The effect of model uncertainty on maintenance optimization

Cornel Bunea
Delft University of Technology, Delft
Tim Bedford
Strathclyde University, Glasgow
Delft University of Technology, Delft

Key Words - Censored data, Competing risks, Copula, Identifiability, Preventive maintenance.

Summary & Conclusions - Much operational reliability data available, for example in the nuclear industry, is heavily right censored by preventive maintenance. The standard methods for dealing with right censored data (Total Time on Test statistic, Kaplan-Meier estimator, adjusted rank methods) assume the independent competing risk model for the underlying failure process and the censoring process, even though there are many dependent competing risk models that can also interpret the data. It is not possible to identify the "correct" competing risk model from censored data. A natural question is whether this model uncertainty is of practical importance. In this paper we consider the impact of this model uncertainty on maintenance optimization and show that it can be substantial. We present three competing risk model classes which can be used to model the data, and determine an optimal maintenance policy. Given these models, we consider the error that is made when optimizing costs using the wrong model. It is shown that model uncertainty can be expressed in terms of the dependence between competing risks, which can be quantified by expert judgement. This enables us to reformulate the maintenance optimization problem to take into account model uncertainty.

1 Introduction

Acronyms

PM preventive maintenance
RC replacement cost
RT replacement time

The standard methods, assuming independent censoring, used to treat right censored data are non-conservative, in the sense that other dependent censoring models estimate the underlying failure process more pessimistically (see [Bedford and Meilijson, 1995]). Without making non-testable assumptions (such as independence of the failure and censoring processes), the true distribution function is not identifiable from the data. Hence, in addition to the usual uncertainty caused by sampling fluctuation we have the extra problem of model uncertainty.

In this paper we test the effect of model uncertainty on the problem of optimizing maintenance. We assume that data is available which contains censors from an existing PM program, and use this data to estimate an optimal age replacement PM program.

In Section 3, we take three model classes of competing risk. The independent model is used as the most extreme pessimistic model of existing PM. The another extreme model is used for the most optimistic model of existing PM. The dependent competing risk model is used for the general case and the dependence between competing risks is given by a copula. The minimally informative copula with respect to the uniform distribution and Archimedean copula are studied - the later will be use to approximate the first one, due to numerical difficulties in working with the minimally informative copula for strong dependence between risks. We present a method by which expert judgement may be used to quantify model uncertainty. In Section 4 we recall the theory of optimal age replacement policies and in Section 5 we will present three numerical examples to determine the error that is made when optimizing costs using the wrong model. The last section shows that model uncertainty does lead to substantial uncertainty in the estimation of optimal maintenance intervals and to excessive costs. This paper extends and develops results given in [Bedford and Mesina, 2000], in particular by showing how expert judgement may be used to quantify model uncertainty.

Notation

X lifetime, $X \geq 0$; a r.v.
Y PM time, $Y \geq 0$; a r.v.
$f_X(x)$, $F_X(x)$ [pdf, Cdf] of X
$f_Y(y)$, $F_Y(y)$ [pdf, Cdf] of Y
$S_X(x)$ 1 - $F_X(x)$; the survival function of X
$S_Y(y)$ 1 - $F_Y(y)$; the survival function of Y
$S^*_X(x)$ $Pr\{X > t, X < Y\}$: the *sub-survival function* of X
$S^*_Y(t)$ $Pr\{Y > t, Y < X\}$: the *sub-survival function* of Y
$F^*_X(x)$ $Pr\{X < t, X < Y\}$: the *sub-distribution function* of X
$F^*_Y(t)$ $Pr\{Y < t, Y < X\}$: the *sub-distribution function* of Y
c_1 the cost of critical failure
c_2 the cost of planned replacement
$C(x, y)$ copula of X and Y
$\rho(x, y)$ Spearman's rho
$\tau(x, y)$ Kendall's tau
θ age replacement time
γ expected cost over time θ

Other, standard notation is given in "Information for Readers & Authors" at the rear of each issue.

2 Competing Risk

In the competing risk approach we model the data as a renewal process, that is as a sequence of i.i.d. variables Z_1, Z_2,.... Each observable Z is the minimum of two variables and the indicator of which variable was smaller. The lifetime of the component is X: this is the lifetime that the component would reach if it were not preventively maintained. The PM time of the component is Y: this is the time at which the component would be preventively maintained if it didn't fail first. Clearly,

$$Z \equiv [\min(X, Y), I(X < Y)].$$

(In fact, usually X will be the minimum of several variables giving the time to failure by a particular failure mode: we shall just consider the case of one failure mode.) The observable data will allow us to estimate the *sub-survival functions*,

$$S_X^*(t) \equiv Pr\{X > t, X < Y\} \quad \text{and} \quad S_Y^*(t) \equiv Pr\{Y > t, Y < X\},$$

but not the true survivor functions of X and Y. Hence we are not able to estimate the underlying failure distribution for X without making additional, non-testable, model assumptions. A characterization of those distributions for X that are possible for given subsurvival functions was made in [Bedford and Meilijson, 1997].

By specifying a copula for the underlying joint distribution of X and Y one *can* identify the marginals (and the full joint distribution) [Zheng and Klein, 1995]. However the choice of such a copula is difficult to make: Bedford [Bedford, 1998] suggests doing this by specifying the Spearman's rank correlation between X and Y, and then using the copula with minimum information with respect to the independent copula (that is, the "most independent" copula with the given Spearman rank correlation).

