Nonparametric Imputation of Missing Values for Estimating Equation Based Inference[1]

DONG WANG and SONG XI CHEN

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

dwang@iastate.edu and *songchen@iastate.edu*

SUMMARY

We propose a nonparametric imputation procedure for data with missing values and establish an empirical likelihood inference for parameters defined by general estimating equations. The imputation is carried out multiple times via a nonparametric 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 evaluated by simulation and an empirical study on the relationship between eye weight and gene transcriptional abundance of recombinant inbred mice.

Some 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.

[1]Author of Correspondence: Song Xi Chen, Department of Statistics, Iowa State University, Ames, IA 50011-1210. Email: songchen@iastate.edu

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 (Wang, Williams and Manly, 2003). 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 those data with missing values.

Substantial research has been done to deal with missing data problems in survey statistics, longitudinal data analysis, and multivariate analysis. See Little and Rubin (2002) for a comprehensive overview.

Most methods in missing data analysis employ certain parametric models. When a parametric model can be defined for all the variables, the maximum likelihood method can be used for inference, which has been greatly facilitated by the EM algorithm (Dempster, Laird and Rubin, 1977). Multiple imputation is another popular method which creates multiple "complete" datasets by making random draws from the predictive distribution of the missing values. The multiple "complete" datasets are analyzed one by one based on a complete-data based inference method, which are then combined to form a final inference that reflects the uncertainty due to nonresponse (Rubin, 1987; Little and Rubin, 2002). To make inference under more relaxed assumptions, Robins, Rotnitzky and Zhao (1994, 1995) proposed using the parametrically estimated propensity scores to weigh estimating equations that define a regression parameter; and Robins and Rotnitzky (1995) established the semiparametric efficiency bound for parameter estimation. This approach 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 often needed to attain the semiparametric efficiency bound.

Nonparametric methods have also been proposed for missing data. Titterington (1977) and Titterington and Mill (1983) considered kernel estimation of a multivariate density for data with incomplete observations. When the parameter of interest is the mean of a response variable which is subject to missingness, Cheng (1994) proposed using the kernel conditional mean estimator to impute the missing values. Hahn (1998) and Hirano, Imbens and Ridder (2003) studied the estimation of average treatment effects using nonparametrically estimated propensity scores. Since in treatment effect problems, the response of any unit can only be observed under one treatment, it is seen as a missing data problem. In survey statistics, Kim and Fuller (2004) proposed the fractional hot deck imputation method, in which multiple values are drawn from the same imputation cell as the missing observation, and a weight is assigned to each imputed value.

In this paper we consider estimation of parameters defined by a set of estimating equations in the presence of missing values. Estimating equation (Godambe, 1991; Boos, 1992) is a very general framework for statistical inference. When the parameter of interest is not directly related to the mean, the commonly used conditional mean based imputation via either a parametric (Yates, 1933) or nonparametric (Cheng, 1994) regression estimator may results in either biased estimation or reduced efficiency. This is especially the case for data with missing covariates.

We propose in this paper a nonparametric imputation procedure that generates multiple copies of missing values from a kernel estimator of the conditional distribution of the missing variables given the fully observable variables. This model-free (nonparametric) imputation is particularly suited for the wide range of parameters defined by estimating equations. We then employ Owen (1988, 1990)'s empirical likelihood to formulate a nonparametric profile likelihood based on an extended sample which consists of the original data and the nonparametrically imputed values. Our use of empirical likelihood is largely encouraged by its attractive inferential features for estimating equations when there is no missing values by Qin and Lawless (1994), as well as its inference for a mean parameter in the case of missing responses by Wang and Rao (2002) who imputed the missing values were from a kernel estimator of the conditional mean.

We show that the maximum empirical likelihood estimator based on the nonparametric imputation is consistent and more efficient than the estimator based on the complete 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 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. Let θ be a p-dimensional parameter so that $E\{g(Z_i, \theta)\} = 0$. Here $g(Z, \theta)$ represents r estimating functions for an integer $r \geq p$. The interest of this paper is the inference on θ when some Y_i's are missing. Our use of Y_i for the missing variable does not prevent it being either a response or covariates in a regression setting.

