Published in final edited form as:

Biometrika. 2013 ; 100(1): . doi:10.1093/biomet/ass056.

Nuisance parameter elimination for proportional likelihood ratio models with nonignorable missingness and random truncation

Kwun Chuen Gary Chan

Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.
kcgchan@u.washington.edu

Summary

We show that the proportional likelihood ratio model proposed recently by Luo & Tsai (2012) enjoys model-invariant properties under certain forms of nonignorable missing mechanisms and randomly double-truncated data, so that target parameters in the population can be estimated consistently from those biased samples. We also construct an alternative estimator for the target parameters by maximizing a pseudo-likelihood that eliminates a functional nuisance parameter in the model. The corresponding estimating equation has a U-statistic structure. As an added advantage of the proposed method, a simple score-type test is developed to test a null hypothesis on the regression coefficients. Simulations show that the proposed estimator has a small-sample efficiency similar to that of the nonparametric likelihood estimator and performs well for certain nonignorable missing data problems.

Keywords

Double truncation; Nonignorable missingness; Pairwise pseudolikelihood; U-statistic

1. Introduction

Regression modelling is popular in data analysis. However, challenges arise when the response has a finite range with a substantial portion of data taking boundary values, when its distribution is multimodal and asymmetric, or when the relationship between response and covariates is nonlinear. To handle regression modelling with an unknown response distribution and other such practical concerns, Luo & Tsai (2012) proposed a semiparametric proportional likelihood ratio model, which is a parsimonious alternative to existing nonparametric models. The semiparametric model relates a response Y and a $p \times 1$ covariate vector X in the following way:

$$f(y|x) = \frac{\exp(\beta^T xy)g(y)}{\int \exp(\beta^T xy)dG(y)}, \quad (1)$$

where β is a Euclidean parameter of interest and $G(y)$ is an unspecified baseline distribution function with density function $g(y)$ with respect to some dominating measure. For arbitrary y_1 and y_2 within the support of G , we have

$$\log \frac{f(y_2|x)}{f(y_1|x)} = \log \frac{g(y_2)}{g(y_1)} + \beta^T x(y_2 - y_1). \quad (2)$$

If X is one-dimensional and the response Y is binary, then

$$\beta = \log \frac{f(1|x+1)/f(0|x+1)}{f(1|x)/f(0|x)},$$

which is exactly the log odds ratio. Therefore, β is a generalization of the log odds ratio for an arbitrary response. For vector-valued X , it follows from (2) that the j th component of β is the loglikelihood ratio that the response increases by one unit given that the j th component of x increases by one unit, while all the other components of x are held fixed. There is no intercept term in this model, because an intercept would be contained in G .

Luo & Tsai (2012) discussed extensively the connection of model (1) with generalized linear models (Nelder & Wedderburn, 1972), density ratio models (Qin & Zhang, 1997), biased sampling models (Gilbert et al., 1999), single-index models (Ichimura, 1993) and exponential tilt regression models (Rathouz & Gao, 2009). Huang & Rathouz (2012) explicitly modelled the mean function for a proportional likelihood ratio model with discrete covariates. We will show that model (1) is a robust extension of the generalized linear model for certain missing data problems and doubly truncated data. For model (1), Luo & Tsai (2012) proposed a nonparametric maximum likelihood method that estimates β and G simultaneously by an iterative algorithm, as usual likelihood arguments do not lead to elimination of the nuisance function $G(y)$. We present an alternative method for estimating β , which eliminates the unknown function G based on conditioning on the rank of responses (Kalbfleisch, 1978) in a pairwise fashion (Liang & Qin, 2000).

2. Invariance properties under missingness and truncation

Missing data are common in practice for reasons that are often beyond the control of investigators. Suppose that we fail to observe response Y for some observations and covariates X for some possibly different observations. Let R_Y and R_X be the indicators that Y and X are being observed, and let $R = R_X R_Y$ be the indicator of both being observed. We refer to an observation with $R = 1$ as a complete case; for such an observation, the conditional density is

$$f(y_i|x_i, r_i=1) = \frac{\text{pr}(r_i=1|y_i, x_i) f(y_i|x_i)}{\int \text{pr}(r_i=1|y, x_i) f(y|x_i) dy}. \quad (3)$$

As seen from (3), inference for β from complete case observations often requires correct specification of the conditional probability of R given (Y, X) . However, when $\text{pr}(R=1|Y, X)$ depends on the unobserved data, the missing mechanism is nonignorable (Rubin, 1976). When $\text{pr}(R|Y, X)$ is arbitrary, inference based on (3) requires untestable assumptions on the missing data mechanism.

