}(\gamma_1)$. Repeat Step 4 for other configurations of ρ and γ_1.

Figure 1 plots the bias when $\beta_1 = \rho = 0$. The results on SDs and MSEs can be found in Web Supplement Tables 1 and 2 and Figures 1 and 2. As shown there, the SDs of the adjusted estimates are almost the same as the SDs of the MAR estimates and, as a result, the differences in the MSE among these estimates are mainly determined by the differences in the size of the bias. Therefore, in Figure 1 we plot only the bias for the purpose of comparison. The plots for other values of ρ lead to qualitatively similar conclusions and are reported as the Web Supplement Figures 3 to 6. As shown in the figures, the MAR estimate $\hat{\beta}_1(0)$ is biased. A general pattern is that the larger the size of γ_1, the larger the size of the bias in $\hat{\beta}_1(0)$. This can be readily seen from the V-shaped bias function of $\hat{\beta}_1(0)$ (as a function of γ_1) in Figure 1. When $\eta_{\gamma_0}(x)$ is a linear function, all three adjusted estimates, $\hat{\beta}_{1L}(\gamma_1), \hat{\beta}_{1P}(\gamma_1)$, and $\hat{\beta}_{1G}(\gamma_1)$, are capable of removing the bias of the MAR estimate $\hat{\beta}_1(0)$ under Cases 1 and 2. This can be seen in Figure 1, as the bias functions of the three adjusted estimates are all flat at a close-to-zero value over γ_1 values for the linear functions. This indicates that ISNI is an accurate sensitivity index and can effectively reduce the bias of the MAR estimate when the missing data mechanism is correctly specified.

Though all three adjusted estimates remove the bias of the MAR estimator when $\eta_{\gamma_0}(x)$ is linear in x, the effectiveness in doing this can be very different for the other forms of $\eta_{\gamma_0}(x)$. We study the simulation results in the three key aspects. (1) If $\eta_{\gamma_0}(x)$ is actually quadratic or cubic, $\hat{\beta}_{1L}(\gamma_1)$ has a significant amount of bias under Case 2. This can be seen from the V-shaped bias function of $\hat{\beta}_{1L}(\gamma_1)$ for Quadratic and Cubic in Figure 1(b). In contrast, both $\hat{\beta}_{1P}(\gamma_1)$ and $\hat{\beta}_{1G}(\gamma_1)$ perform much better in removing the bias, as shown by their flat bias functions at a close-to-zero value in these figures. This shows that the misspecification of the missing data model can lead to large bias for the adjusted parameter estimates in the complete data model, and it can be important to choose a proper missing data model. (2) Interestingly, in Case 1 the bias of $\hat{\beta}_{1L}(\gamma_1)$ is almost the same as those of $\hat{\beta}_{1P}(\gamma_1)$ and $\hat{\beta}_{1G}(\gamma_1)$, even for the quadratic and cubic forms of $\eta_{\gamma_0}(x_i)$. This shows that $\hat{\beta}_{1L}(\gamma_1)$ has a certain degree of robustness with respect to the misspecification of the missing data model. (3) When $\eta_{\gamma_0}(x_i)$ follows a sine form, both $\hat{\beta}_{1L}(\gamma_1)$ and $\hat{\beta}_{1P}(\gamma_1)$ have sizable biases
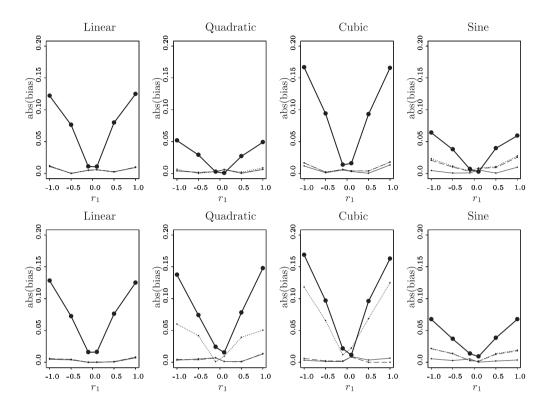
Figure 1. (a) Upper Panel: Plot of bias of four estimates for case 1 and $\beta_1 = \rho = 0$.