Let $\delta_i = 1$ if Y_i is observed and $\delta_i = 0$ if Y_i is missing. Like Cheng (1994), Wang and Rao (2002) 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 (1983). As a result,

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

where $p(x)$ is called the propensity score and prescribes selection bias in the missingness if $p(x)$ is not a constant function.

Let $F(y|X_i)$ be the conditional distribution of Y given $X = X_i$. Let

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

be a kernel estimator of $F(y|X_i)$ based on the completely observed portion (no missing values) of the sample. Here $W(\cdot)$ is a d_x-dimensional kernel function; h is a smoothing bandwidth satisfying $\sqrt{n}h^2 \to 0$ and $nh^{d_x} \to \infty$ as $n \to \infty$; and $I(\cdot)$ is the d_y-dimensional indicator function. Here we concentrate on the case that both X and Y are continuous random variables. Extension to discrete random variables can be readily made; see Section 5 for a case of binary random variable.

We propose to impute a missing Y_i with a $\tilde{Y}_i$ 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

$$\frac{W\{(X_l - X_i)/h\}}{\sum_{j=1}^{n} \delta_j W\{(X_j - X_i)/h\}}. \tag{2}$$

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

$$\tilde{g}(\tilde{Z}_i, \theta) = \delta_i g(Z_i, \theta) + (1 - \delta_i)\kappa^{-1} \sum_{\nu=1}^{\kappa} g(X_i, \tilde{Y}_{i\nu}, \theta) \tag{3}$$

as the estimation function for the i-th observation.

A popular method of imputation is to impute a missing Y_i by the conditional mean of Y given $X = X_i$ as proposed in Yates (1933) under a parametric regression model and in Cheng (1994) and Wang and Rao (2002) via the Nadaraya-Watson kernel estimator for the conditional mean. However, it may not work for other parameters, for instance, quantiles or correlation coefficients. Nor is it generally applicable to the case of missing covariates in a regression context. The proposed nonparametric imputation is generally applicable for any parameter defined by estimating equations. We are to show that when θ is a mean related parameter, the proposed imputation method leads to a parameter estimator that has the same efficiency as that obtained by conditional mean imputation.

"Curse of dimensiona" 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, 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 $\sqrt{n}h^2 \to 0$. There is one effect of the dimension though: when $d_x \geq 4$, controlling the bias requires a higher order kernel. Using a higher order kernel causes $\hat{F}(y|X_i)$ not being a bona fide conditional distribution and can cause a minor problem for the imputation. Wang and Chen (2006) propose a refinement of the imputation procedure to allow high order kernels and hence $d_x \geq 4$. To simplify our exposition, we confine ourself in this paper to $d_x \leq 3$.

3. EMPIRICAL LIKELIHOOD

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

$$\tilde{Z}_i = \begin{cases} Z_i, & \text{if } \delta_i = 1; \\ (X_i, \{\tilde{Y}_{i\nu}\}_{\nu=1}^{\kappa})^\tau, & \text{if } \delta_i = 0. \end{cases} \tag{4}$$

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 (3). This is the formulation of Qin and Lawless (1994) on adjusted estimating functions. By following the standard derivation of empirical likelihood (Qin and Lawless, 1994), 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

$$\frac{1}{n}\sum_i \frac{\tilde{g}(\tilde{Z}_i, \theta)}{1 + t^\tau(\theta)\tilde{g}(\tilde{Z}_i, \theta)} = 0. \tag{5}$$

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 $L(\theta)$.

The efficiency of $\hat{\theta}$ is studied in the next section which also includes a proposal for constructing confidence regions for θ based on the empirical likelihood ratio. The latter is largely motivated by the attractive features (natural shape and orientation as well as range respecting) of empirical likelihood confidence regions in the absence of missing values; see Hall and La Scala (1990) and Chen and Cui (2006).

4. MAIN RESULTS

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 $\kappa \to \infty$,*

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

in distribution, 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 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 Qin and Lawless (1994). 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 (1994) and Wang and Rao (2002) 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 in Chen, Hong and Tarozzi (2004).