We assume the following factorization of the missing data model:

$$\text{pr}(R=1|Y=y, X=x) = h_1(x)h_2(y) \quad (4)$$

for arbitrary functions $h_1(\cdot)$ and $h_2(\cdot)$. This assumption will be satisfied for the following examples, which include cases where missingness is nonignorable.

Example 1. Suppose that response Y is subject to missingness but covariate X is completely observed; that is, $R_X = 1$.

- a. Suppose that the missingness depends only on unobserved Y , $\text{pr}(R_Y=1|Y, X) = 1_A(Y)$; then (4) is satisfied with $h_1(x) = 1$ and $h_2(y) = 1_A(y)$.

- b. Suppose that the missingness depends only on observed X , $\text{pr}(R_Y = 1 \mid Y, X) = h_1(X)$; then (4) is satisfied with $h_1(x) = h_1(x)$ and $h_2(y) = 1$.

Example 2. Suppose that response Y is completely observed but covariate X is subject to missingness; that is, $R_Y = 1$.

- a. Suppose that the missingness depends only on observed Y , $\text{pr}(R_X = 1 \mid Y, X) = h_2(Y)$; then (4) is satisfied with $h_1(x) = 1$ and $h_2(y) = h_2(y)$.
- b. Suppose that the missingness depends only on unobserved X , $\text{pr}(R_X = 1 \mid Y, X) = h_1(X)$; then (4) is satisfied with $h_1(x) = h_1(x)$ and $h_2(y) = 1$.

Example 3. Here both the response Y and the covariate X may be missing.

- a. Suppose that the missing data mechanism for both Y and X depends only on response Y , i.e. $\text{pr}(R_X = 1, R_Y = 1 \mid Y, X) = h_1(Y)$; then (4) is satisfied with $h_1(x) = 1$ and $h_2(y) = h_1(y)$.
- b. Suppose that the missing data mechanism for both Y and X depends only on covariate X , i.e., $\text{pr}(R_X = 1, R_Y = 1 \mid Y, X) = h_1(X)$; then (4) is satisfied with $h_1(x) = h_1(x)$ and $h_2(y) = 1$.
- c. Suppose that the missing data mechanism for Y depends only on Y and the missing data mechanism for X depends only on X , i.e., $\text{pr}(R_Y = 1 \mid Y, X) = h_2(Y)$ and $\text{pr}(R_X = 1 \mid Y, X) = h_1(X)$. Suppose further that R_X and R_Y are conditionally independent given (Y, X) . Then (4) is satisfied with $h_1(x) = h_1(x)$ and $h_2(y) = h_2(y)$.

Except for Examples 1(b) and 2(a), all the cases have nonignorable missing data. We now state an invariance property for the proportional likelihood ratio model.

Property 1. If (4) is satisfied, then

$$f(y|x, r=1) = \frac{\exp(\beta^T xy) dG_1(y)}{\int \exp(\beta^T xy) dG_1(y)}$$

where $G_1(y) = \int_{-\infty}^y h_2(y') dG(y')$ and β is the same as in (1).

Property 1 can be shown by combining (1), (3) and (4), and it generalizes the model-invariant property of the logistic regression model under case-control sampling (Prentice & Pyke, 1979). Suppose that a binary response Y follows a logistic regression model $\text{logit}\{\text{pr}(Y = 1 \mid X = x)\} = \beta^T x$. This model is a special case of model (1) where the baseline distribution follows a Bernoulli distribution with success probability $\exp(\beta^T x) / \{1 + \exp(\beta^T x)\}$. Case-control sampling is a special case of Example 2(a) with $\text{pr}(R_X = 1 \mid Y = y) = h_2(y)$ ($y = 0, 1$). It is well known that case-control data also follow a logistic regression model with the same coefficients as in the population model but a different intercept $\beta^* = \beta + \log(h_2(1)/h_2(0))$.

The second invariance property is for randomly truncated data, which are often collected in medical studies. Suppose that the response Y is observed only when $T_1 \leq Y \leq T_2$, where T_1 and T_2 are random variables indicating the left and right truncation times. When Y is a time-to-event response, left truncation is often observed in cross-sectional sampling, where prevalent subjects have to be surviving at the recruitment time to be eligible to enter the study. Right truncation often occurs in retrospective sampling, where subjects are only observed if the failure event happens before the sampling time. Let $f_{T_1, T_2}(t_1, t_2)$ be the joint

distribution of (T_1, T_2) , and let $f_S(y|x)$ be the sampling conditional density for the randomly truncated data.

