I N S T I T U T D E S T A T I S T I Q U E
B I O S T A T I S T I Q U E E T
S C I E N C E S A C T U A R I E L L E S
(I S B A)

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

2014/34

Inference in a Survival Cure Model with Mismeasured Covariates using a SIMEX Approach

BERTRAND, A., LEGRAND, C. CARROLL, R. DE MEESTER, C. and I. VAN KEILEGOM

Inference in a Survival Cure Model with Mismeasured Covariates using a SIMEX Approach

Aurélie Bertrand* Catherine Legrand* Raymond J. Carroll†

Christophe de Meester‡ Ingrid Van Keilegom*

September 2, 2014

Abstract

In many situations in survival analysis, it may happen that a fraction of individuals will never experience the event of interest: they are considered to be cured. The promotion time cure model is one of the survival models taking this feature into account. We consider the case where one or more explanatory variables in the model are subject to measurement error. This error should be taken into account in the estimation of the model, to avoid biased estimators. A general approach that exists in the literature is the SIMEX algorithm, a method based on simulations which allows one to estimate the effect of measurement error on the bias of the estimators and to reduce this bias. We extend the SIMEX approach to the promotion time cure model. We explain how the algorithm works, and we show that the proposed estimator is consistent and

*Institute of Statistics, Biostatistics and Actuarial Sciences, Université catholique de Louvain, Louvain-la-Neuve, Belgium (aurelie.bertrand@uclouvain.be, catherine.legrand@uclouvain.be, ingrid.vankeilegom@uclouvain.be)

†Texas A&M University, College Station, USA (carroll@stat.tamu.edu)

‡Cardiovascular Research group, Institute of Experimental and Clinical Research, Université catholique de Louvain, Brussels, Belgium (christophe.demeester@uclouvain.be)

asymptotically normally distributed. We also show via simulations that the suggested method performs well in finite samples. Finally, we analyze a database in cardiology: among the explanatory variables of interest is the ejection fraction, which is known to be measured with error. There are supplementary materials online for this paper.

KEY WORDS: Bias correction; Cure fraction; Measurement error; Promotion time cure model; Semiparametric method.

1 INTRODUCTION

When analyzing time-to-event data, it often happens that a certain proportion of subjects will never experience the event of interest. For example, in medical studies where one is interested in the time until recurrence of a certain disease, it is known that, for some diseases, some patients will never suffer a relapse. In studies in econometrics on duration of unemployment, some unemployed people will never find a new job, and in sociological studies on the age at which a person marries, some people will stay unmarried for their whole life. Other examples can be found in finance, marketing, demography, and education, among others, where each time there is a certain proportion of subjects whose time to event is infinite and hence they are said to be cured. Since classical survival models implicitely assume that all individuals will eventually experience the event of interest, they cannot be used in such contexts. They would in fact lead to incorrect results such as, among others, an overestimation of the survival of the non-cured subjects. This is why specific models, called cure models, have been developped.

In order to model the impact of a set of covariates on the time-to-event variable, two main streams of cure models (and proposals that overarch both types of models) can be found in the literature. The first one is the so-called mixture cure model,

which supposes that the conditional survival function is given by $S(t|\boldsymbol{x}_1, \boldsymbol{x}_2) = P(T > t|\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_2 = \boldsymbol{x}_2) = p(\boldsymbol{x}_2) + \{1 - p(\boldsymbol{x}_2)\} S_u(t|\boldsymbol{x}_1)$, where $p(\boldsymbol{x}_2)$ is the probability of being cured for a given vector of covariates $\boldsymbol{x}_2$, and $S_u(t|\boldsymbol{x}_1)$ is the conditional survival function of the non-cured subjects, where $\boldsymbol{x}_1$ is another set of covariates (possibly with common components). This model has been studied by, among others, Boag (1949), Berkson and Gage (1952), Farewell (1982), Kuk and Chen (1992), Taylor (1995), Peng and Dear (2000), Sy and Taylor (2000), Peng (2003) and Lu (2008). A second class of models is based on an adaptation of the Cox model (Cox 1972) to allow for a cure fraction. It is called the class of promotion time cure models and supposes that

$$S(t|\boldsymbol{x}) = \exp\{-\theta(\boldsymbol{x})F(t)\}, \tag{1}$$

where $F(\cdot)$ is a proper baseline cumulative distribution function (cdf) and $\theta(\cdot)$ captures the effect of the covariates on the conditional survival function. One often chooses $\theta(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\boldsymbol{\beta})$, where the first component of the P-dimensional covariate $\boldsymbol{x}$ is supposed to be 1, in order to include an intercept in the model. Note that the classical Cox model (without cure fraction) does not include an intercept, since it supposes that $F(t)$ tends to infinity when t tends to infinity, and an intercept would therefore not be identifiable. References on the promotion time cure model include Yakovlev and Tsodikov (1996), Tsodikov (1998a,b, 2001), Chen et al. (1999), Ibrahim et al. (2001), Tsodikov et al. (2003), Zeng et al. (2006) and Carvalho Lopes and Bolfarine (2012).

