Published in final edited form as:

Stat Med. 2014 May 10; 33(10): 1713–1722. doi:10.1002/sim.6057.

Use empirical likelihood to calibrate auxiliary information in partly linear monotone regression models

Baojiang Chen^{1,*} and Jing Qin^{2,*}

¹Department of Biostatistics, University of Nebraska Medical Center, Omaha, Nebraska, US 68198

²National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD 20892, US

Abstract

In statistical analysis, a regression model is needed if one is interested in finding the relationship between a response variable and covariates. When the response depends on the covariate, then it may also depends on the function of this covariate. If one has no knowledge of this functional form but expect for monotonic increasing or decreasing, then the isotonic regression model is preferable. Estimation of parameters for isotonic regression models is based on the pool-adjacent-violators algorithm (PAVA), where the monotonicity constraints are built in. With missing data, people often employ the augmented estimating method to improve estimation efficiency by incorporating auxiliary information through a working regression model. However, under the framework of the isotonic regression model, the PAVA does not work as the monotonicity constraints are violated. In this paper, we develop an empirical likelihood based method for isotonic regression model to incorporate the auxiliary information. Since the monotonicity constraints still hold, the PAVA can be used for parameter estimation. Simulation studies demonstrate that the proposed method can yield more efficient estimates, and in some situations, the efficiency improvement is substantial. We apply this method to a dementia study.

Keywords

Auxiliary information; empirical likelihood; isotonic regression; missing data

1. Introduction

The partly linear regression model

$$Y = \mathbf{X}^T \beta + h(Z) + \epsilon, \quad (1)$$

provides a convenient way to model the complex relationship between the response and covariates, where Y is a response variable, $\mathbf{X}$ is a $p \times 1$ vector, Z is a scalar, β is a $p \times 1$ regression parameter vector, $h(Z)$ is an unknown function of Z , and we assume that the error term ϵ is independent of (X, Z) and has zero mean. This model specification is appealing when the effects of $\mathbf{X}$ are of major interest and the effects of Z are nuisance. This allows us to make inference of the effects $\mathbf{X}$ while making minimal assumptions on Z . In practice, monotonic dependencies between responses and the covariate Z are common. For instance,

the probability of dementia can be monotonic function of age; cancer rates can increase with age and exposure, etc. Under the monotonic assumption of function h , model (1) becomes the isotonic regression model.

Previous studies on partly linear model (1) usually assume a smooth function h , and the penalized least square method is often used to estimate β and h . In the estimation procedure, a smoothing parameter or penalty parameter must be specified, and in applications, it may not be easy to specify an optimal one. Here we consider the isotonic regression model. One advantage of the isotonic regression method is that, there is no need to make a smooth assumption for h , allowing more flexibility relationship between the response and covariates. As a result, for the isotonic approach, there is no need to specify the smoothing parameter or penalty parameter. Ayer et al. [1] developed the pool-adjacent-violators algorithm (PAVA) to estimate the nonparametric function h , where the monotonicity constraints are built in, allowing for choosing the “smoothing” parameter automatically. Some related work includes [2], [3], [4], [5], et al.

Missing data are common in many settings, and in general, analyses based only on individuals with complete data can lead to invalid inferences when data are missing at random (MAR) or missing not at random (MNAR) [6]. Horvitz and Thompson [7] developed an inverse probability weighting method for cross-sectional data which can yield consistent estimates when data are MAR. The weights are obtained from models for the missing data process, and these models must be correctly specified for the resulting estimators to be consistent.

A disadvantage of the inverse probability weighting method is that the resultant estimate may not be efficient. To improve estimation efficiency, recent works on adjusting weighted by calibration or estimation could improve efficiency significantly (e.g., [8], [9], [10], [11]). Another popular method to improve estimation efficiency is to use an augmented estimating method (e.g., [8]), where we add an augmented estimating equation for the observed data. This method works well for generalized linear regression models. However, under the framework of the isotonic regression model, the augmented estimating method will bring trouble for the estimation, as the monotonicity constraints will no longer be met, thus the PAVA will not work, and new estimation procedure must be developed. In the framework of isotonic regression model, this is not an easy job.