Property 2. If (T_1, T_2) is independent of (Y, X) , then

$$f_S(y|x) = \frac{\exp(\beta^T xy) dG_2(y)}{\int \exp(\beta^T xy) dG_2(y)}$$

where $G_2(y) = \int_{-\infty}^y \int_{-\infty}^{y'} \int_{-\infty}^{\infty} f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 dG(y')$ and is the same as in (1).

A derivation of Property 2 is given in the Appendix. If the population follows a proportional likelihood ratio model, randomly truncated data will also follow such a model, with the same target parameters as in the population model. Randomly truncated data can be used directly for estimating β , even though the sample is biased. Property 2 holds under an independent truncation condition and is not true in general when the truncation distribution is covariate-dependent. We will discuss covariate-dependent truncation in § 5.

3. Pairwise conditioning and estimation

In this section, we propose an alternative method for estimating β in model (1), which eliminates the functional nuisance parameter $G(y)$. By the invariance properties 1 and 2 discussed in § 2, the method also estimates the target parameters β in certain nonignorable missing data problems and for randomly truncated responses.

Suppose that we observe $(y_i, x_i)(i = 1, \dots, n)$, which are independent and identically distributed as the random variables (Y, X) where the conditional density function of Y given X follows model (1). Let $y = (y_1, \dots, y_n)$ and $x = (x_1, \dots, x_n)$, and let $y_{(\cdot)} = (y_{(1)}, \dots, y_{(n)})$ be the order statistics for $(y_1, \dots, y_n)$. Kalbfleisch (1978) showed that

$$f(y|x, y_{(\cdot)}) = \frac{\prod_{i=1}^n f(y_i|x_i)}{\sum_{j \in \sigma} \prod_{l=1}^n f(y_{j_l}|x_{j_l})},$$

where $j = (j_1, \dots, j_n)$ and σ is the set of permutations of the integers $\{1, \dots, n\}$. For model (1), we get

$$f(y|x, y_{(\cdot)}) = \frac{\exp(\beta^T \sum_{i=1}^n x_i y_i)}{\sum_{j \in \sigma} \exp(\beta^T \sum_{l=1}^n x_{j_l} y_{j_l})}, \quad (5)$$

which contains only the target parameters β but not the nuisance parameter G . The conditional likelihood (5) provides a legitimate way of constructing likelihood inference for β , but the computational burden is of order $n!$. By applying (5) to a parametric generalized linear model with missing data, Liang & Qin (2000) proposed the pairwise pseudolikelihood

$$\prod_{i < k} \frac{f(y_i|x_i) f(y_k|x_k)}{f(y_i|x_i) f(y_k|x_k) + f(y_i|x_k) f(y_k|x_i)},$$

which applies the conditioning argument of Kalbfleisch (1978) to every pair of observations. The computational burden is of order n^2 . Upon applying this argument to model (1), we have the pairwise likelihood

$$L_{\text{PL}}(\beta) \propto \prod_{i < k} \frac{1}{1 + \exp\{-\beta^T(x_i - x_k)(y_i - y_k)\}},$$

which is again free of the functional parameter G . We denote by $\hat{\beta}_{\text{PL}}$ the value of β that maximizes $L_{\text{PL}}(\beta)$, which is equivalent to the solution of the pseudo-score equation

$$\frac{2}{n(n-1)} \sum_{i < k} - (x_i^T - x_k^T)(y_i - y_k) \frac{\exp\{-\beta^T(x_i - x_k)(y_i - y_k)\}}{1 + \exp\{-\beta^T(x_i - x_k)(y_i - y_k)\}} = 0. \quad (6)$$

The factor $2/\{n(n-1)\}$ is the reciprocal of the number of terms in the summation.

