Nonparametric Imputation of Missing Values for Estimating Equation Based Inference

DONG WANG

Department of Statistics, University of Nebraska, Lincoln, NE 68583

dwang3@unlnotes.unl.edu

SONG XI CHEN

Department of Statistics, Iowa State University, Ames, IA 50011

songchen@iastate.edu

Summary. We consider an empirical likelihood inference for parameters defined by general estimating equations when some components of the random observations are subject to missingness. As the nature of the estimating equations is wide ranging, we propose a nonparametric imputation of the missing values from a kernel estimator of the conditional distribution of the missing variable given the always observable variable. The empirical likelihood is used to construct a profile likelihood for the parameter of interest. We demonstrate that the proposed nonparametric imputation can remove the selection bias in the missingness and the empirical likelihood leads to more efficient parameter estimation. The proposed method is further evaluated by simulation and an empirical study on a genetic dataset on recombinant inbred mice.

Key words: Empirical likelihood; Estimating equations; Kernel estimation; Missing at random; Nonparametric imputation.

1. Introduction. Missing data are encountered in many statistical applications. A major undertaking in biological research is to integrate data generated by different experiments and technologies. Examples include the effort by *genenetwork.org* and other data depositories to combine genetics, microarray data and phenotypes in the study of recombinant inbred mouse lines [33]. One problem in using measurements from multiple experiments is that different research projects choose to perform experiments on different subsets of mouse strains. As a result, only a portion of the strains have all the measurements, while other strains have missing measurements. The current practice of using only those complete measurements is undesirable since the selection bias in the missingness can cause the parameter estimators to be inconsistent. Even in the absence of the selection bias (missing completely at random), the complete measurements based inference is generally not efficient as it throws away data with missing values. Substantial research has been done to deal with missing data problems; see [15] for a comprehensive overview.

Inference based on estimating equations [8, 3] is a general framework for statistical inference, accommodating a wide range of data structure and parameters. It has been used extensively for conducting semiparametric inference in the context of missing values. Robins, Rotnitzky and Zhao [23, 24] proposed using the parametrically estimated propensity scores to weigh estimating equations that define a regression parameter; and Robins and Rotnitzky [22] established the semiparametric efficiency bound for parameter estimation. The approach based on the general estimating equations has the advantage of being more robust against model misspecification, although a correct model for the conditional distribution of the missing variable given the observed variable is needed to attain the semiparametric efficiency. See

[30] for a comprehensive review.

In this paper we consider an empirical likelihood based inference for parameters defined by general estimating equations in the presence of missing values. Empirical likelihood introduced by Owen [17, 18] is a computer-intensive statistical method that facilitates a likelihood-type inference in a nonparametric or semiparametric setting. It is closely connected to the bootstrap as the empirical likelihood effectively carries out the resampling implicitly. On certain aspects of inference, empirical likelihood is more attractive than the bootstrap, for instance its ability of internal studentizing so as to avoid explicit variance estimation and producing confidence regions with natural shape and orientation; see [19] for an overview. In an important development, Qin and Lawless [21] proposed an empirical likelihood for parameter defined by a set of general estimating equations and established the Wilks theorem for the empirical likelihood ratio. Chen and Cui [5] show that the empirical likelihood of [21] is Bartlett correctable, indicating that the empirical likelihood has this delicate second order property of the conventional likelihood under the general setting of estimating equations. In the context of missing responses, Wang and Rao [32] studied empirical likelihood for the mean with imputed missing values from a kernel estimator of the conditional mean, and demonstrated that some of those attractive features of the empirical likelihood continue to hold.

When the parameter of interest defined by the general estimating equations is not directly related to a mean, or a regression model is not assumed as the model structure, the commonly used conditional mean based imputation via either a parametric [35] or nonparametric [6] regression estimator may results in either biased estimation or reduced efficiency; for instance when the parameter of interest is a quantile (conditional or unconditional) or some covariates are subject to missingness. To suit the general nature of parameters defined by general estimating equations and to facilitate a nonparametric likelihood inference in the presence of missing values, we propose a nonparametric imputation procedure that imputes missing values repeatedly from a kernel estimator of the conditional distribution of the missing variables given the fully observable variables. To control the variance of the estimating functions with imputed values, the estimating functions are averaged based on the multiple imputed values for each missing value. We show that the maximum empirical likelihood estimator based on the nonparametric imputation is consistent and is more efficient than the estimator based on the completely observed portion of the data only. In particular, when the number of the estimating equations is the same as the dimension of the parameter, the proposed empirical likelihood estimator attains the semiparametric efficiency bound.

The paper is structured as follows. The proposed nonparametric imputation method is described in Section 2. The formulation of the empirical likelihood is outlined in Section 3. Section 4 gives theoretical results of the proposed empirical likelihood estimator. Results from simulation studies are reported in Section 5. Section 6 analyzes a genetic dataset on recombinant inbred mice. All technical details are provided in the appendix.

2. Nonparametric imputation. Let $Z_i = (X_i^\tau, Y_i^\tau)^\tau$, $i = 1, \cdots, n$, be a set of independent and identically distributed random vectors, where X_i's are d_x-dimensional and are always observable, and Y_i's are d_y-dimensional and are subject to missingness. In practice, the missing components may vary among incomplete observations. For ease of presentation, we assume the missing components occupy the same components of Z_i. Extensions to the general case can be readily made. Furthermore, our use of Y_i for the missing variable does not prevent it being either a response or covariates in a regression setting.

Let θ be a p-dimensional parameter so that $E\{g(Z_i,\theta)\}=0$. Here $g(Z,\theta)=(g_1(Z,\theta),\ldots,g_r(Z,\theta))^T$ represents r estimating functions for an integer $r\geq p$. The interest of this paper is in the inference on θ when some Y_i's are missing.

Define $\delta_i=1$ if Y_i is observed and $\delta_i=0$ if Y_i is missing. Like in [6], [32] and others, we assume that δ and Y are conditionally independent given X, namely the strongly ignorable missing at random proposed by Rosenbaum and Rubin [25]. As a result,

$$P(\delta=1\mid Y,X)=P(\delta=1\mid X)=:p(X)$$

where $p(x)$ is the propensity score and prescribes a pattern of selection bias in the missingness.

Let $F(y|X_i)$ be the conditional distribution of Y given $X=X_i$. A kernel estimator of $F(y|X_i)$ based on the completely observed portion (no missing values) of the sample is

$$\hat{F}(y|X_i)=\sum_{l=1}^{n}\frac{\delta_l W(\frac{X_l-X_i}{h})I(Y_i\leq y)}{\sum_{j=1}^{n}\delta_j W(\frac{X_j-X_i}{h})}. \tag{1}$$

Here $W(\cdot)$ is a d_x-dimensional kernel function, h is a smoothing bandwidth and $I(\cdot)$ is the d_y-dimensional indicator function which is defined as $I(Y_i\leq y)=1$ if all components of Y_i are less than or equal to the corresponding components of y respectively, and $I(Y_i\leq y)=0$ otherwise. The property of the kernel estimator when there are no missing values is well understood in the literature, for instance in [10]. Its properties in the context of the missing values can be established in a standard fashion. An important property that mirrors one for unconditional multivariate distribution estimators given in [13] is that the efficiency of $\hat{F}(y|X_i)$ is not influenced by the dimension of Y_i. Here we concentrate on the case that X_i is a continuous random vector. Extension to discrete random variables can be readily made; see Section 5 for an implementation with binary random variables.

We propose to impute a missing Y_i with a $\tilde{Y}_i$ which is randomly generated from the estimated conditional distribution $\hat{F}(y|X_i)$. Effectively $\tilde{Y}_i$ has a discrete distribution where the probability of selecting a Y_l with $\delta_l = 1$ is

$$\frac{W\{(X_l - X_i)/h\}}{\sum_{j=1}^{n} \delta_j W\{(X_j - X_i)/h\}}.$$

To control the variability of the estimating functions with imputed values, we make κ independent imputations $\{\tilde{Y}_{i\nu}\}_{\nu=1}^{\kappa}$ from $\hat{F}(y|X_i)$ and use

$$\tilde{g}(\tilde{Z}_i, \theta) = \delta_i g(Z_i, \theta) + (1 - \delta_i)\kappa^{-1} \sum_{\nu=1}^{\kappa} g(X_i, \tilde{Y}_{i\nu}, \theta) \tag{2}$$

as the estimating function for the i-th observation. Like the conventional multiple imputation procedure [15], to attain the best efficiency, κ is required to converge to ∞. Our numerical experience indicates that setting $\kappa = 20$ worked quite well in our simulation experiments reported in Section 5.

The way we impute missing values depends critically on the nature of the parameter and model. A popular imputation method is to impute a missing Y_i by the conditional mean of Y given $X = X_i$ as proposed in [35] under a parametric regression model and in [6] and [32] via the kernel estimator for the conditional mean. However, this mean imputation may not work for a general parameter and a general model structure other than the regression model; for instance when the parameter is a correlation coefficient, or a conditional or unconditional quantile [1] where the estimating equation is based on a kernel smoothed distribution function. Nor is it generally applicable to missing covariates in a regression context. In contrast, the proposed nonparametric imputation is applicable for any parameter defined by estimating equations.

