

NBER WORKING PAPER SERIES

A SEMIPARAMETRIC APPROACH FOR ANALYZING NONIGNORABLE MISSING  
DATA

Hui Xie  
Yi Qian  
Leming Qu

Working Paper 16270  
<http://www.nber.org/papers/w16270>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
August 2010

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Hui Xie, Yi Qian, and Leming Qu. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

A Semiparametric Approach for Analyzing Nonignorable Missing Data  
Hui Xie, Yi Qian, and Leming Qu  
NBER Working Paper No. 16270  
August 2010  
JEL No. C01,J16

### **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 which 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 non-linear relationship between missingness and the potentially missing outcome, whereas the existing literature commonly assumes a more restricted linear relationship. To avoid the 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 shows 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.

Hui Xie  
Department of Biostatistics  
School of Public Health  
University of Illinois at Chicago  
huixie@uic.edu

Leming Qu  
Department of Statistics  
Boise State University  
lqu@boisestate.edu

Yi Qian  
Department of Marketing  
Kellogg School of Management  
Northwestern University  
2001 Sheridan Road  
Evanston, IL 60208  
and NBER  
yiqian@kellogg.northwestern.edu

# A Semiparametric Approach for Analyzing Nonignorable Missing Data

Hui Xie\*, Yi Qian,<sup>†</sup> and Leming Qu<sup>‡</sup>

July 1, 2010

## SUMMARY

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 which 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 non-linear relationship between missingness and the potentially missing outcome, whereas the existing literature commonly assumes a more restricted linear relationship. To avoid the 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

---

\*Department of Epidemiology and Biostatistics, University of Illinois, Chicago, IL 60612.  
Email: huixie@uic.edu.

<sup>†</sup>Northwestern University. 2001 Sheridan Rd, Evanston, IL 60208.

<sup>‡</sup>Boise State University. 1910 University Dr., Boise ID 83725.

with the commonly used parametric approaches. The simulation study shows 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: Generalized Additive Model; MNAR; Semiparametric Joint Selection Model; Nonignorability; Sensitivity Analysis.

## 1. Introduction

Missing data arise frequently in studies across different disciplines, including public health, medicine, economics, and 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 where this assumption can be questionable. In the analysis of potentially nonignorable missing data, the selection model is a popular class of models, where 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. In situations where 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 very heavy and requires specialized programming. More importantly, such a joint selection model is often weakly or non-identified (Little 1995, Troxel 1998, Troxel et al. 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 et al. 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 computation burden, we utilize an index approach to local sensitivity (Troxel et al. 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 et al. 2005, Xie (2008, 2009), Qian and Xie 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. Furthermore, 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 (1) simple to perform by avoiding excessive additional computation and (2) 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 of the sensitivity index in the Appendix when the relationship follows a quadratic form. In Section 5, we conduct 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 bias of the MAR estimates. In Section 6, we apply the methodology to an application on estimating wage offer function. We conclude with a discussion in Section 7.

## 2. Selection Model for Nonignorable Missingness

Consider that we have data  $(Y_i, X_i, Z_i, G_i)$  from the unit  $i$ ,  $i = 1, \dots, n$ . The underlying ideal outcome,  $Y_i$ , arises independently from a distribution with a probability density function  $f_\theta(Y_i|X_i)$ , where  $X_i$  contains a set of fully observed covariates. We are interested in drawing inferences on  $\theta$  or a subset of it. Here for the purpose of this presentation, we restrict our attention to the case where  $Y_i$  is a 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 following missing data model  $f_\gamma(g_i|y_i, z_i)$ , where  $G_i|Y_i, Z_i \sim \text{Bernoulli}(P_\gamma(G_i = 1|Y_i, Z_i))$  and