3 Three Models for Competing Risk

In this section we present three competing risk models in which the marginal distribution functions are identifiable. Two of them are the "extreme" cases - independent model and high correlated censoring model and the third one assumes that the dependence between competing risks is given by a copula.

3.1 Model 1: Independence

If F_X has a density function $f_X(t)$, then the failure rate $r_X(t)$ of X is

$$r_X(t) = f_X(t)/S_X(t) = -(dS_X(t)/dt)/S_X(t).$$

Since

$$d[\log(S_X)] = dS_X/S_X,$$

we have

$$S_X(t) = \exp\{-\int_0^t r_X(s)ds\}.$$

But from competing risk data we observe a different rate of failure for X. The observed failure rate for X is defined as

$$obr_X(t) \equiv \lim_{\delta \to 0} Pr\{X > t, X < Y, X \in (t, t+\delta) | Z > t\}/\delta = -\frac{dS_X^*(t)/dt}{S_X^*(t) + S_Y^*(t)}.$$

For the most frequently made assumption in the literature, that of probabilistic independence between X and Y, we have

$$S_X^*(t) + S_Y^*(t) = Pr\{X > t, Y > t\} = Pr\{X > t\}Pr\{Y > t\} = S_X(t)S_Y(t).$$

Using the above results Cooke [Cooke, 1996] showed that if the competing risks X and Y are independent with differentiable survival functions, then the failure rate is equal with the observed failure rate

$$r_X(s) = obr_X(s).$$

Now, the underlying marginal distributions of X and Y can be identified in terms of the observable subsurvivor functions,

$$S_X(t) = \exp\left(\int_0^t \frac{dS_X^*(s)}{S_X^*(s) + S_Y^*(s)}\right). \qquad (1)$$

3.2 Model 2: Highly Correlated Censoring

Clearly, independent censoring does not capture the notion that preventive maintenance is carried out when the equipment has given some sign of future failure.

The most extreme case is described as follows: Preventive maintenance aims to prevent the failure of the component at a time immediately before failure. If that aim is not achieved then the PM action is applied immediately after failure. The PM is unsuccessful with probability p and successful with probability $1-p$, independently of the time at which the failure occurs. We model this by taking $Y = X + \delta\varepsilon$, where $\varepsilon > 0$ is very small but depends on X, and $\delta = \{1, -1\}$ with probability p respectively $1-p$ is independent of X. For very small ε Model 2 gives the following relationships:

$$S_X^*(t) \equiv Pr\{X > t, X < Y\} = Pr\{X > t, \delta = 1\} = Pr\{\delta = 1\}Pr\{X > t\} = pS_X(t),$$

and

$$S_Y^*(t) \equiv Pr\{Y > t, Y < X\} = Pr\{\delta = -1\}Pr\{Y > t\} \approx (1-p)Pr\{X > t\} = (1-p)S_X(t).$$

Hence the *normalized* subsurvivor functions (normalized so that they take that value 1 at $t = 0$) are approximately equal,

$$\frac{S_X^*(t)}{p} \approx \frac{S_Y^*(t)}{1-p}, \qquad (2)$$

and both are equal to $S_X(t)$. Now, this condition can be checked from the data. If it does not hold then Model 2 is not correct. An example is shown in Figure 1 where we have taken

$$S_X(t) = \exp(-t^{0.5}).$$

We took a sample 1000 times for the model above with $p = 1/3$ and then we plot the empirical functions $\frac{\widehat{S}_X^*(t)}{p}$, $\frac{\widehat{S}_Y^*(t)}{1-p}$ and the theoretical $S_X(t)$.

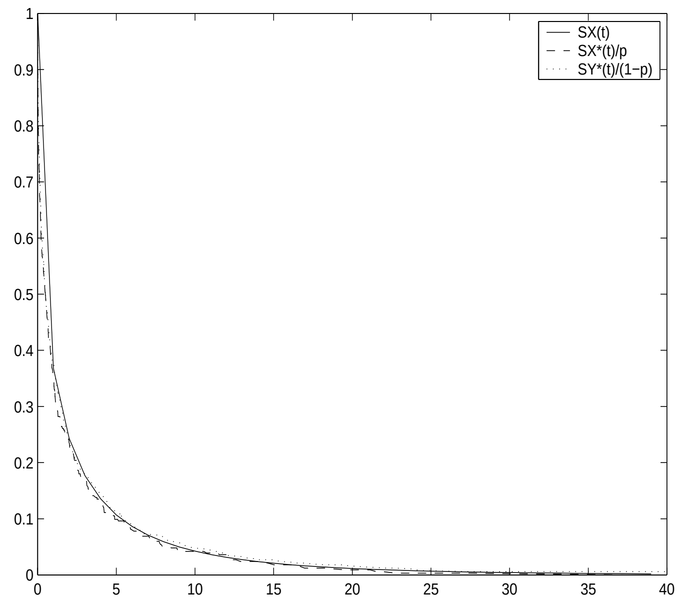

Figure 1. Highly correlated censoring

If Equation 2 does hold then the model might be correct, but the independent model might also hold with the same observable data. Assuming Model 1 (independence) when Model 2 holds would lead to an incorrect assessment of the marginals. The following proposition is obtained by using Equation 1 [Bedford and Mesina, 2000]:

Proposition 1 *Suppose X and Y have a joint distribution described by Model 2. Let $\tilde{X}$ and $\tilde{Y}$ be independent with*

$$S_X^*(t) = S_{\tilde{X}}^*(t) \quad \text{and} \quad S_Y^*(t) = S_{\tilde{Y}}^*(t).$$

Then

$$S_{\tilde{X}}(t) = [S_X(t)]^p \quad \text{and} \quad S_{\tilde{Y}}(t) = [S_X(t)]^{1-p}.$$

Model 2 is a special case of the random signs model of Cooke [Cooke, 1993]. This model can be used when the subsurvivor functions satisfy the inequality

$$\frac{S_X^*(t)}{p} \geq \frac{S_Y^*(t)}{1-p} \quad t \geq 0. \tag{3}$$

The random signs model says that $Y = X - \xi$, where ξ is a random variable, $\xi \leq X$, $Pr\{\xi = 0\} = 0$, whose sign is independent of X. The failure is observed with probability $Pr\{X < Y\} = Pr\{\xi > 0\} = p$.

3.3 Model 3: Dependent Competing risks

In this model we assume that the dependence structure between X and Y is given by a copula. As defined by [Schweizer and Wolff, 1981] the copula of two random variables X and Y is the distribution C on the unit square $[0,1]^2$ of the pair $(F_X(X), F_Y(Y))$ (recall that for a continuous random variable X with pdf F_X, the random variable $F_X(x)$ is always uniformly distributed on $[0,1]$). The functional form of $C : [0,1]^2 \to \mathbb{R}$ is

$$C(u,v) \equiv H(F_X^{-1}(u), F_Y^{-1}(v)),$$

where H is the joint distribution function of (X,Y) and F_X^{-1} and F_Y^{-1} are the right-continuous inverses of F_X and F_Y. Under independence of X and Y the copula is $C(u,v) = uv \equiv \Pi$, and any copula must fall between $M(u,v) \equiv min(u,v)$ and $W(u,v) \equiv max(u+v-1,0)$, the copulas of the upper and lower Fréchet bounds [Nelsen, 1995]. As we saw in the first model, under the assumption of independence of X and Y, the marginal distribution functions of X and Y are uniquely determined by the sub-survival functions of X and Y. Zheng and Klein [Zheng and Klein, 1995] showed the more general result that, if the copula of (X,Y) is known, then the marginal distributions functions of X and Y are uniquely determined by the competing risk data. This result is captured in the following theorem:

Theorem 1 *Suppose the marginal distribution functions of (X,Y) are continuous and strictly increasing in $(0,\infty)$. Suppose the copula C is known and the corresponding probability measure for any open set of the unit square is positive. Then F_X and F_Y, the marginal distribution functions of X and Y, are uniquely determined by the subdistribution functions.*

In the Appendix we show briefly why the marginals are identifiable when the densities and subdensities exist.

We now discuss the problem of choosing a copula. There are many measures of association for the pair (X,Y), which are symmetric in X and Y. The best known measures of association are Kendall's tau and Spearman's rho (we will use the more modern term "measure of association" instead of the term "correlation coefficient" for a measure of dependence between random variables).

Kendall's tau for a vector (X,Y) of continuous random variables with joint distribution function H is defined as follows: Let (X_1,Y_1) and (X_2,Y_2) be i.i.d. random vectors, each with joint distribution H, then Kendall's tau is defined as the probability of concordance minus the probability of discordance:

$$\tau(X,Y) \equiv Pr\{(X_1-X_2)(Y_1-Y_2)>0\} - Pr\{(X_1-X_2)(Y_1-Y_2)<0\}$$

or

$$\tau(X,Y) \equiv Pr\{sgn(X_1-X_2) = sgn(Y_1-Y_2)\} - Pr\{sgn(X_1-X_2) \neq sgn(Y_1-Y_2)\}.$$

The other measure of association (Spearman's rho) is defined as follows: Let X and Y be continuous random variables then the Spearman's rho is defined as the product moment correlation of $F_X(X)$ and $F_Y(Y)$:

$$\rho_r(x,y) \equiv \rho(F_X(X),F_Y(Y)) = \frac{Cov\{F_X(X),F_Y(Y)\}}{\sqrt{Var\{F_X(X)\}Var\{F_Y(Y)\}}}.$$

Simple formulae relating the measures of association to copula density are given in the Appendix.

Since the measure of association is to be treated as a primary parameter, it is necessary to choose a family of copulae which are as "smooth" as possible and which model all possible measures of association in a simple way. Bedford [1997] proposed using the unique copula with the given Spearman's rho that has minimum information with respect to the independent distribution, and also he gave a method to calculate numerical this copula. Now, due to the difficulty of the interpretation of Spearman's rho by a non-specialist and due to the difficulty of quantifying it, we will use as a primary parameter Kendal's tau. Kendall's tau has the advantage of a definition which can be explained to a non-specialist, but the value can not be estimated using only the competing risk data , because of "identifiability problem". Thus we need to use some prior knowledge or subjective information to obtain information about the value of tau. To model the uncertainty over tau we will use expert judgement. This will be discuss later in this paper, but for now it remains to clarify the way that we obtain the copula.

Work of Zheng and Klein [Zheng and Klein, 1995] suggests that the important factor for an estimate of the marginal survival function is a reasonable guess at the strength of the association between competing risks and not the functional form of the copula. For this reason we will choose a class of copula with which it is easy to work from the mathematical point of view. A such class is Archimedean copula. First, we recall some definitions about the Archimedean copula and some properties of Kendall's tau for a certain Archimedean family of copula.