The curse of dimension is an issue with kernel estimators. Indeed, the estimation accuracy of $\hat{F}(y|X_i)$ deteriorates as d_x increases. However, as demonstrated in Section 4, as the target of the inference is a finite dimensional θ, the curse of dimension does not pose any leading order effect on the estimation of θ as long as the bias of the kernel estimator is controlled by letting $\sqrt{n}h^2 \to 0$ while $nh^{d_x} \to \infty$ to ensure the consistency of the conditional distribution estimation. When $d_x \geq 4$, controlling the bias requires a higher order, say $q-th$ order kernel, so that $\sqrt{n}h^q \to 0$ instead of $\sqrt{n}h^2 \to 0$. Using a higher order kernel may cause $\hat{F}(y|X_i)$ not being a proper conditional distribution and creates a minor problem for the imputation. See [31] for ways to get around it.

3. Empirical likelihood. The nonparametric imputation produces an extended sample $\{\tilde{Z}_i\}_{i=1}^n$ where

$$\tilde{Z}_i = \begin{cases} Z_i, & \text{if } \delta_i = 1; \\ (X_i, \{\tilde{Y}_{i\nu}\}_{\nu=1}^{\kappa})^\tau, & \text{if } \delta_i = 0. \end{cases} \tag{3}$$

With the imputed estimating equations, usual estimating equation approach can be used to make inference on θ. The variance of the general estimating equation based estimator for θ can be estimated using a sandwich estimator and the confidence regions can be obtained by asymptotic normal approximation. In this article, we would like to carry out a likelihood type inference using empirical likelihood, encouraged by its attractive performance for estimating equations without missing values as demonstrated by Qin and Lawless [21] and the work of Wang and Rao [32] for inference on a mean with missing responses. An advantage of empirical likelihood is that it has no pre-determined shape of the confidence region, instead it produces regions that reflect the features of the data set. Our proposal of using empirical likelihood in conjunction with nonparametric imputation is especially attractive, since it requires very few assumptions for both imputation and inference procedures while also has the flexibility inherent to empirical likelihood and estimating equations.

Let p_i represents the probability weight allocated to $\tilde{Z}_i$. The empirical likelihood for θ is

$$L(\theta) = \sup\Big\{\prod_{i=1}^{n} p_i \Big| p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} \tilde{g}(\tilde{Z}_i, \theta) = 0\Big\},$$

where $\tilde{g}$ is the adjustment to the original estimating function given in (2). By the standard derivation of empirical likelihood [21], the optimal p_i is

$$p_i = \frac{1}{n}\frac{1}{1 + t^{\tau}(\theta)\tilde{g}(\tilde{Z}_i, \theta)},$$

where $t(\theta)$ is the Lagrange multiplier that satisfies

$$\frac{1}{n}\sum_i \frac{\tilde{g}(\tilde{Z}_i,\theta)}{1+t^\tau(\theta)\tilde{g}(\tilde{Z}_i,\theta)}=0. \tag{4}$$

Let $\ell(\theta) = -\log\{L(\theta)/n^{-n}\}$ be the log empirical likelihood ratio and $\hat{\theta}$ be the maximum empirical likelihood estimator that maximizes $L(\theta)$.

4. Main results. The efficiency of $\hat{\theta}$ is studied in this section which also includes a proposal for constructing confidence regions for θ based on the empirical likelihood ratio.

Let θ_0 denote the true parameter value. Write $g(Z) =: g(Z, \theta_0)$. We define

$$\begin{aligned}\tilde{\Gamma} &= E\left[p(X)Cov\{g(Z)|X\} + E\{g(Z)|X\}E\{g^\tau(Z)|X\}\right],\\ \Gamma &= E\left[p^{-1}(X)Cov\{g(Z)|X\} + E\{g(Z)|X\}E\{g^\tau(Z)|X\}\right]\end{aligned}$$

and $V = \{E\left(\frac{\partial g}{\partial\theta}\right)^\tau \tilde{\Gamma}^{-1}E\left(\frac{\partial g}{\partial\theta}\right)\}^{-1}$ at $\theta = \theta_0$.

Theorem 1. *Under the conditions given in the Appendix, as $n \to \infty$ and $\kappa \to \infty$,*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma)$$

with $\Sigma = VE(\frac{\partial g}{\partial\theta})^\tau\tilde{\Gamma}^{-1}\Gamma\tilde{\Gamma}^{-1}E(\frac{\partial g}{\partial\theta})V$.

The estimator $\hat{\theta}$ is consistent and asymptotically normal for θ_0 and the potential selection bias in the missingness as measured by the propensity score $p(x)$ has been filtered out. If there is no missing values, $\tilde{\Gamma} = \Gamma = E(gg^\tau)$, which means that

$$\Sigma = \left\{E(\frac{\partial g}{\partial\theta})^\tau(Egg^\tau)^{-1}E(\frac{\partial g}{\partial\theta})\right\}^{-1}.$$

This is the asymptotic variance of the maximum empirical likelihood estimator based on full observations given in [21]. Comparing the forms of Σ with and without missing values shows that the efficiency of the maximum empirical likelihood estimator based on the proposed imputation will be close to that based on full observations if either the proportion of missing data is low, that is when $p(X)$ is close to 1, or if $E\{p^{-1}(X)Cov(g|X)\}$ is small relative to $E\{E(g|X)E(g^\tau|X)\}$, namely when X is highly "correlated" with Y.

In the case of $\theta = EY$, $\Sigma = E\{\sigma^2(X)/p(X)\} + Var\{m(X)\}$, where $\sigma^2(X) = Var(Y|X)$ and $m(X) = E(Y|X)$. Thus in this case, $\hat{\theta}$ is asymptotically equivalent to the estimator proposed by Cheng [6] and Wang and Rao [32] based on the conditional mean imputation.

When $r = p$, namely the number of estimating equations is the same as the dimension of θ,

$$\Sigma = \left\{ E(\frac{\partial g}{\partial \theta})^\tau \Gamma^{-1} E(\frac{\partial g}{\partial \theta}) \right\}^{-1},$$

which is the semiparametric efficiency bound for the estimation of θ as given by Chen, Hong and Tarozzi [4].

To appreciate the proposal of letting the number of imputation $\kappa \to \infty$, we note that when κ is fixed, the Γ and $\tilde{\Gamma}$ matrices used to define Σ have forms:

$$\begin{aligned} \Gamma &= E\Big[\big\{p^{-1}(X) + \kappa^{-1}(1 - p(X))\big\}Cov(g|X) + E(g|X)E(g^\tau|X)\Big] \quad \text{and} \\ \tilde{\Gamma} &= E\Big[\big\{p(X) + \kappa^{-1}(1 - p(X))\big\}Cov(g|X) + E(g|X)E(g^\tau|X)\Big]. \end{aligned}$$

Hence, a larger κ will reduce the terms in Γ and $\tilde{\Gamma}$ which are due to a single nonparametric imputation. Our numerical experience suggests that $\kappa = 20$ is sufficient for most situations.

Let us now turn our attention to the log empirical likelihood ratio

$$\mathcal{R}(\theta_0) = 2\ell(\theta_0) - 2\ell(\hat{\theta}).$$

Let I_r be the r-dimensional identity matrix. The next theorem shows that the log empirical likelihood ratio converges to a linear combination of independent chi-square distributions.

Theorem 2. *Under the conditions given in the Appendix, as $n \to \infty$ and $\kappa \to \infty$,*

$$\mathcal{R}(\theta_0) \xrightarrow{\mathcal{L}} Q^\tau \Omega Q,$$

where $Q \sim N(0, I_r)$ and $\Omega = \Gamma^{1/2}\tilde{\Gamma}^{-1}E\left(\frac{\partial g}{\partial \theta}\right)VE\left(\frac{\partial g}{\partial \theta}\right)^\tau \tilde{\Gamma}^{-1}\Gamma^{1/2}$.

When there is no missing values, $\Gamma = \tilde{\Gamma} = E(gg^\tau)$ and

$$\Omega = E(gg^\tau)^{-1/2}E\left(\frac{\partial g}{\partial \theta}\right)\left[E(\frac{\partial g}{\partial \theta})^\tau\{E(gg^\tau)\}^{-1}E(\frac{\partial g}{\partial \theta})\right]^{-1}E\left(\frac{\partial g}{\partial \theta}\right)^\tau E(gg^\tau)^{-1/2},$$

which is symmetric and idempotent with $tr(\Omega) = p$. This means that $\mathcal{R}(\theta_0) \xrightarrow{\mathcal{L}} \chi^2_p$, which is the nonparametric version of Wilks theorem established in [21].

When there are missing values, Wilks Theorem for empirical likelihood is no longer available due to a mis-match between the variance of $n^{-1/2}\sum_{i=1}^n \tilde{g}(\tilde{Z}_i, \theta_0)$ and the probability limit of $n^{-1}\sum_{i=1}^n \tilde{g}(\tilde{Z}_i, \theta_0)\tilde{g}^\tau(\tilde{Z}_i, \theta_0)$. This phenomenon also appears when a nuisance parameter is replaced by a plugged-in estimator as revealed by Hjort, McKeague and Van Keilegom [11].

When $\theta = EY$, $\mathcal{R}(\theta_0) \xrightarrow{\mathcal{L}} \{V_1(\theta_0)/V_2(\theta_0)\}\chi^2_1$, where

$$V_1(\theta_0) = E\{\sigma^2(X)/p(X)\} + Var\{m(X)\}$$

and $V_2(\theta_0) = E\{\sigma^2(X)p(X)\} + Var\{m(X)\}$. This is the limiting distribution given in [32].