In this paper we consider the promotion time cure model (1) in which we leave F completely unspecified. We suppose that the survival time T is subject to random right censoring, i.e. instead of observing T we observe $Y = \min(T, C)$ and $\delta = I(T \leq C)$, where the censoring time C is independent of T given $\boldsymbol{X}$. An immediate consequence of the presence of right censoring is that for the censored observations, we

do not observe whether they are cured or not cured, the latter also called susceptible.

In addition to being exposed to censoring, the data can also be subject to another type of incompleteness. As is often the case in practice, we suppose that some (or all) continuous covariates are subject to measurement error. For instance, in medical studies the error can be caused by imprecise medical instruments, and in econometric studies economic variables like welfare or income can often not be measured in a precise way, in which case one has to work with approximate measures, including some error. Although this measurement error is rarely taken into account, ignoring it leads to several issues, including incorrect conclusions drawn from biased estimators (Carroll et al. 2006). In order to deal with this measurement error, some assumptions about its form are necessary. We consider a classical additive measurement error model for the continuous covariates, so that we have, for the whole vector of covariates,

$$\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}, \tag{2}$$

where $\boldsymbol{W}$ is the vector of observed covariates and $\boldsymbol{U}$ is the vector of measurement errors. We further assume that $\boldsymbol{U}$ is independent of $\boldsymbol{X}$ and $\boldsymbol{U}$ follows a continuous distribution with mean zero and known covariance matrix $\boldsymbol{V}$, where the elements of $\boldsymbol{V}$ corresponding to covariates with no measurement error (the non-continuous covariates, and possibly some continuous ones) are set to 0. It is also assumed that (T, C) and $\boldsymbol{W}$ are independent given $\boldsymbol{X}$. When $\boldsymbol{U}$ is assumed to be normally distributed, this is the model studied, for example, by Cook and Stefanski (1994) and Ma and Yin (2008).

The methods designed to deal with measurement error in the covariates can be classified into two families: structural modeling and functional modeling approaches (Carroll et al. 2006). In structural modeling, the distribution of the unobservable covariates $\boldsymbol{X}$ must be modeled (usually parametrically), while in functional modeling, no assumptions are made regarding the distribution of $\boldsymbol{X}$. When the distributional

assumptions are met, the approaches of the first group yield higher efficiency. However, an obvious advantage of the second type of approaches is their robustness with respect to a possible misspecification of the distribution of $\boldsymbol{X}$.

In this paper we choose to work with a functional approach, the so-called SIMEX (Simulation-Extrapolation) approach for correcting for the measurement error. The basic idea of SIMEX consists of two steps. In the first step we consider several increasing amounts of measurement error, and simulate a large number of data sets for each of these levels of measurement error. At each level we estimate the vector $\boldsymbol{\beta}$ of regression coefficients ignoring the presence of measurement error. Next, in the second step we extrapolate the so-obtained estimators corresponding to the different levels of measurement error to the situation where the covariates are observed without error. This very intuitive algorithm was proposed by Cook and Stefanski (1994) in the context of generalized linear models, and has a number of important advantages that will be summarized in the Discussion (Section 6). The method has increasing popularity and has been considered in many different contexts. In survival analysis, it has been used in, e.g., the Cox model (Carroll et al. 2006), the Cox model with nonlinear effect of mismeasured covariates (Crainiceanu et al. 2006), the multivariate Cox model (Greene and Cai 2004) and the frailty model (Li and Lin 2003), but, as far as we know, not in cure models.

To the best of our knowledge, the problem considered in this paper has been addressed in only one other paper in the literature: Ma and Yin (2008) also studied a promotion time cure model with right censored responses and mismeasured covariates. But instead of using the SIMEX approach, they introduced a corrected score approach in order to deal with the measurement error in the covariates. Their approach yields consistent and asymptotically normal estimators when the measurement error variance is known and the error is normally distributed. However, their

method only works for the specification $\theta(\boldsymbol{x}) = \exp(\boldsymbol{x}^T\boldsymbol{\beta})$, while the SIMEX algorithm can be used for any parametric version of $\theta(\boldsymbol{x})$. Moreover, they mention the case of non-gaussian measurement error, but they do not study it in detail.

