Statistica Sinica **27** (2017), 1175-1192
doi:https://doi.org/10.5705/ss.202014.0141

A GENERAL APPROACH TO GOODNESS OF FIT FOR U-PROCESSES

Youngjoo Cho and Debashis Ghosh

University of Rochester and University of Colorado

Abstract: Goodness of fit procedures are essential tools for assessing model adequacy. In this work, we present a general theory and approach to goodness of fit techniques based on U-processes for the accelerated failure time (AFT) model. Many of the examples focus on U-statistics of order 2. While many authors have proposed goodness of fit tests for U-statistics of order one, less has been developed for higher order U-statistics. We propose goodness of fit tests for U-statistics of order 2 by using theoretical results from Nolan and Pollard (1987) and Nolan and Pollard (1988). We advance a resampling approach that is a generalization of the one proposed in Lin, Robins and Wei (1996). Simulation studies are used to illustrate the proposed methods.

Key words and phrases: U-statistics, Gaussian process, perturbation method, survival analysis.

1. Introduction

Goodness of fit is fundamental for assessing the appropriateness of a model. Methodology for model checking for parametric regression has been well developed (Lin, Wei and Ying (2002); (Klein and Moeschberger, 2003, Chap. 12)). Assessing adequacy in parametric models is based on studying residuals that capture the difference between the observed and predicted parts of a model (Lin, Wei and Ying (2002)). Residuals are important in that they enable graphical and numerical summaries for assessing model fit.

We consider the linear model

$$T = \mathbf{Z}^T \boldsymbol{\eta}_0 + \epsilon, \tag{1.1}$$

where T is the response variable, possibly log-transformed, $\mathbf{Z}$ is a $p \times 1$ vector of covariates, $\boldsymbol{\eta}_0$ is a $p \times 1$ vector of regression coefficients and ϵ is an error term. The distribution of ϵ is unspecified, so semiparametric methods are used to estimate $\boldsymbol{\eta}_0$.

U-statistics (Hoeffding (1948)) occupy an important role. For a parameter vector $\boldsymbol{\theta}$ and sample $X_1, \ldots X_n$, a U-statistic of order K is defined as

$$\mathbf{U}_n(\boldsymbol{\theta}) = \binom{n}{K}^{-1} \sum_{1 \le i_1 \ldots i_K \le n} h(X_{i_1}, \ldots X_{i_K}),$$

where $h(\cdot)$ is called the kernel, usually symmetric in $(X_{i_1}, \ldots X_{i_K})$.

U-statistics surface in the accelerated failure time model in survival analysis, when observations are right-censored. In this context, an estimating function was given by Tsiatis (1990), and a rank estimator was proposed by Fygenson and Ritov (1994). The first of these is essentially a U-statistic of order 1, while the second is a U-statistics order 2.

Model checking techniques for censored data and uncensored data have been studied in many settings. For censored data, Therneau, Grambsch and Fleming (1990) developed a graphical approach of checking the Cox model by using martingale residuals. Lin, Wei and Ying (1993) proposed model checking based on cumulative sums of martingale residuals for the Cox proportional hazard model. Lin, Robins and Wei (1996) proposed model checking procedures for the accelerated failure time (AFT) model in overall fit. For uncensored data, Lin, Wei and Ying (2002) proposed a cumulative residual approach to check the functional form and link function in generalized linear models. Arbogast and Lin (2004) developed a goodness of fit method for matched case-control studies. León and Cai (2012) proposed checking the form of covariates using 'robust residuals' based on work from León, Cai and Wei (2009). They argued that when a random variable of interest and other covariates have high correlation, in the uncensored case, the approach of Lin, Wei and Ying (2002) clearly fails to detect misspecification because of the high correlation.

The above-mentioned methodology for goodness of fit is based on U-statistics of order 1. Many rank-based estimators arise from U-statistics of order 2, and performing model checking based on the U-statistic of the wrong order may lead to bias. We propose methodology for goodness of fit for U-statistics of order 2 under linear models for censored and uncensored data. Theoretical justification is based on U-process theory from Nolan and Pollard (1987) and Nolan and Pollard (1988). In Section 2, we describe the method of goodness of fit for U-statistic of order 2. Section 3 outlines the results of some simulation studies, while an application to data from an HIV clinical trial is given in Section 4. Some discussion concludes in Section 5.

2. Checking Overall Model Fit

2.1. Independent censoring

Here we consider censored data and assume that failure times are independently censored. Let $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i^T)^T, i = 1, \ldots, n$ and $\boldsymbol{\eta}$ be parameter of interest, $\boldsymbol{\eta}_0$ is true value. We assume that ϵ_i are i.i.d which has $\mathbf{V}_i$ i.i.d. General U-statistics of order 2 with standardization to estimate $\boldsymbol{\eta}_0$ are given by

$$\mathbf{U}_n(\boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} \mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}), \tag{2.1}$$

where $\mathbf{h}(\cdot, \cdot, \boldsymbol{\eta})$ is a kernel function such that $E\{n^{-1/2}\mathbf{U}_n(\boldsymbol{\eta}_0)\} = 0$.

Under mild conditions, the estimator $\hat{\boldsymbol{\eta}}$, the solution of $\mathbf{U}_n(\boldsymbol{\eta}) = 0$, is strongly consistent and asymptotically normal (Jin, Ying and Wei (2001); Honoré and Powell (1994)). Using the assumptions of Honoré and Powell (1994),

$$\mathbf{U}_n(\boldsymbol{\eta}) = \mathbf{U}_n(\boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Psi}_0(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1 + n^{1/2}\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|), \tag{2.2}$$

where $\boldsymbol{\Psi}_0$ is the derivative of $E\{n^{-1/2}\mathbf{U}_n(\boldsymbol{\eta})\}$ evaluated at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. To assess the overall fit of the model, we define

$$\mathbf{U}_n(t; \boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} \mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) I\{g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) \leq t\},$$

where g is a function that belongs to the Euclidian class (Nolan and Pollard (1988)). One natural choice of g is the maximum function. For example, in the AFT model, $g(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta})$, where $a \vee b$ denotes the maximum of a and b. Then (2.2) leads to the following expansion (Lin, Robins and Wei (1996)):

$$\mathbf{U}_n(t; \boldsymbol{\eta}) = \mathbf{U}_n(t; \boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Psi}_0(t)(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1 + n^{1/2}\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\|), \tag{2.3}$$

where $\boldsymbol{\Psi}_0(t)$ is expectation of the slope matrix of $\mathbf{U}_n(t; \boldsymbol{\eta}_0)$ at time t. Note that when $t = \infty$, (2.3) is equal to (2.2). Since the solution of the estimating function $\mathbf{U}_n(\boldsymbol{\eta})$, is strongly consistent, we have that

$$\mathbf{U}_n(t; \hat{\boldsymbol{\eta}}) = \mathbf{U}_n(t; \boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Psi}_0(t)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + o_p(1).$$

If the model is correct, then $\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$ fluctuates around 0. $\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$ contains information about the model behavior, analogous to the martingale residuals in Lin, Robins and Wei (1996) and Lin, Wei and Ying (1993).