(b): Lower panel: Plot of bias of four estimates for case 2 and $\beta_1 = \rho = 0$.

The thick solid line: the MAR estimate $\hat{\beta}_1(0)$.

The dotted line: the adjusted estimate using the linear predictor $\hat{\beta}_{1L}(\gamma_1)$.

The dashed line: the adjusted estimate using the polynomial predictor $\hat{\beta}_{1P}(\gamma_1)$.

The thin solid line: the adjusted estimate using the smoothing spline predictor $\hat{\beta}_{1G}(\gamma_1)$.

and $\hat{\beta}_{1G}(\gamma_1)$ performs best. This shows that it can be important to use a data-driven approach, such as a GAM, to model the continuous predictors in the missing data model.

These findings from the simulation studies suggest that the adjusted estimator based on ISNI_L, which assumes a linear logistic regression, has a certain degree of robustness to misspecification of the missing data mechanism. There can, however, be situations in which ISNI_L is seriously affected by the misspecification of functional forms in the missing data model. In this case, both ISNI_P and ISNI_G can protect one from having a misleading assessment of the potential changes in the estimates. In particular, ISNI_G performs better because of

its more robust and automated process to discover the complex missing data process. Because of the availability of standard software for fitting a GAM, the success of ISNI_G in reducing the bias of the MAR estimates depends less on the experience of the data analyst to detect model misspecification than does ISNI_P.

6. An Application

Mroz (1987), used a wage offer dataset to demonstrate the sensitivity of empirical econometric analysis to various economic and statistical assumptions. Many of these assumptions, though useful, are often untestable, and thus it is insufficient to base conclusions solely on a single analysis. A more prudent approach is to compare the analysis with those obtained under alternative assumptions. If conclusions are reasonably robust, one has more confidence in the conclusions drawn. To demonstrate our method, we mainly focus on the potential misspecification of functional forms in the missing data mechanism.

The aim is to estimate the wage offered to a married woman as a function of education level and experience, after controlling for her other observed characteristics. That is, we are interested in estimating the linear regression model

$$lwage_i = (Intecept, educ, exper, expersq, age, nwifeinc, kidslt6, kidsge6)_i^t \beta + \epsilon_i,$$
(6.1)

where $i = 1, \ldots, 753$, $lwage$ is the logarithm of the wife's wage, $educ$ is the wife's years of schooling, $exper$ is the wife's labor market experience and $expersq$ is the square term of $exper$, age is the wife's age; $nwifeinc$ is the non-wife family income, $kidslt6$ is the number of children less than 6 years, and $kidsge6$ is the number of children between 6 and 18 inclusive.

Not every married woman in the sample had her wage outcome observed. Among the 753 married women in the sample, 325 did not participate in the labor force and as a result their potential wage outcomes, if employed, were missing. It is possible that the participation of a married woman in the labor force depends on her potential wage outcome, even after conditioning on the other observed variables. In order to account for this potential nonignorability, we assume the following model for self-selection to employment:

$$logit(P(G_i = 1)) = \gamma_{00} + \eta_1(educ_i) + \eta_2(exper_i) + \eta_3(age_i) + \eta_4(nwifeinc_i)$$
$$+ kidslt6_i\gamma_{05} + kidsge6_i\gamma_{06} + lwage_i\gamma_1,$$

where G_i is the indicator variable for participation in the labor force. As a comparison, we also consider the linear logistic labor participation model

$$logit(P(G_i = 1)) = (intercept, educ, exper, age, nwifeinc, kidslt6, kidsge6)_i^T \gamma_0$$
$$+ lwage_i\gamma_1.$$

Table 1. ISNIs for Parameter Estimates in the Wage Offer Dataset.