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:

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

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 chisquare distributions.

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

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

in distribution, 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 $\mathcal{R}(\theta_0) \to \chi^2_p$ in distribution, which is the nonparametric version of Wilks' theorem established in Qin and Lawless (1994).

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 (2004).

When $\theta = EY$, $\mathcal{R}(\theta_0) \to \{V_1(\theta_0)/V_2(\theta_0)\}\chi_1^2$ in distribution 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 Wang and Rao (2002).

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 chisquares. 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 Shao and Sitter (1996) 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 (4).

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 $I_{\alpha} = \{\theta \mid R(\theta) \leq q_{\alpha}^{\star}\}$. 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* $\boldsymbol{\chi}_{n}$,

$$\mathcal{R}^{*}(\hat{\theta}) \rightarrow Q^{\tau}\Omega^{*}Q, \tag{6}$$

in distribution, with $Q \sim N(0, I_{r})$, *and* $\Omega^{*} \rightarrow \Omega$ *in probability as* $n \rightarrow \infty$ *and* $\kappa \rightarrow \infty$.

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 are 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 Chen et al. (2004). 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, Hansen, 1982), 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 Chen et al. (2004) is that we used a kernel based estimator for $p(X_i)$, instead of the sieve estimator described in Chen et al. (2004). The bandwidth used to construct $\hat{p}(X_i)$ is the one that gives the smallest empirical mean square error among several bandwidths that we experimented, including the theoretically optimal one.

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 (Azzalini and Capitanio, 2003) 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 (Owen, 1990), $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2)$ are treated as nuisance parameters.

Table 1 contains the bias and standard error 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 $B = 2000$. 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 nonparametric imputation method significantly reduces the bias compared to inference based only on complete observations when the data are missing at random but not missing completely at random. The estimator based on the completely observed data suffers quite severe bias under missing mechanisms (a) and (c). The proposed imputation reduces the variance relative to the estimator based on only complete observations under all three missing mechanisms, including the case of missing completely at random. The weighted-GMM method can also reduce the bias compared to analysis based on complete observations only, but tends to have larger variance than that of the proposed estimator. Confidence intervals based on the complete observations only and the weighted-GMM method can have severe under-coverage: the former is due to the selection bias and the latter is due to the normal approximation. The proposed confidence intervals have 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 Ibrahim, Chen, Lipsitz and Herring (2005). Empirical likelihood for GLM's with no missing data was first studied by Kolaczyk (1994). Application of empirical likelihood method to GLM's can help overcome difficulties with parametric likelihood, especially in the aspect of overdispersion.

We consider a logistic regression model with binary response variable X_2 and covariates $S = (X_1, Y)^\tau$. We choose $\text{logit}\{P(X_{2i} = 1)\} = -1 + X_{1i} - 1.5Y_i$, $X_{1i} \sim N(3, 0.5^2)$, and Y_i being binary with $\text{logit}\{P(Y_i = 1)\} = -1 + 0.5X_{1i}$. Here X_{1i} and X_{2i} are always observable while the binary Y_i is subject to missingness with $\text{logit}\{P(Y_i \text{ is missing})\} = 0.5 + X_{1i} - 3X_{2i}$.

When no missing data are involved, the empirical likelihood analysis for the logistic model simply involves the estimating equations $\sum_{i=1}^n S_i\{X_{2i} - \pi(S_i^\tau \beta)\} = 0$ with 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_{2i} values can be easily accommodated by splitting the data into two parts according to the value of X_{2i} (binning), and then applying the proposed imputation scheme to each part by smoothing on the continuous X_{1i}. The maximum empirical likelihood estimator for β uses a modified version of the fitting procedure described in Chapter 2 of McCullagh and Nelder (1983).

The results of the simulation study with $n = 150$ and 250 are shown in Table 2. For parameters β_0 and β_1, 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 square error by as much as 20% relative to the weighted-GMM. All three methods give similar mean squared errors for the parameter β_2. 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 (Williams, Gu, Qi and Lu, 2001). 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 (Pierce et al., 2004).