In this case, the key issue is to show that the process $\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$ converges to a mean-zero Gaussian process. We cannot use the empirical process results from Lin, Wei and Ying (1993) and Lin, Robins and Wei (1996), because we do not have a sum of independent and identically distributed random variables in the estimating function. However, by using the U-process theory of Nolan and Pollard (1987) and Nolan and Pollard (1988), the following result can be obtained.

Theorem 1. *If model* (1.1) *holds,* $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ *converges to a Gaussian process with mean zero and covariance function given in the online Supplementary Materials.*

The proof of the result is in the online Supplementary Materials. The idea of the proof is to use the fact that the class of indicator functions is Euclidean and to show the tightness of each term in $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$.

The next issue is to find the null distribution of $\mathbf{U}_n(t;\boldsymbol{\eta})$. Since the structure of the process $\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta})I\{g(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta})\le t\}$ is unknown, it is very difficult to tackle the process directly. One way to solve this problem is to approximate the process by a known distribution (Lin, Wei and Ying (1993)). Since $\mathbf{U}_n(t;\boldsymbol{\eta})$ is nonsmooth, approximation through Taylor expansion does not work. To find an expression for the approximate distribution of $n^{1/2}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0)$, a resampling approach (Parzen, Wei and Ying (1994)) is used. Resampling has been used in a variety of covariance matrix estimation settings for rank regression estimators (e.g. Parzen, Wei and Ying (1994); Lin, Robins and Wei (1996); Peng and Fine (2006); Jin, Ying and Wei (2001)). In this approach,

$$\mathbf{U}_n(\boldsymbol{\eta}) = -\mathbf{u}_r, \tag{2.4}$$

with $\mathbf{u}_r$ simulated from a normal distribution whose mean is 0 and covariance matrix is $\hat{\boldsymbol{\Sigma}}$, where $\hat{\boldsymbol{\Sigma}}$ is estimated covariance matrix of $\mathbf{U}_n(\boldsymbol{\eta})$. Let the solution of (2.4) be $\boldsymbol{\eta}^*$. Under mild conditions, given the observed data, $n^{1/2}(\boldsymbol{\eta}^*-\hat{\boldsymbol{\eta}})$ has the same asymptotic distribution as the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}_0)$ (Parzen, Wei and Ying (1994)). Let $Q_1,\ldots,Q_n$ be standard normal random variables.

Theorem 2. *If model* (1.1) *holds,*

$$\hat{\mathbf{U}}_n(t;\boldsymbol{\eta}^*) = \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j;\hat{\boldsymbol{\eta}})\le t\}Q_i+\mathbf{U}_n(t;\boldsymbol{\eta}^*)-\mathbf{U}_n(t;\hat{\boldsymbol{\eta}}),$$

conditional on the observed data, converges weakly to the same Gaussian Process limit as the $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ *of Theorem 1.*

The proof of this result is also in the online Supplementary Materials. These processes, called bootstrapped processes, are fundamental for checking the overall fit of model. We can adopt the approach of Lin, Robins and Wei (1996) for graphical and numerical summaries. For a graphical summary, we randomly choose 20 or 30 observations from $\hat{\mathbf{U}}_n(\cdot)$ and plot them with the observed process. Lack of fit can be checked by examining the behavior of observed process and observation from resampling processes graphically. In addition to the graphical approach, it is possible to perform a more formal test as in the case of U-statistics of order one. Similar to assessing proportional hazards (Wei (1984); Lin, Wei and Ying (1993)), the test statistic for evaluating overall fit is $D = \sup_t \|\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})\|$. Larger values of D indicate stronger evidence for lack of fit. Let $\boldsymbol{\eta}^{i*}$ be ith value from resampling and suppose there are M resampling values. We can compute a p-value as Hsieh, Ding and Wang (2011) by

$$p = \frac{1}{M} \sum_{i=1}^{M} I\{\sup_t \|\hat{\mathbf{U}}_n(t; \boldsymbol{\eta}^{i*})\| \geq D\}.$$

Let Fygenson and Ritov (1994) estimating function with $n^{1/2}$ standardization be

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}],$$

where $e_i(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T \boldsymbol{\eta}$. The test statistic is $\sup_t \|\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})\|$, where

$$\begin{aligned} \mathbf{U}_n^{FR}(t; \boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)} \sum_{i \neq j} & (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} \\ & - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}] I(e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta}) \leq t). \end{aligned}$$

Now it is necessary to find the null distribution of $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta})$. By the arguments in Fygenson and Ritov (1994), $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has an asymptotically normal distribution with mean 0 and covariance matrix $\boldsymbol{\Gamma}_0^{-1} \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^{-1}$, where $\boldsymbol{\Gamma}_0$ is nonsingular and $\boldsymbol{\Omega}_0$ is an asymptotic covariance matrix of $\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0)$. They proposed to use numerical derivatives for estimating $\boldsymbol{\Gamma}_0$, but these involved unknown hazard functions of the event of the interest and can be numerically unstable.

We again use the approach from Parzen, Wei and Ying (1994) to simulate from the null distribution. The empirical influence function for the asymptotic distribution of $\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0)$ is given by

$$\hat{\mathbf{v}}_i = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}].$$

Then we construct

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = -n^{-1/2} \sum_{i=1}^{n} \hat{\mathbf{v}}_i Q_i. \tag{2.5}$$

Let the solution of (2.5) be $\boldsymbol{\eta}^*$. By Parzen, Wei and Ying (1994), the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has the same limiting distribution as the conditional distribution of $n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}})$ given the data. Then the bootstrapped processes are given by

$$\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*) = \frac{n^{1/2}}{2n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$$
$$\times I(e_i(\boldsymbol{\eta}) \vee e_j(\boldsymbol{\eta}) \le t)Q_i + \mathbf{U}_n^{FR}(t;\boldsymbol{\eta}^*) - \mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}).$$

These bootstrapped processes are random processes whose asymptotic distribution is identical to $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. As described before, a graphical test can be performed by plotting 20 or 30 realized values of $\hat{\mathbf{U}}_n^{FR}(\cdot;\cdot)$ with the observed process $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. A p-value can be computed from replications of $\boldsymbol{\eta}^*$.