As confidence regions can be readily transformed to test statistics for testing a hypothesis regarding θ, we shall focus on confidence regions. There are potentially several methods for the construction of a confidence region for θ. One is based on an estimation of the covariance matrix Σ and the asymptotic normality given in Theorem 1. Another method is to estimate the matrix Ω in Theorem 2 and then use Fourier inversion or a Monte Carlo method to simulate the distribution of the linear combinations of chi-squares. Despite the loss of Wilks theorem, confidence regions based on the empirical likelihood ratio $R(\theta)$ still have the attractions of likelihood based confidence regions in terms of having natural shape and orientation and respecting the range of θ.

We propose the following bootstrap procedure to approximate the distribution of $R(\theta_0)$. Bootstrap for imputed survey data has been discussed in [27] in the context of ratio and regression imputations. We use the following bootstrap procedure in which the bootstrap data set is imputed in the same way as the original data set was imputed:

1. Draw a simple random sample $\boldsymbol{\chi}_n^* = \{(\tilde{Z}_i^*, \delta_i^*) : i = 1, \ldots, n\}$ with replacement from the extended sample $\boldsymbol{\chi}_n = \{(\tilde{Z}_i, \delta_i) : i = 1, \ldots, n\}$ defined in (3).

2. Let $\boldsymbol{\chi}_{nc}^* = \{(Z_i^*, \delta_i^*) : \delta_i^* = 1\}$ be the portion of $\boldsymbol{\chi}_n^*$ without imputed values and $\boldsymbol{\chi}_{nm}^* = \{(\tilde{Z}_i^*, \delta_i^*) : \delta_i^* = 0\}$ be the set of vectors in the bootstrap sample with imputed values. Then replace all the imputed Y values in $\boldsymbol{\chi}_{nm}^*$ using the proposed imputation method where the estimation of the conditional distribution is based on $\boldsymbol{\chi}_{nc}^*$.

3. Let $\ell^*(\hat{\theta})$ be the empirical likelihood ratio based on the re-imputed data set $\boldsymbol{\chi}_n^*$, $\hat{\theta}^*$ be the corresponding maximum empirical likelihood estimator, and $\mathcal{R}^*(\hat{\theta}) = 2\ell^*(\hat{\theta}) - 2\ell^*(\hat{\theta}^*)$.

4. Repeat the above steps B-times for a large integer B and obtain B bootstrap values $\{\mathcal{R}_b^*(\hat{\theta})\}_{b=1}^B$.

Let $q_\alpha^\star$ be the $1-\alpha$ sample quantile based on $\{\mathcal{R}_b^*(\hat{\theta})\}_{b=1}^B$. Then, an empirical likelihood confidence region with nominal coverage level $1-\alpha$ is $I_\alpha = \{\theta \mid R(\theta) \le q_\alpha^\star\}$. The following theorem justifies that this confidence region has correct asymptotic coverage.

Theorem 3. *Under the conditions given in the Appendix and conditioning on the original sample* $\boldsymbol{\chi}_n$*,*

$$\mathcal{R}^*(\hat{\theta}) \xrightarrow{\mathcal{L}} Q^\tau \Omega^* Q$$

with $Q \sim N(0, I_r)$*, and* $\Omega^* \to \Omega$ *in probability as* $n \to \infty$ *and* $\kappa \to \infty$*.*

5. Simulation results. We report results from two simulation studies in this section. In each study, the proposed empirical likelihood inference based on the proposed nonparametric imputation is compared with the empirical likelihood inference based on (1) the **complete observations only** by ignoring data with missing values and (2) the **full observations** since the missing values are known in a simulation. When there is a selection bias in the missingness, the complete observations based estimator may not be consistent. The proposed imputation will remove the selection bias in the missingness and improve estimation efficiency due to utilizing more data information. Obtaining the full observations based estimator allows us to gauge how far away the proposed imputation based estimator is from the ideal case.

We also compare the proposed method with a version of the inverse probability weighted generalized method of moments (IPW-GMM) described in [4]. In particular, it is based on the fact that

$$E\Big\{g(Z_i,\theta_0)\frac{P(\delta_i=1)}{p(X_i)}\Big|\delta_i=1\Big\}=0.$$

Based on the usual formulation of the generalized method of moments [GMM, 9], the weighted-GMM estimator for θ_0 considered in our simulation is

$$\tilde{\theta}=\arg\min_{\theta}\Big\{\frac{1}{n_c}\sum_{i=1}^{n}\delta_i g(Z_i,\theta)\frac{1}{\hat{p}(X_i)}\Big\}^{\tau}A_T\Big\{\frac{1}{n_c}\sum_{i=1}^{n}\delta_i g(Z_i,\theta)\frac{1}{\hat{p}(X_i)}\Big\},$$

where n_c is the number of complete observations, A_T is a nonnegative definite weighting matrix, and $\hat{p}(X_i)$ is a consistent estimator for $p(X_i)$. The difference between the weighted-GMM method we use and that of [4] is that we used a kernel based estimator for $p(X_i)$, instead of the sieve estimator described in [4]. The bandwidth used to construct $\hat{p}(X_i)$ is obtained by the cross-validation method. Cross-validation method is also used to choose the smoothing bandwidth in the kernel estimator $\hat{F}(y|X)$ given in (1) for the proposed nonparametric imputation. To satisfy the requirement $\sqrt{n}h^2\to 0$, we use half of the bandwidth produced by the cross-validation procedure. The kernel function $W(\cdot)$ is taken to be the Gaussian or product Gaussian kernel for the two simulation studies.

5.1 *Correlation coefficient.* In the first simulation, the parameter θ is the correlation coefficient between two random variables X and Y where X is always observed, but Y is subject to missingness. We first generate bivariate random vector $(X_i, U_i)^\tau$ from a skewed bivariate t-distribution [2] with five degrees of freedom, mean $(0,0)^\tau$, shape parameter $(4,1)^\tau$, and dispersion matrix

$$\bar{\Omega} = \begin{bmatrix} 1 & .955 \\ .955 & 1 \end{bmatrix}.$$

Then we let $Y_i = U_i - 1.2X_iI(X_i < 0)$. The vector $(X_i, Y_i)^\tau$ has mean $(0, 0.304)$ and correlation coefficient 0.676.

We consider three missing mechanisms:

(a): $p(x) = (0.3 + 0.175|x|)I(|x| < 4) + I(|x| \geq 4)$;

(b): $p(x) \equiv 0.65$ for all x;

(c): $p(x) = 0.5I(x > 0) + I(x \leq 0)$.

The missing mechanism (b) is missing completely at random; whereas the other two are missing at random and prescribe selection bias in the missingness.

Let μ_x and μ_y be the means, and σ_x^2 and σ_y^2 be the variances of X and Y, respectively. In the construction of the empirical likelihood for θ [18], $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ are treated as nuisance parameters.

Table 1 contains the bias and standard deviation of the four estimators considered based on 1000 simulations with the sample size $n = 100$ and 200 respectively. It also contains the empirical likelihood confidence intervals using the full observations, complete observations only, and the proposed nonparametric imputation method at a nominal level of 95% . They are all based on the proposed bootstrap calibration method with $B = 1000$. When using the nonparametric imputation method, $\kappa = 20$ imputations were made for each missing Y_i. The confidence intervals based on the weighted-GMM are calibrated using the asymptotic normal approximation with the covariance matrix estimated by the kernel method.

The results in Table 1 can be summarized as follows. The proposed empirical likelihood estimator based on the nonparametric imputation method significantly reduced the bias compared to inference based only on complete observations when the data were missing at random but not missing completely at random. The estimator based on the completely observed data suffered severe bias under missing mechanisms (a) and (c). The proposed estimator had smaller standard deviations than the complete observation based estimator under all three missing mechanisms, including the case of missing completely at random. The weighted-GMM method also performed better than the complete observation based estimator. However, it had larger variance than the proposed estimator. Most strikingly, the standard deviations of the empirical likelihood estimator based on the proposed imputation method were all quite close to the full observation based estimator, which confirmed its good theoretical properties. Confidence intervals based on the complete observations only and the weighted-GMM method could have severe undercoverage: the former is due to the selection bias and the latter is due to the normal approximation. The proposed confidence intervals had satisfactory coverages which are quite close to the nominal level 0.95.

5.2 *Generalized linear models with missing covariates.* In the second simulation study we consider missing covariates in a generalized linear model (GLM). We also take the opportunity to discuss an extension of the proposed imputation procedure to binary random variables. Commonly used methods in dealing with missing data for GLM are reviewed in [12]. Empirical likelihood for GLM's with no missing data was first studied by Kolaczyk [14]. Application of empirical likelihood method to GLM's can help overcome difficulties with parametric likelihood, especially in the aspect of overdispersion.

To demonstrate how to extend the proposed method to discrete component in X_i, we consider a logistic regression model with binary response variable X_3 and covariates X_1, X_2 and Y. We choose $\text{logit}\{P(X_{3i} = 1)\} = -1 + X_{1i} + X_{2i} - 1.5Y_i$, $X_{1i} \sim N(0, 0.5^2)$, $X_{2i} \sim N(3, 0.5^2)$, and Y_i being binary with $\text{logit}\{P(Y_i = 1)\} = -1 + X_{1i} + 0.5X_{2i}$. Here X_{1i}, X_{2i}, and X_{3i} are always observable while the binary Y_i is subject to missingness with $\text{logit}\{P(Y_i \text{ is missing})\} = 0.5 + 2X_{1i} + X_{2i} - 3X_{3i}$. This model with $d_x = 3$ also allows us to see if there is a presence of the curse of dimension due to the use of the kernel estimator in the proposed imputation procedure.