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

$h$  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 unspecified now; Let  $\gamma = (\gamma_0, \gamma_1)$  where  $\gamma_0$  is a vector of parameters that associates the probability of missingness with observed data, and  $\gamma_1$  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 stacked data over all the units. We rewrite  $Y$  as  $(Y_{\text{obs}}, Y_{\text{mis}})$ , where  $Y_{\text{obs}}$  refers to the observed components of  $Y$  and  $Y_{\text{mis}}$  refers to the missing components of  $Y$ . The covariates  $X$  and  $Z$  are considered as fixed, and the conditioning on them in  $f_{\theta}(y_i|x_i)$  and  $f_{\gamma}(g_i|y_i, z_i)$  will be suppressed for notational simplicity. The data to be modeled are thus  $(Y_{\text{obs}}, G)$ . The correct log-likelihood of the model parameters is

$$L_C(\theta, \gamma; y_{\text{obs}}, g) \propto \ln f_{\theta, \gamma}(y_{\text{obs}}, g) = \ln \int f_{\gamma}(g|y_{\text{obs}}, y_{\text{mis}}) f_{\theta}(y_{\text{obs}}, y_{\text{mis}}) dy_{\text{mis}}. \quad (2)$$

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

$$L_I(\theta; y_{\text{obs}}) \propto \ln f_{\theta}(y_{\text{obs}}) = \ln \int f_{\theta}(y_{\text{obs}}, y_{\text{mis}}) dy_{\text{mis}}. \quad (3)$$

In practice, the simpler log-likelihood  $L_I$  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 thus 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  $L_C(\theta, \gamma; y_{\text{obs}}, g)$ .

We define:

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

One can then vary  $\gamma_1$  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 missing data process, that is, a restriction that  $\eta_{\gamma_0}(z_i) = \gamma_0^T z_i$  and  $\eta_{\gamma_1}(y_i) = \gamma_1 y_i$ . In this selection model, the likelihood  $L_C(\theta; \gamma_1)$  is proportional to the likelihood  $L_I(\theta)$  for all  $\theta \in \Omega_\theta$  when  $\gamma_1 = 0$ .  $\hat{\theta}(0)$  is then the MLE of  $\theta$  in the ignorable model. The difference between  $\hat{\theta}(0)$  and  $\hat{\theta}(\gamma_1)$  is a sensible measure of the sensitivity of the MLE when  $\gamma_1$  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 follows:

$$\hat{\theta}(\gamma_1) \approx \hat{\theta}(0) + \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1} \right|_{\gamma_1=0} \times \gamma_1, \quad (4)$$

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

$$\text{ISNI} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1} \right|_{\gamma_1=0} = -\nabla^2 L_{\theta, \theta}^{-1} \nabla^2 L_{\theta, \gamma_1}, \quad (5)$$

with

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

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  $\theta$  and  $\gamma_1$ . One limitation of ISNI, as a local sensitivity method, is that the above local approximation 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  $\theta$ , the parameters 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. That is, 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):

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

where  $Z_i$  is composed of  $m$  missingness predictors,  $(Z_1, \dots, Z_m)$ , and  $\eta_{0j}$  is an arbitrary smooth and mean zero function for the  $j$ th covariate  $Z_j$ ,  $j = 1, \dots, m$ . Because of the additivity in the nonparametric component, the model is termed as 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 will 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:

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

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  $\gamma_1$ . For any given value of  $\gamma_1$ , we must obtain  $(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1))$  by maximizing  $L_C^P$  over  $(\theta, \gamma_0)$ . An algorithm such as the EM algorithm can be used to maximize the above likelihood and obtain  $(\hat{\theta}(\gamma_1), \hat{\gamma}_0(\gamma_1))$ . The optimization can be heavy (e.g. taking a long time to converge) and it also requires specialized programming. Moreover, this optimization needs to be repeatedly performed for a range of  $\gamma_1$  values, which further compounds the computational burden.

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

$$\begin{aligned} \nabla^2 L_{\theta, \theta} &= \left. \frac{\partial^2 L_C^P}{\partial \theta \partial \theta^T} \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0} = \left. \frac{\partial \ln f_\theta(Y_{\text{obs}})}{\partial \theta \partial \theta^T} \right|_{\hat{\theta}(0)} \\ \nabla^2 L_{\theta, \gamma_1} &= \left. \frac{\partial^2 L_C^P}{\partial \theta \partial \gamma_1^T} \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0} = -h_i \left( \left. \frac{\partial E_\theta(Y_i)}{\partial \theta}, \dots, \frac{\partial E_\theta(Y_i^Q)}{\partial \theta} \right) \right|_{\hat{\theta}(0), \hat{\gamma}_0(0), \gamma_1=0}, \quad (6) \end{aligned}$$

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 following log-likelihood:

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

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 following penalized log-likelihood:

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

The conventional algorithm for the estimation of a GAM is the local scoring procedure (Hastie and Tibshirani 1990), which maximizes the above PELL. Equation (6) 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 missing). 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

$$\text{ISNI}_r = \left. \frac{\partial \hat{\theta}(\gamma_{11}, r)}{\partial \gamma_{11}} \right|_{\gamma_{11}=0} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} \left. \frac{\partial \gamma_1}{\partial \gamma_{11}} \right|_{\gamma_1=\gamma_{11} \cdot r} = \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} r, \quad (7)$$

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

$$\hat{\theta}(\gamma_1) - \hat{\theta}(0) \approx \left. \frac{\partial \hat{\theta}(\gamma_1)}{\partial \gamma_1^T} \right|_{\gamma_1=0} \gamma_1 = \text{ISNI}_r * \gamma_{11}. \quad (8)$$

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 conduct simulation studies to compare the performance of the parametric and semiparametric approaches for modeling the missing data process. Specifically, we simulate data from both linear and nonlinear missing data models, and then investigate if the ISNI based on a GAM missing data model provides a more faithful and robust adjustment of an MAR estimate than those based on various linear logistic missing data models. We follow the steps below to perform the simulation studies.

**Step 1:** Generate the hypothetical complete data,  $(Y_i, X_i)$ , independently from the following 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  $i = 1, \dots, n$  and the sample size  $n = 500$ . The parameter  $\rho$  takes the value of -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 following missing data model:

$$\text{logit}(P_\gamma(G_i = 1|Y_i = y_i, X_i = x_i)) = \eta_{\gamma_0}(x_i) + \gamma_1 y_i, \quad (9)$$

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  $\gamma_1$  to be  $(-1, -0.5, -0.1, 0.1, 0.5, 1)$ . The proportion of missingness in  $Y$  ranges from 15% to 65% in the generated datasets.

**Step 3:** With each generated dataset, we 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 Equation (4), we calculate three ISNI-adjusted estimates as follows:

$$\begin{aligned}
\hat{\beta}_{1L}(\gamma_1) &= \hat{\beta}_1(0) + \text{ISNI}_L \times \gamma_1, & \hat{\beta}_{1P}(\gamma_1) &= \hat{\beta}_1(0) + \text{ISNI}_P \times \gamma_1 \\
\hat{\beta}_{1G}(\gamma_1) &= \hat{\beta}_1(0) + \text{ISNI}_G \times \gamma_1. & &
\end{aligned} \tag{10}$$

The three ISNIs,  $\text{ISNI}_L$ ,  $\text{ISNI}_P$ , and  $\text{ISNI}_G$ , are listed in order of increasing generality to model the missing data process. The most constrained one is  $\text{ISNI}_L$ , whose calculation assumes *a priori* that  $\eta_{01}(x_i)$  in Equation (9) is linear in  $x_i$  as  $\eta_{01}(x_i) = \gamma_{00} + \gamma_{01}x_i$ . A more general one,  $\text{ISNI}_P$ , increases modeling flexibility by manually adding higher-order polynomial terms for  $x_i$  (i.e., quadratic term, cubic term,  $\dots$ ). This process stops when adding the next higher-order term of  $x_i$  into 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 seek more acceptable models for the missing data mechanism. The most general one, ISNI<sub>G</sub>, 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  $\eta_{\gamma_{01}}(x_i)$  and lets data tell the functional form of  $x_i$ . As compared with ISNI<sub>P</sub>, ISNI<sub>G</sub> enjoys two advantages. GAM is more general as it applies to arbitrary smooth functions whereas the parametric additive model applies to *a priori* specified parametric family (e.g. a polynomial family). Another important benefit with ISNI<sub>G</sub> is the automation of the procedure which avoids manually increasing the model complexity