As a nonparametric method, the empirical likelihood method was proposed by Owen [12] to construct confidence intervals for the mean or functional of mean. Recently, the empirical likelihood method has been widely used to calibrate auxiliary information to improve estimation efficiency. Among others, Chen and Sitter [13], Chen, Sitter and Wu [14] and Wu [15] have found excellent applications of this method in survey sampling. Cheng, Zhao, and Li [16] considered empirical likelihood method for additive isotonic regression models without missing data. In this paper, we explore the empirical likelihood method to the application in the isotonic regression model in the presence of missing data. This method can overcome the problem (monotonicity constraints do not hold) arising from the augmented estimating method by putting the working regression model into the empirical likelihood constraints. As a result, the monotonicity constraints still hold, and the PAVA still works. This method not only provides a simple estimation procedure but also yields an efficient estimate. In some situations, the efficiency improvement is substantial.

This paper is organized as follows. In Section 2, we give some notation and models. In Section 3, we develop the empirical likelihood based method for the isotonic regression model with missing data. Simulation studies are conducted in Section 4 to evaluate the

performance of the proposed method. Data from a dementia study are used in Section 5. We conclude the paper in Section 6.

2. Notation and models

In the following, without any explanation, we consider the isotonic regression model (1) by assuming h is a monotonic function. Let $\mu(X, Z) = X^T\beta + h(Z)$.

Suppose a sample of size n observations is collected where (Y, X, Z) are completely observed in a subsample, but (Y, Z) are completely observed in the remaining subsample. Let R be the missing indicator variable, which equals 1 if X is observed completely, and 0 otherwise. We consider a missing at random (MAR) mechanism, where we assume

$$P(R=1|Y, X, Z) = P(R=1|Y, Z) \equiv \pi(Y, Z).$$

Let $\pi(Y, Z, \alpha) = P(R=1|Y, Z, \alpha)$ be a specified probability distribution for a given $q \times 1$ parameter α . Typically, we build a generalized linear model for π , say, logistic regression, as follows:

$$\text{logit } \pi = \mathbf{u}^T \alpha, \quad (2)$$

where $\mathbf{u}$ is a function of $\{Y, Z\}$.

Let (y_i, x_i, z_i, r_i) be a realization of random vector (Y, X, Z, R) , $i = 1, \dots, n$. For simplicity, let the first n_1 subjects be completely observed.

3. Isotonic Regression with Missing Covariates

In the absence of missing data, the least-squares estimator $(\hat{\beta}, \hat{h})$ of (β, h) is obtained through minimizing

$$\sum_{i=1}^n (y_i - h(z_i) - x_i^T \beta)^2$$

subject to the constraints that β belongs to some convex subset in its domain, and $h(z_1) \leq h(z_2) \leq \dots \leq h(z_n)$ for $z_1 \leq z_2 \leq \dots \leq z_n$. Specifically, given an initial value of β_0 , we obtain the estimate of β and h through the following iterations,

- a. Regress $y - x^T \beta_0$ on covariate z for the estimate of h using the PAVA and Active Set Methods ([17]). Denote $\hat{h}(z)$ the estimate of $h(z)$.
- b. Regress $y - \hat{h}(z)$ on covariate x for the estimate of β , denote it as β_1 .
- c. Let $\beta_0 = \beta_1$, then repeat (a) and (b) until convergence.

Let $\hat{\beta}$ be the estimate of β obtained from the above iteration procedure. Under some regularity conditions, $\hat{\beta}$ is asymptotically normally distributed ([18]).

In the presence of missing data, following the idea of inverse probability weighting method, to obtain a consistent estimator of (β, h) , we may consider the weighted least-squares estimator by minimizing

$$\sum_{i=1}^n \frac{r_i}{\pi_i} (y_i - h(z_i) - x_i^T \beta)^2, \quad (3)$$

based on the same iteration algorithm. Under some regularity conditions, and if the missing data model is correctly specified, the resultant estimator is still consistent. However, this method only use the observed data, and (y_i, z_i) is not used when $r_i = 0$. As a result, this will reduce the estimation efficiency of the parameter, especially when the missing proportion is high.

