2015

A general approach to goodness of fit for U processes

Debashis Ghosh, *University of Colorado Denver*
Youngjoo Cho, *University of Rochester*

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 statistics. 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 will 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. In this paper, 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 propose a resampling approach which is a generalization of that proposed in Lin et al. (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 model. Methodology for model checking for parametric regression has been well developed (Lin et al. 2002; Klein and Moeschberger, 2003, Chapter 12, pp. 409-423). Assessing adequacy in parametric models is based on studying residuals, which capture the difference between observed and predicted part from a model (Lin et al. 2002). Residuals are an important element in model checking. They enable statisticians to perform graphical and numerical summaries for assessing model fit.

The model considered in this paper is the linear model, which is given by

$$T = \mathbf{Z}^T \boldsymbol{\eta}_0 + \epsilon.$$

where T is response variable, $\mathbf{Z}$ is $p \times 1$ vector of covariates, $\boldsymbol{\eta}_0$ is $p \times 1$ vector of regression coefficients and ϵ is an error term. Note that the distribution of ϵ is unspecified, so to estimate $\boldsymbol{\eta}_0$, nonparametric methods are used.

U-statistics, initially proposed by Hoeffding (1948), occupy an important role in the theory of statistics. For 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 kernel. $h(\cdot)$ is usually symmetric on $(X_{i_1}, \ldots X_{i_K})$. U-statistics are a critical element in semiparametric models. For censored data, we only observe subjects until certain time points. Denote $Y = T \wedge C$ and $\Delta = I(T \le C)$. The observed data are n i.i.d copies of $(Y, \Delta, \mathbf{Z})$, $i = 1, \ldots, n$. One estimating equation for $\boldsymbol{\eta}_0$ is (Tsiatis, 1990) given by

$$\mathbf{S}_n(\boldsymbol{\eta}) = n^{-1/2} \sum_{i=1}^{n} \Delta_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \ge e_i(\boldsymbol{\eta})\} \mathbf{Z}_j}{\sum_{j=1}^{n} I\{e_j(\boldsymbol{\eta}) \ge e_i(\boldsymbol{\eta})\}} \right], \tag{1.1}$$

where $e_i(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T \boldsymbol{\eta}$. Another rank estimator, proposed by Fygenson and Ritov (1994), for the AFT model is given by the solution for the following estimating equation:

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (\mathbf{Z}_i - \mathbf{Z}_j) \Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\}.$$

which can be expressed as

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = \frac{1}{2n(n-1)} \sum_{i \ne j} (\mathbf{Z}_i - \mathbf{Z}_j) [\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]. \tag{1.2}$$

Note that (1.1) is U-statistic of order 1, i.e., U-statistic with $K = 1$ and (1.2) is U-statistic order of 2, i.e., U-statistic with $K = 2$. Let $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i^T)^T, \quad i = 1, \ldots, n$. For (1.1), $\mathbf{h}(\mathbf{V}_i, \boldsymbol{\eta}) = \mathbf{Z}_i - \dfrac{\sum_{j=1}^n I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{e_j(\boldsymbol{\eta}) \geq e_i(\boldsymbol{\eta})\}}$ and for (1.2), $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = \dfrac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$. Note also that if $\Delta_i = 1$ for all $i = 1, \ldots, n$, i.e., all observations are not censored, then $\mathbf{U}_n^{FR}(\boldsymbol{\eta})$ is usual Wilcoxon-type rank test statistic for linear regression (Jin et al. 2001).

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

However, 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. Clearly, $\mathbf{U}_n^{FR}(\boldsymbol{\eta})$ is U-statistic of order 2. In this case, performing model checking based on U-statistic order may lead to bias. In this paper, we propose methodology for goodness of fit for U-statistic order 2 principles using linear model 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, 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 Section 5.

2. Checking overall fit of model

2.1 Independent censoring

In this subsection, we consider censored data and assume that failure times are independently censored. As can be seen in (1.2), the estimating equation proposed by Fygenson and Ritov (1994) is a U-statistic of order 2. Let $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i^T)^T, i = 1, \ldots, n$ and $\boldsymbol{\eta}$ be parameter of interest and $\boldsymbol{\eta}_0$ be true value. General U-statistics of order 2 with standardization to estimate $\boldsymbol{\eta}_0$ in (2.1) is

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

where $\mathbf{h}(\cdot, \cdot, \boldsymbol{\eta})$ is a kernel function such that $E\{n^{-1/2}\mathbf{U}_n(\boldsymbol{\eta}_0)\} = 0$. Note that the kernel of estimating equation in (1.2) is $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}) = \frac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}) > e_i(\boldsymbol{\eta})\} - \Delta_j I\{e_i(\boldsymbol{\eta}) > e_j(\boldsymbol{\eta})\}]$.

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

$$\mathbf{U}_n(\boldsymbol{\eta}) = \mathbf{U}_n(\boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Psi}_0(\boldsymbol{\eta} - \boldsymbol{\eta}_0) + o_p(1 + n^{1/2}||\boldsymbol{\eta} - \boldsymbol{\eta}_0||), \tag{2.1}$$

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 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 maximum of a and b. Then (2.2) leads to the following expansion (Lin et al., 1996) :

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

where $\boldsymbol{\Psi}_0(t)$ is 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 equation $\mathbf{U}_n(\boldsymbol{\eta})$, 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 on the model behavior, analogous to the martingale residuals for in Lin et al. (1996) and Lin et al. (1993).

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

Theorem 1. *Assuming that the model (2.1) is true,* $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$ *converges to Gaussian process with mean 0 and covariance function.*

The proof of this theorem is in the Appendix. The idea of the proof is to use the fact that indicator function is Euclidean and show tightness of each term in $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$.

The next issue is to find null distribution of $\mathbf{U}_n(t;\boldsymbol{\eta})$. Since the structure of 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 et al., 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 et al. 1994) is used. Resampling has been used in a variety of covariance matrix estimation settings for rank regression estimators (e.g. Parzen et al. 1994; Lin et al. 1996; Peng and Fine 2006; Jin et al. 2001). In this approach,

$$\mathbf{U}_n(\boldsymbol{\eta}) = -\mathbf{u}_r, \tag{2.3}$$

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 observed data, $n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}})$ has the same asymptotic distribution as $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ (Parzen et al., 1994). Define $Q_1, \ldots, Q_n$ to be standard normal random variables.

Theorem 2. *Assuming that the model (2.1) is true,*

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

converges weakly to the same Gaussian Process limit as $\mathbf{U}_n(t; \hat{\boldsymbol{\eta}})$.