For our simulation model, the formula for ISNI is derived to be

$$-\hat{\sigma}^2 \left( \sum_{i:g_i=1} \tilde{x}_i \tilde{x}_i^T \right)^{-1} \sum_{i:g_i=0} h_i \tilde{x}_i,$$

where  $\hat{\sigma}^2 = \frac{\sum_{i:g_i=1} (y_i - \hat{\beta}_0(0) - \hat{\beta}_1(0)x_i)^2}{\sum_i 1(g_i=1)}$  is the MAR estimate of the residual variance, and  $\hat{\beta}_0(0)$  and  $\hat{\beta}_1(0)$  are the MAR estimates of the  $\beta_0$  and  $\beta_1$ , respectively;  $\tilde{x}_i = [1, x_i]^T$  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<sub>L</sub>,  $h_i = h(\hat{\gamma}_{00}(0) + \hat{\gamma}_{01}(0)x_i)$ . For ISNI<sub>P</sub>,  $h_i = h(\hat{\gamma}_{00}(0) + \sum_{j=1}^J \hat{\gamma}_{0j}(0)x_i^j)$ , where  $J$  is the selected order of the polynomial function of  $x_i$ . For ISNI<sub>G</sub>,  $h_i = h(\hat{\gamma}_{00}(0) + \hat{\eta}_{\gamma_{01}}(x_i))$ , 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.

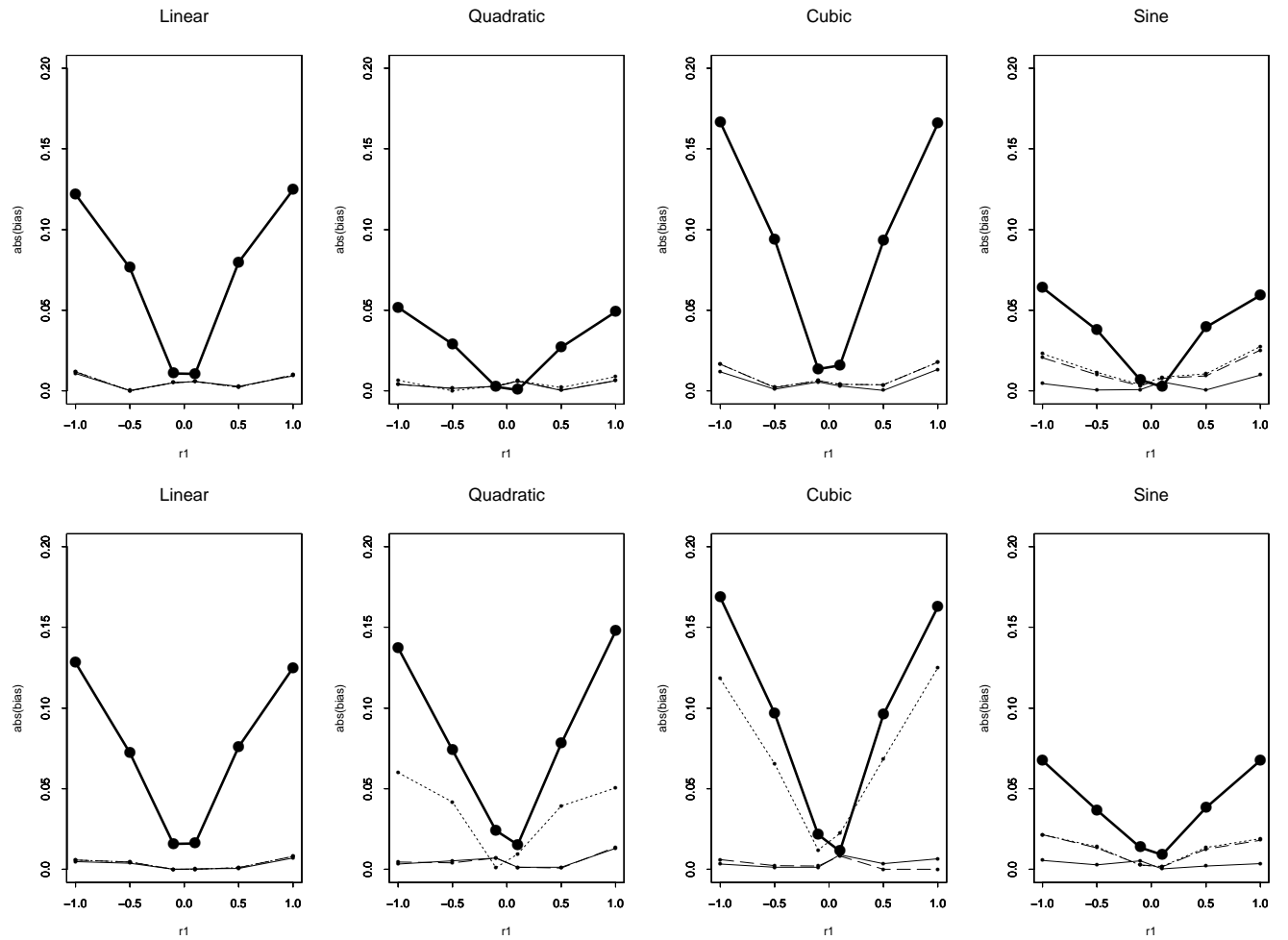
In practical applications, one can calculate the ISNI-adjusted estimates in Equation (10) for a plausible range of  $\gamma_1$  values, and investigate the sensitivity of the MAR estimates to nonignorable missingness. In the simulation studies, we plug in the true value of  $\gamma_1$ . The performance 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 for 300 times for the same values of  $\rho$  and  $\gamma_1$ . Using the resulting sample of estimates, we compute the mean squared error (MSE), bias, and standard deviation (SD) for each of the four estimators of  $\beta_1$ :  $\hat{\beta}_1(0)$ ,  $\hat{\beta}_{1L}(\gamma_1)$ ,  $\hat{\beta}_{1P}(\gamma_1)$ ,  $\hat{\beta}_{1G}(\gamma_1)$ . We then repeat Step 4 for other configurations of  $\rho$  and  $\gamma_1$ .