Minimizing (3) with respect to β is equivalent to solving the estimating equation

$$\sum_{i=1}^n \frac{r_i x_i (y_i - x_i^T \beta - h(z_i))}{\pi_i} = 0. \quad (4)$$

When the missing data model π_i is correctly specified, the weighted estimating equation (4) yields a consistent estimate for β conditional that the marginal mean μ is correctly specified, as $E(r_i/\pi_i) = 1$. However, the estimate may not be efficient as only the observed data are used. To improve estimation efficiency, following the idea of the augmented estimating equation method ([8]), we add a function of β in the tangent space of the conditional distribution of $\mathbf{R}_i$, $\psi(y_i, z_i, \beta)$, so that information from the incomplete data will be used. Specifically, we may consider solving for the following estimating equation

$$\sum_{i=1}^n \left[\frac{r_i x_i (y_i - x_i^T \beta - h(z_i))}{\pi_i} - \frac{r_i - \pi_i}{\pi_i} \psi(y_i, z_i, \beta) \right] = 0, \quad (5)$$

for the estimation of β . In practice, there are many choice of $\psi(y_i, z_i, \beta)$, and the optimal one (the most efficient one) is proved to has the form

$\psi(y_i, z_i, \beta) = E[x_i(y_i - x_i^T \beta - h(z_i)) | y_i, z_i]$ ([8]). When the missing data model π_i is correctly specified, the augmented estimating equation (5) still yields a consistent estimate for β conditional that the marginal mean μ is correctly specified, as $E(r_i/\pi_i) = 1$ and $E\{(r_i - \pi_i)/\pi_i\} = 0$. In practical implementation, we may replace a “working regression model” for $\psi(y_i, z_i, \beta)$. Solving (5) for (β, h) is not easy as the monotonicity constraints will not hold, hence the PAVA algorithm does not work. Developing new estimation procedures for the augmented estimating equation under the framework of isotonic regression models may not be easy.

To overcome the difficulty arising from the augmented estimation equation method, we propose the following empirical likelihood approach to incorporate auxiliary information for the estimation of β , instead of solving for the augmented estimating equation (5). We consider the complete data likelihood ([19])

$$\prod_{i=1}^{n_1} \frac{\pi(y_i, z_i) dF(y_i, x_i, z_i)}{\int \pi(y, z) dF(y, x, z)},$$

where $F(y_i, x_i, z_i)$ is the distribution function of (y_i, x_i, z_i) . There are many parametric or nonparametric model specification of the distribution F ; here we focus on an empirical likelihood approach ([20]) by replacing the density dF in the likelihood by probability

masses $\{p_i: i = 1, \dots, n_1\}$ on the complete data. For the missing data probability distribution $\pi(y_i, z_i)$, we replace it by the parametric distribution $\pi(y_i, z_i, \alpha)$. With a known parameter α , the resulting empirical log likelihood function is

$$\ell(\theta, p) = \sum_{i=1}^{n_1} [\log(\pi(y_i, z_i, \alpha)) + \log p_i - \log(\theta)], \quad (6)$$

where $\theta = \int \pi(y, z) dF(y, x, z)$. Each (β, h) satisfies the constraints

$$\begin{aligned} p_i &\geq 0, \sum_{i=1}^{n_1} p_i = 1, \\ \sum_{i=1}^{n_1} p_i \{\pi(y_i, z_i, \alpha) - \theta\} &= 0, \\ \sum_{i=1}^{n_1} p_i \{\psi(y_i, z_i, \beta) - \bar{\psi}\} &= 0, \end{aligned}$$

where $\bar{\psi} = n^{-1} \sum_{i=1}^n \psi(y_i, z_i, \beta)$. Thus, the missing data and working regression models are incorporated in. By introducing Lagrange multipliers and profiling p_i 's, we obtain

$$p_i = \frac{1}{n_1} \frac{1}{1 + \lambda_1 \{\pi(y_i, z_i, \alpha) - \theta\} + \lambda_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}},$$

where $\lambda = (\lambda_1, \lambda_2^T)^T$ is a vector of the Lagrange multipliers ([20],[21]).