These processes, which are called bootstrapped processes, are fundamental for checking the overall fit of model. We can adopt the approach of Lin et al. (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 formal test as in the case of U-statistics of order one. Similar to assessing proportional hazards (Wei, 1984; Lin et al., 1993), the test statistic of 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 is M resampling values. We can compute a p-value by (Hsieh et al. 2011)

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

According to approach above, for the estimating equation (1.2), the test statistic is $\sup_t ||\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})||$, where

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

Now it is necessary to find the null distribution of $\mathbf{U}_n^{FR}(t;\boldsymbol{\eta})$. By arguments in Ferguson 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 numerical derivatives involved unknown hazard functions of the event of the interest and can be numerically unstable.

We instead use the approach from Parzen et al. (1994). 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

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}) = -n^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{v}}_i Q_i. \tag{2.4}$$

Let the solution of the equation (2.5) be $\boldsymbol{\eta}^*$. By Parzen et al. (1994), the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ has the same limiting distribution as the conditional distribution of $n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}})$. 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, the 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 by replications of $\boldsymbol{\eta}^*$.

Now it is important to show that the proposed test procedure is consistent. A consistent test is one whose power approaches 1 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 $\mathbf{W} = (\mathbf{Z}^T, \mathbf{Z}^{*T})^T$. 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

$$T = \mathbf{W}^T\boldsymbol{\eta}_0 + \epsilon.$$

where ϵ is an i.i.d error term. Suppose that model is fitted using $\mathbf{Z}$ only, i.e., there is misspecification on model fitting. We need next theorem before proving the consistency of the test.

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*

$$\begin{aligned}\lambda^*(\boldsymbol{\eta}) = \frac{1}{2}E\bigg[&(\mathbf{Z}_1 - \mathbf{Z}_2)\int_0^\infty \bar{G}(t + \mathbf{W}_1^T\boldsymbol{\eta}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)\bar{G}(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)\\ &\times\{\bar{F}(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)f(t + \mathbf{W}_1^T\boldsymbol{\eta}_0 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)\\ &- \bar{F}(t + \boldsymbol{\beta}_0^T\mathbf{W}_1 - \mathbf{Z}_1^T\boldsymbol{\eta}|\mathbf{Z}_1)f(t + \mathbf{W}_2^T\boldsymbol{\eta}_0 - \mathbf{Z}_2^T\boldsymbol{\eta}|\mathbf{Z}_2)\}dt\bigg].\end{aligned}$$

where f is an error density, $\bar{F}$ is survival function of error and $\bar{G}$ is survival function of $C - \boldsymbol{W}^T\boldsymbol{\eta}_0$.

Now we can propose following theorem for consistency of the test.

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.*

2.2 Dependent censoring

In the previous section, independent censoring was assumed. However, it is common that this assumption is violated when the event of interest is disease occurrence with existence of death. This data structure is called 'semicompeting risks data' (Fine et al. 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. Define $\tilde{X} = X \wedge D \wedge C$, $\tilde{D} = D \wedge C$, $\xi = I(D \leq C)$, $\delta = I(X \leq D \wedge C)$. 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 et al. 1996; Peng and Fine, 2006):

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

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, from approach by Tsiatis (1990), estimator for $\boldsymbol{\eta}_0$ is obtained by solving $\mathbf{S}_n(\boldsymbol{\eta}) = 0$. 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 artificial censoring technique. Define

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

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

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

To evaluate model fit, Peng and Fine (2006) adapted the approach of Lin et al. (1996).

Let $N_{1i}(t;\boldsymbol{\eta}) = \xi_i I(\tilde{D}_i^*(\boldsymbol{\eta}) \leq t)$ and $N_{2i}(t;\boldsymbol{\gamma}) = \tilde{\delta}_i^*(\boldsymbol{\gamma}) I\{\tilde{X}_i^*(\boldsymbol{\gamma}) \leq t\}$, where

$$\tilde{D}_i^*(\boldsymbol{\eta}) = \tilde{D}_i - \mathbf{Z}_i^T\boldsymbol{\eta},$$

$$d(\boldsymbol{\gamma}) = \max\{0, \mathbf{Z}_i^T(\boldsymbol{\theta} - \boldsymbol{\eta})\},$$

$$\tilde{X}_i^*(\boldsymbol{\gamma}) = (X_i - \mathbf{Z}_i^T\boldsymbol{\theta}) \wedge (D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d(\boldsymbol{\gamma})),$$

$$\tilde{\delta}_i^*(\boldsymbol{\gamma}) = I\{X_i - \mathbf{Z}_i^T\boldsymbol{\theta} \leq (D_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d(\boldsymbol{\gamma})) \wedge (C_i - \mathbf{Z}_i^T\boldsymbol{\eta} - d(\boldsymbol{\gamma}))\}.$$

In this case, $\tilde{X}_i^*(\boldsymbol{\gamma})$ is the transformed time to adjust dependent censoring and $\tilde{\delta}_i^*(\boldsymbol{\gamma})$ is a new censoring indicator for disease occurrence (Lin et al., 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

$$\hat{M}_{1i}(t;\boldsymbol{\eta}) = N_{1i}(t;\boldsymbol{\eta}) - \int_{-\infty}^t I\{\tilde{D}_i^*(\boldsymbol{\eta}) \geq u\} d\hat{\Lambda}_{10}(u;\boldsymbol{\eta}),$$

$$\hat{M}_{2i}(t;\boldsymbol{\gamma}) = N_{2i}(t;\boldsymbol{\gamma}) - \int_{-\infty}^t I\{\tilde{X}_i^*(\boldsymbol{\gamma}) \geq u\} d\hat{\Lambda}_{20}(u;\boldsymbol{\gamma}),$$

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

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

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

To derive a p-value, as in the previous section, a resampling approach is used to derive the null distribution. Let $\hat{\boldsymbol{\gamma}}$ be estimator of $\boldsymbol{\gamma}_0$. Then by Theorem 2 in Peng and Fine (2006), $n^{1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$ has an asymptotically normal distribution with mean 0 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

$$\begin{aligned}
\mathbf{J}_i^{(1)} = \xi_i \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_i^*(\hat{\boldsymbol{\eta}})\}}\right] - \sum_{l=1}^n \frac{\xi_l I\{\tilde{D}_i^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}} \\
\times \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}\mathbf{Z}_j}{\sum_{j=1}^n I\{\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq \tilde{D}_l^*(\hat{\boldsymbol{\eta}})\}}\right], \\
\mathbf{J}_i^{(2)} = \frac{2}{n-1}\sum_{j=1}^n (\mathbf{Z}_i - \mathbf{Z}_j)\phi_{ij}(\hat{\boldsymbol{\gamma}}).
\end{aligned}$$

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

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

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 et al, 1994). By using a similar approach as in section 2.1, we can show that joint process