It is important to show that the proposed test procedure is consistent, which implies that power approaches one when sample size goes to infinity. Since the power is closely related to rejecting the misspecified model, the estimator under a misspecified model should converge to some constant value (Struthers and Kalbfleisch (1986); Lin and Wei (1989)). Before proving consistency of the proposed test, it is necessary to prove the consistency of estimator under a misspecified model. Let T be the time to failure and C be independent censoring. Let $Y = T \wedge C$, $\Delta = I(T \le C)$ and covariates be $\mathbf{W} = (\mathbf{Z}^T, \mathbf{Z}^{*T})^T$. Let $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ be the true parameter values corresponding to $\mathbf{Z}$ and $\mathbf{Z}^*$, respectively. Let $\boldsymbol{\tau}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$. The observed data are n i.i.d replicates of $(Y, \Delta, \mathbf{W})$. As before, all times are log-transformed. Assume that the true model is

$$T = \mathbf{W}^T\boldsymbol{\tau}_0 + \epsilon,$$

where ϵ is an i.i.d error term. Suppose that model is fitted using $\mathbf{Z}$ only.

Theorem 3. *Let $\hat{\boldsymbol{\eta}}^{mis}$ be the estimator from the misspecified model. Then $\hat{\boldsymbol{\eta}}^{mis}$ is a consistent estimator of $\boldsymbol{\eta}^{mis}$, which is a solution of*

$$\lambda^*(\boldsymbol{\eta}) = \frac{1}{2}E\bigg[(\mathbf{Z}_1 - \mathbf{Z}_2)\int_{-\infty}^{\infty}\bar{G}(t + \mathbf{W}_1^T\boldsymbol{\tau}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)\bar{G}(t + \mathbf{W}_2^T\boldsymbol{\tau}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)$$
$$\times\{\bar{F}(t + \mathbf{W}_2^T\boldsymbol{\tau}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)f(t + \mathbf{W}_1^T\boldsymbol{\tau}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)$$
$$-\bar{F}(t + \mathbf{W}_1^T\boldsymbol{\tau}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)f(t + \mathbf{W}_2^T\boldsymbol{\tau}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)\}dt\bigg],$$

where f is a (true) density of error term ϵ, $\bar{F}$ is (true) survival function of error and $\bar{G}$ is (true) survival function of $C - \boldsymbol{W}^T\boldsymbol{\tau}_0$.

Theorem 4. *The test $D = \sup_t \|\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})\|$ is consistent against the alternative hypothesis that violates the null hypothesis.*

Proofs of Theorems 3 and 4 can be found in the online Supplementary Materials.

2.2. Dependent censoring

It is common that independent censoring does not hold when the event of interest is disease occurrence and death is the dependent censoring mechanism. This is called 'semicompeting risks data' (Fine, Jiang and Chappell (2001); Peng and Fine (2006)). Let X be the time to event of interest, D be time to dependent censoring, C be time to independent censoring and $\mathbf{Z}$ be $p \times 1$ vector of covariates. We assume that all times are log-transformed. Take $\tilde{X} = X \wedge D \wedge C$, $\tilde{D} = D \wedge C$, $\xi = I(D \leq C)$, and $\delta = I(X \leq D \wedge C)$. The observed data is $(\tilde{X}_i, \tilde{D}_i, \xi, \delta, \mathbf{Z}_i), i = 1, \ldots, n$. Now the model is a bivariate AFT model (Lin, Robins and Wei (1996); Peng and Fine (2006)):

$$\begin{pmatrix} X_i = \mathbf{Z}_i^T \boldsymbol{\theta}_0 + \epsilon_i^X \\ D_i = \mathbf{Z}_i^T \boldsymbol{\eta}_0 + \epsilon_i^D \end{pmatrix} \quad i = 1 \ldots n, \tag{2.6}$$

where $\boldsymbol{\gamma}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$ is $2p \times 1$ vector of true value $\boldsymbol{\gamma} = (\boldsymbol{\eta}^T, \boldsymbol{\theta}^T)^T$ and $\epsilon_i = (\epsilon_i^X, \epsilon_i^D)^T$ are independent and identically distributed with unspecified survival function F. Since D only depends on independent censoring, the Tsiatis (1990) estimator for $\boldsymbol{\eta}_0$ is obtained by solving $\mathbf{S}_n(\boldsymbol{\eta}) = 0$ where $\mathbf{S}_n(\boldsymbol{\eta})$ is defined by

$$\mathbf{S}_n(\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^n \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^n I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\}} \right].$$

For the event of the interest, it is necessary to adjust for the effect of dependent censoring to remove bias. To adjust for it, Peng and Fine (2006) used an artificial censoring technique. Let

$$\begin{aligned}
d_{ij}(\boldsymbol{\gamma}) &= \max\{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta}), \mathbf{Z}_j^T(\boldsymbol{\theta} - \boldsymbol{\eta})\}, \\
\tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) &= (X_i - \mathbf{Z}_i^T \boldsymbol{\theta}) \wedge (D_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})), \\
\tilde{\delta}_{i(j)}^*(\boldsymbol{\gamma}) &= I\{(X_i - \mathbf{Z}_i^T \boldsymbol{\theta}) \leq (D_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d_{ij}(\boldsymbol{\gamma}))\}, \\
\psi_{ij}(\boldsymbol{\gamma}) &= \tilde{\delta}_{i(j)}^*(\boldsymbol{\gamma}) I\{\tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) \leq \tilde{X}_{j(i)}^*(\boldsymbol{\gamma})\} - \tilde{\delta}_{j(i)}^*(\boldsymbol{\gamma}) I\{\tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) \geq \tilde{X}_{j(i)}^*(\boldsymbol{\gamma})\}.
\end{aligned}$$