Let X and Y be continuous random variables with joint distribution H and marginal distribution F_X and F_Y. When X and Y are independent, we have $H(x,y) = F_X(x)F_Y(y)$, and this is the only case when the joint distribution is write into a product of F_X and F_Y. But, there are some families of distributions in which we have $\lambda(H(x,y)) = \lambda(F_X(x))\lambda(F_Y(y))$, see [Nelsen (1999)]. Using the function $\varphi(t) = -\log\lambda(t)$ (λ must be positive on the interval (0,1)), we can also write H as a sum of the marginals F_X and F_Y, $\varphi(H(x,y)) = \varphi(F_X(x)) + \varphi(F_Y(y))$, or in terms of copula $\varphi(C(u,v)) = \varphi(u) + \varphi(v)$. Copulas of this form are called *Archimedean copulas*. The function φ is called an *additive generator* of the copula. If $\varphi(0) = \infty$, φ is a strict generator and $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ is a strict Archimedean copula. For our goal we choose an one-parameter family of copulae which has a strict generator. The Gumbel family is defined as follows:

$$C_\alpha(u,v) \equiv \exp(-[(-\log u)^\alpha + (-\log v)^\alpha]^{1/\alpha}) \quad \text{for} \quad \alpha \in [1,\infty).$$

The generator is the function $\varphi_\alpha(t) = (-\log t)^\alpha$.

As shown in the Appendix, we can directly write α as a function of Kendall's tau,

$$\alpha_\tau = 1/(1-\tau).$$

It remains now to quantify the uncertainty in Kendall's tau using expert opinion. Experts can not be directly asked to quantify their uncertainty over tau, instead they are asked to give uncertainties over physically realizable quantities [Bedford and Cooke, 2001]. Consider two sockets with failure times X_1 and X_2 and the PM times Y_1 and Y_2. The expert can be asked for the probability that an attempt to preventively maintain socket one would occur before the PM for socket two, given that the failure of socket one occurs before the failure of socket two. Let this probability be q. By symmetry we have the same probability for the occurrence of the PM for socket two before the PM for socket one if the failure time of socket one is greater than the failure time of socket two. Also we get that the probability of occurrence of the PM for socket two before the PM for socket one if the failure time of socket one is smaller than the failure time of socket two is equal to $1-q$.

If the experts can give a distribution over $q \equiv Pr\{Y_1 > Y_2 | X_1 > X_2\}$, then we can this translate to a distribution over Kendall's tau. Indeed, after a little simple algebra we have:

$$Pr\{(X_1 - X_2)(Y_1 - Y_2) > 0\} = Pr\{(X_1 > X_2) \cap (Y_1 > Y_2)\} + Pr\{(X_1 < X_2) \cap (Y_1 < Y_2)\} =$$

$$= Pr\{(X_1 > X_2)\} Pr\{(Y_1 > Y_2 | X_1 > X_2\} + Pr\{(X_1 < X_2)\} Pr\{(Y_1 < Y_2 | X_1 < X_2\} = q.$$

Similarly we find:

$$Pr\{(X_1 - X_2)(Y_1 - Y_2) < 0\} = 1 - q$$

and so

$$\tau = 2q - 1.$$

Note that q can be considered an observable quantity because q is the approximate average rate for which $\{Y_1^{(n)} > Y_2^{(n)} | X_1^{(n)} > X_2^{(n)}\}$ holds when a large sample of pairs $(X_1^{(n)}, Y_1^{(n)})$, $(X_2^{(n)}, Y_2^{(n)})$ is observed.

Now for each replacement time θ and measure of association τ we can calculate the long term specific cost and furthermore we can optimize this replacement cost finding the minimal one. This is discussed in the next section.

4 Maintenance Optimization

We consider the effect of uncertainty about the underlying lifetime distribution on the selection of the maintenance policy. To keep things simple we just consider the *age replacement policies*. Recall that an age replacement policy is one for which replacement occurs at failure or at age θ, whichever occurs first. Unless otherwise specified, θ is taken to be a constant.

In the finite time span replacement model we will try to minimize expected cost $C(\theta)$ experienced during $[0, \theta]$, where cost may be computed in money units, time, or some appropriate combination. For a infinite time span, an appropriate objective function is expected cost per unit of time, expressed as

$$\gamma(\theta) \equiv \lim_{\theta \to \infty} \frac{C(\theta)}{\theta}.$$

Letting $N_1(\theta)$ denote the number of failures during $[0, \theta]$ and $N_2(\theta)$ denote the number of planned preventive maintenance during $[0, \theta]$, we may express the expected cost during $[0, \theta]$ as

$$C(\theta) \equiv c_1 E\{N_1(\theta)\} + c_2 E\{N_2(\theta)\},$$

where c_1 is the cost of critical failure and c_2 is the cost for planned replacement. We only consider non-random age replacement in seeking the policy minimizing the specific cost $\gamma(\theta)$ for an infinite time span.

Starting from the definition of the specific cost

$$\gamma(\theta) \equiv \lim_{\theta \to \infty} [c_1 \frac{E\{N_1(\theta)\}}{\theta} + c_2 \frac{E\{N_2(\theta)\}}{\theta}]$$

[Barlow and Proschan, 1965] showed that

$$\gamma(\theta) \equiv \frac{c_1 F(\theta) + c_2 S(\theta)}{\int_0^\theta S(t)\,dt},$$

where F and S are the lifetime distribution function respectively the lifetime survival function.