$[\{\mathbf{S}_n(t;\hat{\boldsymbol{\eta}})\}^T, \{\mathbf{U}_n^P(s;\hat{\boldsymbol{\gamma}})\}^T]^T$ is approximated by $[\{\hat{\mathbf{S}}_n(u;\boldsymbol{\eta}^*)\}^T, \{\hat{\mathbf{U}}_n^P(s;\boldsymbol{\gamma}^*)\}^T]^T$, where

$$\hat{\mathbf{S}}_n(t;\boldsymbol{\eta}^*) = n^{-1/2}\sum_{i=1}^n \int_{-\infty}^u \left[\mathbf{Z}_i - \frac{\sum_{j=1}^n I(\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v)\mathbf{Z}_j}{\sum_{j=1}^n I(\tilde{D}_j^*(\hat{\boldsymbol{\eta}}) \geq v)}\right] d\hat{M}_i(v;\hat{\boldsymbol{\eta}})Q_i + \mathbf{S}_n(t;\boldsymbol{\eta}^*) - \mathbf{S}_n(t;\hat{\boldsymbol{\eta}})$$

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

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

Remark. As can be seen in this section, 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 derivation of joint processes of failure of interest and dependent censoring for evaluation of the model fit. However, numerical summaries (test statistic and p-value) can be computed for failure of interest and dependent censoring, respectively.

2.3 Uncensored case

In uncensored case, for usual linear model (with transforming response variable by log), the method proposed Section 2.1 still holds because the estimating equation for parameter $\boldsymbol{\eta}$ is same except $\Delta_i = 1$ 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, theorems of asymptotic theory for censored data in Section 2.1 are also still valid for uncensored data.

3. Simulation Studies

We first considered simulation studies using the estimating equation from Fygenson and Ritov (1994). The error term is distributed as $\epsilon \sim N(0, 1)$. For covariates, in scenario 1, we generate random variable $\mathbf{W} = (Z_1, Z_2)^T$ from bivariate normal distribution with

sample size \ p-values	Censored data		Uncensored data	
	Cutoff values		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

Table 3.1: Size of the proposed method

mean 0 and covariance matrix $\begin{pmatrix} 1 & 1 \\ 1 & 25 \end{pmatrix}$. In other words, Z_1 and Z_2 have normal distribution with variance 1 and variance 25, respectively, and they are correlated. True parameter values are $\boldsymbol{\eta}_0 = (0.2, 1)^T$. We both consider censored and uncensored cases. We run 400 simulation runs and 200 resamplings. In calculation, due to computational expense, we only consider time points using $5[j]\%$ quantiles for t. Sample size is $n = 50$ and $n = 100$. Censoring variable C follows logarithm of uniform distribution with minimum 0 and maximum 200. For censored case, censoring rate is approximately 20% on average.

Table 3.1 shows simulation result of size from censored data and uncensored data for the proposed method. Then we compute power of the proposed method and Lin et al. (1996) method. For power comparison, we fit the model using only Z_1. Table 3.2 shows simulation result of power comparison on censored data and uncensored data between the Lin et al. (1996) method and the proposed method.

The proportion of rejections from the proposed method is higher than that from Lin et al. (1996) method. Figure 3.1 and 3.2 show the power corresponding to threshold values of p-value. The plot shows that our proposed method performs better than Lin et al. (1996)'s method. Table 3.1 shows power comparison between new method and Lin et al. (1996) method. Numerical results indicate that the proposed approach has higher power than that of Lin et al. (1996) method. Moreover, the difference of rejection rate between two methods is higher in censored case than one in uncensored case.

Censored data				
p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (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 et al. (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 et al. (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 et al. (1996)	0.0125	0.0175	0.0275	0.0575
Proposed method	0.055	0.1275	0.19	0.245

Table 3.2: Power comparison between new method and Lin et al. (1996)'s method for independent censoring case

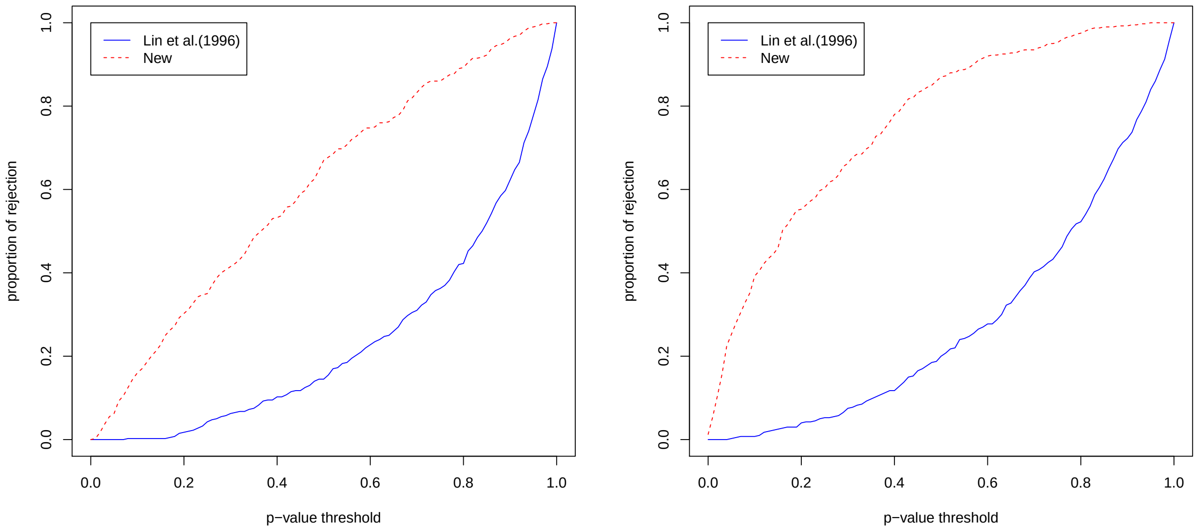

Figure 3.1: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for independent censoring case in the scenario 1

