Inverse probability weighted estimation in a reliability model with missing data

A. Mezaouer[1], J.-F. Dupuy[2], K. Boukhetala[3]

[1] *Faculté des Sciences, Université Saad Dahleb, Blida, Algeria* (a_mezaouer@univ-blida.dz)

[2] *Laboratoire Mathématiques, Image et Applications, Université de La Rochelle, France* (jean-francois.dupuy@univ-lr.fr)

[3] *Université des Sciences et Technologie Houari Boumédienne, Bab-Ezzouar, Algeria* (kboukhetala@usthb.dz)

Abstract

The linear transformation model is a useful regression model for the analysis of reliability data. In this paper, we consider the problems of estimation and testing in this model in a context of missing data. Precisely, we consider the situation where explanatory variables are available for every unit in the experiment sample, while the event times and censoring indicators can only be observed on a subset of the sample. We rely on an inverse probability weighted-type estimation approach for approximating the regression parameter of interest. The theoretical and numerical properties of the resulting estimator are investigated. The proposed approach appears to outperform the classical complete-case analysis.

Key words: Inverse probability weighted estimation, Large-sample properties, Linear transformation model, Missing data, Simulation study

Short title: Estimation in a reliability model with missing data

1 Introduction

Reliability is defined as the "probability that an item can perform its intended function for a specified interval under stated conditions" (see Dovich (1990) for example). More generally, the statistical approach to reliability includes the statistical methods and models designed to the analysis of failure time data. Among them, the regression models have received much attention over the past decades. Regression models in reliability relate the time to failure (the response variable) to a set of explanatory variables or covariates (such as the temperature, pressure, voltage, various stresses,...). Usual objectives of a regression analysis of reliability data are to: i) estimate the effects of the explanatory variables on the risk of failure, ii) examine various hypotheses about these effects.

A number of regression models have been proposed for that purpose (accelerated failure time models, generalized proportional hazards models, additive models, multiplicative models,...). Numerous examples and applications can be found, for example, in Klein and Moeschberger (1997), Meeker and Escobar (1998), Bagdonavičius and Nikulin (2002), and Martinussen and Scheike (2006). In this paper, we consider a general class of semi-parametric regression models for failure time data, which is called the class of linear transformation models. This class includes as particular cases the well-known proportional hazards and proportional odds models. The transformation models for failure time data have given rise to an extremely rich literature so far; see, among others, Chen *et al.* (2002), Slud and Vonta (2004), Ma and Kosorok (2005), Kosorok and Song (2007). See also Fleming and Lin (2000), Bagdonavičius and Nikulin (2002), Martinussen and Scheike (2006), and Dupuy (2008) for recent reviews and numerous references.

In particular, Cheng *et al.* (1995) have proposed simple estimating equations for estimating the regression parameter of interest in this class of models. The estimation proceeds from a sample of n independent observations (X_1, Δ_1, Z_1), ..., (X_n, Δ_n, Z_n) of the vector (X, Δ, Z), where X denotes the (eventually right-censored) failure time, Δ is the censoring indicator, and Z is a vector of explanatory variables. The proposed estimators are consistent and asymptotically normally distributed, with a covariance that can be consistently estimated. Using Cheng *et al.*'s approach, it is straightforward to carry out the statistical analysis of reliability data (estimation and testing).

However, in reliability, cost constraints and unexpected technical issues arising during the course of the study often prevent engineers from observing the full vector (X, Δ, Z) on all the study items. In some situations, it may happen that the explanatory variables Z can be observed on all the items (at the beginning of the experiment, say) while X and Δ can only be observed on a subset of the initial sample. The following two-stage procedure provides a simple way of making statistical inference with such data: i) remove from the analysis the observations i ($i \in \{1, \ldots, n\}$) such that (X_i, Δ_i) is unobserved, ii) use Cheng *et al.*'s estimating equations on the resulting data set (the so-called complete data). As long as the missing-data mechanism is random (MAR, see Tsiatis (2006) for example), we may expect this complete-case analysis (CC therafter) to provide an unbiased estimator for the regression coefficient of interest. However, having discarded part of the sample, we may also expect an increase in the variance of this estimator, resulting in a loss of power of say, the Wald test of nullity of the regression parameter. This, in turn, may affect our interpretation of the true relationship between the covariates and the risk of failure.

Therefore in this paper, we rely on an alternative estimation approach for the

linear transformation model with missing values of the couple (X, Δ). This approach uses the full observed data that is, it also takes account of the Z_i's for those items with unobserved (X_i, Δ_i). We investigate the properties of this method (both theoretically and numerically), and we compare its performances with the CC approach. It appears from our simulation study that the proposed approach outperforms the CC analysis.
The rest of the paper is organized as follows. In Section 2, we briefly describe the general class of linear transformation models, we recall the estimating equation proposed by Cheng *et al.* (1995), and we describe our problem. In Section 3, we describe an inverse probability weighted estimation procedure adapted to this problem, and we investigate the asymptotic properties of the resulting estimator. Technical details are postponed to the Appendix. Section 4 describes a simulation study. A conclusion and a discussion are given in Section 5.