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*, which are identified by the corresponding probe names in the microarray dataset. 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 second partial derivatives, and $\inf_x p(x) \geq c_0$ for some $c_0 > 0$.

C2: The estimating function $g(x, y, \theta_0)$ has bounded second partial derivative with regard to x, 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 non-negative, symmetric and bounded probability density function with finite second moments.

C5: The smoothing bandwidth h satisfies $nh^{d_x} \to \infty$, $\sqrt{n}h^2 \to 0$ as $n \to \infty$, and $d_x \leq 3$.

Assuming $p(x)$ 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 required for empirical likelihood based inference with estimating equations. Conditions C4 and C5 are standard in kernel estimation, and that $\sqrt{n}h^2 \to 0$ is to control the bias induced by the kernel smoothing.

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

$$n^{-1/2} \sum_{i=1}^{n} \tilde{g}(\tilde{Z}_i, \theta_0) \to N(0, \Gamma), \tag{A1}$$

in distribution, 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 $n^{-1/2} \sum_{i=1}^{n} \tilde{g}_u(\tilde{Z}_i, \theta_0) \to N(0, u^\tau \Gamma u)$ in distribution, 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

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

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

$$\check{T}_{n1} = \sum_{j=1}^n E\{T_{n1} \mid (X_j,Y_j,\delta_j)\} = \sum_{j=1}^n \delta_j E\{T_{n1} \mid (X_j,Y_j,\delta_j = 1)\} \tag{A2}$$

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

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

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

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

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

$$\check{T}_{n1}=\frac{1}{n}\sum_{j=1}^{n}\delta_j\{g_u(X_j,Y_j,\theta)-m_{gu}(X_j)\}\frac{1-p(X_j)}{p(X_j)}+o_p(n^{-\frac{1}{2}}). \tag{A3}$$

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,

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

The last step is obtained by an argument similar to one used in proving the consistency of Nadaraya-Watson estimators in Stone (1977) and Devroye and Wagner (1980). 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

$$\sqrt{n}T_n \to N\big[0, E\{(1-p(X))^2\sigma^2_{g_u}(X)/p(X)\}\big], \tag{A4}$$

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

Also note $\sqrt{n}S_n \to N[0, E\{p(X)\sigma^2_{g_u}(X)\}]$ and $\sqrt{n}R_n \to N[0, Var\{m_{g_u}(X)\}]$ both in distribution. 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

$$\sqrt{n}(S_n + T_n + R_n) \to N\big[0, E\{\sigma^2_{g_u}(X)/p(X)\} + Var\{m_{g_u}(X)\}\big], \tag{A5}$$

in distribution.

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 $\hat{\gamma}_{g_u}(x) = \sum_{j=1}^{n}\delta_j W_h(x - X_j) g^2_u(x, Y_j, \theta_0)/\hat{\eta}(x)$ is a kernel estimator of $\gamma_{g_u}(x) = E\{g^2_u(X, Y, \theta_0)|X = x\}$. By verifying Lyapounov's condition, we can show that conditioning on the original observations,

$$\sqrt{n}A_n \to N\Big[0, (n\kappa)^{-1}\sum_{i=1}^{n}(1-\delta_i)\{\hat{\gamma}_{g_u}(X_i) - \hat{m}^2_{g_u}(X_i)\}\Big], \tag{A6}$$

in distribution. The conditional variance

$$(n\kappa)^{-1}\sum_{i=1}^{n}(1-\delta_i)\{\hat{\gamma}_{g_u}(X_i) - \hat{m}^2_{g_u}(X_i)\} \xrightarrow{p} \kappa^{-1}E[\{1-p(X)\}\sigma^2_{g_u}(X)]. \tag{A7}$$

By Lemma 1 of Schenker and Welsh (1988), 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

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

Then Lemma 1 is proved by using the Cramèr-Wold device. $\square$

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,

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

$$:= \quad T_{n1a} + T_{n1b} + T_{n1c} + T_{n1d}.$$

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 (A4), $T_{n2b} \xrightarrow{p} 0$ 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 (5).