The estimating function proposed by Peng and Fine (2006) is

$$\mathbf{U}_n^P(\gamma) = \frac{2n^{1/2}}{n(n-1)} \sum_{1 \leq i < j \leq n} (\mathbf{Z}_i - \mathbf{Z}_j) \psi_{ij}(\boldsymbol{\gamma}).$$

To evaluate model fit, they adapted the approach of Lin, Robins and Wei (1996). Let $N_{1i}(t; \boldsymbol{\eta}) = \xi_i I(\tilde{D}_i^*(\boldsymbol{\eta}) \leq t)$ and $N_{2i}(t; \gamma) = \tilde{\delta}_i^*(\gamma) I\{\tilde{X}_i^*(\gamma) \leq t\}$, where

$$\begin{aligned}
\tilde{D}_i^*(\boldsymbol{\eta}) &= \tilde{D}_i - \mathbf{Z}_i^T \boldsymbol{\eta}, \\
d(\boldsymbol{\gamma}) &= \max\{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta})\}, \\
\tilde{X}_i^*(\boldsymbol{\gamma}) &= (X_i - \mathbf{Z}_i^T \boldsymbol{\theta}) \wedge (D_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d(\boldsymbol{\gamma})), \\
\tilde{\delta}_i^*(\boldsymbol{\gamma}) &= I\{X_i - \mathbf{Z}_i^T \boldsymbol{\theta} \leq (D_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T \boldsymbol{\eta} - d(\boldsymbol{\gamma}))\}.
\end{aligned}$$

In this case, for $i = 1, \ldots, n$, $\tilde{X}_i^*(\boldsymbol{\gamma})$ is the transformed time to adjust dependent censoring and $\tilde{\delta}_i^*(\boldsymbol{\gamma})$ is a new censoring indicator for disease occurrence (Lin, Robins and Wei (1996)). Observed processes for the dependent censoring and the event of interest are defined by

$$\mathbf{S}_n(t;\hat{\boldsymbol{\eta}}) = n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i \hat{M}_{1i}(t;\hat{\boldsymbol{\eta}}),$$
$$\mathbf{U}_n^L(t;\hat{\boldsymbol{\gamma}}) = n^{-1/2}\sum_{i=1}^n \mathbf{Z}_i \hat{M}_{2i}(t;\hat{\boldsymbol{\gamma}}),$$

where

$$\hat{M}_{1i}(t;\boldsymbol{\eta}) = N_{1i}(t;\boldsymbol{\eta}) - \int_{-\infty}^t I\{\tilde{D}_i^*(\boldsymbol{\eta}) \geq u\} d\hat{\Lambda}_{10}(u;\boldsymbol{\eta}),$$
$$\hat{M}_{2i}(t;\boldsymbol{\gamma}) = N_{2i}(t;\boldsymbol{\gamma}) - \int_{-\infty}^t I\{\tilde{X}_i^*(\boldsymbol{\gamma}) \geq u\} d\hat{\Lambda}_{20}(u;\boldsymbol{\gamma}),$$
$$\hat{\Lambda}_{10}(u;\boldsymbol{\eta}) = \int_{-\infty}^u \frac{\sum_{i=1}^n dN_{1i}(t;\boldsymbol{\eta})}{\sum_{j=1}^n I\{\tilde{D}_j^*(\boldsymbol{\eta}) \geq t\}},$$
$$\hat{\Lambda}_{20}(u;\boldsymbol{\gamma}) = \int_{-\infty}^u \frac{\sum_{i=1}^n dN_{2i}(t;\boldsymbol{\gamma})}{\sum_{j=1}^n I\{\tilde{X}_j^*(\boldsymbol{\gamma}) \geq t\}}.$$

Peng and Fine (2006) used a martingale approach to check model fit. However, their estimating function does not have a martingale structure. Moreover, the artificial censoring applied in the assessment of model fit is one by Lin, Robins and Wei (1996), which differs from that in Peng and Fine (2006). Thus applying a model assessment method using the Lin, Robins and Wei (1996) approach for $\mathbf{U}_n^P(\boldsymbol{\gamma})$ is problematic. By using a similar approach to what is done in Fygenson and Ritov (1994), we can define the score process y

$$\mathbf{U}_n^P(t;\hat{\boldsymbol{\gamma}}) = \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\hat{\boldsymbol{\gamma}}) I\{\tilde{X}_{i(j)}^*(\hat{\boldsymbol{\gamma}}) \vee \tilde{X}_{j(i)}^*(\hat{\boldsymbol{\gamma}}) \leq t\}.$$

As before, a resampling approach is used to derive the null distribution. Let $\hat{\boldsymbol{\gamma}}$ be the estimator of $\boldsymbol{\gamma}_0$ from Peng and Fine (2006). Then by their Theorem 2, $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotically normal distribution with mean zero and covariance matrix $\boldsymbol{\Upsilon}_0^{-1}\boldsymbol{\Xi}_0\boldsymbol{\Upsilon}_0^{-1}$, where $\boldsymbol{\Upsilon}_0$ is nonsingular matrix and $\boldsymbol{\Xi}_0$ is covariance matrix of $\lim_{n\to\infty}\mathbf{W}_n^P(\boldsymbol{\gamma}_0)$, where $\mathbf{W}_n^P(\boldsymbol{\gamma}_0) = [\mathbf{S}_n^T(\boldsymbol{\eta}_0), \{\mathbf{U}_n^P(\boldsymbol{\gamma}_0)\}^T]^T$. By Peng and Fine (2006), the empirical distribution for the asymptotic distribution of $\mathbf{U}_n^P(\boldsymbol{\gamma}_0)$ is

$$\mathbf{J}_i^{(1)} = \xi_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}}\right] - \sum_{l=1}^n \frac{\xi_l I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}$$

$$\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}} \right],$$

$$\mathbf{J}_i^{(2)} = \frac{2}{n-1} \sum_{j=1}^n (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\hat{\gamma}).$$

Let $\mathbf{S}_n(\boldsymbol{\eta}) = n^{1/2}\mathbf{S}_n^{TS}(\boldsymbol{\eta})$ and $\mathbf{J}_i = [\{\mathbf{J}_i^{(1)}\}^T, \{\mathbf{J}_i^{(2)}\}^T]^T$. To apply the resampling approach of Parzen, Wei and Ying (1994), perturbed terms need to be generated. The perturbed term is generated by constructing linear combinations of $\mathbf{J}_i$s and Q_is. Then $\boldsymbol{\gamma}^*$ can be obtained by solving the equations

$$\begin{pmatrix} \mathbf{S}_n(\boldsymbol{\eta}) = -n^{-1/2} \sum_{i=1}^n \mathbf{J}_i^{(1)} Q_i \\ \mathbf{U}_n^P(\gamma) = -n^{-1/2} \sum_{i=1}^n \mathbf{J}_i^{(2)} Q_i \end{pmatrix}.$$