2 The linear transformation model and a problem of missing data

We provide a brief introduction to the linear transformation model for reliability data. We refer to Bagdonavičius and Nikulin (2002), Martinussen and Scheike (2006), and Dupuy (2008) among others, for detailed treatments and reviews.
Let T be some random failure time and Z be a p-dimensional vector of covariates. The class of linear transformation models relates T to Z via the following equation:

$$e(T) = -\beta_0' Z + \epsilon, \tag{1}$$

where e is an unknown strictly increasing transformation function, β_0 is a p-dimensional unknown regression parameter of interest (β_0 reflects the amount of dependence between the covariates and the risk of failure), and ϵ is a random error variable with known distribution function F_ϵ (ϵ is assumed independent of Z). It is convenient to reparameterize the model (1) as $H(T) = e^{-\beta_0' Z} e^{\epsilon}$, where $H(u) = \exp(e(u))$ is a strictly increasing positive function such that $H(0) = 0$ and $\lim_{u\to\infty} H(u) = \infty$. Letting h be the derivative of H, some simple algebra shows that the hazard function $\lambda(t) = \lim_{\delta\downarrow 0} \delta^{-1}\mathbb{P}(t \leq T < t + \delta | T \geq t, Z)$ of T given Z can be expressed as

$$\lambda(t) = \lambda_{e^\epsilon}(e^{\beta_0' Z} H(t)) e^{\beta_0' Z} h(t), \tag{2}$$

where λ_{e^ϵ} is the hazard function of $\exp(\epsilon)$. From this, we deduce the following well-known examples of transformation models (several other examples can be found in Kosorok and Song (2007) and Ma and Kosorok (2005)):

Exemple 1. Let ϵ have the extreme value distribution that is, $F_\epsilon(u) = 1 - \exp(-e^u)$. Then $\exp(\epsilon)$ is distributed as a standard exponential random variable and thus $\lambda_{e^\epsilon}(u) = 1$. It follows that $\lambda(t) = e^{\beta_0' Z} h(t)$, and (2) reduces to the hazard function of a Cox proportional hazards model with baseline hazard rate h and cumulative baseline hazard function H.

Exemple 2. Let ϵ have the standard logistic distribution that is, $F_\epsilon(u) = \exp(u)/(1 + \exp(u))$. Then $\lambda_{e^\epsilon}(u) = (1 + u)^{-1}$ and (2) reduces to $\lambda(t) = \frac{h(t)}{H(t)+e^{-\beta_0' Z}}$, which is known as the proportional odds model.

The usual objectives of the statistical inference in model (1) are to estimate the unknown β_0 and to examine various hypotheses about its value. Assume that we observe n independent copies $(X_1, \Delta_1, Z_1), \ldots, (X_n, \Delta_n, Z_n)$ of the random vector (X, Δ, Z), where $X = T \wedge C$ ($\wedge$ denotes the minimum), T denotes the failure time of interest, C is a random right-censoring time, $\Delta = 1(T \leq C)$, and Z is a p-dimensional vector of explanatory variables. If $s \geq 0$, let $Y(s) = 1(X \geq s)$ denote the at-risk indicator. Several procedures have been developed to estimate β_0. We refer to Martinussen and Scheike (2006) for a detailed exposition. In particular, Cheng *et al.* (1995) have proposed a fairly simple method, which estimates β_0 by the solution $\widehat{\beta}$ of the estimating equation:

$$U(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} \omega(Z_{ij}'\beta) Z_{ij} \left\{ \frac{\Delta_j Y_i(X_j)}{\widehat{G}^2(X_j)} - \xi(Z_{ij}'\beta) \right\} = 0, \tag{3}$$

where $Z_{ij} = Z_i - Z_j$, $\omega(\cdot)$ is a weight function (Cheng *et al.* (1995) proposed to use either $\omega(\cdot) = 1$ or $\omega(\cdot) = \dot{\xi}(\cdot)/\{\xi(\cdot)(1 - \xi(\cdot))\}$, where $\xi(s) = \int_{-\infty}^{+\infty} \{1 - F_\epsilon(t + s)\} dF_\epsilon(t)$ and $\dot{\xi}(\cdot)$ is the derivative of $\xi(\cdot)$), and $\widehat{G}$ is the Kaplan-Meier estimator of the survival function G of the censoring variable C (one motivation for the equation (3) is that $\mathbb{E}[Y_i(T_j)|Z_i, Z_j] = \xi(Z_{ij}'\beta_0)$). Cheng *et al.* (1995) show that $\widehat{\beta}$ is a consistent estimator of β_0 and that the distribution of $n^{\frac{1}{2}}(\widehat{\beta} - \beta_0)$ can be approximated, for large n, by a normal distribution with mean 0 and a covariance matrix that can be consistently estimated. Inference for model (1) can then be made, based on these large-sample results.

As explained in the introduction, we consider the situation where the explanatory variables Z can be observed on all the items under study (at the beginning of the experiment, say) while the couple (X, Δ) can only be observed on a random subset of the initial sample. Let R be the random variable which equals 1 if (X, Δ) is observed and 0 otherwise. Then we observe n independent vectors $(X_i R_i, \Delta_i R_i, R_i, Z_i)$, $i = 1, \ldots, n$. We consider the case where (X_i, Δ_i), if missing, is missing-at-random that is, R_i is independent of (T_i, C_i) given

Z_i (we refer to Tsiatis (2006), for example, for a detailed description of the various missing data mechanisms).
The statistical problem is to estimate β_0 in model (1) from this incomplete data set. An obvious estimation procedure consists in solving the estimating equation (3), based on the subset $\mathcal{S} = \{i \in \{1, \ldots, n\} : R_i = 1\}$ that is, to remove from the analysis the items i such that (X_i, Δ_i) is missing. This however will cause a loss of information since the covariates $\{Z_i : i \notin \mathcal{S}\}$ will be ignored. In the next section, we describe an alternative estimation procedure.