Proof of Theorem 1: Using argument similar to that of Qin and Lawless (1994), 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,

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

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, $\sqrt{n} Q_{1n}(\theta_0, 0) \to N(0, \Gamma)$, and the theorem follows. $\square$

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

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

Note that

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

and by Lemma 1, $\sqrt{n} Q_{1n}(\theta_0, 0) \rightarrow N(0, \Gamma)$ in distribution, 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

$$\sqrt{n}\{S_n^*(\hat{\theta}) + T_n^*(\hat{\theta}) + R_n^*(\hat{\theta}) - S_n(\hat{\theta}) - T_n(\hat{\theta}) - R_n(\hat{\theta})\} \quad \text{(A8)}$$

$$\to \quad N\big[0, E_*\{\sigma^2_{g_u}(X,\hat\theta)/p(X)\} + Var_*\{m_{g_u}(X,\hat\theta)\}\big],$$

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

$$\hat m_{gu}(x,\hat\theta) = \frac{\sum_{i=1}^n \delta_i W(\frac{x-X_i}{h}) g_u(x,Y_i,\hat\theta)}{\sum_{i=1}^n \delta_i W(\frac{x-X_i}{h})} \text{ and } \hat m^*_{g_u}(x,\hat\theta) = \frac{\sum_{i=1}^n \delta^*_i W(\frac{x-X^*_i}{h}) g_u(x,Y^*_i,\hat\theta)}{\sum_{i=1}^n \delta^*_i W(\frac{x-X^*_i}{h})}.$$

Then

$$\begin{aligned}
& S^*_n(\hat\theta) + T^*_n(\hat\theta) + R^*_n(\hat\theta) - S_n(\hat\theta) - T_n(\hat\theta) - R_n(\hat\theta) \\
= \ & \frac{1}{n}\sum_{i=1}^n \Big[\delta^*_i\{g_u(Z^*_i,\hat\theta) - m_{gu}(X^*_i,\hat\theta)\} - \frac{1}{n}\sum_{j=1}^n \delta_j\{g_u(Z_j,\hat\theta) - m_{g_u}(X_j,\hat\theta)\}\Big] \\
+ \ & \frac{1}{n}\sum_{i=1}^n [(1-\delta^*_i)\{\hat m^*_{g_u}(X^*_i) - \hat m_{g_u}(X^*_i)\}] \\
+ \ & \frac{1}{n}\sum_{i=1}^n \Big[(1-\delta^*_i)\{\hat m_{g_u}(X^*_i,\hat\theta) - m_{g_u}(X^*_i,\hat\theta)\} - \frac{1}{n}\sum_{j=1}^n (1-\delta_j)\{\hat m_{g_u}(X_j,\hat\theta) - m_{g_u}(X_j,\hat\theta)\}\Big] \\
+ \ & \frac{1}{n}\sum_{i=1}^n \Big\{m_{g_u}(X^*_i,\hat\theta) - \frac{1}{n}\sum_{j=1}^n m_{g_u}(X_j,\hat\theta)\Big\} \\
:= \ & B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

For both B_1 and B_4, we can apply the central limit theorem for bootstrap samples (e.g. Shao and Tu, 1985) to derive

$$\sqrt{n}B_1 \to N\big[0, E_*\{p(X)\sigma^2_{g_u}(X,\hat\theta)\}\big] \text{ and } \sqrt{n}B_4 \to N\big[0, Var_*\{m_{g_u}(X,\hat\theta)\}\big], \qquad \text{(A9)}$$

in distribution. Also it can be shown $B_2 = o_p(n^{-1/2})$. Use similar argument to (A3) 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 (A8).

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

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

in distribution. Similar to the proof of Lemma 1, by employing Lemma 1 of Schenker and Welsh (1988)

$$\frac{1}{\sqrt{n}}\Big\{\sum_{i=1}^{n} g_u(\tilde{Z}_i^*, \hat{\theta}) - n^{-1}\sum_{j=1}^{n} g_u(\tilde{Z}_j, \hat{\theta})\Big\} \to N\big[0, E_*\{\sigma_{g_u}^2(X, \hat{\theta})/p(X)\} + Var_*\{m_{g_u}(X, \hat{\theta})\}\big],$$

in distribution. The bootstrap version of Lemma 1 is justified by noting

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

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

□

REFERENCES

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Ser. B 65*, 367–389.