When no missing data are involved, the empirical likelihood analysis for the logistic model simply involves the estimating equations $\sum_{i=1}^n S_i\{X_{3i} - \pi(S_i^\tau \beta)\} = 0$ with $S_i = (1, X_{1i}, X_{2i}, Y_i)^\tau$, β being the parameter and $\pi(z) = \exp(z)/\{1 + \exp(z)\}$. Although our proposed imputation in Section 2 is formulated directly for continuous random variables, binary response X_{3i} values can be easily accommodated by splitting the data into two parts according to the value of X_{3i} (binning), and then applying the proposed imputation scheme to each part by smoothing on the continuous X_{1i} and X_{2i}. The maximum empirical likelihood estimator for β uses a modified version of the fitting procedure described in Chapter 2 of [16].

The results of the simulation study with $n = 150$ and 250 are shown in Table 2(a) and 2(b) respectively. Despite that the dimension of X_i is increased to 3, there was no sign of the curse of dimension as the standard deviations of the proposed estimator were still quite close to the full observation based empirical likelihood estimator. This was very encouraging. For parameters β_0, β_1 and β_2, the mean squared error of the proposed estimator are several folds smaller than that based on the complete observations only; the proposed method also leads to a reduction in the mean squared error by as much as one fold relative to the weighted-GMM. All three methods give similar mean squared errors for the parameter β_3 while the proposed estimator had the smallest mean squared error. The confidence intervals based on only complete observations or the weighted-GMM tend to show notable undercoverage, while the proposed confidence intervals have satisfactory coverage levels for all parameters.

6. Empirical study. Microarray technology provides an powerful tool in molecular biology by measuring the expression level of thousands of genes simultaneously. One problem of interest is to test whether the expression level of genes is related to a traditional trait like body weight, food consumption, or bone density. This is usually the first step in uncovering roles that a gene plays in important pathways. The BXD recombinant inbred strains of mouse were derived from crosses between C57BL/6J (B6 or B) and DBA/2J (D2 or D) strains [34]. Around one hundred BXD strains have been established by researchers at University of Tennessee and the Jackson Laboratory. A variety of phenotype data are accumulated for BXD mouse over the years [20].

The trait that we consider is the fresh eye weight measured on 83 BXD strains by Zhai, Lu, and Williams (ID 10799, BXD phenotype data base). The Hamilton Eye Institute Mouse Eye M430v2 RMA Data Set contains measures of expression in the eye on 39,000 transcripts. It is of interest to test whether the fresh eye weight is related to the expression level of certain genes. However, the microarray data are only available for 45 out of the 83 BXD mouse strains for which fresh eye weights are all available. The most common practice is to use only complete observations and ignore missing values in the statistical inference. As demonstrated in our simulation, this approach can lead to inconsistent parameter estimators if there is a selection bias in the missingness. Even in the absence of selection bias, the estimators are not efficient as only those complete observations are used.

We conduct four separate simple linear regression analysis of the eye weight on the expression level of four genes respectively. The genes are *H3071E5*, *Slc26a8*, *Tex9*, and *Rps16*. Here we have missing covariates in our analysis. The missing gene expression levels are imputed from a kernel estimator of the conditional distribution of the gene expression level given the fresh eye weight. The smoothing bandwidths were selected based on the cross-validation method, which is 1.5 for the first three genes in Table 3 and 1.8 for gene *Rps16*.

Table 3 reports empirical likelihood estimates of the intercept and slope parameters and their 95% confidence intervals based on the proposed nonparametric imputation and empirical likelihood. It also contains results from a conventional parametric regression analysis using only the complete observations, assuming independent and identically normally distributed residuals. Table 3 shows that these two inference methods can produce quite different parameter estimates and confidence intervals. The difference in parameter estimates is as large as 50% for the intercept and 25% for the slope parameter. Table 3 also reports estimates and confidence intervals of the correlation coefficients using the proposed method and Fisher's z transformation. The latter is based on the complete observations only and is the method used by *genenetwork.org*. We observe again differences between the two methods despite not being significant at 5% level. The largest difference of about 30% is registered at gene *H3071E5*. As indicated earlier, part of the differences may be the estimation bias of the complete observations based estimators as they are unable to filter out selection bias in the missingness.

APPENDIX

Let $f(x)$ be the probability density function of X and $m_g(x) = E\{g(X, Y, \theta_0)|X = x\}$. The following conditions are needed in the proofs of the theorems.

C1: The functions $p(x)$, $f(x)$ and $m_g(x)$ all have bounded partial derivatives up to order q with $q \geq 2$ and $2q > d_x$, and $\inf_x p(x) \geq c_0$ for some $c_0 > 0$.

C2: The estimating function $g(x, y, \theta_0)$ has bounded partial derivative with regard to x up to order q, and $E\|g(Z, \theta_0)\|^4 < \infty$. In addition, $\partial^2 g(z, \theta)/\partial\theta\partial\theta^\tau$ is continuous in θ in a neighborhood of the true value θ_0; $\|\partial g(z, \theta)/\partial\theta\|$, $\|g(z, \theta)\|^3$, and $\|\partial^2 g(z, \theta)/\partial\theta\partial\theta^\tau\|$ are all bounded by some integrable functions in the neighborhood.

C3: The matrices Γ and $\tilde{\Gamma}$ are, respectively, positive definite with the smallest eigenvalue bounded away from zero, and $E[\partial g(z, \theta)/\partial\theta]$ has full column rank p.

C4: The kernel function W is a d_x dimensional kernel of order q, namely, $\int W(s_1, \ldots, s_{d_x}) ds_1 \ldots ds_{d_x} = 1$, and for any $i = 1, \ldots, d_x$,

$$\int s_i^l W(s_1, \ldots, s_{d_x}) ds_1 \ldots ds_{d_x} = 0$$

for any $1 \le l < q$, and $\int s_i^q W(s_1, \ldots, s_{d_x}) ds_1 \ldots ds_{d_x} \neq 0$.

C5: The smoothing bandwidth h satisfies $nh^{d_x} \to \infty$ and $\sqrt{n}h^q \to 0$ as $n \to \infty$.

Assuming $p(x)$ being bounded away from zero in C1 implies that data cannot be missing with probability 1 anywhere in the domain of the X variable. Conditions C2 and C3 are standard assumption for empirical likelihood based inference with estimating equations. Conditions C4 and C5 are standard in kernel estimation, and that $\sqrt{n}h^q \to 0$ is to control the bias induced by the kernel smoothing. To simplify the exposition, we will only deal with the case that $q = 2$ in the proof.

Lemma 1. *Assume that conditions C1-C5 are satisfied, then as* $n \to \infty$ *and* $\kappa \to \infty$,

$$n^{-1/2} \sum_{i=1}^{n} \tilde{g}(\tilde{Z}_i, \theta_0) \xrightarrow{\mathcal{L}} N(0, \Gamma),$$

where $\Gamma = E\left\{p^{-1}(X) Cov(g|X) + E(g|X)E(g^\tau|X)\right\}$.

Proof of Lemma 1: Let $u \in \mathbb{R}^r$ and $\|u\| = 1$. Also let $g_u(Z, \theta_0) = u^\tau g(Z, \theta_0)$ and $\tilde{g}_u(\tilde{Z}, \theta_0) = u^\tau \tilde{g}(\tilde{Z}, \theta_0)$. First we need to show that $n^{-1/2} \sum_{i=1}^{n} \tilde{g}_u(\tilde{Z}_i, \theta_0) \xrightarrow{\mathcal{L}} N(0, u^\tau \Gamma u)$, and then use the Cramér-Wold device to prove Lemma 1. Define

$$m_{g_u}(x) = E(g_u(X, Y, \theta_0)|X = x) \;\text{ and }\; \hat{m}_{g_u}(x) = \frac{\sum_{i=1}^{n} \delta_i W(\frac{x - X_i}{h}) g_u(x, Y_i, \theta_0)}{\sum_{i=1}^{n} \delta_i W(\frac{x - X_i}{h})}.$$

Now we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i g_u(X_i, Y_i, \theta_0) + (1 - \delta_i) \kappa^{-1} \sum_{\nu=1}^{\kappa} g_u(X_i, \tilde{Y}_{i\nu}, \theta_0) \right\} \\ = \;& \frac{1}{n} \sum_{i=1}^{n} \delta_i \{ g_u(X_i, Y_i, \theta_0) - m_{g_u}(X_i) \} \end{aligned}$$

$$
\begin{aligned}
& +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\kappa^{-1}\sum_{\nu=1}^{\kappa} g_u(X_i,\tilde{Y}_{i\nu},\theta_0)-\hat{m}_{g_u}(X_i)\right\} \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_{g_u}(X_i)-m_{g_u}(X_i)\}+\frac{1}{n}\sum_{i=1}^{n} m_{g_u}(X_i) \\
:= \ & S_n + A_n + T_n + R_n.
\end{aligned}
$$