3 The proposed estimation method with missing data

We propose to estimate β_0 by adapting to our setting the general method of inverse weighting of complete cases. Let us denote by $\eta_i = \mathbb{P}(R_i = 1|Z_i)$ the probability of observing a complete case. The basic intuition underlying the method is as follows. Consider an item i, randomly sampled from a population with covariate value Z_i. Then the probability that this item will have a complete observation (X_i, Δ_i, Z_i) is η_i. Therefore, any item with covariate Z_i and complete observation can be thought of as representing $\frac{1}{\eta_i}$ items from the population.
This suggests that we may estimate β_0 by the solution $\widehat{\beta}_n$ of the following estimating equation:

$$
\begin{aligned}
U_n(\beta) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \omega(Z_{ij}^{'}\beta) Z_{ij} \left\{ \frac{\Delta_j Y_i(X_j) R_i R_j}{G^2(X_j)\eta_i\eta_j} - \xi(Z_{ij}^{'}\beta) \right\} \\
&:= \sum_{i=1}^{n}\sum_{j=1}^{n} U_{ij}(\beta) = 0 \qquad (4)
\end{aligned}
$$

To motivate this equation, note that $\mathbb{E}[U_{ij}(\beta_0)] = 0$. To see this, remark that

$$
\begin{aligned}
&\mathbb{E}\left[\omega(Z_{ij}^{'}\beta_0) Z_{ij} \left\{ \frac{\Delta_j Y_i(X_j) R_i R_j}{G^2(X_j)\eta_i\eta_j} \right\} \middle| Z_i, Z_j \right] \\
&= \mathbb{E}\left[\omega(Z_{ij}^{'}\beta_0) Z_{ij}\, \mathbb{E}\left[\frac{1(T_j \le C_j) 1(T_i \wedge C_i \ge T_j) R_i R_j}{G^2(T_j)\eta_i\eta_j} \middle| T_j, Z_i, Z_j \right] \middle| Z_i, Z_j \right] \\
&= \omega(Z_{ij}^{'}\beta_0) Z_{ij} \xi(Z_{ij}^{'}\beta_0).
\end{aligned}
$$

It follows that $\mathbb{E}[U_{ij}(\beta_0)|Z_i, Z_j] = 0$ and therefore $\mathbb{E}[U_{ij}(\beta_0)] = 0$. It is thus natural to estimate β_0 by the solution to the empirical counterpart of this latter equation, which is $\sum_{i=1}^{n}\sum_{j=1}^{n} U_{ij}(\beta) = 0$. However, in practice, the probabilities η_i and $G^2(T_j)$ are unknown, and thus have to be estimated. Following Cheng *et al.* (1995), G can be estimated by the Kaplan-Meier estimator $\widehat{G}$

(see Andersen *et al.* (1993)), and we propose to estimate η_i by fitting a regression model (such as the logistic regression model, for example) to the data $(R_1, Z_1), \ldots, (R_n, Z_n)$, which results in an estimate $\widehat{\eta}_i$. Finally, we propose to estimate β_0 by the solution of the following approximated version of the equation (4):

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\omega(Z'_{ij}\beta)Z_{ij}\left\{\frac{\Delta_j Y_i(X_j)R_iR_j}{\widehat{G}^2(X_j)\widehat{\eta}_i\widehat{\eta}_j}-\xi(Z'_{ij}\beta)\right\}=0. \tag{5}$$

This yields an approximation $\widetilde{\beta}_n$ of the theoretical estimator $\widehat{\beta}_n$. In the following, we establish the asymptotic properties of $\widehat{\beta}_n$. If the censoring survival function G and the η_i are reasonably estimated, we may expect the approximated estimator $\widetilde{\beta}_n$ to inherit the large-sample properties of $\widehat{\beta}_n$. This point will be investigated in the simulation study.
Before turning to the asymptotic properties of $\widehat{\beta}_n$, we introduce some additional notations and regularity conditions. If v is a p-dimensional column vector, we note $v^{\otimes 2} = vv'$. Then, define

$$\begin{aligned}
\widehat{e}_{ij}(\beta) &= \frac{\Delta_j Y_i(X_j)R_iR_j}{\widehat{G}^2(X_j)\widehat{\eta}_i\widehat{\eta}_j} - \xi(Z'_{ij}\beta)\\
\widehat{\Lambda}_n^{-1} &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega(Z'_{ij}\widehat{\beta}_n)Z_{ij}Z'_{ij}\dot{\xi}(Z'_{ij}\widehat{\beta}_n)\\
\widehat{\Gamma}_n &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}\left\{\omega(Z'_{ij}\widehat{\beta}_n)\widehat{e}_{ij}(\widehat{\beta}_n)-\omega(Z'_{ji}\widehat{\beta}_n)\widehat{e}_{ji}(\widehat{\beta}_n)\right\}\\
&\qquad\times\left\{\omega(Z'_{ik}\widehat{\beta}_n)\widehat{e}_{ik}(\widehat{\beta}_n)-\omega(Z'_{ki}\widehat{\beta}_n)\widehat{e}_{ki}(\widehat{\beta}_n)\right\}Z_{ij}Z'_{ik}\\
&\quad-\frac{4}{n^3}\sum_{l=1}^{n}\frac{R_l(1-\Delta_l)}{(\sum_{k=1}^{n}R_kY_k(X_l))^2}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\omega(Z'_{ij}\widehat{\beta}_n)Z_{ij}\frac{\Delta_jY_i(X_j)R_iR_j}{\widehat{G}^2(X_j)\widehat{\eta}_i\widehat{\eta}_j}Y_j(X_l)\right\}^{\otimes 2}.
\end{aligned}$$

We assume that the covariates are bounded that is, there exists a finite constant c such that $|Z_{ij}| \leq c$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, p$. We also assume that the censoring is independent of Z. This condition is reasonable in many settings (note that this was also assumed in Cheng *et al.* (1995)'s paper). Some ways to relax it are however discussed in Section 5. Finally, in order to identify β_0, we assume that $G(\cdot) > 0$ over the experiment time interval and that $\mathbb{P}(R = 1|Z = z) > 0$ for every z. We are now in position to state the following theorem, whose proof is given in the appendix.