Then $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has the same asymptotic distribution as $n^{1/2}(\boldsymbol{\gamma}^* - \hat{\boldsymbol{\gamma}})$ (Parzen, Wei and Ying (1994)). By using a similar approach as in Section 2.1, we can show that the joint process $\mathbf{W}_n(t, s; \hat{\boldsymbol{\gamma}}) \equiv [\{\mathbf{S}_n(t; \hat{\boldsymbol{\eta}})\}^T, \{\mathbf{U}_n^P(s; \hat{\boldsymbol{\gamma}})\}^T]^T$ is approximated by $\hat{\mathbf{W}}_n(t, s) = [\{\hat{\mathbf{S}}_n(t; \boldsymbol{\eta}^*)\}^T, \{\hat{\mathbf{U}}_n^P(s; \boldsymbol{\gamma}^*)\}^T]^T$, where

$$\begin{aligned} \hat{\mathbf{S}}_n(t; \boldsymbol{\eta}^*) &= n^{-1/2} \sum_{i=1}^n \int_{-\infty}^t \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v)\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v)} \right] d\hat{M}_i(v; \hat{\boldsymbol{\eta}}) Q_i \\ &\quad + \mathbf{S}_n(t; \boldsymbol{\eta}^*) - \mathbf{S}_n(t; \hat{\boldsymbol{\eta}}), \\ \hat{\mathbf{U}}_n^P(s; \boldsymbol{\gamma}^*) &= \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\hat{\gamma}) I\{\tilde{X}_{i(j)}^*(\hat{\gamma}) \vee \tilde{X}_{j(i)}^*(\hat{\gamma}) \leq s\} Q_i \\ &\quad + \mathbf{U}_n^P(s; \boldsymbol{\gamma}^*) - \mathbf{U}_n^P(s; \hat{\boldsymbol{\gamma}}). \end{aligned}$$

Both $[\mathbf{S}_n(t; \hat{\boldsymbol{\eta}})^T, \{\mathbf{U}_n^P(s; \hat{\boldsymbol{\gamma}})\}^T]^T$ and $[\hat{\mathbf{S}}_n(t; \boldsymbol{\eta}^*)^T, \{\hat{\mathbf{U}}_n^P(s; \boldsymbol{\gamma}^*)\}^T]^T$, conditional on the observed data, converge weakly to the same bivariate Gaussian process. The testing procedure based on this bivariate process is the same as in the case of independent censoring.

Remark. Unlike modeling in the independent censoring, joint modeling of failure of interest and dependent censoring is required when there exists dependence between failure of interest and censoring. This leads to derivation of joint processes for the failure of interest and dependent censoring for evaluation of the model fit. However, numerical summaries (test statistic and p-value) can be computed separately for the failure of interest and dependent censoring.

2.3. Uncensored case

In the uncensored case, for the usual linear model (possibly with transforming

Table 1. Size of the proposed method in scenario 1.

sample size \ p-values	Censored data Cutoff values		Uncensored data Cutoff values	
	0.01	0.05	0.01	0.05
$n = 50$	0.005	0.0275	0.0075	0.0275
$n = 100$	0.01	0.085	0.0025	0.0425

response variable by log), the method proposed in Section 2.1 still holds because the estimating function for parameter $\boldsymbol{\eta}$ is the same except that $\Delta_i = 1$ for $i = 1, \ldots, n$. Thus the test statistic and bootstrapped processes for overall fit are equal to $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta})$, except $\Delta_i = 1$ for $i = 1, \ldots, n$. Moreover, the asymptotic theory for censored data in Section 2.1 also hold for uncensored data.

3. Simulation Studies

We first considered simulation studies using the estimating function from Fygenson and Ritov (1994). The error term was distributed as $\epsilon \sim N(0, 1)$. For covariates, in scenario 1, we generated random variable $\mathbf{W} = (Z_1, Z_2)^T$ from a bivariate normal distribution with mean 0 and covariance matrix $\begin{pmatrix} 1 & 1 \\ 1 & 25 \end{pmatrix}$. True parameter values were $\boldsymbol{\eta}_0 = (0.2, 1)^T$. In this setting, say scenario 1, we considered both the censored and uncensored cases. We computed size and power in this setting. We generated 400 simulated datasets and tried 200 resamplings for both size and power calculation. In calculation, due to computational expense, we only considered time points transformed by $\{e_i(\hat{\boldsymbol{\eta}})\}_{i=1}^n$ using $5[j]\%$ quantiles for t. Sample sizes used here were $n = 50$ and $n = 100$. The censoring variable C, when exponentiated, was uniform [0, 200]. For the censored case, the censoring rate was approximately 20% on average.

Table 1 shows the simulation results for type I error rates for the censored and uncensored data scenarios using the proposed method. We computed the powers of the proposed method and the Lin, Robins and Wei (1996) method. For power comparison, we fit the model using only Z_1. Table 2 shows the simulation results comparing the Lin, Robins and Wei (1996) and our method.

The proportion of rejections from the proposed method is higher than that from Lin, Robins and Wei (1996). Figures 1 and 2 show the power corresponding to threshold values of p-value. The plot shows that our proposed method performs better than the Lin, Robins and Wei (1996) method. Table 1 shows the power comparison between the new method and the Lin, Robins and Wei (1996)

Table 2. Comparison of empirical power between new method and Lin, Robins and Wei (1996)'s method in scenario 1.