Figure 1 plots the bias when  $\beta_1 = \rho = 0$ . The results on SDs and MSEs can be found in Online Supplement Tables 1,2 and Figures 1 and 2. As shown there, the SDs for 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  $\rho$  lead to qualitatively similar conclusions and are reported as the online 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  $\gamma_1$ , 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  $\gamma_1$ ) in Figure 1. When  $\eta_{\gamma_{01}}(x)$  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 both 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  $\gamma_1$  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  $\eta_{\gamma_{01}}(x)$  is linear in  $x$ , the effectiveness in doing this can be very different for the other forms of  $\eta_{\gamma_{01}}(x)$ . We study the simulation results in the following three key aspects. (1) If  $\eta_{\gamma_{01}}(x)$  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 comparison, both  $\hat{\beta}_{1P}(\gamma_1)$  and  $\hat{\beta}_{1G}(\gamma_1)$  perform much better in removing the bias,

[hp]



**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 linear predictor,  $\hat{\beta}_{1L}(\gamma_1)$ .

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

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



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 form of  $\eta_{01}(x_i)$ . 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  $\eta_{01}(x_i)$  follows a sine form, both  $\hat{\beta}_{1L}(\gamma_1)$  and  $\hat{\beta}_{1P}(\gamma_1)$  have sizable biases, and  $\hat{\beta}_{1G}(\gamma_1)$  performs best. This shows that  $\text{ISNI}_G$  is most general, and 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  $\text{ISNI}_L$ , which assumes a linear logistic regression, has a certain degree of robustness to misspecification of missing data mechanism. There can be, however, situations where  $\text{ISNI}_L$  is seriously affected by the misspecification of functional forms in the missing data model. In this case, both  $\text{ISNI}_P$  and  $\text{ISNI}_G$  are useful to protect one from having a misleading assessment of the potential change of the estimates. In particular,  $\text{ISNI}_G$  performs better due to its modeling generality and more robust and automated process to discover the complex missing data process. Due to the availability of standard software for fitting a GAM, the success of  $\text{ISNI}_G$  in reducing the bias of the MAR estimates depends less on the experience of the data analyst to detect model misspecification, as compared with  $\text{ISNI}_P$ .

## 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 the conclusions are reasonably robust, one can have more confidence about the conclusions drawn. To demonstrate our method, we will mainly focus on the potential misspecification of functional forms in the missing data mechanism.