Theorem. *Under the regularity conditions stated above, $\widehat{\beta}_n$ is a consistent estimator of β_0, and $n^{\frac{1}{2}}(\widehat{\beta}_n - \beta_0)$ is asymptotically distributed as a Gaussian vector with mean* 0 *and a covariance matrix that can be consistently estimated by $\widehat{\Lambda}_n^{-1}\widehat{\Gamma}_n\widehat{\Lambda}_n'^{-1}$.*

We now investigate, via simulations, the numerical properties of the approximated estimator $\widetilde{\beta}_n$ (bias, variance). We investigate the quality of the Gaussian approximation of its large-sample distribution. We also investigate the power and level of the Wald-type test based on $\widetilde{\beta}_n$ (the Wald test is widely used to test the null hypothesis that a given covariate has no effect on the response variable).

4 A simulation study

The simulation setting is as follows. We simulate right-censored failure times from the linear transformation model $e(T) = -\beta_0 Z + \epsilon$, where e is the logarithm function, ϵ is distributed according to the extreme value distribution, and Z is normally distributed with mean 0 and variance 1. The values log(1.5) and 0 are considered for β_0. An exponential distribution with parameter λ is used to simulate censoring (λ is successively chosen to yield 15% and 30% of censoring). We consider small ($n = 75$) and moderate ($n = 150$) sample sizes. The missingness indicator R is simulated from a Bernoulli random variable with parameter $\mathbb{P}(R = 1|Z) = \exp(\theta_0 + \theta_1 Z)/(1 + \exp(\theta_0 + \theta_1 Z))$, with θ_0 and θ_1 chosen to yield various missingness percentages (15% and 30%). The ideal case (referred to as FD for full-data in the sequel) where all the (X_i, Δ_i) are observed is also considered. In this case, β_0 is estimated using Cheng *et al.* (1995)'s estimating equation (3) (with $\omega(\cdot) = 1$, as in Cheng *et al.* (1995)'s numerical study). The resulting estimates indeed provide a natural benchmark for evaluating the performance of our proposed estimator $\widetilde{\beta}_n$.

For each combination: `sample size` $\times$ `censoring percentage` $\times$ `missingness percentage` $\times$ `parameter value` of the design parameters, we simulate $N = 500$ samples. As mentioned above, the unknown survival function G of the censoring variable is estimated by the classical Kaplan-Meier estimator, and a logistic regression model is fitted to the data (R_i, Z_i) $(i = 1, \ldots, n)$ (for each of the N samples), which allows us to estimate the unknown individual probabilities η_i by $\widehat{\eta}_i = \exp(\widehat{\theta}_{0,n} + \widehat{\theta}_{1,n} Z_i)/(1 + \exp(\widehat{\theta}_{0,n} + \widehat{\theta}_{1,n} Z_i))$.

Based on the N replications, and for each of the combinations mentioned above, we obtain an averaged value $N^{-1} \sum_{j=1}^{N} \widetilde{\beta}_n^{(j)}$ of the estimates of β_0, where $\widetilde{\beta}_n^{(j)}$ is the estimate obtained from the j-th simulated sample (formula (5) with $\omega(\cdot) = 1$). We also obtain the average value mean($\widehat{\text{s.e.}}$) of the asymptotic standard error estimates, and the variance of the $\widetilde{\beta}_n^{(j)}$. When $\beta_0 \neq 0$ (respectively $\beta_0 = 0$), we obtain the empirical power (respectively the empirical size) of the Wald test at the 5% level for testing $\beta_0 = 0$. The results are summarized in the Table 1. For comparison, the method which applies Cheng *et al.* (1995)'s estimating equation (3) to the subset of complete cases (CC) only is also evaluated.

Table 1 about here

From these results, both the proposed estimator $\widetilde{\beta}_n$ (referred to as IPW for Inverse Probability Weighted, in the sequel) and the CC estimator approximate reasonably well the unknown value β_0. Also, as expected, the accuracy of both estimators degradates as the missingness percentage increases. The magnitude of the degradation is however larger for the CC than for the IPW estimator. This is particularly clear when the sample size is small. We note also that the power of the Wald-type tests based on the CC and IPW estimators decreases when the missingness percentage increases. But in all cases, the power of the IPW-based Wald test is higher than the power of the CC-based Wald test. The magnitude of the difference between powers is particularly large when the sample size is small. For moderate sample size, the IPW-based Wald test maintains a high power even when the missingness percentage is high. When $\beta_0 = 0$, both IPW- and CC-based tests globally satisfy the prescribed significance level $\alpha = 0.05$, but again, the averaged standard error estimates reveal better performance for the proposed IPW approach, over the simpler CC method.

For each simulation scenario, we plot the histograms of the N values $\widetilde{\beta}_n^{(j)}$ and of the N CC estimates, along with the associated Q-Q plots. Figures 1 to 8 display the results.

Figures 1 to 8 about here

From these figures, it appears that the normal approximation stated in the theorem is reasonably satisfied by the IPW estimator while the finite sample distribution of the CC estimator appears to be skewed when the sample size is small and/or when the percentage of missingness is high.

These results indicate that a reliable statistical inference in the linear transformation model with missing durations and censoring indicators may be based on the IPW principle, provided that some reasonable conditions (such as $n \geq 75$, and a moderately large missingness percentage) are met.