Predictor	MAR Est.	S.E.	Linear Logistic			GAM + Linear $\eta_{\gamma_1}(y)$		
			ISNI	c	MAR Est. +ISNI/σ	ISNI	c	MAR Est. +ISNI/σ
Intercept	-0.36000	0.32000	-0.26000	0.87	-0.72000	-0.19000	1.20	-0.6200
educ	0.10000	0.01500	0.01900	0.57	0.12600	0.01800	0.61	0.1240
exper	0.04000	0.01300	0.02400	0.40	0.07300	0.01600	0.57	0.0640
expersq	-0.00075	0.00039	-0.00050	0.60	-0.00150	-0.00022	1.29	-0.0010
nwifeinc	0.00570	0.00330	-0.00350	0.67	0.00079	-0.00170	1.43	0.0034
kidslt6	-0.05600	0.08800	-0.12000	0.53	-0.22100	-0.12000	0.53	-0.2200
kidsge6	-0.01800	0.02800	0.00700	2.76	-0.00770	0.00400	4.92	-0.0120
age	-0.00350	0.00500	-0.00600	0.67	-0.01100	-0.00700	0.57	-0.0130

Table 1 presents the MAR analysis that shows that education level and experience have statistically significant positive effects on the wage offer after adjustment by the other variables. Table 1 also presents the ISNI values that evaluate the effect of nonignorable missingness of wage outcome on the model estimates. When Y is continuous, the ISNI value depends on the scale of Y. Thus, to better gauge the sensitivity, we use the sensitivity transformation statistic $c = |S.E./\{\text{ISNI}/\sigma\}|$, where $S.E.$ is the standard error of a parameter estimate under the MAR model and σ is the standard deviation of Y. Thus, for ISNI to be equal to one $S.E.$, $|\gamma_1|$ needs to be at least c/σ which, under the logit link, corresponds to a magnitude of nonignorability such that a change of σ/c in Y is associated with a change of $e^1 = 2.7$ in the odds of being observed. Thus, the c statistic represents the critical magnitude of nonignorability above which the bias due to nonignorable missingness is larger than the sampling error and therefore causes concern. The smaller the c statistic, the larger the sensitivity to nonignorability. Following Troxel, Ma and Heitjan (2004), we suggest using $c = 1$ as a cutoff value for important sensitivity as this implies that the bias is the same size as the sampling error for a nonignorability where a change of one σ in Y is associated with a change of 2.7 in the odds of being observed.

The c statistics summarized in Table 1 show that the MAR estimates of both *educ* and *exper* are sensitive to nonignorable missingness in the outcome. Both MAR estimates have c statistics less than 1, and this is so when the missing data model is either logistic or GAM. Thus, the conclusions regarding the sensitivity of these two estimates are robust to the choice of missing data model. The conclusions regarding *expersq* and *nwifeinc*, however, depend on the missing data model chosen. With the linear logistic model, we find that the MAR estimates for *expersq* and *nwifeinc* have c statistics less than 1, indicating that both MAR estimates are sensitive to nonignorable selection for labor force participation. With

the GAM model, these MAR estimates have c statistics larger than 1, indicating that these two MAR estimates are not sensitive to nonignorability.

Using ISNI, we can also calculate the adjusted estimates when γ_1, the parameter for nonignorable selection, is perturbed from zero. A positive value of γ_1 is plausible because it is highly unlikely that one would decline a job offer when the offered wage was high. Here we consider $\gamma_1 = 1/\sigma$. This corresponds to a magnitude of nonignorability where a change of one standard deviation in *lwage* corresponds to the odds ratio of labor force participation of 2.7. In the wage offer dataset, the MAR estimate of σ is 0.72. Therefore, as the offered wage changes by a factor of $e^{0.72} = 2.1$, the odds of labor force participation change by 2.7. This seems to be a moderate nonignorability. The resulting adjusted estimates for this moderate nonignorability are reported under the column "MAR Est. + ISNI/σ" in Table 1. With this γ_1 value, we see that the adjusted estimates for *educ* and *exper* become larger than the corresponding MAR estimates, which implies that the MAR estimates likely underestimate the true effects of education and experience. It is also important to note that the adjusted estimate (-0.0015) for *expersq* under the linear logistic model is almost 50% larger in size than that (-0.0010) under the GAM for missing data process. This can be a significant difference in estimating the effect of working experience on the wage outcome.

To explore possible reasons for the discrepancy in ISNI values between the linear logistic and the GAM labor participation model we plot, in Figure 2, the smooth fitted functions of the four continuous predictors for labor force participation obtained from the GAM model. The figure shows that the relationship between experience and labor force participation is nonlinear. A chi-square test shows that this nonlinear trend is statistically significant (p-value= 0.01). It is plausible that this nonlinear relationship between experience and labor force participation drives the difference in the ISNI values.