Then $\gamma(0) = \infty$ and $\gamma(\infty) = c_1 / \int_0^\theta S(t)\,dt$. Differentiating γ to find the optimum, $\frac{d\gamma(\theta)}{d\theta} = 0$, we obtain the equation

$$r(\theta) \int_0^\theta S(t)\,dt - F(\theta) = \frac{c_2}{c_1 - c_2}.$$

When $F_X(x)$ has an increasing failure rate, the optimal replacement time θ_0 is the unique solution of the above equation. For a r.v. with constant failure rate or decreasing failure rate the specific cost has not an optimum ($sign(\frac{d\gamma(\theta)}{d\theta})$ is constant), thus this type of maintenance policy is not appropriate for a such r.v.

When we have as primary parameter Kendall's tau and the information over τ is given by a distribution function $F_\tau(\tau)$ with density $f_\tau(\tau)$, the specific cost is dependent on τ and θ:

$$\gamma(\tau, \theta) \equiv \frac{c_1 F(\tau, \theta) + c_2 S(\tau, \theta)}{\int_0^\theta S(\tau, t)\,dt}.$$

So the long term specific cost given θ is

$$\gamma(\theta) = \int_0^1 C(\tau, \theta) f_\tau(\tau) d\tau$$

and the optimal replacement time θ_0 is obtaining minimizing $\gamma(\theta)$.

5 Numerical Examples

We now give the results of three sets of numerical experiments to show the effect of using Model 1 when Model 2 actually holds, to show the dependence of replacement cost with the measure of association (Kendall's tau) and finally to find the optimal replacement time of the average specific cost.

For the first part of numerical computations, we consider two underlying distributions for X. The first, Distribution 1, is that X has failure rate $r_X(t) = t^{3/2}$. The second, Distribution 2, assumes a failure rate of $r_X(t) = t^2$, while Distribution 3 assumes a failure rate $r_X(t) = t^3$. Both failure rates are continuous and increasing, and correspond to Weibull distributions.

Since the costs of critical failure can be much higher than those of planned maintenance (because of other consequences to the system beyond the need simply to replace the failed unit), we assume that c_1 is much larger than c_2. Since actual plant data shows a considerable number of preventive maintenance actions we assume that p is small. Specifically we take $c_1/c_2 = 10$ and $c_1/c_2 = 20$, and also $p = 0.3$ and $p = 0.1$, thus giving us 4 different cases on which the two models are compared. Both replacement times (RT)

c_1/c_2	10				20			
p	0.3		0.1		0.3		0.1	
	RT	RC	RT	RC	RT	RC	RT	RC
Dist 1, Model 1	0.8280	10.9217	1.2849	9.8756	0.6127	18.8850	0.9508	21.6988
Dist 1, Model 2	0.3860	7.5921	0.3860	7.5921	0.2574	9.9433	0.2574	9.9433
Dist 2, Model 1	0.8687	4.9674	1.5047	6.5283	0.5950	7.2682	1.0305	10.1017
Dist 2, Model 2	0.4758	4.2823	0.4758	4.2823	0.3259	6.1916	0.3259	6.1916
Dist 3, Model 1	0.3393	3.3011	0.3353	3.3280	0.2322	4.6522	0.2303	4.6821
Dist 3, Model 2	0.3556	3.2006	0.3556	3.2006	0.2393	4.5459	0.2393	4.5459

Table 1: Optimal maintenance times and costs

and replacement costs (RC) are given in Table 1. The replacement times are the optimal replacement times calculated under the assumption that the model under consideration is correct. For Model 2 the replacement costs are equal to the optimal replacement costs. For Model 1 they are equal to the replacement costs of Model 2 (which is actually the correct model), evaluated with the optimal replacement time calculated for Model 1. Hence the costs given for Model 1 are always higher than those of the true Model 2. Table 2 gives the ratio of the two model outcomes (Model 1 divided by Model 2) for the time and costs of each of the distributions.

For the second part, we consider three sub-survival functions for X which for the extreme cases (independence and high correlation) take the same failure rates for X as in the first part, and for every sub-survival function of X, we take other three sub-survival functions Y in such a way that Inequality 3 is satisfied (for Weibull distributions with the same shape parameter of S^*_X and S^*_Y, the scale parameter of S^*_Y, a_Y, must be greater then the scale parameter of S^*_X, a_X). Specifically we take $a_X/a_Y = 1/2$, $a_X/a_Y = 1/4$, $a_X/a_Y = 1/8$ and for p and c_1/c_2 we take the same values as in the first numerical example.

Figure 2, 3 and 4 show the way in which the RC (normalized by RC for the independent case) depends on Kendall's tau.

To obtain a distribution for Kendall's tau we ask an expert to give quantiles for the probability q defined in the previous section. If the expert gives 5% and 95% quantiles then we can fit a beta distribution. Specifically if $Pr\{q \leq 0.7\} = 0.05$ and $Pr\{q \leq 0.95\} = 0.95$, then the 5% and 95% quantiles for τ are $Pr\{\tau \leq 0.4\} = 0.05$ and $P\{\tau \leq 0.9\} = 0.95$. Taking the beta distribution as appropriate for τ, we easily obtain the parameters of this distribution a and b given the above quantiles by Newton's method as: $a = 5.6705$ and $b = 2.7322$.