5 Conclusion and discussion

In this paper, we have considered the problem of estimation in the linear transformation model for failure time data, when the failure times and censoring indicators are missing at random for some units in the experiment sample. We have proposed an estimating approach for this problem, which relies on the principle of inverse probability weighting of complete cases. We have derived the asymptotic properties of the resulting estimator, and we have compared its numerical performances to the complete case estimator. It appears that the proposed IPW estimator outperforms the simpler CC estimator, under

various realistic conditions of sample size, censoring, and missingness.
The proposed procedure is valid under a set of reasonable conditions. The assumption that the censoring is independent of Z may not be satisfied in some experiments. In such a case, one may simply modify the proposed estimating equation (5) by using an appropriate estimator for the conditional survival function $G(t|Z) = \mathbb{P}(C \geq t|Z)$ (a non-parametric estimator may be used if Z is continuous, while a stratified estimator $\widehat{G}(\cdot|Z = j)$ may be used when $Z \in \{1, \ldots, J\}$ is discrete). We have also assumed that the missing data mechanism is random. The case when the missingness is non-ignorable deserves its own attention, and constitutes a topic for future research.
Finally, the IPW approach is useful if we are able to correctly specify the model for R given Z. The robustness of this approach to a misspecification of this model is also of interest, and constitutes a further topic for investigation.

Appendix

We give an outline of the proof of the theorem stated in Section 3. The arguments are similar to those given in Cheng *et al.* (1995), but the technical details are different. Note first that the following holds in probability, as $n \to \infty$:

$$\frac{1}{n^2} U'_n(\beta)(\beta - \beta_0) \to \mathbb{E}\left[\omega(Z'_{12}\beta) Z'_{12}(\beta - \beta_0) \left\{ \frac{\Delta_2 Y_1(X_2) R_1 R_2}{G^2(X_2)\eta_1\eta_2} - \xi(Z'_{12}\beta) \right\}\right].$$

Letting $z_{12} = z_1 - z_2$ and H be the distribution function of Z, this limit can be written as

$$\int_{z_1, z_2} \omega(Z'_{12}\beta) Z'_{12}(\beta - \beta_0) \left\{ \xi(Z'_{12}\beta_0) - \xi(Z'_{12}\beta) \right\} dH(z_1) dH(z_2),$$

which is zero only when $\beta = \beta_0$ and therefore, by classical arguments of Z-estimation (see van der Vaart (1998) for example), $\widehat{\beta}_n$ converges in probability to β_0.

To prove the asymptotic normality of $\widehat{\beta}_n$, first write a Taylor series expansion of $U_n(\widehat{\beta}_n)$ around β_0 as $0 = U_n(\widehat{\beta}_n) = U_n(\beta_0) + (\widehat{\beta}_n - \beta_0)\dot{U}_n(\beta_n^*)$, where β_n^* lies on the line segment between $\widehat{\beta}_n$ and β_0, and $\dot{U}_n$ denotes the derivative of U_n with respect to the components of β. Next, some algebraic manipulations yield $n^{\frac{1}{2}}(\widehat{\beta}_n - \beta_0) =$

$$n^{-\frac{3}{2}} \mathbb{E}\left[\omega(Z'_{12}\beta_0) Z_{12} Z'_{12} \dot{\xi}(Z'_{12}\beta_0)\right] U_n(\beta_0) + n^{-\frac{3}{2}} U_n(\beta_0) \cdot o_p(1),$$

where $o_p(1)$ denotes a sequence which converges to 0 in probability as $n \to \infty$. By Slutsky's theorem, the asymptotic normality of $n^{\frac{1}{2}}(\widehat{\beta}_n - \beta_0)$ will be proved if we can prove that $n^{-\frac{3}{2}} U_n(\beta_0)$ is asymptotically normal. $n^{-\frac{3}{2}} U_n(\beta_0)$ is a U-statistic, and it follows from the central limit theorem for U-statistics that it is asymptotically normally distributed with mean 0 (see Kowalski and Tu (2007) for example). To calculate the asymptotic variance, we first use the martingale integral representation of $(\widehat{G} - G)/G$ given by Gill (1980), and we re-write $n^{-\frac{3}{2}} U_n(\beta_0)$ as

$$n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega(Z'_{ij}\beta_0) Z_{ij} e_{ij}(\beta_0) + 2n^{-\frac{1}{2}} \sum_{k=1}^{n} \int_0^\infty \frac{q(t)}{\pi(t)} dM_k(t) + o_p(1) \tag{6}$$

where $e_{ij}(\beta) = \frac{\Delta_j Y_i(X_j) R_i R_j}{G^2(X_j)\eta_i \eta_j} - \xi(Z'_{ij}\beta)$, Λ_G is the cumulative hazard function of C, $M_k(t) = 1(X_k \le t, \Delta_k = 0) - \int_0^t Y_k(u) d\Lambda_G(u)$ denotes the martingale associated to the censoring counting process $1(X_k \le t, \Delta_k = 0)$, $\pi(t) = \mathbb{P}(X \ge t)$, and

$$q(s) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega(Z'_{ij}\beta_0) Z_{ij} \frac{\Delta_j Y_i(X_j) R_i R_j}{G^2(X_j)\eta_i \eta_j} Y_j(s).$$

The first term in (6) is also a U-statistic, with limiting variance given by

$$\mathbb{E}\Bigg[n^{-3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \ne j} \left\{ \omega(Z'_{ij}\beta_0) e_{ij}(\beta_0) - \omega(Z'_{ji}\beta_0) e_{ji}(\beta_0) \right\} \\ \times \left\{ \omega(Z'_{ik}\beta_0) e_{ik}(\beta_0) - \omega(Z'_{ki}\beta_0) e_{ki}(\beta_0) \right\} Z_{ij} Z'_{ik} \Bigg].$$