The interest of the empirical application is to estimate the wage offer outcome as a function of education level and experience, after controlling for other observed characteristics of an woman. That is, one is interested in estimating the following linear regression model:

$$lwage_i = (\text{Intecept}, educ, exper, expersq, age, nwifeinc, kidslt6, kidsge6)_i^T \beta + \epsilon_i, \quad (11)$$

where  $i = 1, \dots, 753$ ,  $lwage$  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 married women did not participate in the labor force and as a result their 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:

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

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

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

Table 1 presents the MAR analysis, which shows that education level and experience have statistically significant positive effects on the wage offer after the adjustment by the other variables. Table 1 also presents the ISNI values, which 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 a sensitivity transformation statistic,  $c = \left| \frac{S.E.}{ISNI/\sigma} \right|$ , where  $S.E.$  is the standard error of a parameter estimate under the MAR model and  $\sigma$  is the standard deviation of  $Y$ . The  $c$  value implies that for ISNI to be equal to one  $S.E.$ ,  $|\gamma_1|$  needs to be at least  $c/\sigma$ , which under the logit link corresponds to a magnitude of nonignorability such that a change of  $\sigma/c$  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 is, the larger the sensitivity to nonignorability is. Following Troxel et al. (2004), we suggest using  $c = 1$  as a cutoff value for important sensitivity as this implies that the bias will be in the same size as the sampling error for a moderate nonignorability where a change of one— $\sigma$  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 important estimates are robust to the choice of missing data model. The conclusions regarding *expersq* and *nwifeinc*, however, depend on the choice of the missing data model. With the linear logistic model, we find that the MAR estimates for *expersq* and *nwifeinc* have  $c$  statistics of 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 of larger than

[hp]

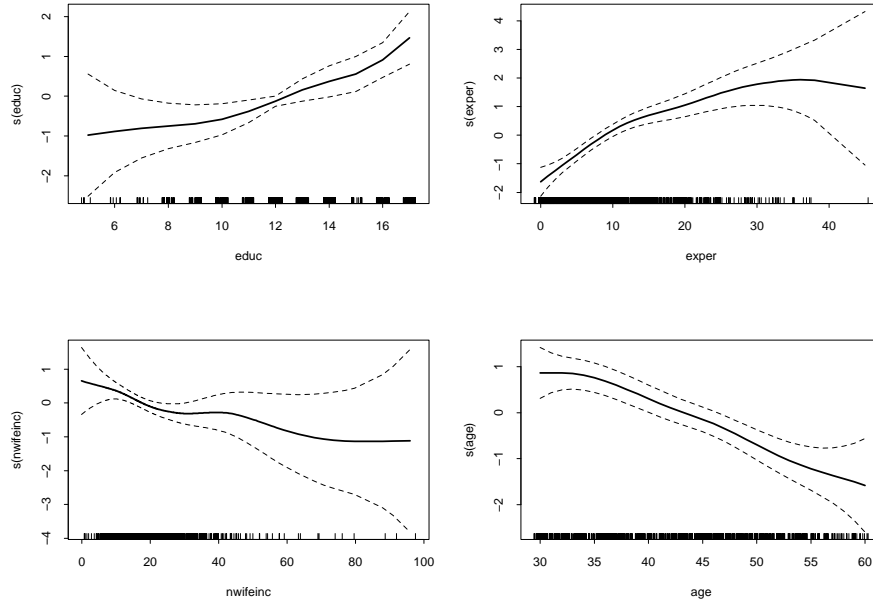
**Table 1**  
*ISNIs for Parameter Estimates in the Wage Offer Dataset.*

Predictor	MAR Est.	S.E.	Linear Logistic			GAM + Linear $\eta_{\gamma_1}(y)$		
			ISNI	$c$	MAR Est. + ISNI/ $\sigma$	ISNI	$c$	MAR Est. + ISNI/ $\sigma$
Intercept	-0.36	0.32	-0.26	0.87	-0.72	-0.19	1.20	-0.62
educ	0.10	0.015	0.019	0.57	0.126	0.018	0.61	0.124
exper	0.04	0.013	0.024	0.40	0.073	0.016	0.57	0.064
expersq	-0.00075	0.00039	-0.00050	0.60	-0.0015	-0.00022	1.29	-0.0010
nwifeinc	0.0057	0.0033	-0.0035	0.67	0.00079	-0.0017	1.43	0.0034
kidslt6	-0.056	0.088	-0.12	0.53	-0.221	-0.12	0.53	-0.220
kidsge6	-0.018	0.028	0.007	2.76	-0.0077	0.004	4.92	-0.012
age	-0.0035	0.005	-0.006	0.67	-0.011	-0.007	0.57	-0.013

1, indicating that these two MAR estimates are not sensitive to nonignorability.