Censored data				
p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0	0.0025	0.0025	0.0175
Proposed method	0.0625	0.16	0.2275	0.3025
p-values / $n = 100$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.0025	0.0075	0.025	0.04
Proposed method	0.2525	0.3925	0.46	0.5525
Uncensored data				
p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0	0.015	0.0275	0.045
Proposed method	0.0475	0.1175	0.1575	0.2175
p-values / $n = 100$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.0125	0.0175	0.0275	0.0575
Proposed method	0.055	0.1275	0.19	0.245

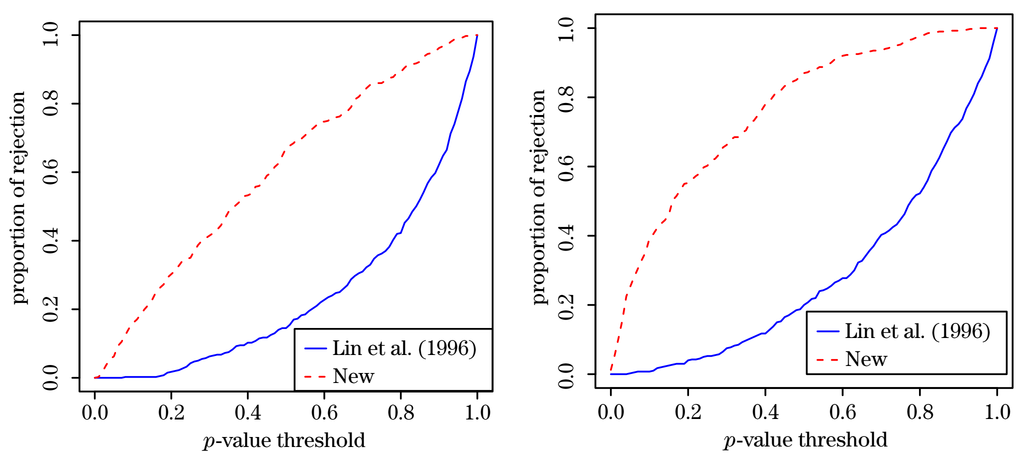


Figure 1. Plot of empirical powers according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for independent censoring case in the scenario 1.

method. Numerical results indicate that the proposed approach has higher power. Moreover, the power difference between the two methods is higher in censored case than in the uncensored case.

In scenario 2, we simulated variable whose variability was larger than in the first scenario and we omitted it in the model fitting. Here we only considered

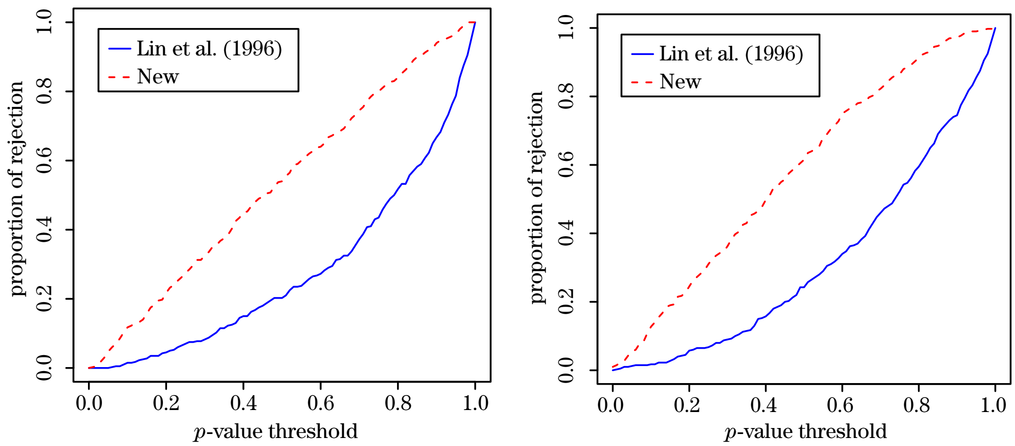


Figure 2. Plot of empirical powers according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for uncensored data in the scenario 1.

censored data. We first generated $(A_1, A_2)^T$ from a bivariate normal distribution with mean 0 and covariance matrix $\begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$.

We let $Z_1 = A_1$ and $Z_2 = \sum_j jA_2^2 I(b[j-1] < A_2 \leq b[j]), j = 1, \ldots 21$, where $b[j]$ was the $5(j-1)\%$ quantile of W_2. $b[0] = -\infty$ and $b[1]$ was the minimum of $b[\cdot]$ and $b[21]$ was the maximum of $b[\cdot]$. The censoring variable was uniform on [0,150]. On average, the censoring rate was between 7% and 8%. True regression coefficient values were $\boldsymbol{\eta}_0 = (0.2, 1)^T$. We simulated 400 datasets with sample size $n = 50, 100,$ and 200. In each simulation run, 200 resampling runs were performed. We fitted the model by using only Z_1 and, for comparison, the testing procedure was compared to that of Lin, Robins and Wei (1996). We only computed the empirical power from each method in this case.

Table 3 shows the numerical results of comparing the two methods and Figure 3 and Figure 4 is a graphical comparison of rejection rates between the methods. As in scenario 1, our method has higher power than the method of the Lin, Robins and Wei (1996). In scenarios, as the sample size increases, the rejection rate of both methods increases and the difference of proportion of rejection between the two methods decreases. When sample size goes to infinity, the power by the proposed method approaches one, supporting Theorem 4.

We applied the proposed method to the dependent censoring case. The steps for data generation are shown below:

Table 3. Comparison of empirical power between new method and Lin, Robins and Wei (1996)'s method for independent censoring case in scenario 2.

p-values / $n = 50$	Cutoff values 0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.0475	0.11	0.18	0.2425
Proposed method	0.1375	0.28	0.37	0.48
p-values / $n = 100$	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.2925	0.37	0.445	0.4925
Proposed method	0.34	0.485	0.6125	0.68
p-values / $n = 200$	0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.5425	0.6475	0.71	0.7525
Proposed method	0.595	0.71	0.8025	0.845

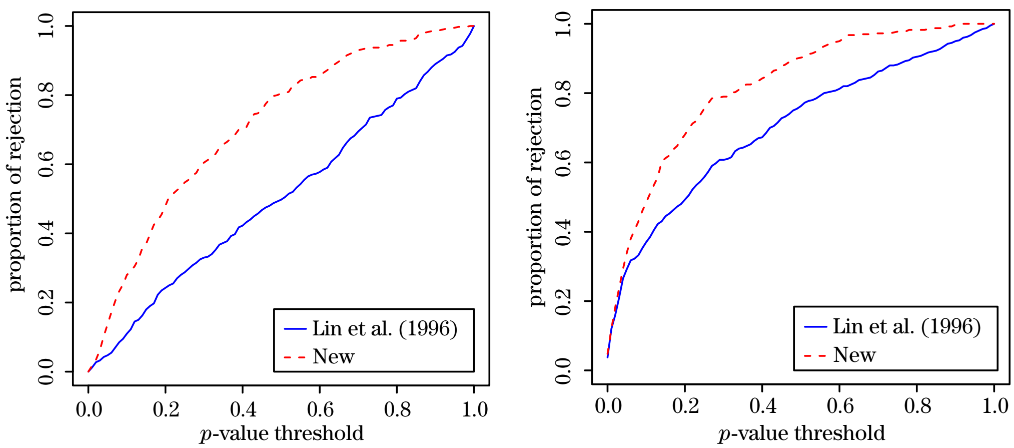


Figure 3. Plot of empirical powers according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for independent censoring case in scenario 2.