Before we discuss the statistical properties of $\hat{\beta}_{\text{PL}}$ and a score test based on (6), let us introduce some notation to simplify the discussion. Let

$$\psi_{ik}(\beta) = - (x_i^T - x_k^T)(y_i - y_k) \frac{\exp\{-\beta^T(x_i - x_k)(y_i - y_k)\}}{1 + \exp\{-\beta^T(x_i - x_k)(y_i - y_k)\}},$$

and denote the score function by $s_{\text{PL}}(\beta) = 2/\{n(n-1)\} \sum_{i < k} \psi_{ik}(\beta)$. The score function is a second-order U-statistic (Serfling, 1980). It follows from Proposition 1 in Liang & Qin (2000) that $n^{1/2}(\hat{\beta}_{\text{PL}} - \beta_0)$ converges weakly to a multivariate normal distribution with

covariance matrix $V = \sum_1^{-1} \sum_2 \sum_1^{-1}$, where $\sum_1 = E\{\partial \psi_{12}(\beta_0)/\partial \beta\}$ and $\sum_2 = 4E\{\psi_{12}(\beta_0)\psi_{13}^T(\beta_0)\}$. To estimate the asymptotic variance, we could use a sandwich-type estimator $\hat{V} = \hat{\sum}_1^{-1} \hat{\sum}_2 \hat{\sum}_1^{-1}$ with

$$\hat{\sum}_1 = \frac{2}{n(n-1)} \sum_{i < k} \frac{\partial \psi_{ik}(\hat{\beta}_{\text{PL}})}{\partial \beta}$$

and

$$\hat{\sum}_2 = \frac{4}{n-1} \sum_{i=1}^n \left[\left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \psi_{ij}(\hat{\beta}_{\text{PL}}) \right\} - s_{\text{PL}}(\hat{\beta}_{\text{PL}}) \right]^{\otimes 2} = \frac{4}{n-1} \sum_{i=1}^n \left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \psi_{ij}(\hat{\beta}_{\text{PL}}) \right\}^{\otimes 2},$$

where $a^{\otimes 2} = aa^T$ for a column vector a . This estimator follows from the variance estimator of U-statistics proposed by Sen (1960).

An added advantage of the proposed method is that a simple score-type test can be constructed to test a null hypothesis $H_0: \beta = \beta_0$ based on the pseudo-score function (6).

Under the null hypothesis, $s_{\text{PL}}(\beta_0)$ is a zero-mean U-statistic. Suppose that

$E\{\psi_{12}(\beta_0)\}^{\otimes 2} < \infty$ and $E\{\psi_{12}(\beta_0)\psi_{13}^T(\beta_0)\}$ is nonsingular; it then follows from Serfling (1980, p. 192) that $n_{1/2}s_{\text{PL}}(\beta_0)$ converges weakly to a normal distribution with mean zero and covariance matrix $\sum_2$. Under the null hypothesis, we could estimate $\sum_2$ by

$$\tilde{\Sigma}_2(\beta_0) = \frac{4}{n-1} \sum_{i=1}^n \left[\left\{ \frac{1}{n-1} \sum_{j=1, j \neq i}^n \psi_{ij}(\beta_0) \right\} - s_{PL}(\beta_0) \right]^{\otimes 2},$$

and the test statistic

$$T = n s_{PL}(\beta_0)^T \tilde{\Sigma}_2(\beta_0)^{-1} s_{PL}(\beta_0) \quad (7)$$

has a limiting χ_p^2 distribution.

4. Simulations

We conducted simulation studies to evaluate the finite-sample performance of the proposed estimator. First, we compared the performance of the proposed estimator $\hat{\beta}_{PL}$ with the nonparametric likelihood estimator $\hat{\beta}_{NL}$ of Luo & Tsai (2012) under the same simulation setting as in that paper. Next, we compared the performance of the proposed estimator and the least-squares estimator for a Gaussian-distributed response with nonignorable missing data and doubly truncated data. We also evaluated the empirical Type I error and power for the score test (7). We generated 1000 independent datasets in each scenario.

A continuous covariate X_2 was generated from a normal distribution with mean zero and standard deviation 0.5. For a given X_2 , X_1 was generated from a Bernoulli distribution with probability of success being $\exp(1 - X_2) / \{1 + \exp(1 - X_2)\}$. The finite-dimensional parameters are $(\beta_1, \beta_2) = (-1, -1)$, and we considered two baseline densities $g(y) = dG(y)$. In scenario 1, $g(y) = (2/\pi)^{1/2} (0.5)^{-1} \exp\{-2(y - 1/4)^2\}$; the response is a positive continuous random variable. In scenario 2, $g(y) = (1 + y)3^y \exp(-3)/(4y!)$; the response is a nonnegative integer-valued random variable. The simulation results for both scenarios are shown in Table 1. The proposed estimator had a slightly higher sampling variability than did the maximum likelihood estimator and a similar small-sample bias. We conclude that even though the proposed estimator does not maximize a full likelihood, its performance is very close to that of the nonparametric maximum likelihood estimator for small samples. We also compared the computation times for simulation studies in scenario 1. For $n = 50, 100$ and 200 , it took on average 0.2, 0.9 and 3.4 seconds, respectively, to compute the proposed estimator, and 2.1, 11.6 and 84.7 seconds to compute the maximum likelihood estimator. Thus, using the proposed estimator led to a substantial gain in computational efficiency; we will discuss this further in § 5.