Note that S_n and R_n are sums of independent and identically distributed random variables. Define $\eta(x) = p(x)f(x)$ and $\hat{\eta}(x) = \frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j - x)$ as its kernel estimator, where $W_h(u) = h^{-d_x}W(u/h)$. Then,

$$
\begin{aligned}
T_n \ = \ & \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j-X_i)\{g_u(X_i,Y_j,\theta_0)-m_{g_u}(X_j)\}}{\eta(X_i)} \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_{g_u}(X_i)-m_{g_u}(X_i)\}\frac{\eta(X_i)-\hat{\eta}(X_i)}{\eta(X_i)} \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j-X_i)(m_{g_u}(X_j)-m_{g_u}(X_i))}{\eta(X_i)}\right\} \\
:= \ & T_{n1}+T_{n2}+T_{n3}.
\end{aligned}
$$

Define

$$
\check{T}_{n1} = \sum_{j=1}^{n} E\{T_{n1} \mid (X_j, Y_j, \delta_j)\} = \sum_{j=1}^{n} \delta_j E\{T_{n1} \mid (X_j, Y_j, \delta_j = 1)\}
$$

to be a projection of T_{n1}. Then write $T_{n1} = \check{T}_{n1} + (T_{n1} - \check{T}_{n1})$. As

$$
\begin{aligned}
T_{n1} \ = \ & \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j-X_i)\{g_u(X_i,Y_j,\theta_0)-m_{g_u}(X_j)\}}{\eta(X_i)} \\
= \ & \frac{1}{n}\sum_{j=1}^{n}\delta_j\{g_u(X_i,Y_j,\theta)-m_{g_u}(X_j)\}\left\{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{W_h(X_i-X_j)}{\eta(X_i)}\right\},
\end{aligned}
$$

$$
\begin{aligned}
&\check{T}_{n1}\\
= \;& \frac{1}{n}\sum_{j=1}^{n}\delta_j E\left[\{g_u(X_i,Y_j,\theta_0)-m_{gu}(X_j)\}\frac{(1-\delta_i)W_h(X_i-X_j)}{\eta(X_i)}\middle| X_j,Y_j\right]\\
= \;& \frac{1}{n}\sum_{j=1}^{n}\delta_j \int\left[\{g_u(x,Y_j,\theta_0)-m_{gu}(X_j)\}\frac{\{1-p(x)\}W_h(x-X_j)}{\eta(x)}\right]f(x)dx\\
= \;& \frac{1}{n}\sum_{j=1}^{n}\delta_j \int\left[\{g_u(x,Y_j,\theta_0)-m_{gu}(X_j)\}\frac{\{1-p(x)\}}{p(x)}W_h(x-X_j)\right]dx\\
= \;& \frac{1}{n}\sum_{j=1}^{n}\delta_j \int\left[\{g_u(X_j+hs,Y_j,\theta_0)-m_{gu}(X_j)\}\frac{\{1-p(X_j+hs)\}}{p(X_j+hs)}W(s)\right]ds.
\end{aligned}
$$

Since both g_u and $\rho(x)=\{1-p(x)\}/p(x)$ has bounded seconded derivative on x, and $\sqrt{n}h^2\to 0$ as $n\to\infty$, a Taylor expansion around X_j leads to

$$
\text{(A1)}\qquad \check{T}_{n1}=\frac{1}{n}\sum_{j=1}^{n}\delta_j\{g_u(X_j,Y_j,\theta)-m_{gu}(X_j)\}\frac{1-p(X_j)}{p(X_j)}+o_p(n^{-\frac{1}{2}}).
$$

Now we show $T_{n1}-\check{T}_{n1}=o_p(n^{-1/2})$. Let

$$
\begin{aligned}
T_{n1i} &= (1-\delta_i)\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j-X_i)\{g_u(X_i,Y_j,\theta_0)-m_{g_u}(X_j)\}}{\eta(X_i)} \quad\text{and}\\
\check{T}_{n1i} &= \sum_{j=1}^{n}E\{T_{n1i}\mid (X_j,Y_j,\delta_j=1)\}.
\end{aligned}
$$

Then by straight forward computation,

$$
\begin{aligned}
\text{(A2)}\qquad & nE(T_{n1}-\check{T}_{n1})^2\\
= \;& \frac{1}{n}\sum_{i=1}^{n}E(T_{n1i}-\check{T}_{n1i})^2+\frac{2}{n}\sum_{i\neq j}E\{(T_{n1i}-\check{T}_{n1i})(T_{n1j}-\check{T}_{n1j})\}\\
= \;& E(T_{n1i}-\check{T}_{n1i})^2=ET_{n1i}^2-E\check{T}_{n1i}^2\leqslant ET_{n1i}^2\\
\leqslant \;& E\left\{\frac{\frac{1}{n}\sum_{j=1}^{n}\delta_j W_h(X_j-X_i)\{g_u(X_i,Y_j,\theta_0)-m_{g_u}(X_j)\}}{\eta(X_i)}\right\}^2\to 0.
\end{aligned}
$$

The last step is obtained by an argument similar to one used in proving the consistency of Nadaraya-Watson estimators in [29] and [7]. This suggests that $T_{n1} = \check{T}_{n1} + o_p(n^{-1/2})$. By standard argument, we can show that $T_{n2} = o_p(n^{-\frac{1}{2}})$. Derivations similar to those for T_{n1} can be used to establish $T_{n3} = o_p(n^{-1/2})$. Thus, we have

$$\text{(A3)} \qquad \sqrt{n}T_n \xrightarrow{\mathcal{L}} N\big[0, E\{(1-p(X))^2\sigma^2_{g_u}(X)/p(X)\}\big],$$

where $\sigma^2_{g_u}(X) = Var\{g_u(X,Y,\theta) \mid X\}$.

Also note $\sqrt{n}S_n \xrightarrow{\mathcal{L}} N[0, E\{p(X)\sigma^2_{g_u}(X)\}]$ and $\sqrt{n}R_n \xrightarrow{\mathcal{L}} N[0, Var\{m_{g_u}(X)\}]$. Further, it is straight forward to show that

$$nCov(S_n, T_n) = E\{(1-p(X))\sigma^2_{g_u}(X)\} + o(1),$$

$nCov(R_n, S_n) = 0$ and $nCov(R_n, T_n) = o(1)$. It readily follows that

$$\text{(A4)} \qquad \sqrt{n}(S_n + T_n + R_n) \xrightarrow{\mathcal{L}} N\big[0, E\{\sigma^2_{g_u}(X)/p(X)\} + Var\{m_{g_u}(X)\}\big].$$

Now we consider the asymptotic distribution of

$$A_n = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\Big\{\kappa^{-1}\sum_{\nu=1}^{\kappa} g_u(X_i, \tilde{Y}_{i\nu}, \theta_0) - \hat{m}_{g_u}(X_i)\Big\}.$$

Given all the original observations, $n^{-1/2}(1-\delta_i)\big\{\kappa^{-1}\sum_{\nu=1}^{\kappa} g_u(X_i, \tilde{Y}_{i\nu}, \theta) - \hat{m}(X_i)\big\}$, $i = 1, 2, \ldots, n$, are independent with conditional mean zero and conditional variance $(n\kappa)^{-1}(1-\delta_i)\{\hat{\gamma}_{g_u}(X_i)-\hat{m}^2_{g_u}(X_i)\}$. Here

$$\hat{\gamma}_{g_u}(x)=\sum_{j=1}^{n}\delta_j W_h(x-X_j)g_u^2(x,Y_j,\theta_0)/\hat{\eta}(x)$$

is a kernel estimator of $\gamma_{g_u}(x)=E\{g_u^2(X,Y,\theta_0)|X=x\}$. By verifying Lyapounov's condition, we can show that conditioning on the original observations,

$$\text{(A5)}\qquad \sqrt{n}A_n \xrightarrow{\mathcal{L}} N\Big[0,(n\kappa)^{-1}\sum_{i=1}^{n}(1-\delta_i)\{\hat{\gamma}_{g_u}(X_i)-\hat{m}^2_{g_u}(X_i)\}\Big].$$

The conditional variance

$$\text{(A6)}\quad (n\kappa)^{-1}\sum_{i=1}^{n}(1-\delta_i)\{\hat{\gamma}_{g_u}(X_i)-\hat{m}^2_{g_u}(X_i)\}\xrightarrow{p}\kappa^{-1}E[\{1-p(X)\}\sigma^2_{g_u}(X)].$$

By Lemma 1 of [26], as $n\to\infty$ and $\kappa\to\infty$, $\sqrt{n}(S_n+T_n+R_n+A_n)$ converges to a normal distribution with mean 0 and variance

$$Var\{m_{g_u}(Z,\theta)\}+E\{p^{-1}(X)\sigma^2_{g_u}(X)\}=u^\tau\Gamma u.$$

Then Lemma 1 is proved by using the Cramèr-Wold device. □

Lemma 2. *Under the conditions C1-C5, as $n\to\infty$ and $\kappa\to\infty$,*

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{g}(\tilde{Z}_i,\theta_0)\tilde{g}^\tau(\tilde{Z}_i,\theta_0)\xrightarrow{p}\tilde{\Gamma},$$

where $\tilde{\Gamma}=E\left\{p(X)Cov(g|X)+E(g|X)E(g^\tau|X)\right\}$.