The above ISNI analysis assumes that a logit transformation of the probability of missingness depends on *lwage* in a linear form. In Section S.1 of the Web Supplement, we conduct additional analyses in which the missingness depends on *lwage* in a quadratic form. The analysis shows smaller assessments of sensitivity for some parameter estimates.

7. Discussion

In this article, we propose using a semiparametric approach to adaptively choose the functional form of the continuous predictors for missingness. We have investigated the consequences of misspecifying a nonignorable missing data model through a simulation study and the analysis of a wage offer dataset. Specifically, we investigated the performance of ISNI, a recently proposed local sensitivity index of nonignorability, under the misspecification of missing data model. We
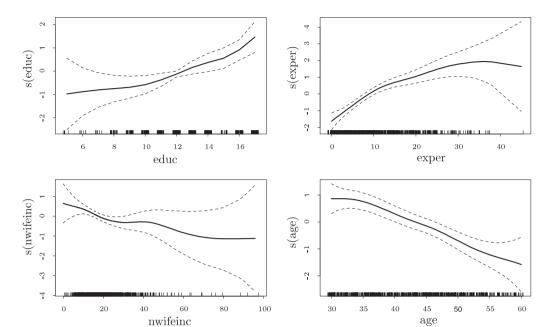
Figure 2. Plots of smooth terms in a generalized additive model for labor force participation in the Wage Offer data. The dashed lines are 95% pointwise confidence intervals.

found that it has some robustness to misspecification of the functional form of the predictors for missingness. There are, however, important situations in which the consequences of misspecification in the missing data mechanism are significant. In these cases, using more flexible missing data models can help protect the analysis. We recommend the semiparametric sensitivity index that uses a GAM approach for modeling the missing data process because of its modeling generality and the automatic nature of the procedure. In particular, the automation of the procedure is also an important benefit, especially when many continuous predictors for missingness exist and the way they affect missingness is not well understood. In these situations, it is cumbersome, if not infeasible, to manually choose proper higher-order terms and/or transformation for each continuous predictor. The more automated fitting of the missing data mechanism that uses a GAM substantially reduces the time and effort invested. This is particularly helpful in light of the fact that modeling missing data mechanism is usually not of primary interest, but has to be properly dealt with in order to draw correct conclusions about the main interest of a study.

The sample sizes in our analyses are reasonably large, of the sort commonly observed in practice. When data are sparse, GAM, as a nonparametric method, might not perform well. Then, one might consider using recently developed sparse

additive model techniques (Ravikumar et al. (2009)) that combine the ideas from sparse modeling and additive non-parametric regression.

The proposed semiparametric index is substantially easier to compute than global sensitivity because there is no need to fit any nonignorable model. Thus, it can be ideal for quickly and robustly measuring the sensitivity of a standard analysis to nonignorable missingness. If the sensitivity is small, then the standard analysis is considered trustworthy; otherwise, one might need to collect more data to better understand the missing data mechanism (Hirano, Ridder and Rubin (2001), Qian (2007)). The semiparametric index can be useful to robustly identify situations in which one may need to take that route.

We have extended ISNI to situations in which missingness depends on the missing outcome through a polynomial function, and have derived explicit ISNI formulas when the nonignorable missingness follows a quadratic form. This extension makes the index applicable to a broader range of applications in which investigators suspect that the nonignorable missingness might be of a complex relationship and would like to investigate sensitivity under such a belief.

The proposed method can be generalized to multivariate outcomes with non-ignorable missingness. Xie and Qian (2010) develop local sensitivity methods for various types of longitudinal data with both dropout and intermittent missingness, resulting in a general pattern of missingness. In their application, the predictors for the missingness are categorical variables. In other longitudinal applications in which the missingness predictors contain continuous variables, a linear logistic missing data model may lead to erroneous conclusions. In this case, the proposed semiparametric index method can be extended to provide a more robust way to measure the impact of nonignorable missingness in longitudinal data analysis.