Substituting all p_i 's into (6), the loglikelihood becomes

$$\ell(\theta, \lambda_1, \lambda_2) = \sum_{i=1}^{n_1} \log \{\pi(y_i, z_i, \alpha)\} - \sum_{i=1}^{n_1} \log \{1 + \lambda_1 \{\pi(y_i, z_i, \alpha) - \theta\} + \lambda_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}\} - n_1 \log(\theta). \quad (7)$$

Differentiating $\ell(\theta, \lambda_1, \lambda_2)$ with respect to $\theta, \lambda_1, \lambda_2$ and setting them to 0, we obtain

$$-\frac{n_1}{\theta} + \sum_{i=1}^{n_1} \frac{\lambda_1}{1 + \lambda_1 \{\pi(y_i, z_i, \alpha) - \theta\} + \lambda_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}} = 0 \quad (8)$$

$$\sum_{i=1}^{n_1} \frac{1}{n_1} \frac{\{\pi(y_i, z_i, \alpha) - \theta\}}{1 + \lambda_1 \{\pi(y_i, z_i, \alpha) - \theta\} + \lambda_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}} = 0 \quad (9)$$

$$\sum_{i=1}^{n_1} \frac{1}{n_1} \frac{\{\psi(y_i, z_i, \beta) - \bar{\psi}\}}{1 + \lambda_1 \{\pi(y_i, z_i, \alpha) - \theta\} + \lambda_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}} = 0 \quad (10)$$

From (8), and $\sum_{i=1}^{n_1} p_i = 1$ we obtain that

$$\lambda_1 = \frac{1}{\theta}. \quad (11)$$

Let $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2^T)^T$ and $\hat{\theta}$ be the estimator of λ and θ by solving above estimating equations (9)-(11) with a replaced estimate $\hat{\alpha}$ of α . In general, the maximum likelihood estimate $\hat{\alpha}$ is often used by maximizing the binomial likelihood

$\sum_{i=1}^n [r_i \log \pi(y_i, z_i, \alpha) + (1 - r_i) \log \{1 - \pi(y_i, z_i, \alpha)\}]$ with respect to α . Then the estimated probability

$$\hat{p}_i = \frac{1}{n_1} \frac{1}{1 + \hat{\lambda}_1 \{\pi_i(y_i, z_i, \hat{\alpha}) - \hat{\theta}\} + \hat{\lambda}_2^T \{\psi(y_i, z_i, \beta) - \bar{\psi}\}}, i=1, \dots, n_1.$$

Note that, to obtain the estimated probability $\hat{p}_i$, the parameter β and the function $h(z_i)$ must be specified. In applications, this can be replaced by a consistent estimate, say, obtained from minimizing (3).

Finally, we minimize

$$\sum_{i=1}^{n_1} \hat{p}_i (y_i - h(z_i) - x_i^T \beta)^2$$

subject to the constraints that β belongs to some convex subset in its domain, and $h(z_1) \leq h(z_2) \leq \dots \leq h(z_n)$ for $z_1 \leq z_2 \leq \dots \leq z_n$. This can be easily implemented using the `gpava` function in the software R. Since the missing data and working regression models are incorporated in the probability masses $\hat{p}_i$, we expect that the resultant estimate of (β, h) should be consistent and more efficient than the estimate obtained from the model without

incorporating auxiliary information. In fact, as $\sum_{i=1}^{n_1} \hat{p}_i I(y_i \leq y, x_i \leq x, z_i \leq z)$ is a consistent estimator of distribution $F(y, x, z)$, as a general result of isotonic regression, our method can produce consistent estimation of β . However the asymptotic distribution is complicated in general. It is out of our scope. In applications, we suggest using a bootstrapping method to estimate the standard error of β . Empirical studies demonstrate that this method works well.