We then considered a scenario in which there were missing data and where the missing mechanism was nonignorable. We considered a Gaussian response and compared $\hat{\beta}_{PL}$ with the least-squares estimator $\hat{\beta}_{LS}$. Covariate data followed the same distribution as in the previous simulation, and $g(y)$ was taken to be the standard normal density function. The response Y was observed with probability $\text{pr}(R = 1 \mid Y = y, X = x) = \exp(-2y) / \{1 + \exp(-2y)\}$. Missingness was nonignorable and the least-squares estimator was inconsistent. The results are shown in Table 2(a). We further considered a scenario for doubly truncated data. We generated T_1 from a normal distribution with mean -1 and variance 1 and let $T_2 = T_1 + 2$. Data are observed only when $T_1 \leq Y \leq T_2$. The results are shown in Table 2(b). In both scenarios, we observed that the least-squares estimator was biased and its confidence interval had a much lower coverage probability than the nominal value. The proposed estimator was more variable, but performed well in the sense of having a small bias and a coverage probability close to the nominal value.

Table 3 summarizes results from the score test (7) for testing $H_0: (\beta_1, \beta_2) = (0, 0)$. A 5% significance level was chosen. The null hypothesis was rejected if T exceeded a critical value based on the χ^2_2 distribution. Data were generated following scenario 1 with a varying n . The results showed that the test had an empirical Type I error that was close to the nominal value even for small samples. Under alternative hypotheses, the power of the tests increased quickly with sample size.

5. Discussion

In this paper, we have discussed invariance properties and an alternative method for estimating the target parameters for a proportional likelihood ratio model that was proposed recently by Luo & Tsai (2012). The invariance properties are useful in situations where the missing data mechanism is probably nonignorable and where the analysts may not have enough knowledge to model the missing data mechanism or the truncation distribution. Such unknown quantities become part of the functional nuisance parameter. Any biased sampling (Gilbert et al., 1999) in the form $f_S(y|x) = b(y, x) f(y|x) / \int b(y, x) \exp(\beta^T xy) dG(y)$, where the data are doubly truncated with covariate-dependent truncation probability, fits into this general situation. In this case, a pairwise pseudolikelihood can be constructed when $b(y, x)$ is known and does not follow a factorization as in (4):

$$L_{PL}(\beta) \propto \prod_{i < k} \frac{1}{1 + R(y_i, y_k, x_i, x_k) \exp\{-\beta^T (x_i - x_k)(y_i - y_k)\}}$$

where

$$R(y_i, y_k, x_i, x_k) = \frac{b(y_i, x_k)b(y_k, x_i)}{b(y_i, x_i)b(y_k, x_k)}.$$

The proposed pseudolikelihood method has a lower computational burden than the maximum likelihood estimator. As can be seen from equation (4) in Luo & Tsai (2012), the computation is dominated by the last term in that expression. The computation of the constraint maximization involves a profiling estimator (Luo & Tsai, 2012, equation 5), so the dominating term will involve summation over three indices. When both Y and X are continuous the computational burden will be $O(n^3)$, while the estimator proposed in this paper has a computational burden of $O(n^2)$.

Acknowledgments

The author thanks the editor, an associate editor, two reviewers and Drs Patrick Heagerty, Ali Shojaie and Mary Lou Thompson for suggestions which have greatly improved this paper.

Appendix

Derivation of Property 2

Conditioning on $T_1 = t_1$ and $T_2 = t_2$, the sampling density for the randomly truncated data is $f(y|x)$ if $t_1 \leq y \leq t_2$ and 0 otherwise. Averaging over the joint distribution of truncation times (T_1, T_2) , the joint probability of $Y = y, T_1 \leq y \leq T_2$ is $\int_{-\infty}^y \int_y^{\infty} f(y|x) f_{T_1, T_2}(t_1, t_2) dt_2 dt_1$, and the probability $\text{pr}(T_1 \leq Y \leq T_2 | X = x)$ is $\int_{-\infty}^y \int_y^{\infty} f(y|x) f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 dy$. For randomly truncated data, $f_S(y|x) = f(y|x, T_1 \leq Y \leq T_2)$, and therefore