Acknowledgements

Appendix: ISNI for GLM when $\eta_{\gamma_1}(y)$ Is of a Quadratic Form

In this Appendix, we derive explicit ISNI formulas when $\eta_{\gamma_1}(y) = \gamma_{11}y + \gamma_{12}y^2$. Specifically, we develop these formulas when the outcome Y_i follows a generalized linear model (GLM) that assumes that Y_i is independent with density

$$f_\theta(y_i) = \exp\left\{\frac{y_i\lambda_i(\beta) - b(\lambda_i(\beta))}{a(\tau)} + c(y_i, \tau)\right\},$$

where λ_i is the canonical parameter, functions $b(\cdot)$ and $c(\cdot, \cdot)$ determine a particular distribution in the exponential family, and $a(\tau) = \tau/w$, where τ is the dispersion parameter and w is a known weight. Because the quadratic does not apply to binary outcomes, we derive the ISNI formulas for other common cases of GLM. In the derivation, we write $\eta_{\gamma_1}(y) = \gamma_{11}(y + r_2 y^2)$, where $r_2 = \gamma_{12}/\gamma_{11}$. The ISNI formula for a linear $\eta_{\gamma_1}(y)$ can be obtained by setting $r_2 = 0$.

Normal Distribution

For a normal linear model, $Y_i \sim N(x_i^T \beta, \tau)$, where $\tau = \sigma^2$. Then $E(Y^2) = E^2(Y) + \tau$ and, according to (4.2), for a given value of r_2 the index for the regression parameter is

$$\text{ISNI}_r = -\hat{\tau}(\sum_{i:g_i=1} x_i x_i^T)^{-1} \sum_{i:g_i=0} (1 + 2r_2 \hat{\mu}_i) x_i h_i,$$

where $\hat{\mu}_i = x_i^T \hat{\beta}$, and $\hat{\beta}$ and $\hat{\tau}$ are the MAR estimates of β and τ, respectively.

Poisson Distribution

For Poisson outcome, we have $E(Y^2) = E^2(Y) + E(Y)$. With the canonical log link: $\ln E(Y_i) = \ln \mu_i = x_i^T \beta$, and the dispersion parameter $\tau = 1$, according to (4.2) for a given value of r_2, the index for the regression parameter is

$$\text{ISNI}_r = -(\sum_{i:g_i=1} \exp(x_i^T \hat{\beta}) x_i x_i^T)^{-1} \sum_{i:g_i=0} (1 + r_2 + 2r_2 \hat{\mu}_i) \exp(x_i^T \hat{\beta}) x_i h_i,$$

where $\hat{\mu}_i = \exp(x_i^T \hat{\beta})$, and $\hat{\beta}$ is the MAR estimate of β.

Gamma Distribution

If the dispersion parameter $\tau = \nu^{-1}$, then τ is also the constant coefficient of variation. For the Gamma distribution, we have $E(Y^2) = ((\nu + 1)/\nu)E(Y)^2$. With the canonical reciprocal link $(E(y_i))^{-1} = \mu_i^{-1} = x_i^T \beta$ then, according to (4.2), for a given value of r_2, the index for the regression parameter is

$$\text{ISNI}_r = \frac{1}{\hat{\nu}}(\sum_{i:g_i=1} (x_i^T \hat{\beta})^{-2} x_i x_i^T)^{-1} \sum_{i:g_i=0} (1 + 2r_2 \hat{\mu}_i \frac{\hat{\nu} + 1}{\hat{\nu}})(x_i^T \hat{\beta})^{-2} x_i h_i,$$

where $\hat{\mu}_i = 1/x_i^T \hat{\beta}$, and $\hat{\beta}$ and $\hat{\nu}$ are the MAR estimates of β and ν, respectively.

Inverse Gaussian Distribution

For the inverse Gaussian distribution, $E(Y^2) = E(Y)^2 + \tau E(Y)^3$, where $a(\tau) = 1/\tau$. With the canonical link $E(Y_i)^{-2} = \mu_i^{-2} = x_i^T \beta$ then, according to (4.2), for a given value of r_2, the index for the regression parameter is

$$\text{ISNI}_r = 2\hat{\tau}(\sum_{i:g_i=1} (x_i^T \hat{\beta})^{-3/2} x_i x_i^T)^{-1} \sum_{i:g_i=0} (1 + 2r_2 \hat{\mu}_i + 3r_2 \hat{\tau} \hat{\mu}_i^2)(x_i^T \hat{\beta})^{-3/2} x_i h_i,$$

where $\hat{\mu}_i = (x_i^T \hat{\beta})^{-2}$, and $\hat{\beta}$ and $\hat{\tau}$ are the MAR estimates of β and τ, respectively.