4. Simulation Study

In this section, we evaluate the performance of the proposed method estimate by comparing the estimate obtained from minimizing (3). We call this method as the weighted PAVA.

Throughout the simulation study, we use the following regression model:

$$y = \beta x + h(z) + \epsilon,$$

where $\epsilon \sim N(0, 0.25)$, $z \sim N(0, 1)$. We consider four $h(z)$ functions: $h(z) = \sin(z)$ if $z \in [-\pi/6, \pi/6]$, -1 if $z < -\pi/6$ and 1 if $z > \pi/6$; $h(z) = \text{round}(z)$; $h(z) = z$; $h(z) = z^{1/3}$. The true value is $\beta = 1$. We consider two situations: covariate x is continuous or binary.

For the missing data model, we consider

$$\text{logit } \pi_i = \alpha_0 + \alpha_1 y_i + \alpha_2 z_i.$$

The true values are $\alpha_1 = 1$, $\alpha_2 = 1$, and we change α_0 to vary the missing proportions.

We first consider that x is continuous, where we assume $x \sim N(0, 1)$. For the proposed method, we consider “working regression models”

$$E(x_i | y_i, z_i) = (1, y_i, z_i, y_i * z_i) \gamma_1$$

and

$$E(x_i^2 | y_i, z_i) = (1, y_i, z_i, y_i * z_i) \gamma_2.$$

We estimate γ_1 and γ_2 based on the observed data and denote the estimate $\hat{\gamma}_1$ and $\hat{\gamma}_2$. So, the “working regression functions” are

$$\psi(y_i, z_i; \beta, \hat{\gamma}_1, \hat{\gamma}_2) = \hat{x}_i y_i - \hat{x}_i h(z_i) - \hat{x}_i^2 \beta,$$

where $\hat{x}_i = (1, y_i, z_i, y_i * z_i) \hat{\gamma}_1$ and $\hat{x}_i^2 = (1, y_i, z_i, y_i * z_i) \hat{\gamma}_2$

To evaluate the performance of the proposed method, we consider two cases: $h(z)$ is a non-continuous or continuous function. Table 1 lists the results, where Std is the standard deviation based on 1000 simulations, and RE is the relative efficiency defined as the ratio of the variance between the proposed method estimate and the weighted PAVA estimate. Figure 1 is the pointwise estimate of the nonparametric function $h(z)$ with the true $h(z) = z^{1/3}$. It can be easily seen that both methods provide ignorable biases for the parametric parameter β and nonparametric function $h(z)$; the proposed method yields highly efficient estimates compared with the weighted PAVA method estimate no matter the $h(z)$ function is continuous or not. In some situations, the efficiency improvement is huge, for example, when sample size $n = 1000$, $\alpha_0 = -3$, and $h(z) = \sin(z)$ if $z \in [-\pi/6, \pi/6]$, -1 if $z < -\pi/6$ and 1 if $z > \pi/6$, the variance of the proposed method estimate is only 23.6% of that of the proposed method estimate. The efficiency improvement also depends on the missing proportion and the sample size; the higher the missing proportion (the smaller the α_0), the higher the relative efficiency (smaller RE) improvement; the bigger the sample size, the higher the relative efficiency (smaller RE) improvement. In this simulation situation, it seems that the efficiency improvement also depends on the continuity of the $h(z)$ function. As is seen, the proposed method yields more efficient estimate than the weighted PAVA estimate when $h(z)$ is non-continuous compared with continuous $h(z)$.

Next, We consider that x is binary, where we assume $x \sim \text{Bin}(1, 0.5)$. For the proposed method, we consider a “working regression model”

$$P(x_i=1|y_i, z_i) = \frac{\exp\{(1, y_i, z_i, y_i * z_i)\gamma_1\}}{1 + \exp\{(1, y_i, z_i, y_i * z_i)\gamma_1\}} = \eta(y_i, z_i, \gamma_1).$$

We estimate γ_1 based on the observed data through a logistic regression and denote the estimate $\hat{\gamma}_1$. The “working regression function” is

$$\psi(y_i, z_i; \beta, \hat{\gamma}_1) = \hat{\eta}_i(y_i - h(z_i) - \beta),$$

where $\hat{\eta}_i = \eta(y_i, z_i, \hat{\gamma}_1)$. Table 2 lists the results, where we only consider two functions: $h(z) = \sin(z)$ if $z \in [-\pi/6, \pi/6]$, -1 if $z < -\pi/6$ and 1 if $z > \pi/6$ and $h(z) = \text{round}(z)$. Similar conclusions are drawn as in the case that x is continuous.