$$f s(y|x) = \frac{\int_{-\infty}^y \int_y^{\infty} f(y|x) f_{T_1, T_2}(t_1, t_2) dt_2 dt_1}{\int \int_{-\infty}^y \int_y^{\infty} f(y|x) f_{T_1, T_2}(t_1, t_2) dt_2 dt_1 dy} = \frac{\exp(\beta^T xy) \int_{-\infty}^y \int_y^{\infty} f_{T_1, T_2}(t_1, t_2) dG(y)}{\int \exp(\beta^T xy) \int_{-\infty}^y \int_y^{\infty} f_{T_1, T_2}(t_1, t_2) dG(y)} = \frac{\exp(\beta^T xy) dG_2(y)}{\int \exp(\beta^T xy) dG_2(y)}.$$

References

- Gilbert PB, Lele SR, Vardi Y. Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika*. 1999; 86:27–43.
- Huang A, Rathouz PJ. Proportional likelihood ratio models for mean regression. *Biometrika*. 2012; 99:223–229.
- Ichimura H. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Economet*. 1993; 58:71–120.
- Kalbfleisch JD. Likelihood methods and nonparametric tests. *J. Am. Statist. Assoc.* 1978; 73:167–170.
- Liang K-Y, Qin J. Regression analysis under non-standard situations: a pairwise pseudolikelihood approach. *J. R. Statist. Soc. B.* 2000; 62:773–786.
- Luo X, Tsai WY. A proportional likelihood ratio model. *Biometrika*. 2012; 99:211–222.
- Nelder JA, Wedderburn RWM. Generalized linear models. *J. R. Statist. Soc. A.* 1972; 135:370–384.
- Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979; 66:403–411.
- Qin J, Zhang B. A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika*. 1997; 84:609–618.
- Rathouz PJ, Gao LP. Generalized linear models with unspecified reference distribution. *Biostatistics*. 2009; 10:205–218. [PubMed: 18824517]
- Rubin DB. Inference and missing data. *Biometrika*. 1976; 63:581–592.
- Sen PK. On some convergence properties of U-statistics. *Calcutta Statist. Assoc. Bull.* 1960; 10:1–18.
- Serfling, RJ. *Approximation Theorems of Mathematical Statistics*. New York: Wiley; 1980.

Table 1

Comparison of PL and NL

	NL		NL		PL		PL		PL		PL	
	Bias $\times 10^3$		SSE $\times 10^3$		Bias $\times 10^3$		SSE $\times 10^3$		SEE $\times 10^3$		95% CP	
	1	2	1	2	1	2	1	2	1	2	1	2
	Scenario 1											
$n = 50$	-145	-177	1059	1018	-170	-207	1090	1048	984	918	95	94
$n = 100$	-56	-38	696	604	-66	-48	705	616	661	616	95	96
$n = 200$	-40	-41	464	434	-44	-44	468	437	457	427	96	94
	Scenario 2											
$n = 50$	-125	-136	366	370	-100	-111	384	376	342	333	93	93
$n = 100$	-68	-68	238	227	-56	-61	253	240	235	238	95	94
$n = 200$	-40	-36	164	159	-34	-28	174	166	167	163	94	95

SSE, sampling standard error; SEE, average of standard error estimates; CP, coverage probability.

Table 2

Comparison of PL and LS

		Bias $\times 10^3$		SSE $\times 10^3$		SEE $\times 10^3$		90% CP		95% CP	
		1	2	1	2	1	2	1	2	1	2
(a) Nonignorable missingness											
$n = 100$	PL	-46	-70	451	377	430	367	90	89	96	95
	LS	284	276	248	230	233	261	63	72	72	83
$n = 200$	PL	-53	-42	319	281	294	255	88	89	94	95
	LS	291	269	180	160	162	182	43	56	57	71
(b) Double truncation											
$n = 100$	PL	-102	-109	623	610	601	585	89	88	94	93
	LS	423	471	264	252	235	266	38	50	50	65
$n = 200$	PL	-61	-65	387	397	378	376	90	89	95	93
	LS	428	425	184	178	169	185	16	25	23	38

SSE, sampling standard error; SEE, average of standard error estimates; CP, coverage probability.

Table 3

Percentage of hypotheses rejected by score test (7)

$(\mathbf{1}, \mathbf{2})$	$(\mathbf{0}, \mathbf{0})$	$(\mathbf{1}, \mathbf{1})$	$(\mathbf{2}, \mathbf{2})$
$n = 50$	6	31	86
$n = 100$	5	59	100
$n = 200$	5	89	100