Proof: Consider each element of the matrix $\frac{1}{n}\sum_{i=1}^{n}\tilde{g}(\tilde{Z}_i,\theta_0)\tilde{g}^\tau(\tilde{Z}_i,\theta_0)$, that is,

$$\frac{1}{n}\sum_{i=1}^{n}\tilde{g}_j(\tilde{Z}_i,\theta_0)\tilde{g}_k(\tilde{Z}_i,\theta_0), \quad 0 \le j,k \le r.$$

Write

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\tilde{g}_j(\tilde{Z}_i,\theta_0)\tilde{g}_k(\tilde{Z}_i,\theta_0) \\
= & \frac{1}{n}\sum_{i=1}^{n}\delta_j g_j(Z_i,\theta_0)g_k(Z_i,\theta_0) \\
& +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\left\{\kappa^{-1}\sum_{\nu=1}^{\kappa}g_j(X_i,\tilde{Y}_{i\nu},\theta_0)\right\}\left\{\kappa^{-1}\sum_{\nu=1}^{\kappa}g_k(X_i,\tilde{Y}_{i\nu},\theta_0)\right\} \\
:= & T_{n1}+T_{n2}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
T_{n1} = & \frac{1}{n}\sum_{i=1}^{n}\delta_i\{g_j(Z_i,\theta_0)-m_{g_j}(X_i)\}\{g_k(Z_i,\theta_0)-m_{g_k}(X_i)\} \\
& -\frac{1}{n}\sum_{i=1}^{n}\delta_i m_{g_j}(X_i)m_{g_k}(X_i)+\frac{1}{n}\sum_{i=1}^{n}\delta_i g_j(Z_i,\theta_0)m_{g_k}(X_i) \\
& +\frac{1}{n}\sum_{i=1}^{n}\delta_i g_k(Z_i,\theta_0)m_{g_j}(X_i) \\
:= & T_{n1a}+T_{n1b}+T_{n1c}+T_{n1d}.
\end{aligned}$$

It is obvious that T_{n1a}, T_{n1b}, T_{n1c} and T_{n1d} are all sums of independent and identically distributed random variables. By law of large numbers and the continuous mapping theorem, we can show that

$$T_{n1}\xrightarrow{p}E\left[p(X)Cov\{g_j(Z,\theta_0),g_k(Z,\theta_0)|X\}+p(X)m_{g_j}(X)m_{g_k}(X)\right].$$

Note that

$$
\begin{aligned}
T_{n2} &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\tilde{g}_j(\tilde{Z}_i,\theta_0)\tilde{g}_k(\tilde{Z}_i,\theta_0) - \hat{m}_{g_j}(X_i)\hat{m}_{g_k}(X_i)\} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_{g_j}(X_i)\hat{m}_{g_k}(X_i) - m_{g_j}(X_i)m_{g_k}(X_i)\} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)m_{g_j}(X_i)m_{g_k}(X_i) \\
&:= T_{n2a} + T_{n2b} + T_{n2c}.
\end{aligned}
$$

As $g_j(X_i, \tilde{Y}_{i\nu}, \theta_0)$ has conditional mean $\hat{m}_{g_j}(X_i)$ given the original observations $\mathcal{X}_n$, it can be shown that $T_{n2a} \xrightarrow{p} 0$ as $\kappa \to \infty$. By argument similar to those used for (A3), $T_{n2b} \xrightarrow{p} 0$ as $n \to \infty$. Obviously T_{n2c} is the sum of independent and identically distributed random variables, which leads to $T_{n2c} \xrightarrow{p} E[\{1 - p(X)\}m_{g_j}(X_i)m_{g_k}(X_i)]$. Hence we have $T_{n2} \xrightarrow{p} E[\{1-p(X)\}m_{g_j}(X_i)m_{g_k}(X_i)]$ as $n \to \infty$ and $\kappa \to \infty$. Therefore,

$$
T_{n1} + T_{n2} \xrightarrow{p} E\left[p(X)Cov\{g_j(Z,\theta_0), g_k(Z,\theta_0)|X\} + m_{g_j}(X)m_{g_k}(X)\right].
$$

This completes the proof of Lemma 2. □

Let us define

$$
\begin{aligned}
Q_{1n}(\theta, t) &= \frac{1}{n}\sum_i \frac{1}{1+t^\tau \tilde{g}(\tilde{Z}_i,\theta)}\tilde{g}(\tilde{Z}_i,\theta), \\
Q_{2n}(\theta, t) &= \frac{1}{n}\sum_i \frac{1}{1+t^\tau \tilde{g}(\tilde{Z}_i,\theta)}\left\{\frac{\partial \tilde{g}(\tilde{Z}_i,\theta)}{\partial\theta}\right\}^\tau t,
\end{aligned}
$$

where $t(\theta)$ is the Lagrange multiplier defined in (4).

Proof of Theorem 1: Using argument similar to that of [21], it can be shown that as $n \to \infty$ and $\kappa \to \infty$, with probability tending to 1, $L(\theta)$ attains its maximum value at some point $\hat{\theta}$ within the open ball $\|\theta - \theta_0\| < n^{-1/3}$, and $\hat{\theta}$ and $\hat{t} = t(\hat{\theta})$ satisfy

$$Q_{1n}(\hat{\theta}, \hat{t}) = 0, \quad Q_{2n}(\hat{\theta}, \hat{t}) = 0.$$

Taking the derivatives with regard to θ and t^τ,

$$\begin{aligned}
\frac{\partial Q_{1n}(\theta, 0)}{\partial \theta} &= \frac{1}{n}\sum_i \frac{\partial \tilde{g}(\tilde{Z}_i, \theta)}{\partial \theta}, \quad & \frac{\partial Q_{1n}(\theta, 0)}{\partial t^\tau} &= -\frac{1}{n}\sum_i \tilde{g}(\tilde{Z}_i, \theta)\tilde{g}^\tau(\tilde{Z}_i, \theta), \\
\frac{\partial Q_{2n}(\theta, 0)}{\partial \theta} &= 0, & \frac{\partial Q_{2n}(\theta, 0)}{\partial t^\tau} &= \frac{1}{n}\sum_i \left\{\frac{\partial \tilde{g}(\tilde{Z}_i, \theta)}{\partial \theta}\right\}^\tau.
\end{aligned}$$

Expanding $Q_{1n}(\hat{\theta}, \hat{t})$, $Q_{2n}(\hat{\theta}, \hat{t})$ at $(\theta_0, 0)$, we have

$$\begin{aligned}
0 &= Q_{1n}(\hat{\theta}, \hat{t}) \\
&= Q_{1n}(\theta_0, 0) + \frac{\partial Q_{1n}(\theta_0, 0)}{\partial \theta}(\hat{\theta} - \theta_0) + \frac{\partial Q_{1n}(\theta_0, 0)}{\partial t^\tau}(\hat{t} - 0) + o_p(\zeta_n), \\
0 &= Q_{2n}(\hat{\theta}, \hat{t}) \\
&= Q_{2n}(\theta_0, 0) + \frac{\partial Q_{2n}(\theta_0, 0)}{\partial \theta}(\hat{\theta} - \theta_0) + \frac{\partial Q_{2n}(\theta_0, 0)}{\partial t^\tau}(\hat{t} - 0) + o_p(\zeta_n),
\end{aligned}$$

where $\zeta_n = \|\hat{\theta} - \theta_0\| + \|\hat{t}\|$. Then we can write

$$\begin{pmatrix} \hat{t} \\ \hat{\theta} - \theta_0 \end{pmatrix} = S_n^{-1} \begin{pmatrix} -Q_{1n}(\theta_0, 0) + o_p(\zeta_n) \\ o_p(\zeta_n) \end{pmatrix},$$

where

$$S_n = \begin{pmatrix} \frac{\partial Q_{1n}}{\partial t^\tau} & \frac{\partial Q_{1n}}{\partial \theta} \\ \frac{\partial Q_{2n}}{\partial t^\tau} & 0 \end{pmatrix}_{(\theta_0,0)} \to \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix} = \begin{pmatrix} -\tilde{\Gamma} & E\left(\frac{\partial g}{\partial \theta}\right) \\ E\left(\frac{\partial g}{\partial \theta}\right)^\tau & 0 \end{pmatrix}.$$

Note that $Q_{1n}(\theta_0, 0) = \frac{1}{n}\sum_{i=1}^n \tilde{g}(\tilde{Z}_i, \theta_0) = O_p(n^{-1/2})$, it follows that $\zeta_n = O_p(n^{-1/2})$. After some matrix manipulation, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) = S_{22.1}^{-1} S_{21} S_{11}^{-1} \sqrt{n} Q_{1n}(\theta_0, 0) + o_p(1),$$

where $V = S_{22.1}^{-1} = \left\{E(\frac{\partial g}{\partial \theta})^\tau \tilde{\Gamma}^{-1} E(\frac{\partial g}{\partial \theta})\right\}^{-1}$. By Lemma 1, $\sqrt{n}Q_{1n}(\theta_0, 0) \xrightarrow{\mathcal{L}} N(0, \Gamma)$, and the theorem follows. □