5. Application to a Dementia Study

We apply the proposed method to the National Alzheimer's Coordinating Center (NACC) Uniform Data Set (UDS). One of the goals of the study is to evaluate the association between the depression status and the severity of cognitive impairment. The Mini- Mental State Examination (MMSE) test ([22]) is a test that is used to screen for cognitive impairment, which is commonly used in medicine to screen for dementia. The response of interest is the MMSE score, and the covariate of interest is the depressions status (yes/no). As is widely known, dementia is a progressive disease, and it gets worse over age. To eliminate the age effect, we must accommodate age into the model.

A sample of 9578 subjects enrolled in 2005 was used to study the relationship between the covariate and responses. However, there are some missing data for the covariate, depression status, which is a questionnaire based variable. We think the MAR mechanism may be reasonable because, for example, a patient's willing to complete the questionnaire may depend on his/her disease status. The missing proportion is about 42%.

Consider the following regression model for the response process

$$E(\text{MMSE} | \text{Age, Depression}) = \beta * \text{Depression} + h(\text{Age}).$$

For the missing data process, we build the following regression models:

$$\text{logit } \pi = \mathbf{u}^T \alpha,$$

where $\mathbf{u}$ includes functions of MMSE and Age.

In line with the simulation study, here we also use two methods to analyze the data: the proposed method and the weighted PAVA method. For the proposed method, we consider the “working regression model”

$$P(\text{Depression}=1 | \text{MMSE, Age}) = \frac{\exp\{(1, \text{MMSE, Age, MMSE} * \text{Age})\gamma\}}{1 + \exp\{(1, \text{MMSE, Age, MMSE} * \text{Age})\gamma\}} = \lambda(\text{MMSE, Age, } \gamma),$$

and we estimate γ based on the observed data through a logistic regression.

Table 3 lists the estimate for the response and missing data models. It can be seen that both methods give close estimates, but the proposed method yields a smaller standard error compared with the weighted PAVA method. All methods reveal that the depression status is negatively associated with the severity of the cognitive impairment. Figure 2 is the estimate of the nonparametric function $h(z)$.

For the missing data model, we carry out standard diagnostic tests for the fit of regression models by comparing a model with an expanded model to do a model selection and lists the final model in Table 3. Significance of MMSE and Age indicates that the missing at random mechanism may be reasonable.

6. Summary

In this paper, we develop an empirical likelihood based method to improve estimation efficiency in the framework of isotonic regression models. This proposed method overcomes the difficulty arising from the augmented estimating method that the monotonicity constraints do not hold. Hence, it provides a simple and easy estimation procedure, and it can be easily implemented using the gpava procedure in R. Simulation studies demonstrate that the proposed method not only provides consistent estimates for parameters, but it can yield more efficient estimates through incorporating auxiliary information. In some situations, the efficiency improvement is substantial.

In this paper, we only discuss that the auxiliary information is mainly the information on Y and Z from incomplete observations. The proposed method can also use other variables (auxiliary variables not in the regression model) to improve estimation efficiency by adding additional constraints into the empirical likelihood.

We focussed here primarily on the estimate regarding the covariate x that is subject to missing. The response and/or covariate z subject to missing are also very common in practice. A future research is to extend this method to the missing response and/or missing covariates (x or z) problem. The idea is similar to the proposed method but the techniques are a little different. This is still under investigation.

Acknowledgments

The authors thank the two referees for their constructive comments and suggestions. Dr. Chen was supported in part by National Institute on Aging grant U01AG016976. The NACC database is supported by National Institute on Aging grant U01AG016976.