The specific costs for different values of Kendall's tau are shown in Figure 5 and the average specific cost with optimal replacement time are shown in Figure 6.

c_1/c_2	10				20			
p	0.3		0.1		0.3		0.1	
	RT	RC	RT	RC	RT	RC	RT	RC
Dist 1	2.15	1.44	3.33	1.3	2.38	1.9	3.69	2.18
Dist 2	1.83	1.16	3.16	1.52	1.83	1.17	3.16	1.63
Dist 3	0.95	1.03	0.94	1.04	0.97	1.02	0.96	1.03

Table 2: Ratio's of maintenance times and costs

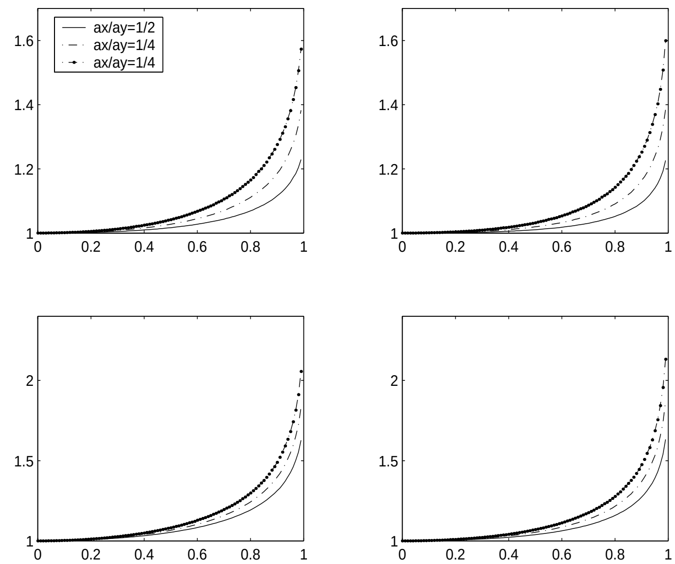

Figure 2. Dependence between RC and the measure of association for the pairs of the sub-survivals for X and Y given by the first sub-survival for X and the other three for Y ; a) $p = 0.3$ and $c_2/c_1 = 0.1$; b) $p = 0.3$ and $c_2/c_1 = 0.05$; c) $p = 0.1$ and $c_2/c_1 = 0.1$; d) $p = 0.1$ and $c_2/c_1 = 0.05$;

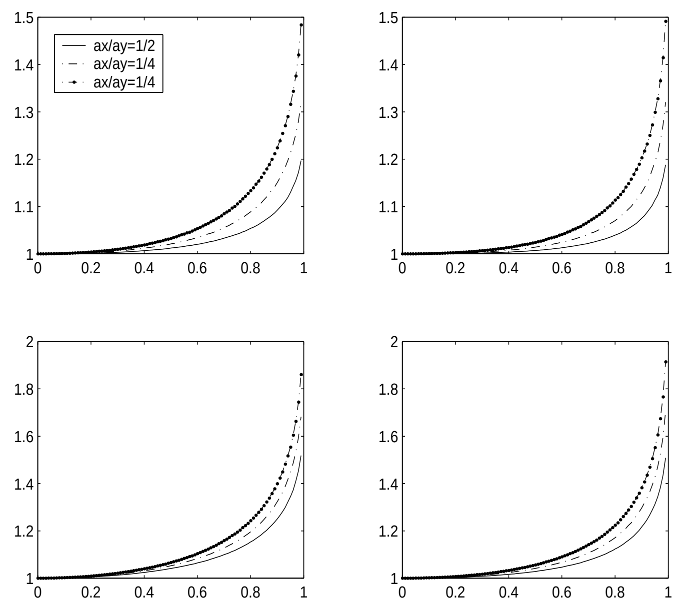

Figure 3. Dependence between RC and the measure of association for the pairs of the sub-survivals for X and Y given by the second sub-survival for X and the other three for Y ; a) $p = 0.3$ and $c_2/c_1 = 0.1$; b) $p = 0.3$ and $c_2/c_1 = 0.05$; c) $p = 0.1$ and $c_2/c_1 = 0.1$; d) $p = 0.1$ and $c_2/c_1 = 0.05$;

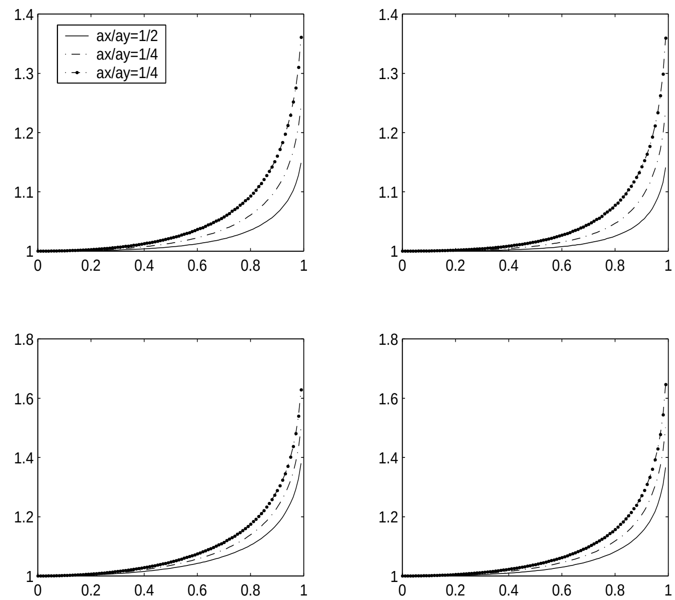

Figure 4. Dependence between RC and the measure of association for the pairs of the sub-survivals for X and Y given by the third sub-survival for X and the other three for Y ; a) $p = 0.3$ and $c_2/c_1 = 0.1$; b) $p = 0.3$ and $c_2/c_1 = 0.05$; c) $p = 0.1$ and $c_2/c_1 = 0.1$; d) $p = 0.1$ and $c_2/c_1 = 0.05$;