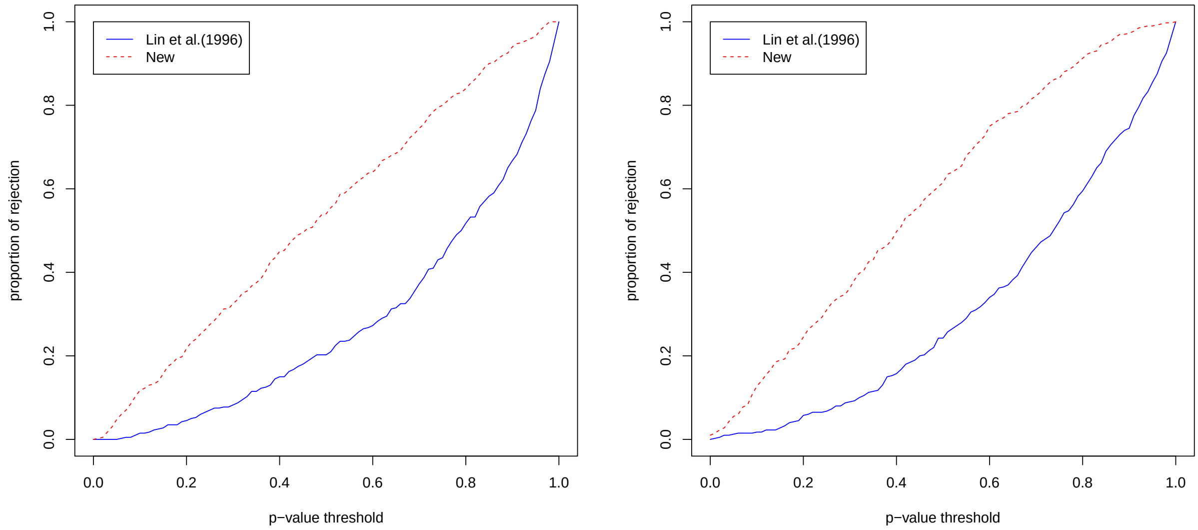

Figure 3.2: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for uncensored data in the scenario 1

In the second scenario, we use variable whose variability is larger than in the first scenario and we omit it in the model fitting. We first generated $(A_1, A_2)^T$ from a bivariate normal distribution with mean 0 and covariance matrix $\begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$.

Next, we define $Z_1 = A_1$ and $Z_2 = jA_2^2 I(b[j-1] < A_2 \leq b[j]), j = 1, \ldots 21$, where $b[j]$ is $5(j-1)\%$ quantile of W_2. $b[0] = -\infty$ and $b[1]$ is minimum of $b[\cdot]$ and $b[21]$ is maximum of $b[\cdot]$. Censoring variable is uniformly distributed with minimum value 0 and maximum value 150. On average, censoring rate is between 7% and 8%. True regression coefficient values are $\boldsymbol{\eta}_0 = (0.2, 1)^T$. We run 400 simulations with sample size $n = 50, 100$ and 200. In each simulation run, 200 resampling runs are performed. As before, we fit the model by using only Z_1 and for comparison, the new testing procedure is compared to that of Lin et al. (1996)

The proportion of rejections from the proposed method is higher than that from Lin et al. (1996) method. Figure 3.1 and 3.2 show the power corresponding to threshold values of p-value. The plot shows that our proposed method performs better than Lin et al. (1996)'s method. Table 3.2 shows power comparison between new method and Lin et al. (1996) method. Numerical results indicate that the proposed approach has higher power than that of Lin et al. (1996) method.

Table 3.2 shows numerical results of comparing two methods and Figure 3.3 and 3.4 describe graphical comparison of rejection rate between the Lin et al. (1996)'s method and the proposed method. As in the scenario 1, the proposed method is higher power than the Lin et al. (1996)'s method. In scenario 1 and scenario 2, as sample size increases, rejection rate of both methods increases and difference of proportion of rejection between Lin et al. (1996)'s method and the proposed method decreases. It implies that when sample size goes infinity, the power by the proposed method approaches 1, which supports Theorem 4.

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

Table 3.3: Power comparison between new method and Lin et al. (1996)'s method for independent censoring case in scenario 2

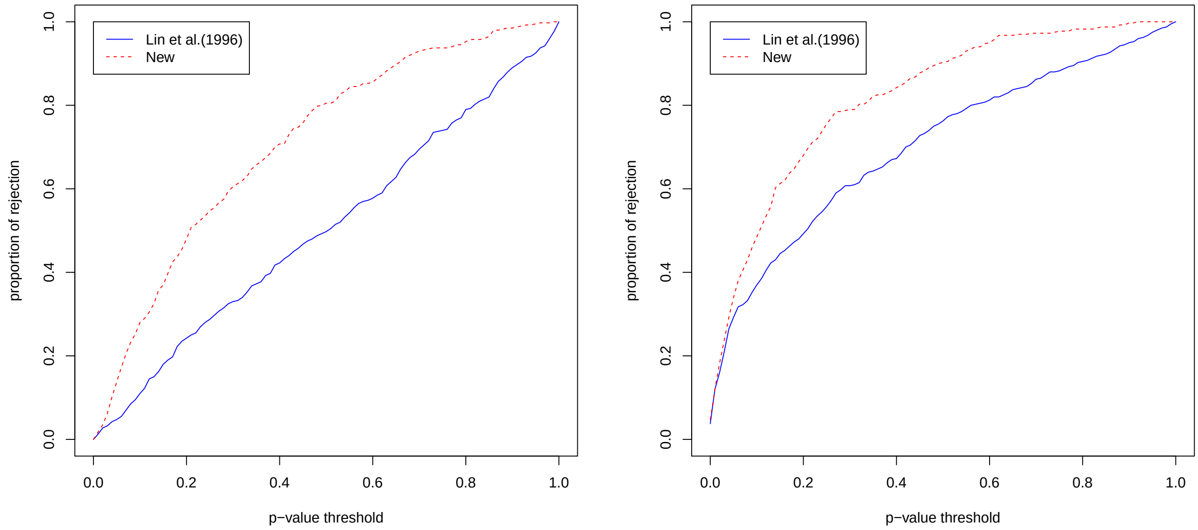

Figure 3.3: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for independent censoring case in scenario 2

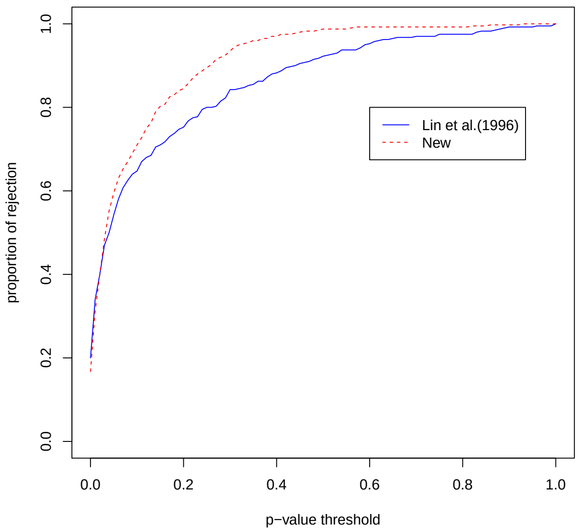

Figure 3.4: Plot of proportion of rejection according to threshold p-values when $n = 200$ for independent censoring case in scenario 2

Next, we applied the proposed method on dependent censoring case. Steps for data generation is shown below:

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 $R_2 = jW_2^2 I(b[j-1] < W_2 \leq b[j]), j = 1, \ldots 21$, 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 $\mathbf{R} = (R_1, R_2)^T$.