References

1. Ayer M, Brunk HD, Ewing GM, Reid WT, Silverman E. An empirical distribution functions for sampling with incomplete information. *The Annals of Mathematical Statistics*. 1955; 26:641–647.
2. Cheng G. Semiparametric additive isotonic regression models. *Journal of Statistical Planning and Inference*. 2009; 139:1980–1991.
3. Rufibach K. An Active Set Algorithm to Estimate Parameters in Generalized Linear Models with Ordered Predictors. *Comput Statist Data Anal*. 2010; 54:1442–1456.
4. Du J, Zhang Z, Xie T. Variable Selection for Semiparametric Isotonic Regression Models. *International Journal of Intelligent Technologies and Applied Statistics*. 2011; 4:87–102.
5. Sun Z, Zhang Z, Du J. Semiparametric analysis of isotonic errors-in-variables regression models with missing response. *Communications in Statistics Theory and Methods*. 2012; 41:2034–2060.

6. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd. John Wiley and Sons, Inc.; 2002.
7. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; 47:663–685.
8. Robins JM, Rotnitzky A, Zhao LP. Estimation of Regression Coefficients When Some Regressor Are Not Always Observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
9. Deville JC, Sarndal CE. Calibration estimators in survey sampling. *Journal of the American Statistical Association*. 1992; 87:376–382.
10. Lawless JF, Kalbfleisch JD, Wild CJ. Semiparametric methods for response-selective and missing data problems in regression. *Journal of Royal Statistical Society Series B*. 1999; 61:413–438.
11. Breslow NE, Lumley T. Semiparametric models and two-phase samples: Applications to Cox regression. *Institute of Mathematical Statistics*. 2013; 9:65–77.
12. Owen A. Empirical likelihood ratio for confidence intervals for a single function. *Biometrika*. 1988; 75:237–249.
13. Chen J, Sitter RR. A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Stat Sinica*. 1999; 9:385–406.
14. Chen J, Sitter RR, Wu C. Using Empirical Likelihood Method to Obtain Range Restricted Weights in Regression Estimators for Surveys. *Biometrika*. 2002; 89:230–237.
15. Wu C. Optimal Calibration Estimators in Survey Sampling. *Biometrika*. 2003; 90:937–951.
16. Cheng G, Zhao Y, Li B. Empirical likelihood inferences for the semiparametric additive isotonic regression. *Journal of Multivariate Analysis*. 2012; 112:172–182.
17. de Leeuw J, Hornik K, Patrick M. Isotone optimization in R: pool-adjacent-violators-algorithm (pava) and active set methods. *Journal of Statistical Software*. 2009; 32:1–24.
18. Huang J. A note on estimating a partly linear model under monotonicity constraints. *Journal of Statistical Planning and Inference*. 2002; 107:343–351.
19. Qin J, Leung D, Shao J. Estimation with survey data under nonignorable nonresponse or informative sampling. *Journal of the American Statistical Association*. 2002; 27:193–200.
20. Owen, A. Empirical likelihood. Boca Raton: Chapman and Hall-CRC; 2001.
21. Qin J, Lawless J. Empirical likelihood and general estimating equations. *Annals of Statistics*. 1994; 22:300–325.
22. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975; 12:189–198. [PubMed: 1202204]

Figure 1. Empirical pointwise estimate for the nonparametric function $\hat{h}(z)$ based on the 1000 simulations: solid line is the true $h(z) = z^{1/3}$ function; circles are the average of the $\hat{h}(z)$ based on the 1000 simulations

Figure 2. Pointwise estimates of the nonparametric function $\hat{h}(\text{age})$ for the National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS)

Empirical biases and standard deviations of β for the proposed method and the weighted pava estimate: x is a continuous covariate; the missing proportions are about 80%, 50% and 30% with $\alpha_0 = -3, 0$ and 3 , respectively; RE is the ratio of the variance of the estimate between the proposed method and the weighted PAVA method.