Proof of Theorem 2: Notice that

$$\mathcal{R}(\theta_0) = 2\left[\sum_i \log\{1 + t_0^\tau \tilde{g}(\tilde{Z}_i, \theta_0)\} - \sum_i \log\{1 + \hat{t}^\tau \tilde{g}(\tilde{Z}_i, \hat{\theta})\}\right]$$

where $t_0 = t(\theta_0)$, and

$$\ell(\hat{\theta}, \hat{t}) = \sum_i \log\{1 + \hat{t}^\tau \tilde{g}(\tilde{Z}_i, \hat{\theta})\} = -\frac{n}{2} Q_{1n}^\tau(\theta_0, 0) A Q_{1n}(\theta_0, 0) + o_p(1)$$

where $A = S_{11}^{-1}(I + S_{12} S_{22.1}^{-1} S_{21} S_{11}^{-1})$. Under H_0,

$$\frac{1}{n}\sum_i \frac{1}{1 + t_0^\tau \tilde{g}(\tilde{Z}_i, \theta_0)} \tilde{g}(\tilde{Z}_i, \theta_0) = 0, \quad t_0 = -S_{11}^{-1} Q_{1n}(\theta_0, 0) S_{11}^{-1} Q_{1n}(\theta_0, 0) + o_p(1),$$

and $\sum_i \log\{1 + t_0^\tau \tilde{g}(\tilde{Z}_i table, \theta_0)\} = -\frac{n}{2} Q_{1n}^\tau(\theta_0, 0) S_{11}^{-1} Q_{1n}(\theta_0, 0) + o_p(1)$. Thus we have

$$\begin{aligned}\mathcal{R}(\theta_0) &= nQ_{1n}^\tau(\theta_0,0)(A-S_{11}^{-1})Q_{1n}(\theta_0,0)+o_p(1)\\ &= \sqrt{n}Q_{1n}^\tau(\theta_0,0)S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1}\sqrt{n}Q_{1n}(\theta_0,0)+o_p(1).\end{aligned}$$

Note that

$$S_{11}^{-1}S_{12}S_{22.1}^{-1}S_{21}S_{11}^{-1} \xrightarrow{p} \tilde{\Gamma}^{-1}E\left(\frac{\partial g}{\partial\theta}\right)VE\left(\frac{\partial g}{\partial\theta}\right)^\tau\tilde{\Gamma}^{-1},$$

and by Lemma 1, $\sqrt{n}Q_{1n}(\theta_0,0)\xrightarrow{\mathcal{L}}N(0,\Gamma)$, the theorem then follows. □

Proof for Theorem 3: The proof for Theorem 3 essentially involves establishing the bootstrap version of Lemma 1 to Theorem 2. We only outline the main steps in proving the bootstrap version of Lemma 1 here.

Let X_i^*, Y_i^*, $\tilde{Y}_{i\nu}^*$, δ_i^* be the counter part to X_i, Y_i, $\tilde{Y}_{i\nu}$, δ_i in the bootstrap sample, $S_n(\hat{\theta})$, $A_n(\hat{\theta})$, $T_n(\hat{\theta})$ and $R_n(\hat{\theta})$ represent the quantities S_n, A_n, T_n and R_n with θ_0 replaced by $\hat{\theta}$ respectively. Let $S_n^*(\hat{\theta})$, $A_n^*(\hat{\theta})$, $T_n^*(\hat{\theta})$ and $R_n^*(\hat{\theta})$ be their bootstrap counterpart. First we will show

(A7)
$$\begin{aligned}&\sqrt{n}\{S_n^*(\hat{\theta})+T_n^*(\hat{\theta})+R_n^*(\hat{\theta})-S_n(\hat{\theta})-T_n(\hat{\theta})-R_n(\hat{\theta})\}\\ \xrightarrow{\mathcal{L}}\ & N\big[0,E_*\{\sigma_{g_u}^2(X,\hat{\theta})/p(X)\}+Var_*\{m_{g_u}(X,\hat{\theta})\}\big],\end{aligned}$$

where $E_*(\cdot)$ and $Var_*(\cdot)$ represent the conditional expectation and variance given the original data respectively. Define

$$\hat{m}_{gu}(x,\hat{\theta}) = \frac{\sum_{i=1}^n\delta_iW(\frac{x-X_i}{h})g_u(x,Y_i,\hat{\theta})}{\sum_{i=1}^n\delta_iW(\frac{x-X_i}{h})} \quad \text{and}$$

$$\hat{m}^*_{g_u}(x,\hat\theta) = \frac{\sum_{i=1}^n \delta_i^* W(\frac{x-X_i^*}{h}) g_u(x, Y_i^*, \hat\theta)}{\sum_{i=1}^n \delta_i^* W(\frac{x-X_i^*}{h})}.$$

Then

$$\begin{aligned}
& S_n^*(\hat\theta) + T_n^*(\hat\theta) + R_n^*(\hat\theta) - S_n(\hat\theta) - T_n(\hat\theta) - R_n(\hat\theta) \\
= & \frac{1}{n}\sum_{i=1}^n \Big[\delta_i^*\{g_u(Z_i^*,\hat\theta) - m_{gu}(X_i^*,\hat\theta)\} - \frac{1}{n}\sum_{j=1}^n \delta_j\{g_u(Z_j,\hat\theta) - m_{g_u}(X_j,\hat\theta)\}\Big] \\
+ & \frac{1}{n}\sum_{i=1}^n [(1-\delta_i^*)\{\hat m^*_{g_u}(X_i^*) - \hat m_{g_u}(X_i^*)\}] \\
+ & \frac{1}{n}\sum_{i=1}^n \Big[(1-\delta_i^*)\{\hat m_{g_u}(X_i^*,\hat\theta) - m_{g_u}(X_i^*,\hat\theta)\} \\
& \qquad -\frac{1}{n}\sum_{j=1}^n (1-\delta_j)\{\hat m_{g_u}(X_j,\hat\theta) - m_{g_u}(X_j,\hat\theta)\}\Big] \\
+ & \frac{1}{n}\sum_{i=1}^n \Big\{m_{g_u}(X_i^*,\hat\theta) - \frac{1}{n}\sum_{j=1}^n m_{g_u}(X_j,\hat\theta)\Big\} \\
:= & B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

For both B_1 and B_4, we can apply the central limit theorem for bootstrap samples [e.g. 28] to derive

$$\sqrt{n} B_1 \xrightarrow{\mathcal{L}} N\big[0, E_*\{p(X)\sigma^2_{g_u}(X,\hat\theta)\}\big] \text{ and } \sqrt{n} B_4 \xrightarrow{\mathcal{L}} N\big[0, Var_*\{m_{g_u}(X,\hat\theta)\}\big].$$

Also it can be shown $B_2 = o_p(n^{-1/2})$. Use similar argument to (A1) to show

$$\begin{aligned}
B_3 = & \frac{1}{n}\sum_{i=1}^n \Big[\delta_i^*\{g_u(Z_i^*,\hat\theta) - m_{g_u}(X_i^*,\hat\theta)\}\frac{1-p(X_i^*)}{p(X_i^*)} \\
& -\frac{1}{n}\sum_{j=1}^n \delta_j\{g_u(Z_j,\hat\theta) - m_{g_u}(X_j,\hat\theta)\}\frac{1-p(X_j)}{p(X_j)}\Big] + o_p(n^{-1/2}).
\end{aligned}$$

Then follow the proof for Lemma 1 and apply the bootstrap central limit theorem to conclude (A7).

For $A_n^*(\hat\theta)$, given the observations in the bootstrap sample that are not imputed, we have

$$\sqrt{n}A_n^*(\hat\theta) \xrightarrow{\mathcal{L}} N\Big[0,(n\kappa)^{-1}\sum_{i=1}^{n}(1-\delta_i^*)\{\hat\gamma^*(X_i^*,\hat\theta)-\hat m^{*2}(X_i^*,\hat\theta)\}\Big],$$

in distribution. Similar to the proof of Lemma 1, by employing Lemma 1 of [26]

$$\begin{aligned} &\frac{1}{\sqrt{n}}\Big\{\sum_{i=1}^{n} g_u(\tilde Z_i^*,\hat\theta)-n^{-1}\sum_{j=1}^{n} g_u(\tilde Z_j,\hat\theta)\Big\} \\ \xrightarrow{\mathcal{L}}\ & N\big[0,E_*\{\sigma_{g_u}^2(X,\hat\theta)/p(X)\}+Var_*\{m_{g_u}(X,\hat\theta)\}\big]. \end{aligned}$$

The bootstrap version of Lemma 1 is justified by noting

$$\begin{aligned} E_*\{\sigma_{g_u}^2(X,\hat\theta)/p(X)\} &\to E\{\sigma_{g_u}^2(X)/p(X)\} \text{ and} \\ Var_*\{m_{g_u}(X,\hat\theta)\} &\to Var\{m_{g_u}(X)\} \end{aligned}$$

as $n\to\infty$, then employ the Cramèr-Wold device.

□

REFERENCES

[1] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, 68:326–328.

[2] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B*, 65:367–389.

[3] Boos, D. D. (1992). On generalized score tests. *Amer. Statist.*, 46:327–333.

[4] Chen, X., Hong, H. and Tarozzi, A. (2004). Semiparametric efficiency in GMM models of nonclassical measurement errors, missing data and treatment effects. Working paper, New York University.

[5] Chen, S. X. and Cui, H. J. (2006). On the second order properties of empirical likelihood for generalized estimating equations. Working paper, Iowa State University.

[6] Cheng, P. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.*, 89:81–87.

[7] Devroye, L. P. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.*, 8:231–239.

[8] Godambe, V. P. (1991). *Estimating Functions.* Oxford University Press, Oxford, U.K.

[9] Hansen, L. (1982). Large sample properties of generalized method of moment estimators. *Econometrica*, 50:1029–1084.