The motivation for considering cure models with right censoring allowing for covariates to be subject to measurement error is multifold. Our research was inspired by a recent study focusing on the link between the ejection fraction and death from heart disease amongst patients suffering from aortic insufficiency. Our data consists of 393 patients with moderate to severe aortic insufficiency followed up for death from heart disease. These patients were taken in charge by a cardiac department, monitored regularly and operated upon if thought necessary by the cardiac surgeon in charge. It is therefore expected that a majority of these patients will actually not die from their heart disease, explaining the presence of cured individuals in our data. Nowadays, the ejection fraction plays a major role in the treatment of these patients, and current practice is to recommend surgery when it goes below a given threshold (Bonow et al. 1998; Vahanian et al. 2007). However, the ejection fraction, as measured by non-invasive techniques (i.e. echocardiography) is known to be measured with error (Otterstad et al. 1997) and it is therefore necessary to take this into account to quantify correctly the impact of this covariate on survival.

The rest of this paper is organized as follows. In the next section, we explain the proposed estimation method for the parameter vector $\boldsymbol{\beta}$ and the baseline distribution F. Section 3 contains the asymptotic properties of these estimators. The finite sample properties of the estimator of $\boldsymbol{\beta}$ are investigated in Section 4 through a simulation study. Section 5 contains the analysis of the aortic insufficiency database, and in Section 6 we discuss the obtained results and mention some ideas for future research. Finally, the Appendix contains the proof of asymptotic normality, while the proof of consistency can be found in the Supplementary Materials.

2 METHODOLOGY

We suppose that we have n independent and identically distributed right-censored observations $(Y_i, \delta_i, \boldsymbol{X}_i)$. We denote by $Y_{(1)}, \ldots, Y_{(m)}$ the m distinct ordered event times, so that $Y_{(1)} < \ldots < Y_{(m)}$. We use model (1), where we consider $\theta(\boldsymbol{x}) = \eta(\boldsymbol{x}^T\boldsymbol{\beta})$ for some given function η. Two examples are $\eta(\cdot) = \exp(\cdot)$ and $\eta(\cdot) = \exp(\cdot)/\{1 + \exp(\cdot)\}$. We present the algorithm to be used when the error $\boldsymbol{U}$ is normally distributed, although it can be applied whatever its distribution.

The general idea of the SIMEX algorithm consists in adding successively increasing (and known) amounts of artificial noise to the covariates subject to measurement error, estimating the model without taking the measurement error into account, and extrapolating back to the case of no measurement error. Two types of parameters have to be chosen: the levels of added noise $\lambda = \lambda_1, \ldots, \lambda_K$ and the number B of simulations for each value of λ. Some common values are $K = 5$ and $B = 50$ (Cook and Stefanski 1994; Carroll et al. 1996).

The SIMEX algorithm for the promotion time cure model is then:

1. For $\lambda = \lambda_1, \ldots, \lambda_K$

   - For $b = 1, \ldots, B$

     – Generate $\boldsymbol{Z}_{b,i} \sim_{iid} N_P(\boldsymbol{0}, \boldsymbol{I}_P)$ independently of the observed data and construct $\boldsymbol{W}_{i,\lambda,b} = \boldsymbol{W}_i + \sqrt{\lambda}\boldsymbol{V}^{1/2}\boldsymbol{Z}_{b,i}$ for each individual $i = 1, \ldots, n$. The variance-covariance matrix of the contaminated $\boldsymbol{W}_{i,\lambda,b}$ is

     $$\text{Var}\left(\boldsymbol{W}_{i,\lambda,b}|\boldsymbol{X}_i\right) = \text{Var}(\boldsymbol{W}_i|\boldsymbol{X}_i) + \lambda\boldsymbol{V} = \boldsymbol{V} + \lambda\boldsymbol{V} = (1+\lambda)\boldsymbol{V},$$

     which converges to the zero matrix as λ converges to -1.