Independent censoring time C is uniformly distributed with minimum value 0 and maximum value 100. On average, approximately 12% of dependent censoring is censored by C and 12% of the event of interest is dependently censored by $\tilde{D}$. We fit the misspecified model from Section 3.2, which only employs R_1 and compute the statistical power using Lin et al. (1996) method and the proposed method for the model of the event of interest X. In each simulation run, 200 resampling runs are tried. Table 3.3 shows the results when $n = 50$ based on 400 simulation runs and for $n = 100$ based on 200 simulation runs. Figure 3.4 shows a plot of proportion of rejection when $n = 50$ and $n = 100$. The plots in Figure 3.4 and numerical summaries from Table 3.3 lead the same conclusion as independent censoring case. Our proposed method performs better than that of Lin et al. (1996).

4. Real Data Analysis

We applied our method to data from AIDS Clinical Trial Study 364 (Albrecht et al.

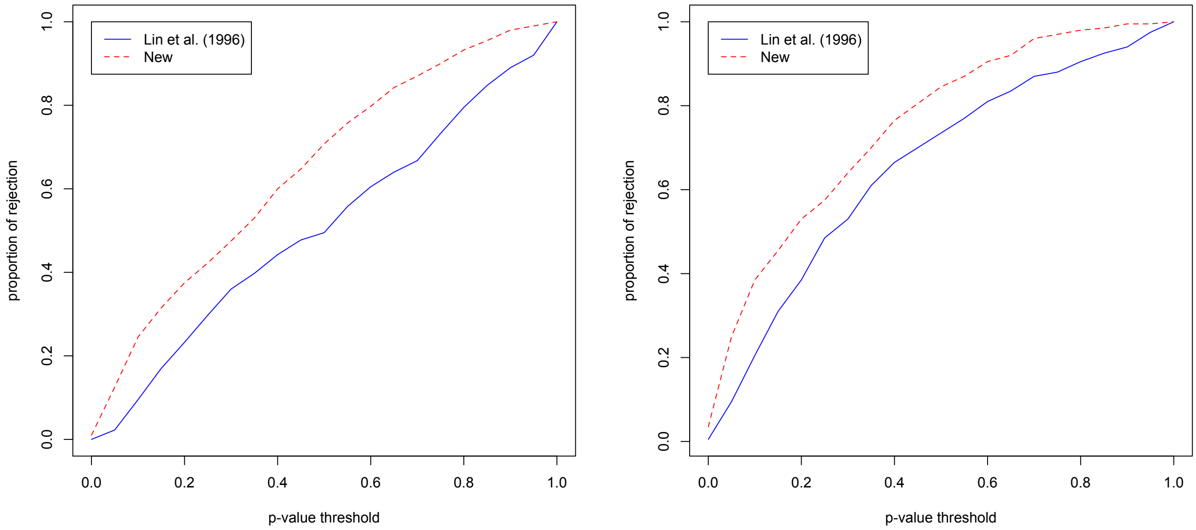

Figure 3.5: Plot of proportion of rejection according to threshold p-values when $n = 50$ (left) and $n = 100$ (right) for model of the event of interest in the presence of dependent censoring

2001), which was previously analyzed by Peng and Fine (2006) and Cho and Ghosh (2015). In this study, 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 to HIV level$\geq$ 2000 at the first time. 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.

p-values / $n = 50$	Cutoff values			
	0.05	0.10	0.15	0.2
Lin et al. (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 et al. (1996)	0.095	0.205	0.31	0.385
Proposed method	0.25	0.385	0.455	0.53

Table 3.4: Power comparison between new method and Lin et al. (1996)'s method for model of the event of interest in the presence of dependent censoring

In this dataset, 3 levels of treatment are considered : nelfinavir (NFV), efavirenz (EFV), and combination of nelfinavir and efavirenz (NFV + EFV). We consider three covariates. Z_1 takes value 1 if treatment assignment of a patient is EFV and 0 otherwise. Z_2 takes value 1 if treatment assignment of a patient is NFV + EFV and 0 otherwise and Z_3 is log(RNA) level. In Cho and Ghosh (2015), the dependent censoring and the event of interest were analyzed using Lin et al. (1996) and Peng and Fine (2006) approaches jointly. For model checking, Cho and Ghosh (2015) used the approach based on Lin et al. (1996) for both Lin et al. (1996) estimator and Peng and Fine (2006) estimator.

Figure 4.1 shows a goodness of fit plot of 20 bootstrapped processes along with the observed process. Observed process is moving around zero and bootstrapped processes suggest that there is no substantial deviation of model fit.

We compare our new approach to that by Cho and Ghosh (2015). The p-value from Cho and Ghosh (2015) is 0.959. The p-value using new approach is 0.51. Although both 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.

5. Discussion

In this paper, we have developed a new goodness of fit approach. Using U-process theory by Nolan and Pollard (1987) and Nolan and Pollard (1988), we adapt the resampling approach from Parzen et al. (1994) and Lin et al. (1996) to derive numerical summaries and graphical tests. Our new approach can be applied to U-statistics of order two estimating equations.

In this paper, our attention has been on checking the overall fit of the model. Other goodness of fit techniques we can consider are checking functional form of covariates and linearity of the model. Lin et al. (1993) proposed method for these scenarios based on the Cox model. However, direct application of Lin et al. (1993)'s approach to the

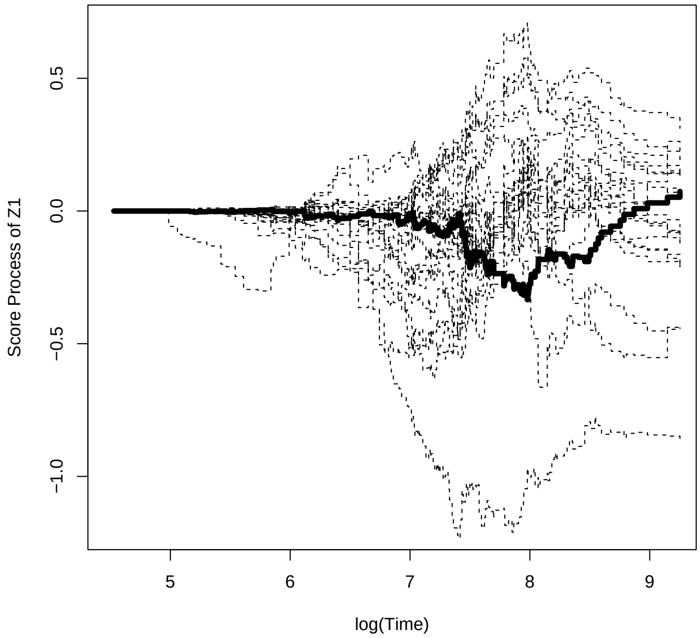

Figure 4.1: Observed process (bold line) and 20 bootstrapped processes (dashed lines) for the first virologic failure

semiparametric AFT model is impossible because the estimating equation is nonsmooth. By mimicking the approach in this paper and Lin et al. (1993), for the procedure of Fygenson and Ritov (1994), one may consider the observed process

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

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

It is also worthwhile to apply idea 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 tool similar to 'robust residuals' can be important. This will be also communicated in separate reports.