References

Chen, Q. and Ibrahim, J. G. (2006). Semiparametric models for missing covariate and response data in regression models. *Biometrics* **62**, 177-184.

Copas, J. B. and Eguchi, S. (2001). Local sensitivity approximations for selectivity bias. *J. Roy. Statist. Soc. Ser. B* **63**, 871-895.

Copas, J. B. and Li, H. G. (1997). Inference for non-random samples. *J. Roy. Statist. Soc. Ser. B* **59**,55-95.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall, London.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19**, 2244-2253.

Hirano, K., Ridder, G. W. and Rubin, D. B. (2001). Combining panels with attrition and refreshment samples. *Econometrica* **69**, 1645-1659.

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statist. Medicine* **17**, 2723-2732.

Little, R. J. A. (1995). Modeling the drop-out mechanism in longitudinal studies. *Journal of the American Statistical Association* **90**, 1112-1121.

Ma, G., Troxel, A. B. and Heitjan, D. F. (2005). An index of local sensitivity to nonignorable dropout in longitudinal modeling. *Statist. Medicine* **24**, 2129-2150.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica* **55**, 765-799.

Qian, Y. (2007). Do national patent laws stimulate domestic innovation in a global patenting environment? A cross-country analysis of pharmaceutical patent protection, 1978-2002. *Rev. Econom. Statist.* **89**, 436-453.

Ravikumar, P., Lafferty, J., Liu, H. and Wasserman, L. (2009). Sparse additive models. *J. Roy. Statist. Soc. Ser. B* **71**, 1009-1030.

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

Scharfstein, D., Rotnizky, A. and Robins, J. M. (1999). Adjusting for nonignorable dropout using semi-parametric models. *J. Amer. Statist. Assoc.* **94**, 1096-1146.

Troxel, A. B. (1998). A comparative analysis of quality of life data from a Southwest Oncology Group randomized trial of advanced colorectal cancer. *Statist. Medicine* **17**, 767-779.

Troxel, A. B., Harrington, D. P. and Lipsitz, S. R. (1998). Analysis of longitudinal measurements with non-ignorable non-monotone missing values. *Appl. Statist.* **47**, 425-438.

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

Vach, W. and Blettner, M. (1995). Logistic regression with incompletely observed categorical covariates - investigating the sensitivity against violation of the missing at random assumption. *Statist. Medicine* **14**, 1315-1329.

Xie, H. (2008). A local sensitivity analysis approach to longitudinal non-Gaussian data with non-ignorable dropout. *Statistics in Medicine* **27**, 3155-3177.

Xie, H. (2009). Bayesian inference from incomplete longitudinal data: A simple method to quantify sensitivity to nonignorable dropout. *Statistics in Medicine* **28**,2725-2747.

Xie, H. and Heitjan, D. F. (2004). Sensitivity analysis of causal inference in a clinical trial subject to crossover. *Clinical Trials* **1**, 21-30.

Xie, H. and Heitjan, D. F. (2009). Local sensitivity to nonignorability: Dependence on the assumed dropout mechanism. *Statist. Biopharmaceutical Res.* **1**, 243-257.

Xie, H. and Qian, Y. (2010). Measuring the impact of nonignorability in panel data with non-monotone nonresponse. *J. Appl. Econom.*, in press.

Zhang, J. and Heitjan, D. F. (2006). A Simple local sensitivity analysis tool for nonignorable coarsening: application to dependent censoring. *Biometrics* **62**, 1260-1268.

Zhang, J. and Heitjan, D. F. (2007). Impact of nonignorable coarsening on Bayesian inference. *Biostatistics* **8**, 722-743.

Division of Epidemiology & Biostatistics and Cancer Center, University of Illinois, Chicago, IL 60612, USA.

E-mail: huixie@uic.edu

Kellogg School of Management, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208, USA.

E-mail: yiqian@northwestern.edu

Department of Mathematics, Boise State University, 1910 University Dr., Boise ID 83725, USA.

E-mail: lqu@boisestate.edu