Table 1

Proposed Method			Weighted PAVA			
n	α_0	Bias	Std	Bias	Std	RE
$h(z) = \sin(z)$ if $z \in [-\pi/6, \pi/6]$, -1 if $z < -\pi/6$ and 1 if $z > \pi/6$						
1000	-3	-0.011	0.025	-0.009	0.051	0.236
1000	0	-0.006	0.014	-0.008	0.026	0.286
1000	3	-0.009	0.012	-0.004	0.021	0.340
400	-3	-0.013	0.035	-0.019	0.057	0.380
400	0	-0.014	0.021	-0.007	0.032	0.424
400	3	-0.009	0.016	-0.006	0.024	0.456
$h(z) = \text{round}(z)$						
1000	-3	-0.013	0.026	-0.015	0.048	0.282
1000	0	-0.016	0.015	-0.008	0.026	0.326
1000	3	-0.011	0.010	-0.002	0.016	0.393
400	-3	-0.012	0.038	-0.018	0.066	0.337
400	0	-0.005	0.022	-0.007	0.035	0.397
400	3	-0.011	0.016	-0.003	0.023	0.521
$h(z) = z^{1/3}$						
1000	-3	-0.012	0.023	-0.014	0.039	0.339
1000	0	-0.005	0.015	-0.008	0.025	0.363
1000	3	-0.011	0.011	-0.004	0.015	0.490
400	-3	-0.010	0.034	-0.016	0.050	0.453
400	0	-0.016	0.022	-0.011	0.031	0.512
400	3	-0.011	0.016	-0.005	0.020	0.637
$h(z) = z$						
1000	-3	-0.012	0.025	-0.018	0.039	0.404
1000	0	-0.007	0.015	-0.010	0.022	0.450
1000	3	-0.002	0.010	-0.003	0.013	0.618

<i>n</i>	α_0	Proposed Method		Weighted PAVA		
		Bias	Std	Bias	Std	RE
400	-3	-0.013	0.042	-0.021	0.052	0.649
400	0	-0.016	0.024	-0.011	0.029	0.686
400	3	-0.011	0.016	-0.003	0.019	0.704

Empirical biases and standard deviations of β for the proposed method and the weighted pava estimate: x is a binary covariate; the missing proportions are about 80%, 50% and 30% with $\alpha_0 = -3, 0$ and 3 , respectively; RE is the ratio of the variance of the estimate between the proposed method and the weighted PAVA method.

Table 2

Proposed Method		Weighted PAVA				
n	α_0	Bias	Std	Bias	Std	RE
$h(z) = \sin(z)$ if $z \in [-\pi/6, \pi/6], -1$ if $z < -\pi/6$ and 1 if $z > \pi/6$						
1000	-3	-0.014	0.037	-0.015	0.056	0.310
1000	0	-0.009	0.032	-0.003	0.051	0.391
1000	3	0.000	0.019	-0.003	0.027	0.493
400	-3	-0.013	0.059	-0.018	0.084	0.487
400	0	-0.010	0.036	-0.003	0.051	0.496
400	3	0.000	0.031	-0.003	0.041	0.561
$h(z) = \text{round}(z)$						
1000	-3	-0.004	0.038	-0.014	0.069	0.303
1000	0	-0.002	0.021	-0.004	0.034	0.392
1000	3	0.001	0.017	-0.003	0.019	0.828
400	-3	-0.003	0.062	-0.015	0.089	0.487
400	0	-0.003	0.036	-0.006	0.050	0.497
400	3	0.002	0.029	-0.002	0.031	0.867

Table 3
 Estimation and inference for the National Alzheimer's Coordinating Center Uniform Data Set (NACC UDS)

Parameter	Proposed Method			Weighted PAVA		
	Estimate	StdErr	p-value	Estimate	StdErr	p-value
Response Model						
Depression	-0.4790	0.1383	0.0005	-0.5432	0.1904	0.0043
Missing Data Model						
(Intercept)	11.887	0.388	<0.001			
MMSE	-0.416	0.011	<0.001			
Age	-0.006	0.003	0.022			