Using ISNI, we can also calculate the adjusted estimates when  $\gamma_1$ , the parameter for nonignorable selection, is perturbed from zero. A positive value of  $\gamma_1$  is plausible because it is highly unlikely that one will decline a job offer when the offered wage is high. Here we consider  $\gamma_1 = 1/\sigma$ , which corresponds to a magnitude of nonignorability where a change of one standard deviation in *lwage* corresponds to the odds ratio of labor force participation being 2.7. In the wage offer dataset, the MAR estimate of  $\sigma$  is 0.72. Therefore, as the offered wage changes by a fold 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/ $\sigma$ ” in Table 1. With this  $\gamma_1$  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

[htp]



**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.

the true effects of education and experience. It is also important to note that the adjusted estimate (-.0015) for *expersq* under the linear logistic model is almost 50% larger than that (-.0010) under the GAM for missing data process. This can lead to a potentially significant difference in estimating the effect of working experience on the wage outcome.

To explore the possible reasons of 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 the section S.1 of the online Supplement, we conduct additional analyses where the missingness depends on *lwage* in a quadratic form. The analysis shows somewhat smaller assessment of sensitivity for some parameter estimates.

## 7. Discussion

It has been recognized that measuring the sensitivity of the inference to alternative missing data assumptions is an important component of data analysis. Such analysis often requires positing a missing data model. There typically exists little information to test the assumptions in the missing data model. Thus, it is desirable to utilize a model that covers a wide range of selection mechanisms. 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 using the simulation study and real data analysis. Specifically, we investigate the performance of ISNI, a recently proposed local sensitivity index of nonignorability, under the misspecification of missing data model. We found that ISNI has some robustness to misspecifying the functional form of the predictors for missingness. There exist, however, important situations where the consequence of misspecification in the missing data mechanism can be significant. In these cases, using more flexible missing data models can help protect the analysis from such misspecification. We recommend the semiparametric sensitivity index that uses a GAM approach for modeling missing data process, due to its modeling generality and the automated feature of the procedure. The semiparametric index enables us to model a larger class of missing data mechanisms than the usual linear logistic model or parametric nonlinear additive model does. The automation of the procedure is also an

important benefit, especially when many continuous predictors for missingness exist, and how 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 efforts invested in such a modeling exercise. This is particularly helpful in light of the fact that modeling the missing data mechanism is usually not of primary interest for a study, but has to be properly dealt with in order to draw correct conclusions about the main interest of the study.

The sample sizes in our analyses are reasonably large and are commonly seen in practice. When data are sparse, GAM, as a non-parametric method, might not perform well. In this scenario, one may consider using recently-developed sparse additive model techniques (Ravikumar et al. 2009), that combines the idea from sparse modeling and additive non-parametric regression.

The proposed semiparametric index is substantially easier to compute than the alternative global sensitivity method 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 et al. 2001, Qian 2007). The semiparametric index can be useful to robustly identify the situations where one may need to take the route.

In this article, we have also extended ISNI to situations where missingness depends on the missing outcome through a polynomial function. We have derived explicit ISNI formulas when the nonignorable missingness follows a quadratic form and illustrated its use in the wage offer dataset. This extension makes the index applicable to a broader range of applications where investigators suspect that the nonignorable missingness might be of a

complex relationship and would like to investigate the sensitivity under such a belief.

The proposed method can be generalized to multivariate outcomes with nonignorable missingness. Qian and Xie (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 all categorical variables. In other longitudinal applications where 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 method to measure the impact of nonignorable missingness in longitudinal data analysis.

#### ACKNOWLEDGEMENTS

We thank the Editor, the Associate Editor and the anonymous referees for many constructive comments that led to substantial improvements in the manuscript.