Acknowledgment

This research is supported by NIH grant R01 CA129102.

Appendix

Before proving main results, we assume conditions C1-C4 in Honoré and Powell (1994). Moreover, the following assumptions are made.

1. There exists $K > 0$ such that $||\mathbf{Z}|| \leq K$, i.e., $\mathbf{Z}$ is uniformly bounded by constant K. In this case, $||\cdot||$ is Euclidean norm.

2. The error distribution has finite Fisher information and the distribution of $\mathbf{Z}$ given $\Delta = 1$ is not concentrated on a proper hyperplane on $\mathbb{R}^p$.

3. The information bound for estimating $\boldsymbol{\eta}_0$ is finite and invertible.

Before proving main results, the following assumptions are made.

1. The parameter space $\boldsymbol{\Theta}$ is compact and the true parameter $\boldsymbol{\eta}_0$ is the interior point of $\boldsymbol{\Theta}$.

2. Let $||\cdot||$ be Euclidean norm. The functions $\mathbf{h}(\cdot,\cdot,\boldsymbol{\eta})$ and

$$\mathbf{u}(\cdot,\cdot,\boldsymbol{\eta},w) = \sup_{||\boldsymbol{\tau}-\boldsymbol{\eta}||\leq w} ||\mathbf{h}(\cdot,\cdot,\boldsymbol{\tau}) - \mathbf{h}(\cdot,\cdot,\boldsymbol{\eta})||$$

   are measurable functions of $\mathbf{V}_{i_1}$ and $\mathbf{V}_{i_2}$ for $1 \leq i_1 \neq i_2 \leq n$ in some open neighborhood of $\boldsymbol{\Theta}$.

3. Let $\lambda(\boldsymbol{\eta}) = E\{n^{-1/2}\mathbf{U}_n^{FR}(\boldsymbol{\eta})\}$. Then $\lambda(\boldsymbol{\eta}_0) = 0$ and $\lambda(\boldsymbol{\eta})$ is differentiable at $\boldsymbol{\eta}_0$ with nonsingular derivative at $\boldsymbol{\eta}_0$.

4. For $1 \leq i_1 \neq i_2 \leq n$, there exist positive constant a_0, b_0 and c_0 such that $E\{\mathbf{u}(\mathbf{V}_{i_1}, \mathbf{V}_{i_2}, \boldsymbol{\eta}, r)\} \leq a_0 r$ and $E\{\mathbf{u}(\mathbf{V}_{i_1}, \mathbf{V}_{i_2}, \boldsymbol{\eta}, r)^2\} \leq b_0 r$ for all $r \leq c_0$ and all $\boldsymbol{\eta}$ in an open neighborhood of $\boldsymbol{\eta}_0$.

5. There exists $K > 0$ such that $E(||\mathbf{Z}^2||) \leq K$.

6. The error distribution has finite Fisher information and the distribution of $\mathbf{Z}$ given $\Delta = 1$ is not concentrated on a proper hyperplane on $\mathbb{R}^p$.

7. The information bound (Bickel et al. 1993, Chapter 2, p23) for estimating $\boldsymbol{\eta}_0$ is finite and invertible.

Moreover, we introduce several definitions, including Euclidean class of functions, which is a crucial to prove tightness.

Definition (Nolan and Pollard, 1987)**.** *Let S be the set equipped with a pseudometric d. The covering number $N(\tau, d, S)$ is defined as the smallest value of N for which there exist N close balls of diameter τ, and centers in S, whose union covers S.*

Definition (Nolan and Pollard, 1987)**.** *Let $\mathcal{F}$ be class of functions and F be envelope of $\mathcal{F}$. Define Q is a measure on space $\mathcal{X} \otimes \mathcal{X}$. If there exist constants A and B such that*

$$N_1(\tau, Q, \mathcal{F}, F) \leq A\tau^{-B}, \text{ for } 0 < \tau \leq 1,$$

if $0 < Q(F) < \infty$, $\mathcal{F}$ is called Euclidean class and A and B are called Euclidean constants for F.

If $\mathcal{F}$ is Euclidean, for each $p > 1$, if $0 < Q(F^p) < \infty$

$$N_p(\tau, Q, \mathcal{F}, F) \leq A2^{pV}\tau^{-pV}, \text{ for } 0 < \tau \leq 1,$$

Now we define metric in the space. Let $\mathcal{F}$ be class of functions and F be envelope of $\mathcal{F}$. Let the metric $d_{Q,p,F}$ which is defined on $\mathcal{F}$ be

$$d_{Q,p,F}(f, g) = \left[\frac{Q|f-g|^p}{Q(G^p)}\right]^{1/p} \quad f, g \in \mathcal{F}$$

where Q is a measure on space $\mathcal{X} \otimes \mathcal{X}$ which satisfies $0 < Q(F^p) < \infty$. We define $N_p(\tau, Q, \mathcal{F}, F)$ is the covering number $N(\tau, d_{Q,p,F}, \mathcal{F})$. Let $y_1, \ldots y_{2n}$ be a sample from distribution P. Define T_n to be the measure which assigns mass one at each of the $4n(n-1)$ pairs of y_v, y_w in function g_{ij} for $u \in \mathcal{F}$, where

$$g_{ij} = u(y_{2i}, y_{2j}) - u(y_{2i}, y_{2j-1}) - u(y_{2i-1}, y_{2j}) + u(y_{2i-1}, y_{2j-1}),$$

This measure takes important role to construct exponential inequality and convergence theorems in U-processes (Nolan and Pollard, 1987; Nolan and Pollard, 1988). Finally, we define

$$J_n(s, Q, \mathcal{F}, F) = \int_0^s \log N_2(x, Q, \mathcal{F}, F)dx,$$

where $N_2(x, Q, \mathcal{F}, F)$ is the covering number $N(x, d_{Q,2,\mathcal{F}}, \mathcal{F})$ (Nolan and Pollard, 1987). Then

$$\sup_n E\{J_n(s, Q, \mathcal{F}, F)\} = \sup_n E\bigg\{\int_0^s \log N_2(x, Q, \mathcal{F}, F)dx\bigg\}.$$

In the proofs, we only prove results for the estimating equation in Fygenson and Ritov (1994). For dependent censoring case, it is very similar to one in Fygenson and Ritov (1994).