The second term in (6) is a martingale with limiting variance $4\int_0^\infty \frac{q(t)q'(t)}{\pi(t)} d\Lambda_G(t)$. Finally, the covariance between the first two terms of (6) can be calculated as

$$\operatorname{cov}\left\{ n^{-\frac{3}{2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \omega(Z'_{ij}\beta_0) Z_{ij} e_{ij}(\beta_0), 2n^{-\frac{1}{2}} \sum_{k=1}^{n} \int_0^\infty \frac{q(t)}{\pi(t)} dM_k(t) \middle| Z_i, Z_j \right\} \\ = \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\int_0^\infty \omega(Z'_{ij}\beta_0) Z_{ij} e_{ij}(\beta_0) \frac{q'(t)}{\pi(t)} \sum_{k=1}^{n} dM_k(t) \middle| Z_i, Z_j \right],$$

which asymptotically, is $-4\int_0^\infty \frac{q(t)q'(t)}{\pi(t)} d\Lambda_G(t)$. Therefore, the asymptotic covariance matrix of $n^{-\frac{3}{2}} U_n(\beta_0)$ is

$$\lim_{n\to\infty} n^{-3}\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}\left\{\omega(Z'_{ij}\beta_0)e_{ij}(\beta_0)-\omega(Z'_{ji}\beta_0)e_{ji}(\beta_0)\right\}\right.$$
$$\left.\times\left\{\omega(Z'_{ik}\beta_0)e_{ik}(\beta_0)-\omega(Z'_{ki}\beta_0)e_{ki}(\beta_0)\right\}Z_{ij}Z'_{ik}-4\int_0^\infty \frac{q(t)q'(t)}{\pi(t)}d\Lambda_G(t)\right].$$

The expression above can be estimated by replacing β_0, G, and η_i by $\widehat{\beta}_n$, $\widehat{G}$, and $\widehat{\eta}_i$ respectively, and by replacing Λ_G by a standard Nelson-Aalen estimator (see Andersen *et al.* (1993)), based on the subset $\mathcal{S}$ (note that under our regularity conditions, this estimator should be consistent for Λ_G). This yields the expression $\widehat{\Gamma}_n$ given in Section 3. Finally, the quantity $\left(\mathbb{E}\left[\omega(Z'_{12}\beta_0)Z_{12}Z'_{12}\dot{\xi}(Z'_{12}\beta_0)\right]\right)^{-1}$ can be consistently estimated by its empirical counterpart $\widehat{\Lambda}_n^{-1}$. This concludes the proof.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., Keiding, N., 1993. Statistical models based on counting processes. New York: Springer.

Bagdonavičius, V., Nikulin, M. S., 2002. Accelerated Life Models. Modeling and Statistical Analysis. Chapman & Hall.

Chen, K., Jin, Z., Ying, Z., 2002. Semiparametric analysis of transformation models with censored data. Biometrika, 89, 659-668.

Cheng, S. C., Wei, L. J., Ying, Z., 1995. Analysis of transformation models with censored data. Biometrika, 82, 835-845.

Dovich, R. A., 1990. Reliability Statistics. Milwaukee: ASQ Quality Press.

Dupuy, J.-F., 2008. Transformation models for failure time data: an overview of some recent developments. Proceedings of the Second International Conference on Accelerated Life Testing in Reliability and Quality Control. 43-47, Bordeaux, France.

Fleming, T. R., Lin, D. Y., 2000. Survival analysis in clinical trials: past developments and future directions. Biometrics, 56, 971-983.

Gill, R. D., 1980. Censoring and Stochastic Integrals. Mathematical Centre Tracts No.124. Amsterdam: Mathematisch Centrum.

Klein, J. P., Moeschberger, M. L., 1997. Survival Analysis: Techniques for Censored and Truncated Data. New York: Springer.

Kosorok, M. R., Song, R., 2007. Inference under right censoring for transformation models with a change-point based on a covariate threshold. Annals of Statistics, 35, 957-989.

Kowalski, J., Tu, X. M., 2007. Modern Applied U-Statistics. Wiley.

Ma, S., Kosorok, M. R., 2005. Penalized log-likelihood estimation for partly linear transformation models with current status data. Annals of Statistics, 33, 2256-2290.

Martinussen, T., Scheike, T. H., 2006. Dynamic Regression Models for Survival Data. New York: Springer.

Meeker, W. Q., Escobar, L. A., 1998. Statistical Methods for Reliability Data. New York: Wiley.

Slud, E. V., Vonta, F., 2004. Consistency of the NPML estimator in the right-censored transformation model. Scandinavian Journal of Statistics, 31, 21-41.

Tsiatis, A. A., 2006. Semiparametric Theory and Missing Data. New York: Springer.

van der Vaart, A. W., 1998. Asymptotic Statistics. Cambridge University Press.

Table 1. Simulation results.