#### Appendix: ISNI for GLM when $\eta_{\gamma_1}(y)$ is of a quadratic form.

In this Appendix, we derive explicit ISNI formulas when  $\eta_{\gamma_1}(y)$  is a quadratic function, i.e.,  $\eta_{\gamma_1}(y) = \gamma_{11}y + \gamma_{12}y^2$ . Specifically, we develop these formulas when the outcome  $Y_i$  follows a generalized linear model (GLM), which 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  $\lambda_i$  is the canonical parameter; functions  $b(\cdot)$  and  $c(\cdot, \cdot)$  determine a particular distribution in the exponential family;  $a(\tau) = \tau/w$ , where  $\tau$  is the dispersion parameter and  $w$  is a known weight. Note that the quadratic function of  $\eta_{\gamma_1}(y)$  does not apply to binary outcomes. Thus, we derive the ISNI formulas for other common cases of GLM. In the derivation below,



we reparameterize  $\eta_{\gamma_1}(y) = \gamma_{11}(y + r_2 y^2)$ , where  $r_2 = \gamma_{12}/\gamma_{11}$ . The ISNI formula for  $\eta_{\gamma_1}(y)$  being a linear function 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 Equation (7), for a given value of  $r_2$ , the index for the regression parameter is:

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

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

### Poisson Distribution

For Poisson outcome, we have  $E(Y^2) = E^2(Y) + E(Y)$ . Assuming the canonical log link:  $\ln E(Y_i) = \ln \mu_i = x_i^T \beta$ , and the dispersion parameter  $\tau = 1$ , then according to Equation (7), for a given value of  $r_2$ , the index for the regression parameter is:

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

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

### Gamma Distribution

Let the dispersion parameter  $\tau = \nu^{-1}$ , and  $\tau$  also denotes the constant coefficient of variation. For Gamma distribution, we have  $E(Y^2) = \frac{\nu+1}{\nu} E(Y)^2$ . Assuming the canonical reciprocal link:  $(E(y_i))^{-1} = \mu_i^{-1} = x_i^T \beta$ , then according to Equation (7), for a given value of  $r_2$ , the index for the regression parameter is:

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

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

### Inverse Gaussian Distribution

For inverse Gaussian distribution,  $E(Y^2) = E(Y)^2 + \tau E(Y)^3$ , where  $a(\tau) = 1/\tau$ . Assuming the canonical link,  $E(Y_i)^{-2} = \mu_i^{-2} = x_i^T \beta$ , then according to Equation (7), for a given value of  $r_2$ , the index for the regression parameter is:

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

where  $\hat{\mu}_i = (x_i^T \hat{\beta})^{-2}$ , and  $\hat{\beta}$  and  $\hat{\tau}$  are the MAR estimates of  $\beta$  and  $\tau$ , 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 Li, H.G.(1997), “Inference for non-random samples,” *Journal of the Royal Statistical Society B*, **59**:55–95.
- Copas, J.B., and Eguchi, S.(2001), “Local sensitivity approximations for selectivity bias,” *Journal of the Royal Statistical Society B*, **63**: 871–895.
- Hastie, T. J., and Tibshirani, R.J.(1990), *Generalized additive models*, London: Chapman & Hall.
- Heitjan, D. F., and Rubin, D.B (1991), “Ignorability and coarse data,” *Annals of Statistics*, **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,” *Statistics in 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,” *Statistics in 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,” *The Review of Economics and Statistics*, **89(3)**, 436–453

- Qian, Y., and Xie, H. (2010), “Measuring the Impact of Nonignorability in Panel Data with Non-monotone Nonresponse,” *Journal of Applied Econometrics*, in press.
- Ravikuma, P., Lafferty, J., Liu, H. and Wasserman, L. (2009), “Sparse Additive Models,” *Journal of the Royal Statistical Society: Series B*, **71**: 1009-1030.
- Rubin, D.B. (1976), “Inference and missing data,” *Biometrika*, **63**: 581-592.
- Scharfstein, D., Rotnitzky, A., and Robins, J.M. (1999) “Adjusting for nonignorable dropout using semi-parametric models,” *Journal of the American Statistical Association*, **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,” *Statistics in 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,” *Applied Statistics*, **47**, 425–438.
- Troxel, A. B., Ma, G., and Heitjan, D. F. (2004), “An Index of Local Sensitivity to Nonignorability,” *Statistica 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,” *Statistics in Medicine*, **14**: 1315–1329.
- Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., and Kenward, M.G. (2001) “Sensitivity analysis for nonrandom dropout: a local influence approach,” *Biometrics*, **57**: 7–14.
- 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. (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. (2009), “Local sensitivity to nonignorability: Dependence on the assumed dropout mechanism,” *Statistics in Biopharmaceutical Research*, **1**(3): 243-257.
- 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.