Appendix A. Proof of tightness of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$

Let N_0 be an open neighborhood of $\boldsymbol{\eta}_0$. By Lemma 2 of Honoré and Powell (1994),

$$\sup_{\boldsymbol{\eta}\in N_0} \frac{||\mathbf{U}_n^{FR}(\boldsymbol{\eta}) - \mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) - n^{1/2}\lambda(\boldsymbol{\eta})||}{1 + n^{1/2}||\lambda(\boldsymbol{\eta})||} = o_p(1).$$

Then by Taylor expansion and consistency of $\hat{\boldsymbol{\eta}}$,

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

where $\boldsymbol{\Gamma}_0(t)$ is slope matrix of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. Clearly, $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges in distribution, so it is tight.

The next step is to show tightness of $\mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0)$. Note that $g(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta}_0) = e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0)$. For each t, a class of functions $g_t\{e_i(\boldsymbol{\eta}_0),e_j(\boldsymbol{\eta}_0)\} = e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0) - t$ is a polynomial class by Lemma 18 of Nolan and Pollard (1987) (Note that for each t, $e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0) - t$ is an element of a finite dimensional vector space of real functions). Then by argument of Nolan and Pollard (1987), a class of functions $g_t\{e_i(\boldsymbol{\eta}_0),e_j(\boldsymbol{\eta}_0)\}$ is Euclidean with envelope 1. Let $\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta}_0,t) = \frac{1}{2}(\mathbf{Z}_i-\mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]I\{e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0)\le t\}$. By assumption, $||\mathbf{Z}_i|| \le K$ for all i. Since $[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$ are bounded and $(\mathbf{Z}_i - \mathbf{Z}_j)$ is difference of two random variables, by Lemma 22 in Nolan and Pollard (1987), $\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta}_0,t)$ is also Euclidean with some envelope $G = G(\cdot,\cdot)$.

Let $\mathcal{G}$ be function space for $\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta}_0,t)$. Note that $\mathcal{G}$ is a class of functions in $\mathcal{L}^2(\mathcal{C}\times[0,a])$, where $\mathcal{C}$ is the space of continuous functions and a is a positive constant. The metric for $\mathcal{G}$ given measure T_n is

$$d_{T_n,2,G}(f^*,g^*) = \left[\frac{T_n|f^*-g^*|^2}{T_n(G^2)}\right]^{1/2} \quad f^*,g^*\in\mathcal{G}.$$

Let $P\mathcal{G}$ be a class of functions of $Pg(x,\cdot)$, where $g\in\mathcal{G}$. Clearly, $P\mathcal{G}$ is the class of functions of $E\{h(\mathbf{v},\mathbf{V},\boldsymbol{\eta}_0,t)\}$. Since $E(\mathbf{Z}^2)$ is bounded, $E\{\mathbf{h}(\mathbf{v},\mathbf{V},\boldsymbol{\eta}_0,t)\}$ is also bounded for all $\mathbf{v}$. By Corollary 21 in Nolan and Pollard (1987), $P\mathcal{G}$ is also Euclidean with envelope, $PG = P[G(x,\cdot)]$. Since $T_n(G^2)$ takes value between 0 and ∞, by argument in Nolan and Pollard (1987) about Euclidean class, there exists positive constant A_1 and B_1 such that $N_2(x,T_n,\mathcal{G},G)\le A_1 4^{B_1}x^{-2B_1}$ for $0<x\le 1$. Then

$$\int_0^1 \log N_2(x,T_n,\mathcal{G},G)dx \le \int_0^1 (\log A_1 + B_1\log 4 - 2B_1\log x)dx$$
$$= \log A_1 + B_1\log 4 - 2B_1(x\log x - x)|_0^1 = \log A_1 + B_1\log 4 + 2B_1 < \infty. \qquad \text{(A.1)}$$

Let μ be measure in $\mathcal{X} \otimes \mathcal{X}$. Define

$$J(t, \mu, \mathcal{G}, G) = \int_0^t \log N_2(x, \mu, \mathcal{G}, G) dx$$

Then by (A.1),

$$\sup_n E\{J(1, T_n, \mathcal{G}, G)\}^2 < \infty. \tag{A.2}$$

Let P_n be an empirical measure on sample $\mathbf{V}_1, \ldots, \mathbf{V}_n$. Thus $0 < P_n(PG^2) < \infty$. Since $P\mathcal{G}$ is also Euclidean, by using similar arguments as the previous paragraph,

$$\sup_n E\{J(1, P_n, P\mathcal{G}, PG)\}^2 < \infty. \tag{A.3}$$

Since $P \otimes P(G^2)$ is also positive and finite, we have,

$$J(1, P \otimes P, \mathcal{G}, G) < \infty. \tag{A.4}$$

Thus it is enough to show that for every $\zeta > 0$ and $\delta > 0$, we can find $\nu > 0$ such that

$$\limsup_{n \to \infty} E\{J(\nu, P_n, P\mathcal{G}, PG) > \zeta\} < \delta. \tag{A.5}$$

Since $P\mathcal{G}$ is also Euclidean and $0 < P_n(PG^2) < \infty$ is satisfied, by similar calculation of (A.1), there exists constant w such that

$$J(t, P_n, P\mathcal{G}, PG) = \int_0^t \log N_2(x, P_n, P\mathcal{G}, PG) dx = w.$$

For $\zeta > 0$, by taking t to be the solution of

$$\int_0^t \log N_2(x, P_n, P\mathcal{G}, PG) dx = \zeta.$$

Thus (A.5) holds. Hence by Theorem 5 of Nolan and Pollard (1988), $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta}_0)$ is tight. Hence $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$ is also tight.

Appendix B. Proof of Theorem 1 and Theorem 2

Let $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0) = \frac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$. Define

$$2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0, t) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t)\}$$

where $\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t) = \mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)I\{g(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0) \leq t\}$ and $2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)\}$. By arguments in the Appendix of Lin et al. (1996) and the Appendix of Peng and Fine (2006),

$$\begin{aligned}\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}}) &= \mathbf{U}_n^{FR}(t; \boldsymbol{\eta}_0) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i + o_p(1), \qquad \text{(B.1)}\end{aligned}$$

where

$$\mathbf{H}_i(t) = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0, t) \quad \mathbf{H}_i = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0).$$

Let $\mathbf{U}_n^{FR*}(t; \boldsymbol{\eta}_0) = n^{-1/2}\sum_{i=1}^{n}\{\mathbf{H}_i(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_i\}$. By the tightness of $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$, $\mathbf{U}_n^{FR*}(t; \boldsymbol{\eta}_0)$ converges to a Gaussian process with mean zero and covariance matrix

$$E[\{\mathbf{H}_1(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}\{\mathbf{H}_1(t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}^T]. \qquad \text{(B.2)}$$

Thus Theorem 1 is proved. Next, we will prove Theorem 2. We only prove the Fygenson and Ritov (1994) case. By Appendix of Lin et al. (1996),