[10] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge, U.K.

[11] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2004). Extending the scope of empirical likelihood. *Ann. Statist.* Under revision.

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

[13] Jin, Z. and Shao, Y. (1999). On kernel estimation of a multivariate distribution function. *Statist. Probab. Lett.*, 2:163–168.

[14] Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica*, 4:199–218.

[15] Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data, 2nd edition.* Wiley, Hoboken, NJ, USA.

[16] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, New York.

[17] Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75:237–249.

[18] Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, 18:90–120.

[19] Owen, A. B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, Boca Raton, Florida.

[20] Pierce, J. L., Lu, L., Gu, J., Silver, L. M. and Williams, R. W. (2004). A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMG Genetics*, 5:http://www.biomedcentral.com/1471–2156/5/7.

[21] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.*, 22:300–325.

[22] Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *J. Amer. Statist. Assoc.*, 90:122–129.

[23] Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, 89:846–866.

[24] Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.*, 90:106–121.

[25] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

[26] Schenker, N. and Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Ann. Statist.*, 16:1550–1566.

[27] Shao, J. and Sitter, R. R. (1996). Bootstrap for imputed survey data. *J. Amer. Statist. Assoc.*, 91:1278–1288.

[28] Shao, J. and Tu, D. (1985). *The Jackknife and Bootstrap.* Springer Verlag, New York.

[29] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, 5:595–620.

[30] Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York.

[31] Wang, D. and Chen, S. X. (2006). Nonparametric imputation of missing values for estimating equation based inference. Technical report, Iowa State University.

[32] Wang, Q. and Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.*, 30:896–924.

[33] Wang, J., Williams, R. W. and Manly, K. F. (2003). WebQTL: Web-based complex trait analysis. *Neuroinformatics*, 1:299–308.

[34] Williams, R. W., Gu, J., Qi, S. and Lu, L. (2001). The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. *Genome Biology*, 2:research0046.1–0046.18.

[35] Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emporium Journal of Experimental Agriculture*, 1:129–142.

Table 1: Inference for the correlation coefficient with missing values. The four methods considered are empirical likelihood using full observations, empirical likelihood using only complete observations (Complete Obs.), inverse probability weighting based generalized method of moments (Weighted-GMM), and empirical likelihood using the proposed nonparametric imputation (N. Imputation). The nominal coverage probability of the confidence interval is 0.95.

$n = 100$					
Methods	Bias	Std. Dev.	MSE	Coverage	Length of CI
Full Observations	-0.0026	0.0895	0.0080	0.936	0.3555
Missing Mechanism *(a)*					
Complete Obs.	0.0562	0.1222	0.0181	0.851	0.4967
Weighted-GMM	0.0108	0.1112	0.0125	0.776	0.2495
N. Imputation	-0.0092	0.1041	0.0109	0.945	0.4875
Missing Mechanism *(b)*					
Complete Obs.	-0.0080	0.1162	0.0136	0.930	0.4482
Weighted-GMM	-0.0150	0.1069	0.0117	0.802	0.2763
N. Imputation	-0.0138	0.0999	0.0101	0.932	0.4173
Missing Mechanism *(c)*					
Complete Obs.	-0.1085	0.1442	0.0326	0.832	0.5593
Weighted-GMM	-0.0266	0.1167	0.0143	0.786	0.2860
N. Imputation	-0.0383	0.1053	0.0125	0.928	0.4322
$n = 200$					
Methods	Bias	Std. Dev.	MSE	Coverage	Length of CI
Full Observations	0.0071	0.0610	0.0038	0.958	0.2484
Missing Mechanism *(a)*					
Complete Obs.	0.0710	0.0776	0.0111	0.824	0.3161
Weighted-GMM	0.0112	0.0734	0.0055	0.799	0.2060
N. Imputation	0.0038	0.0709	0.0050	0.955	0.3180
Missing Mechanism *(b)*					
Complete Obs.	-0.0030	0.0799	0.0064	0.937	0.3091
Weighted-GMM	-0.0031	0.0719	0.0052	0.832	0.2075
N. Imputation	-0.0023	0.0668	0.0045	0.942	0.2797
Missing Mechanism *(c)*					
Complete Obs.	-0.0915	0.0979	0.0179	0.788	0.3919
Weighted-GMM	-0.0107	0.0745	0.0057	0.820	0.2131
N. Imputation	-0.0118	0.0680	0.0048	0.936	0.2860

Table 2: Inference for parameters in a logistic regression model with missing values. The four methods considered are empirical likelihood using full observations (Full Obs.), empirical likelihood using only complete observations (Complete Obs.), inverse probability weighting based generalized method of moments (Weighted-GMM), and empirical likelihood using the proposed nonparametric imputation (N. Imputation). The nominal coverage probability of the confidence interval is 0.95.

Table 2(a): $n = 150$

Methods	Bias	Std. Dev.	MSE	Coverage	Length of CI
		$\beta_0 = -1$			
Full Obs.	-0.0296	1.292	1.669	0.964	5.477
Complete Obs.	-1.715	1.618	5.559	0.920	6.840
Weighted-GMM	-0.7835	1.562	3.053	0.891	5.250
N. Imputation	0.0349	1.317	1.736	0.967	5.549
		$\beta_1 = 1$			
Full Obs.	0.0519	0.4384	0.1949	0.964	1.820
Complete Obs.	0.7898	0.5603	0.9377	0.796	2.510
Weighted-GMM	0.4302	0.5486	0.4860	0.834	1.811
N. Imputation	-0.0605	0.4388	0.1962	0.961	1.851
		$\beta_2 = 1$			
Full Obs.	0.0367	0.4500	0.2038	0.972	2.007
Complete Obs.	0.4205	0.5590	0.4892	0.945	2.599
Weighted-GMM	0.2542	0.5484	0.3653	0.896	1.791
N. Imputation	-0.0110	0.4576	0.2095	0.966	1.993
		$\beta_3 = -1.5$			
Full Obs.	-0.0531	0.4979	0.2507	0.976	2.137
Complete Obs.	-0.0684	0.5713	0.3310	0.975	2.592
Weighted-GMM	-0.0751	0.5843	0.3471	0.838	1.574
N. Imputation	0.0718	0.5521	0.3100	0.966	2.474

Table 2(b): $n = 250$

Methods	Bias	Std. Dev.	MSE	Coverage	Length of CI
			$\beta_0 = -1$		
Full Obs.	-0.0286	0.9651	0.9321	0.956	3.916
Complete Obs.	-1.670	1.212	4.255	0.801	4.790
Weighted-GMM	-0.6393	1.150	1.7304	0.862	3.832
N. Imputation	0.0284	0.9801	0.9615	0.962	3.963
			$\beta_1 = 1$		
Full Obs.	0.0195	0.3332	0.1114	0.953	1.349
Complete Obs.	0.7270	0.4398	0.7220	0.665	1.789
Weighted-GMM	0.3166	0.4223	0.2786	0.782	1.304
N. Imputation	-0.0660	0.3367	0.1177	0.947	1.380
			$\beta_2 = 1$		
Full Obs.	0.0305	0.3374	0.1147	0.958	1.400
Complete Obs.	0.3902	0.4134	0.3232	0.867	1.729
Weighted-GMM	0.1966	0.3993	0.1981	0.874	1.297
N. Imputation	-0.0173	0.3406	0.1163	0.967	1.384
			$\beta_3 = -1.5$		
Full Obs.	-0.0611	0.3818	0.1495	0.950	1.529
Complete Obs.	-0.0351	0.4445	0.1988	0.963	1.797
Weighted-GMM	-0.0419	0.4596	0.2130	0.791	1.165
N. Imputation	0.0762	0.4377	0.1974	0.944	1.759

Table 3: Parameter estimates and confidence intervals (shown in parentheses) based on a simple linear regression model using the parametric method with complete observations only and the empirical likelihood method using the proposed nonparametric imputation. For the parametric inference, the confidence intervals for the intercept and slope are obtained using quantiles of the t-distribution, and the confidence intervals for the correlation coefficient are obtained by Fisher's z transformation. The four different genes are identified by the probe names.

Gene	Complete Observations Only (parametric)		Nonparametric Imputation (with empirical likelihood)	
		Intercept		
H3071E5	-21.99	(-40.97, -2.998)	-15.69	(-37.02, 5.209)
Slc26a8	73.59	(49.45, 97.73)	67.28	(38.34, 95.87)
Tex9	-23.81	(-46.12, -1.507)	-14.66	(-38.57, 8.776)
Rps16	-13.52	(-31.08, 4.041)	-8.090	(-26.76, 10.18)
		Slope		
H3071E5	10.16	(5.720, 14.59)	8.736	(2.688, 14.21)
Slc26a8	-6.352	(-9.294, -3.411)	-5.561	(-9.431, -1.471)
Tex9	5.101	(2.588, 7.613)	4.094	(0.8753, 6.979)
Rps16	6.766	(3.371, 10.16)	5.754	(1.948, 9.236)
		Correlation Coefficient		
H3071E5	0.5757	(0.3395, 0.7436)	0.4426	(0.1321, 0.6977)
Slc26a8	-0.5533	(-0.7285, -0.3102)	-0.4319	(-0.6809, -0.0761)
Tex9	0.5296	(0.2996, 0.7124)	0.4024	(0.1013, 0.6846)
Rps16	0.5256	(0.2744, 0.7097)	0.4151	(0.0755, 0.6613)