Boos, D. D. (1992). On generalized score tests. *The American Statistician 46*, 327–333.

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

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

Cheng, P. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association 89*, 81–87.

Dempster, A. P., A. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B 39*, 1–38.

Devroye, L. P. and T. J. Wagner (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *the Annals of Statistics 8*, 231–239.

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

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica 66*, 315–331.

Hall, P. and B. La Scala (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review 58*, 109–127.

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

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*, 1161–1189.

Hjort, N., I. McKeague, and I. Van Keilegom (2004). Extending the scope of empirical likelihood. *The Annals of Statistics*. Under revision.

Ibrahim, J. G., M. Chen, S. R. Lipsitz, and A. H. Herring (2005). Missing-data methods for generalized linear models: a comparative review. *Journal of the American Statistical Association 100*, 332–346.

Kim, J. K. and W. Fuller (2004). Fractional hot deck imputation. *Biometrika 91*, 559–578.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica 4*, 199–218.

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

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

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

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*, 90–120.

Pierce, J. L., L. Lu, J. Gu, L. M. Silver, and R. W. Williams (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.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*, 300–325.

Robins, J., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association 89*, 846–866.

Robins, J. M. and A. Rotnitzky (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association 90*, 122–129.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association 90*, 106–121.

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

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Schenker, N. and A. H. Welsh (1988). Asymptotic results for multiple imputation. *The Annals of Statistics 16*, 1550–1566.

Shao, J. and R. R. Sitter (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association 91*, 1278–1288.

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

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics 5*, 595–620.

Titterington, D. M. (1977). Analysis of incomplete multivariate binary data by kernel method. *Biometrika 64*, 455–460.

Titterington, D. M. and G. M. Mill (1983). Kernel-based density estimates from incomplete data. *Journal of the Royal Statistical Society, Serie B 45*, 258–266.

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

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

Wang, Q. and J. N. K. Rao (2002). Empirical likelihood-based inference under imputation for missing response data. *The Annals of Statistics 30*, 896–924.

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

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

$n = 100$					
Methods	Bias	Std. Err.	MSE	Coverage	Length of CI
Full Observations	0.0018	0.0899	0.0081	0.937	0.3543
Missing Mechanism *(a)*					
Complete Obs.	0.0597	0.1262	0.0195	0.863	0.4896
Weighted-GMM	0.0057	0.1105	0.0123	0.761	0.3568
N. Imputation	-0.0050	0.1013	0.0103	0.948	0.4888
Missing Mechanism *(b)*					
Complete Obs.	-0.0046	0.1160	0.0160	0.921	0.4562
Weighted-GMM	-0.0097	0.1068	0.0115	0.740	0.3057
N. Imputation	-0.0077	0.0991	0.0099	0.936	0.4239
Missing Mechanism *(c)*					
Complete Obs.	-0.1123	0.1480	0.0345	0.831	0.3594
Weighted-GMM	-0.0226	0.1141	0.0135	0.780	0.3569
N. Imputation	-0.0217	0.1050	0.0115	0.944	0.4264
$n = 200$					
Methods	Bias	Std. Err.	MSE	Coverage	Length of CI
Full Observations	0.0047	0.0605	0.0037	0.949	0.2514
Missing Mechanism *(a)*					
Complete Obs.	0.0727	0.0776	0.0113	0.849	0.3269
Weighted-GMM	0.0072	0.0755	0.0058	0.800	0.2479
N. Imputation	0.0076	0.0695	0.0049	0.953	0.3239
Missing Mechanism *(b)*					
Complete Obs.	-0.0007	0.0753	0.0057	0.957	0.3146
Weighted-GMM	-0.0067	0.0688	0.0048	0.793	0.2338
N. Imputation	-0.0004	0.0648	0.0041	0.957	0.2841
Missing Mechanism *(c)*					
Complete Obs.	-0.0905	0.1000	0.0182	0.782	0.3955
Weighted-GMM	-0.0035	0.0747	0.0060	0.773	0.2751
N. Imputation	-0.0055	0.0677	0.0046	0.946	0.2862

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.