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}^*) = \mathbf{U}_n^{FR}(\hat{\boldsymbol{\eta}}) + n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}) + o_p(1) = n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}) + o_p(1) \qquad \text{(B.3)}$$

Combining (2.5) and (B.3) provides

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

$$-\boldsymbol{\Gamma}(t)n^{1/2}(\boldsymbol{\eta}^*-\hat{\boldsymbol{\eta}})+o_p(1)$$

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

$$-\boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})Q_i+o_p(1).$$

We need to show that given the observed data, limiting distribution of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is Gaussian process and that the limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is same as that of $\mathbf{U}_n^{FR*}(t;\boldsymbol{\eta}_0)$. Let

$$g_{ij} = \mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\le t\} - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})$$

Then

$$\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*) = \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}g_{ij}Q_i+o_p(1)$$

By the multivariate central limit theorem and strong consistency of $\hat{\boldsymbol{\eta}}$, $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ converges to Gaussian process and its asymptotic covariance function is $E(\mathbf{L}\mathbf{L}^T)$, where

$$\mathbf{L} = 2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0,t) - \boldsymbol{\Gamma}(t)\boldsymbol{\Gamma}_0^{-1}2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0).$$

The limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ conditional on the observed data is the same as that of $\mathbf{U}_n^{FR*}(t;\boldsymbol{\eta}_0)$. This concludes the proof.

Appendix C. Proof of consistency in misspecified model

Let $e_i^*(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T\boldsymbol{\eta}$. Then the estimating equation is

$$\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)}\sum_{i\neq j}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j^*(\boldsymbol{\eta}) > e_i^*(\boldsymbol{\eta})\} - \Delta_j I\{e_i^*(\boldsymbol{\eta}) > e_j^*(\boldsymbol{\eta})\}] = 0. \tag{C.1}$$

By Theorem 2.1(i) in Fygenson and Ritov (1994), the solution of equation (C.1) exists. Denote this solution by $\hat{\boldsymbol{\eta}}^{mis}$. By the strong law of large numbers,

$$n^{-1/2}\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \lambda^*(\boldsymbol{\eta}) + o(1).$$

Assume that $\lambda^*(\boldsymbol{\eta})$ has a unique solution $\boldsymbol{\eta}^{mis}$. Without loss of generality, it is assumed that $\boldsymbol{\eta}_0 = 0$. If $\boldsymbol{\eta} \to \boldsymbol{\eta}^{mis}$, by Fygenson and Ritov (1994),

$$\lambda^*(\boldsymbol{\eta}) = \frac{1}{2}E\Bigg[(\mathbf{Z}_1 - \mathbf{Z}_2)(\mathbf{Z}_1 - \mathbf{Z}_2)^T \times \int_{-\infty}^{\infty} -\bar{G}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_1)\bar{G}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_2)$$
$$\times f(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})f(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis})dt + (\mathbf{Z}_1 - \mathbf{Z}_2)\int_{-\infty}^{\infty}\bar{G}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_1)\bar{G}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}|\mathbf{Z}_2)$$
$$\times\{\mathbf{Z}_2^T\bar{F}(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})f'(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis}) - \mathbf{Z}_1^T\bar{F}(t - \mathbf{Z}_2^T\boldsymbol{\eta}^{mis})\times$$
$$f'(t - \mathbf{Z}_1^T\boldsymbol{\eta}^{mis})\}\Bigg](\boldsymbol{\eta} - \boldsymbol{\eta}^{mis}) + o(\boldsymbol{\eta} - \boldsymbol{\eta}^{mis}).$$

By Fygenson and Ritov (1994)'s argument, it is linear in a neighborhood of $\boldsymbol{\eta}^{mis}$. Moreover, since $\mathbf{U}_n^{FRmis}(\boldsymbol{\eta})$ and $\lambda^*(\boldsymbol{\eta})$ are monotone with respect to $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}}^{mis}$ is a consistent estimator of $\boldsymbol{\eta}^{mis}$.

Appendix D. Proof of consistency of the test

Suppose that the alternative hypothesis is that $\boldsymbol{\eta}$ in the AFT model depends on time, i.e.,

$$T = \mathbf{Z}^T\boldsymbol{\eta}(s) + \epsilon. \tag{D.1}$$

Let $\hat{\boldsymbol{\eta}}^{mt}$ be estimator of $\boldsymbol{\eta}$ assuming AFT model that has time independent parameters while in the true model parameters actually depend on time. Then by applying similar arguments for the misspecified AFT model, $\hat{\boldsymbol{\eta}}^{mt}$ converges almost surely to constant vector, say $\boldsymbol{\eta}^{mt}$. To show consistency of test, it suffices to show that $n^{-1/2}\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}^{mt})$ converges to nonzero limit (Lin et al. 1993; Arbogast and Lin, 2004) against the alternative hypothesis. Under the alternative hypothesis, by strong law of large number of U-statistics, $n^{-1/2}\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}^{mt})$ converges almost surely to

$$\begin{aligned}&\tfrac{1}{2}E[(\mathbf{Z}_1-\mathbf{Z}_2)\times\\&E[I\{e_1(\boldsymbol{\eta}^{mt})\vee e_2(\boldsymbol{\eta}^{mt})\le t\}(\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\})|\mathbf{Z}_1,\mathbf{Z}_2].\end{aligned} \tag{D.2}$$

Then given $e_1(\boldsymbol{\eta}^{mt})\vee e_2(\boldsymbol{\eta}^{mt})\le t$, the inner expectation of (5.12) is $P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\}]$. Then,

$$\begin{aligned}&P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt})>e_1(\boldsymbol{\eta}^{mt})\}-\Delta_2 I\{e_1(\boldsymbol{\eta}^{mt})>e_2(\boldsymbol{\eta}^{mt})\}]\\&=P\{(T_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\le(T_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\}\\&-P\{(T_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\le(T_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\}\\&=P[\{\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\le[\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})]\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})]\\&-P[\{\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\le\{\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})\}\wedge(C_1-\mathbf{Z}_1^T\boldsymbol{\eta}^{mt})\wedge(C_2-\mathbf{Z}_2^T\boldsymbol{\eta}^{mt})].\end{aligned} \tag{D.3}$$

Since $\boldsymbol{\eta}(\cdot)$ depends on time and covariates, $\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ and $\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ do not have same distribution, thus the probability in expression (D.3) is not 0. For the expression in (D.3) to be 0, the distribution of $\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ and $\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s)-\boldsymbol{\eta}^{mt})$ should be same. Thus for the function in (D.3) to be 0, $\boldsymbol{\eta}(s)=\boldsymbol{\eta}^{mt}$.

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.

Department of Biostatistics and Computational Biology

University of Rochester

601 Elmwood Avenue, P.O. Box 630

Rochester, NY 14642

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

E-mail: DEBASHIS.GHOSH@ucdenver.edu