1. Generate $W = (W_1, W_2)^T \sim \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$.

2. Set $R_1 = I(W_1 > 0)$ and $R_2 = \sum_j jW_2^2 I(b[j-1] < W_2 \le b[j]), j = 1, \ldots 21$, where $b[j]$ is $5(j-1)\%$ quantile of W_2. $b[0] = -\infty, b[1]$ is minimum of $b[\cdot]$ and $b[21]$ is maximum of $b[\cdot]$.

3. Generate $\epsilon = (\epsilon^X, \epsilon^D) \sim N\left\{ \begin{pmatrix} 0 \\ 1.2 \end{pmatrix}, \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix} \right\}$.

4. Set $\boldsymbol{\theta}_0 = (1, 0.5)$ and $\boldsymbol{\eta}_0 = (0.5, 1)$ and generate $X = \mathbf{R}^T\boldsymbol{\theta}_0 + \epsilon^X$ and $D = \mathbf{R}^T\boldsymbol{\eta}_0 + \epsilon^D$, where $\mathbf{R} = (R_1, R_2)^T$.

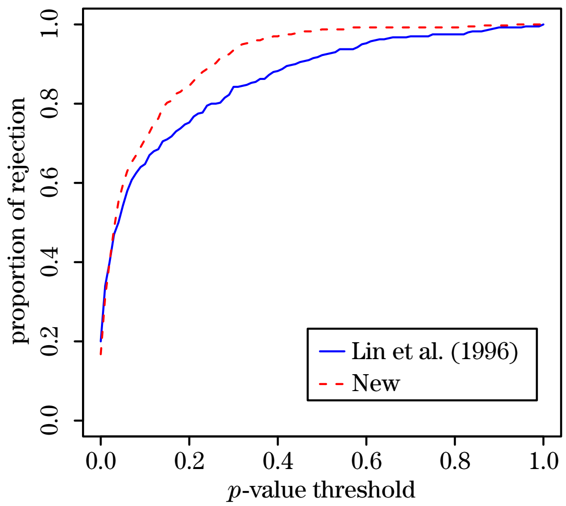
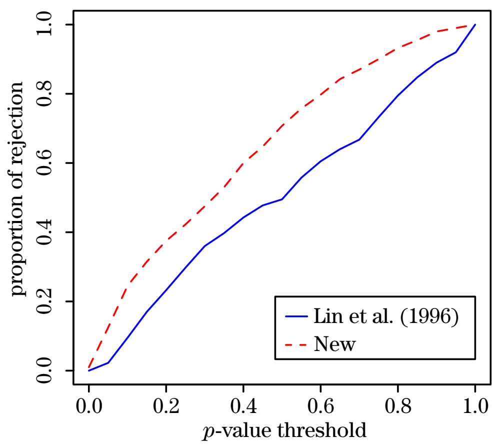


Figure 4. Plot of empirical powers according to threshold p-values when $n = 200$ for independent censoring case in scenario 2.

Table 4. Comparison of empirical power between new method and Lin, Robins and Wei (1996)'s method for model of the event of interest in the presence of dependent censoring.

p-values / $n = 50$	Cutoff values 0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.0225	0.095	0.17	0.2325
Proposed method	0.125	0.245	0.315	0.375
p-values / $n = 100$	Cutoff values 0.05	0.10	0.15	0.2
Lin, Robins and Wei (1996)	0.095	0.205	0.31	0.385
Proposed method	0.25	0.385	0.455	0.53

The independent censoring time C was uniformly distributed [0,100]. On average, approximately 12% of dependent censoring was censored by C and 12% of the event of interest was dependently censored by $\tilde{D}$. We fitted the misspecified model from Section 3.2, which only employs R_1 and computed the statistical power of our method, as well as that of Lin, Robins and Wei (1996), focusing on the event of interest X. In each simulation run, 200 resampling runs were tried. Table 4 shows the results when $n = 50$ based on 400 simulation runs and for $n = 100$ based on 200 simulation runs. Figure 5 shows a plot of empirical power when $n = 50$ and $n = 100$. The plots in Figure 5 and numerical summaries from Table 4 lead the same conclusion as in the independent censoring case. Our method performs better than that of Lin, Robins and Wei (1996).



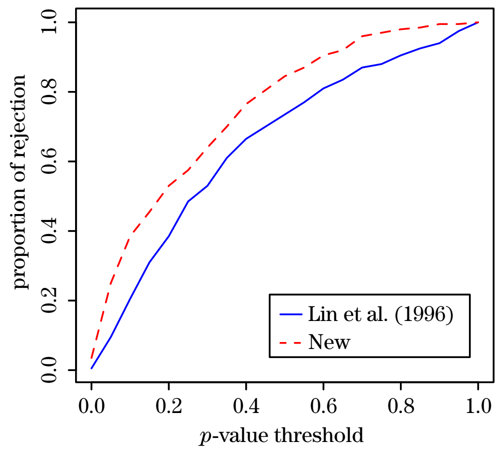


Figure 5. Plot of empirical powers according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for model of the event of interest in the presence of dependent censoring.

4. Real Data Analysis

We applied our method to data from AIDS Clinical Trial Study 364 (Albrecht et al. (2001)), which was previously analyzed by Peng and Fine (2006) and Cho and Ghosh (2015). In this study, the plasma RNA level of every patient is at least 500 copies per ml. The event of interest is time to first viologic failure, which is defined by time of first clinical visit when HIV level $\geq$ 2,000. Patients will leave the study due to deterioration of heath status as time progresses (Peng and Fine (2006)). Hence dependence between failure of interest and censoring (withdrawal) exists.