β_0	n	% censoring	% missing	0	15		30	
			method	FD	IPW	CC	IPW	CC
log(1.5)	75	15	mean($\widetilde{\beta}_n$)	0.386	0.393	0.705	0.382	0.825
			mean($\widehat{s.e.}$)	1.193	1.355	2.712	1.481	4.719
			var	0.019	0.027	0.152	0.031	0.259
			power	0.803	0.672	0.582	0.631	0.141
		30	mean($\widetilde{\beta}_n$)	0.363	0.366	0.636	0.354	0.865
			mean($\widehat{s.e.}$)	1.168	1.312	2.372	1.494	5.174
			var	0.021	0.032	0.135	0.041	0.358
			power	0.756	0.689	0.653	0.545	0.116
	150	15	mean($\widetilde{\beta}_n$)	0.382	0.379	0.639	0.383	0.834
			mean($\widehat{s.e.}$)	1.218	1.382	2.411	1.591	4.564
			var	0.011	0.016	0.062	0.026	0.129
			power	0.968	0.901	0.909	0.794	0.473
		30	mean($\widetilde{\beta}_n$)	0.367	0.367	0.634	0.369	0.841
			mean($\widehat{s.e.}$)	1.199	1.348	2.386	1.551	4.633
			var	0.011	0.018	0.079	0.028	0.150
			power	0.956	0.874	0.885	0.764	0.501
0	75	15	mean($\widetilde{\beta}_n$)	-0.003	0.002	0.003	0.010	0.015
			mean($\widehat{s.e.}$)	1.125	1.351	1.697	1.583	2.476
			var	0.019	0.026	0.064	0.031	0.154
			level	0.028	0.038	0.038	0.037	0.015
		30	mean($\widetilde{\beta}_n$)	0.002	0.003	0.002	-0.002	0.004
			mean($\widehat{s.e.}$)	1.137	1.367	1.726	1.636	2.962
			var	0.022	0.030	0.077	0.047	0.293
			level	0.048	0.056	0.047	0.072	0.016
	150	15	mean($\widetilde{\beta}_n$)	-0.002	0.000	0.000	0.000	-0.007
			mean($\widehat{s.e.}$)	1.143	1.369	1.673	1.609	2.644
			var	0.009	0.012	0.030	0.017	0.105
			level	0.032	0.028	0.032	0.052	0.041
		30	mean($\widetilde{\beta}_n$)	-0.002	-0.004	-0.007	-0.008	-0.028
			mean($\widehat{s.e.}$)	1.138	1.382	1.704	1.649	2.737
			var	0.009	0.015	0.038	0.019	0.129
			level	0.024	0.040	0.044	0.052	0.028

Note: n denotes the sample size. FD stands for Full Data, IPW for Inverse Probability Weighted estimator, CC for Complete Case estimator.

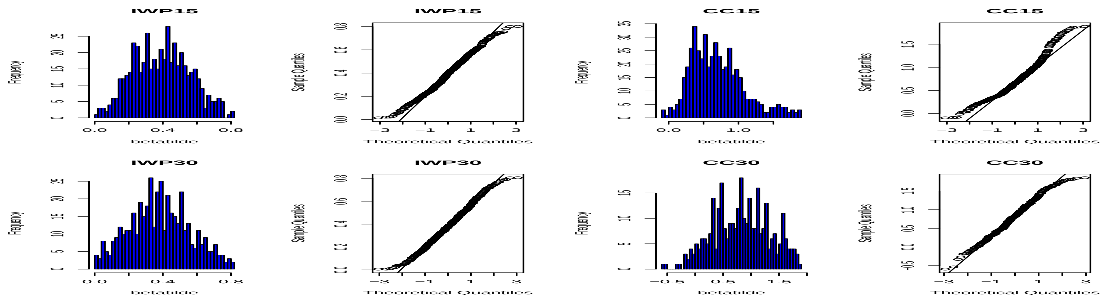

Fig. 1. Histograms and Q-Q plots for $n = 75$, $\beta_0 = \log(1.5)$, and 15% of censoring.

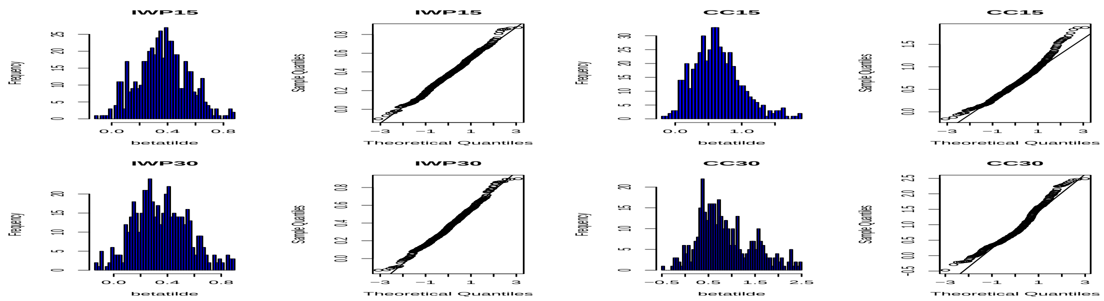

Fig. 2. Histograms and Q-Q plots for $n = 75$, $\beta_0 = \log(1.5)$, and 30% of censoring.

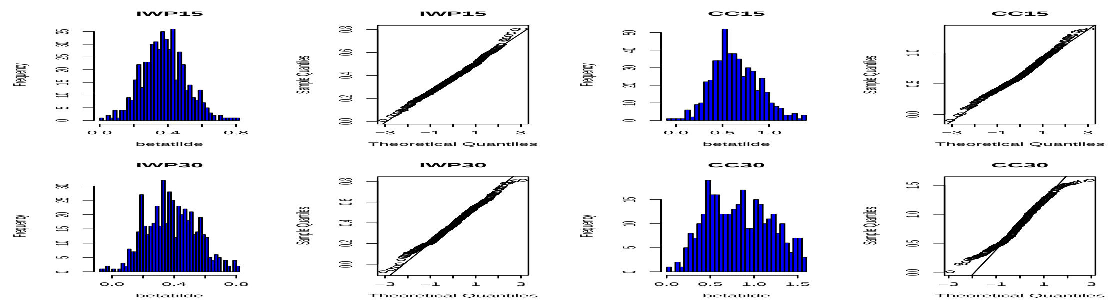

Fig. 3. Histograms and Q-Q plots for $n = 150$, $\beta_0 = \log(1.5)$, and 15% of censoring.