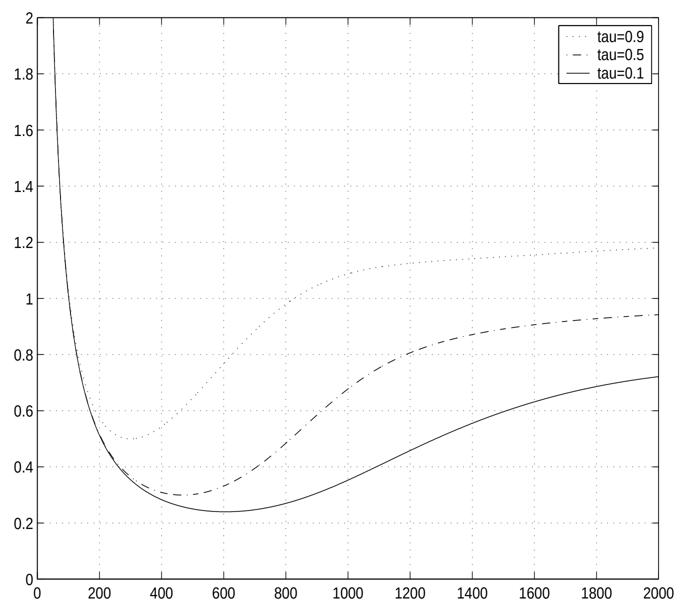

Figure 5. Specific cost for three values of Kendall's tau: 0.1, 0.5 and 0.9

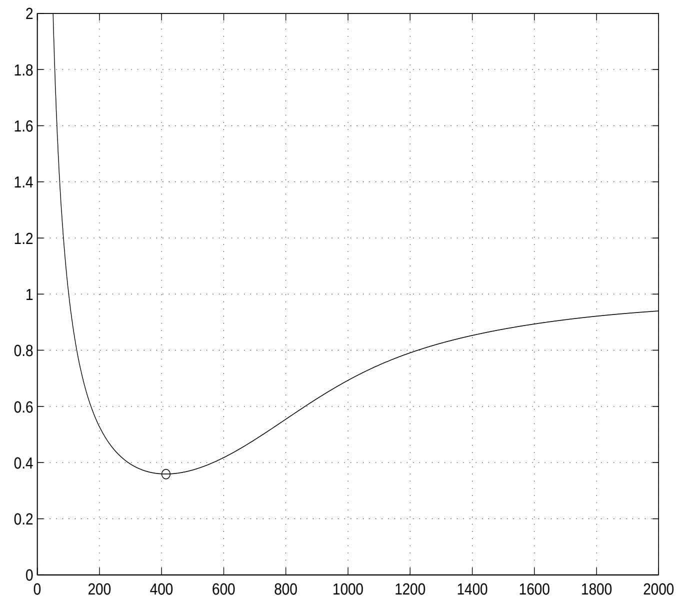

Figure 6. Average specific cost and optimal replacement time

6 Discussion

The results presented in Table 1 and 2 show that the "optimal replacement interval" and "optimal replacement costs" can be dramatically non-optimal when the wrong model is used to estimate the underlying failure distribution from censor data. The difference is least when the failure rate increases quickly. when the failure rate increase more slowly, the difference is larger. For one case calculated here the specific costs obtained by using the independent model would be more than twice the best possible specific costs using the correct model. In the second part we consider the effect of model uncertainty due to impossibility of identifying the "correct" competing risk model from censored data. Using the expert judgement to quantify the dependence between competing risks , we have shown that the replacement cost is highly sensitive to the measure of association Kendall's tau. Figure 1, 2 and 3 show that sensitivity is higher for the first model and for a certain case RC can be twice than RC for independent case. Figure 4 shows also that the difference between optimal replacement costs and optimal replacement time can be more than twice and Figure 5 presents the long term specific cost and the optimal replacement time.

The work carried out here demonstrates the importance of using good expert judgement from experts with insight into the maintenance process. If the experts are able to select the correct correlation level then this will aid model selection considerably.

7 References

[1] R.E. Barlow, F. Proschan, *Mathematical theory of reliability*, 1965, Wiley, New York.

[2] T. Bedford , "On the use of minimally informative copulae in competing risk problems", *Statistical Probabilistic Models in Reliability*, C. Ionescu and N. Limnios (eds), 1998, Birkhauser, Boston.

[3] T. Bedford, R.M. Cooke, *Probabilistic Risk Analysis: Foundations and Methods*, 2001, Cambridge

[4] T. Bedford, I. Meilijson, “The marginal distributions of lifetime variables which right censor each other”, *Analysis of Censored Data*, IMS Lecture Notes Monograph Series, vol 27, 1995, Eds H.L. Koul and J.V. Deshpande.

[5] T. Bedford, I. Meilijson, “A characterization of marginal distributions of (possibly dependent) lifetime variables which right censor each other”, *Annals of Statistics*, **25**, 1997, pp 1622-1645.