$n = 150$					
Methods	Bias	Std. Err.	MSE	Coverage	Length of CI
		$\beta_0 = -1$			
Full Observations	-0.0035	1.244	1.549	0.967	5.380
Complete Obs.	-1.622	1.489	4.847	0.901	6.429
Weighted-GMM	-0.3113	1.402	2.061	0.932	5.107
N. Imputation	0.0645	1.279	1.640	0.953	5.368
		$\beta_1 = 1$			
Full Observations	0.0270	0.4270	0.1831	0.965	1.835
Complete Obs.	0.4070	0.4995	0.4152	0.908	2.308
Weighted-GMM	0.1200	0.4795	0.2443	0.935	1.722
N. Imputation	-0.0030	0.4346	0.1889	0.951	1.828
		$\beta_2 = -1.5$			
Full Observations	-0.0766	0.5009	0.2568	0.976	2.172
Complete Obs.	-0.0664	0.5539	0.3112	0.975	2.506
Weighted-GMM	-0.0663	0.5589	0.3168	0.837	1.513
N. Imputation	-0.0330	0.5512	0.3049	0.940	2.037
$n = 250$					
Methods	Bias	Std. Err.	MSE	Coverage	Length of CI
		$\beta_0 = -1$			
Full Observations	-0.0591	0.9506	0.9073	0.951	3.760
Complete Obs.	-1.634	1.136	3.960	0.774	4.621
Weighted-GMM	-0.2787	1.053	1.187	0.933	3.895
N. Imputation	-0.0196	0.9728	0.9468	0.952	3.907
		$\beta_1 = 1$			
Full Observations	0.0331	0.3252	0.1069	0.955	1.338
Complete Obs.	0.4018	0.3873	0.3115	0.829	1.645
Weighted-GMM	0.0999	0.3663	0.1441	0.925	1.327
N. Imputation	0.0146	0.3378	0.1143	0.950	1.413
		$\beta_2 = -1.5$			
Full Observations	-0.0384	0.3705	0.1387	0.968	1.537
Complete Obs.	-0.0402	0.4153	0.1741	0.961	1.719
Weighted-GMM	-0.0443	0.4220	0.1801	0.850	1.148
N. Imputation	-0.0159	0.4152	0.1726	0.967	1.726

Table 2: Inference for parameters in a logistic regression model 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.

Gene	Probe	Complete Observations Only (parametric)		Nonparametric Imputation (with empirical likelihood)	
		Intercept			
H3071E5	1444597_at	-21.99	(-40.97, -2.998)	-15.69	(-37.02, 5.209)
Slc26a8	1441747_at	73.59	(49.45, 97.73)	67.28	(38.34, 95.87)
Tex9	1453360_a_at	-23.81	(-46.12, -1.507)	-14.66	(-38.57, 8.776)
Rps16	1455835_at	-13.52	(-31.08, 4.041)	-8.090	(-26.76, 10.18)
		Slope			
H3071E5	1444597_at	10.16	(5.720, 14.59)	8.736	(2.688, 14.21)
Slc26a8	1441747_at	-6.352	(-9.294, -3.411)	-5.561	(-9.431, -1.471)
Tex9	1453360_a_at	5.101	(2.588, 7.613)	4.094	(0.8753, 6.979)
Rps16	1455835_at	6.766	(3.371, 10.16)	5.754	(1.948, 9.236)
		Correlation Coefficient			
H3071E5	1444597_at	0.5757	(0.3395, 0.7436)	0.4426	(0.1321, 0.6977)
Slc26a8	1441747_at	-0.5533	(-0.7285, -0.3102)	-0.4319	(-0.6809, -0.0761)
Tex9	1453360_a_at	0.5296	(0.2996, 0.7124)	0.4024	(0.1013, 0.6846)
Rps16	1455835_at	0.5256	(0.2744, 0.7097)	0.4151	(0.0755, 0.6613)

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.