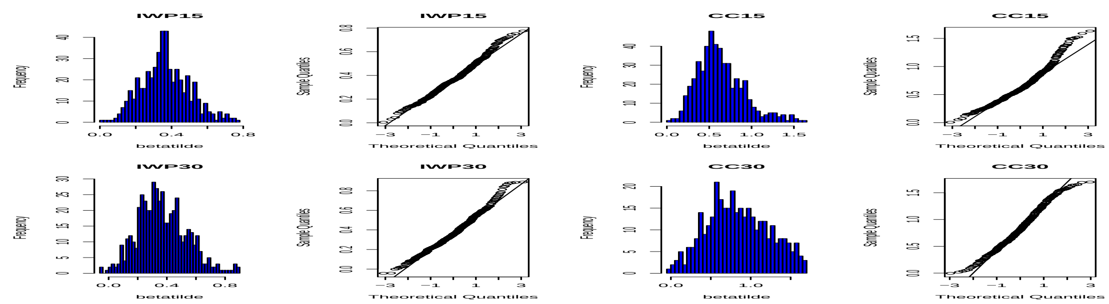

Fig. 4. Histograms and Q-Q plots for $n = 150$, $\beta_0 = \log(1.5)$, and 30% of censoring.

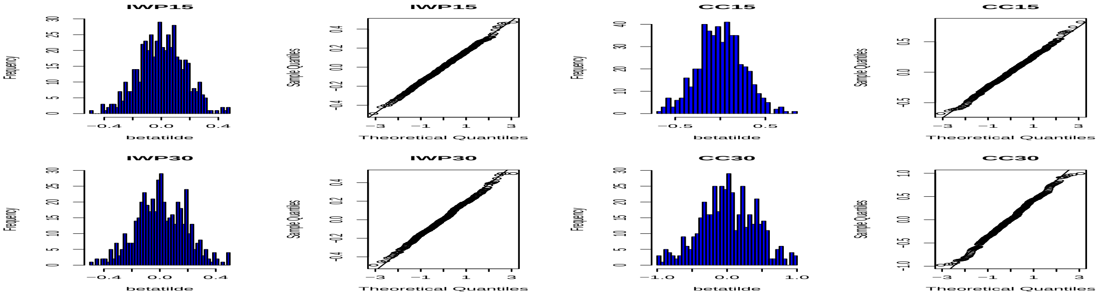

Fig. 5. Histograms and Q-Q plots for $n = 75$, $\beta_0 = 0$, and 15% of censoring.

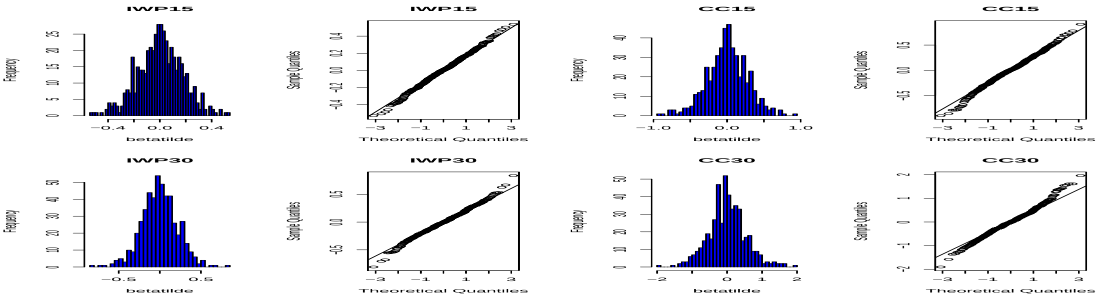

Fig. 6. Histograms and Q-Q plots for $n = 75$, $\beta_0 = 0$, and 30% of censoring.

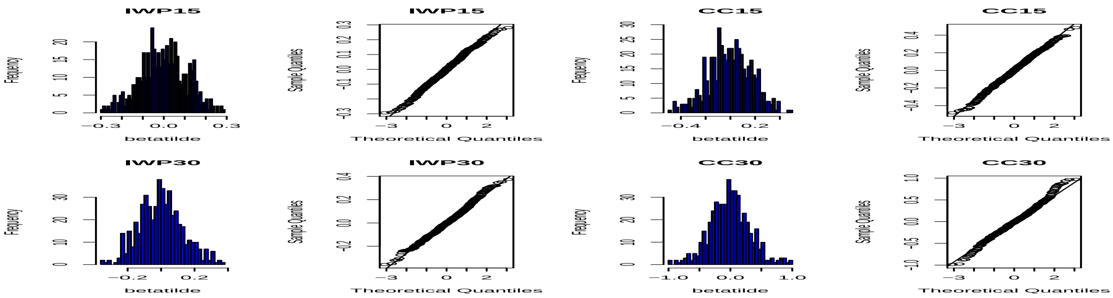

Fig. 7. Histograms and Q-Q plots for $n = 150$, $\beta_0 = 0$, and 15% of censoring.

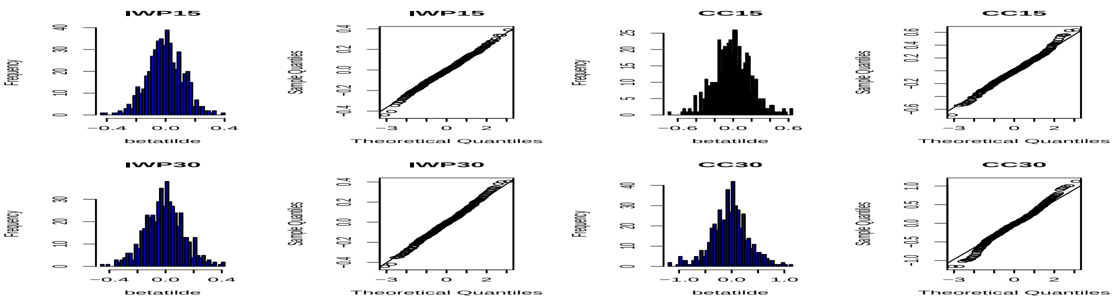

Fig. 8. Histograms and Q-Q plots for $n = 150$, $\beta_0 = 0$, and 30% of censoring.