[6] T. Bedford, C. Mesina, “The impact of modeling assumptions on maintenance optimization”, *Mathematical Methods in Reliability*, N. Limnios, M. Nikulin (eds), 2000, Birkhauser, Boston

[7] R.M. Cooke, “The total time on test statistic and age-dependent censoring” *Statistics and Probability Letters*, vol 18, num 5, 1993.

[8] R.M. Cooke, “The design of reliability data bases, Part II” *Reliability Engineering and System Safety* vol 51, num 2, 1996, pp 209-225

[9] R.B. Nelsen, *An Introduction to copulas*, 1995, Springer

[10] B. Schweizer, and E.F. Wolff, “On nonparametric measures of dependence for random variables” *Annals of Statistics*, **9**, 1981, pp 879-885.

[11] A. Tsiatis, “A non-identifiability aspect in the problem of competing risks”, *Proc. Nat. Acad. Sci. USA*, vol **72**, pp 20-22, 1975.

[12] M. Zheng, J.P. Klein, “Estimates of marginal survival for dependent competing risks based on an assumed copula”, *Biometrika*, vol **82** , 1995, pp 127-138.

8 Authors

MSc. Cornel Bunea: Dep’t of Statistics, Probability Theory and Operations Research; Delft University of Technology - Faculty of Information Technology and Systems; Mekelweg 4, 2628 CD Delft, The Netherlands. Internet (e-mail): c.bunea@its.tudelft.nl

Prof. Tim Bedford: Dep’t of Management Science; Strathclyde University; 40 George Street, Glasgow G1 1QE, UK. Internet (e-mail): tim@mansci.strath.ac.uk

Cornel Bunea (born 1975) received his MSc (1999) in Safety and Reliability of Power Systems from Bucharest University of Technology, and he is now a PhD student in Competing Risk Analysis at Delft University of Technology.

Tim Bedford (born 1960) is Professor at Strathclyde University in Scotland and Associate Professor at TU Delft in the Netherlands. His current interests are in reliability, risk analysis and decision making. Research in these areas is carried out often in combination with private industry and/or government institutes. Together with Roger Cooke he has written a book ”Probabilistic Risk Analysis: Foundations and Methods”, published by Cambridge University Press.

9 Appendix

9.1 Part 1

We show briefly why the marginals are identifiable in the case that densities and subdensities exist. By definition we have that the subdistribution function of X is

$$F_X^*(t) \equiv P\{X \leq t, X < Y\}.$$

After a straightforward calculation we get:

$$F_X^*(t) = \int_0^t \int_x^\infty h(x,y)dxdy = \int_0^t H_X(x,\infty) - H_X(x,x)dx =$$

$$= F_X(t) - \int_0^t H_X(x,x)dx = F_X(t) - \int_0^t C_u(F_X(x), F_Y(x))f_X(x)dx,$$

where $h(x,y)$ is the joint density function of X and Y and $H_X(x,\infty)$ respectively $H_X(x,x)$ denote the first order partial derivative $\frac{\delta}{\delta x}H(x,y)$ calculated in (x,∞) respectively in (x,x). We obtain an analogous formula for F_Y^*. From this formula it follows that the marginal distributions functions F_X and F_Y are solutions of the following system of ordinary differential equations:

$$\begin{cases} \{1 - C_u(F_X(t), F_Y(t))\}F_X'(t) = F_X^{*'}(t) \\ \{1 - C_v(F_X(t), F_Y(t))\}F_Y'(t) = F_Y^{*'}(t) \end{cases}$$

with initial conditions $F_X(0) = F_Y(0) = 0$, where $C_u(F_X(t), F_Y(t))$ and $C_v(F_X(t), F_Y(t))$ denote the first order partial derivatives $\frac{\delta}{\delta u}C(u,v)$ and $\frac{\delta}{\delta v}C(u,v)$ calculated in $(F_X(t), F_Y(t))$.

9.2 Part 2

To see which are the relations between the measures of association and copula we will recall three theorems (see [Nelsen, 1995]).

Theorem 2 *Let X and Y be continuous random variables whose copula is C. Then Kendall's tau for X and Y (which we will denote by either $\tau(X,Y)$ or τ_C) is given by*

$$\tau(X,Y) = 4\int_0^1 \int_0^1 C(u,v)dC(u,v) - 1.$$

Theorem 3 *Let X and Y be continuous random variables with copula C. Then Spearman's rho for X and Y (which we will denote by either $\rho(X,Y)$ or ρ_C) is given by*

$$\rho(X,Y) = 12\int_0^1 \int_0^1 uvdC(u,v) - 3,$$

$$\rho(X,Y) = 12\int_0^1 \int_0^1 C(u,v)dudv - 3.$$

Recall also from [Nelsen, 1995] the following theorem which enables us to determine the parameter α (and implicitly the copula) when we know Kendall's tau.

Theorem 4 *Let X and Y be random variables with an Archimedean copula C generated by $\varphi \in \Omega$. Kendall's tau for X and Y is given by*

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

If C_α is a member of the Gumbel family, then for $\alpha \geq 1$,

$$\frac{\varphi(t)}{\varphi'(t)} dt = \frac{t \log t}{\alpha},$$

so that $\tau(\alpha) = 1 - 1/\alpha$. Now it is easy to see that

$$\alpha_\tau = 1/(1 - \tau).$$