In this dataset, three levels of treatment were considered: nelfinavir (NFV), efavirenz (EFV), and combination of nelfinavir and efavirenz (NFV + EFV). We considered three covariates: Z_1 was 1 if treatment assignment of a patient was EFV and 0 otherwise; Z_2 was value 1 if treatment assignment of a patient was NFV + EFV and 0 otherwise; Z_3 was log(RNA) level. In Cho and Ghosh (2015), the dependent censoring and the event of interest were analyzed using the Lin, Robins and Wei (1996) and Peng and Fine (2006) approaches jointly. For model checking, Cho and Ghosh (2015) used the approach based on Lin, Robins and Wei (1996) for both the Lin, Robins and Wei (1996) estimator and the Peng and Fine (2006) estimator.

We fitted the model (2.6) using covariates Z_1, Z_2 and Z_3. We compared our approach to that by Cho and Ghosh (2015). The p-value from Cho and Ghosh (2015) was 0.959. The p-value using the new approach was 0.51. Although both

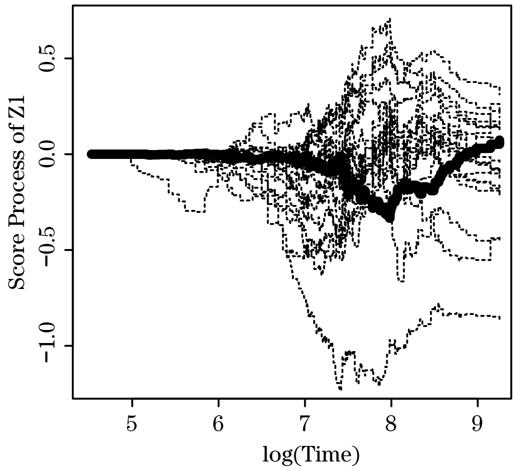


Figure 6. Observed process (bold line) and 20 bootstrapped processes (dashed lines) for the first virologic failure.

p-values show that there is no evidence of lack of fit for the model, substantial decrease is made on the proposed method, suggestive of higher power.

Figure 1 shows a goodness of fit plot of 20 bootstrapped processes along with the observed process. The observed process is moving around zero and bootstrapped processes suggest that there is no substantial deviation of model fit.

5. Discussion

In this paper, our attention has been on checking the overall fit of the model. Other goodness of fit techniques that can be considered are checking functional form of covariates and linearity of the model. Lin, Wei and Ying (1993) proposed methods for these scenarios based on the Cox model. However, direct application of their approaches to the semiparametric AFT model is impossible because the estimating function is nonsmooth. By mimicking our approach and that of Lin, Wei and Ying (1993), for the procedure of Fygenson and Ritov (1994), one can consider the observed process

$$\begin{aligned}\mathbf{U}_{2k}(x;\boldsymbol{\eta}) = \frac{n^{1/2}}{n(n-1)} \sum_{i \neq j} I(Z_{ki} \vee Z_{kj} \leq x)(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} \\ - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]\end{aligned}$$

to check the form of covariates. Developing details about checking functional form of covariates and linearity of the model will be communicated in separate reports.

It is also worthwhile to apply ideas of León and Cai (2012) on checking overall fit in the U-statistics of order 2 case under observational studies. In U-statistics of order 2 case, however, there is no concept of residuals. Thus developing a tool similar to 'robust residuals' can be important. This will be communicated in separate reports. One may be interested in obtaining optimal $g(\cdot,\cdot,\cdot)$ to improve the performance of the test. A possible approach is to combine a set of $g(\cdot,\cdot,\cdot)$s by using weights. This weighting approach is popular (Wei, Lin and Weissfeld (1989); Cho and Ghosh (2015)). However, the way to determine optimal weight for the functions is not straightforward because there is an infinite number of functions. This is also future research.

Supplementary Materials

The proofs of theorems can be found in online Supplementary Materials.

Acknowledgements

This research is supported by NIH grant R01 CA129102.

References

Albrecht, M. A., Bosch, R. J., Hammer, S. M., Liou, S.-H., Kessler, H., Para, M. F., Eron, J., Valdez, H., Dehlinger, M., Katzenstein, D. A. and for The AIDS Clinical Trials Group 364 Study Team. (2001). Nelfinavir, Efavirenz, or both after the failure of Nucleoside treatment of HIV infection. *N. Engl. J. Med.* **345**, 398–407.

Arbogast, P. G. and Lin, D. Y. (2004). Goodness-of-fit methods for matched case-control studies. *Can. J. Stat.* **32**, 373–386.

Cho, Y. and Ghosh, D. (2015). Weighted estimation of the accelerated failure time model in the presence of dependent censoring, *PLOS ONE*, DOI: 10.1371/journal.pone.0124381.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data. *Biometrika* **88**, 907–919.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data, *Ann. Stat.* **22**, 732–746.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Stat.* **19**, 293–325.

Honoré, B. E. and Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *J. Econom.* **64**, 241–278.

Hsieh, J.-J., Ding, A. A. and Wang, W. (2011). Regression analysis for recurrent events data under dependent censoring. *Biometrics* **67**, 719–729.

Jin, Z., Ying, Z. and Wei, L. J. (2001). A simple resampling method by perturbing the minimand, *Biometrika* **88**, 381–390.

Klein, J. P. and Moeschberger, M. L. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*, second edition. Springer-Verlag, New York.

León, L. F., Cai, T. and Wei, L. J. (2009). Robust inferences for covariate effects on survival time with censored linear regression models. *Stat. Biosci.* **1**, 50–64.

León, L. F. and Cai, T. (2012). Model checking techniques for assessing functional form specifications in censored linear regression models. *Stat. Sin.* **22**, 509–530.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *J. Am. Stat. Assoc.* **84**, 1074–1078.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381–393.

Lin, D. Y., Wei, L. J. and Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

Nolan, D. and Pollard, D. (1987). U-processes: rate of convergence. *Ann. Stat.* **15**, 780–799.

Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16**, 1291–1298.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *J. Am. Stat. Assoc.* **101**, 1085–1093.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363–369.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–160.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Stat.* **18**, 354–372.

Wei, L. J. (1984). Testing goodness-of-fit for proportional hazards model with censored observations. *J. Am. Stat. Assoc.* **79**, 649–652.

Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Stat. Assoc.* **84**, 1065–1073.

Department of Biostatistics and Computational Biology, University of Rochester, 601 Elmwood Avenue, P.O. Box 630, Rochester, NY 14642, USA.

E-mail: youngjoo_cho@urmc.rochester.edu

Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, 13001 E. 17th Place, Aurora, CO 80045, USA.

E-mail: DEBASHIS.GHOSH@ucdenver.edu

(Received December 2014; accepted June 2016)