     – Replace $\boldsymbol{X}_i$ by $\boldsymbol{W}_{i,\lambda,b}$ in the promotion time cure model:

     $$S\left(t|\boldsymbol{W}_{i,\lambda,b}\right) = \exp\left\{-F(t)\eta(\boldsymbol{W}_{i,\lambda,b}^T\boldsymbol{\beta}_\lambda)\right\}.$$

– When the $\boldsymbol{W}_{i,\lambda,b}$ are known, this model is the standard promotion time cure model. Obtain the estimates $\widehat{\boldsymbol{\beta}}_{\lambda,b}$ of $\boldsymbol{\beta}_\lambda$, by using a *naive* estimation method, i.e., a method which does not take the measurement error into account.

- Obtain $\widehat{\boldsymbol{\beta}}_\lambda = B^{-1}\sum_{b=1}^{B}\widehat{\boldsymbol{\beta}}_{\lambda,b}$.

2. Choose an extrapolant (linear, quadratic, fractional, etc.) for each parameter (i.e. for each element $\widehat{\beta}_{\lambda,p}$ of the vector $\widehat{\boldsymbol{\beta}}_\lambda$), as a function of the λ's: $\boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\lambda) = \{g_{\beta_1}(\boldsymbol{\gamma}_{\beta_1},\lambda),\ldots,g_{\beta_P}(\boldsymbol{\gamma}_{\beta_P},\lambda)\}^T$ depending on a vector of parameters $\boldsymbol{\gamma}_{\boldsymbol{\beta}} = (\boldsymbol{\gamma}_{\beta_1}^T,\ldots,\boldsymbol{\gamma}_{\beta_P}^T)^T$. In the case of the quadratic extrapolant, one obtains:

$$\widehat{\beta}_{\lambda_k,p} = g_{\beta_p}(\boldsymbol{\gamma}_{\beta_p},\lambda_k) + \eta_{p,k} = \gamma_{\beta_p,1} + \gamma_{\beta_p,2}\lambda_k + \gamma_{\beta_p,3}\lambda_k^2 + \eta_{p,k},$$
$$p = 1,\ldots,P; \quad k = 1,\ldots,K.$$

Fit these parametric models for each $p = 1,\ldots,P$ in order to obtain $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\gamma}}_{\beta_1}^T,\ldots,\widehat{\boldsymbol{\gamma}}_{\beta_P}^T)^T$.

3. Obtain the SIMEX estimated values

$$\widehat{\boldsymbol{\beta}}_{SIMEX} = \lim_{\lambda\to -1}\boldsymbol{g}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}},\lambda).$$

This algorithm is illustrated (for the coefficient of the mismeasured covariate in the cardiology database analyzed in Section 5) in Figure 1, with $B = 50$, $\lambda \in \{0, 0.5, 1, 1.5, 2\}$ and a quadratic extrapolation function.

The naive estimation method that appears in the simulation step can be any method that allows estimation of the parameters of the promotion time cure model. When there is no measurement error in the covariates, Zeng et al. (2006) and Ma and Yin (2008) explain how to estimate the model parameters $\boldsymbol{\beta}$ and F. They show that the log-likelihood of the promotion time cure model without measurement error is

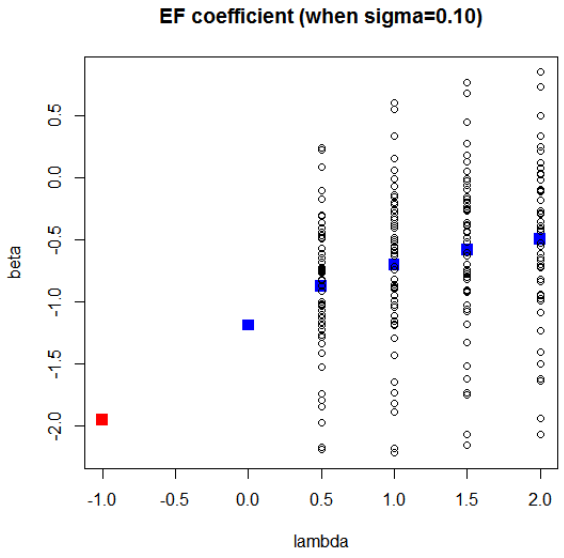

Figure 1: Illustration of the functioning of the SIMEX algorithm for one parameter. The circles are the estimated values (50 for each value of λ). The squares corresponding to $\lambda = 0.5, 1, 1.5, 2$ are the average estimated values. They are used together with the naive estimator corresponding to $\lambda = 0$ to fit the extrapolation curve. The value on the y-axis of the square corresponding to $\lambda = -1$ is the SIMEX estimator.

(we use $\eta(\boldsymbol{X}^T\boldsymbol{\beta})$ instead of the particular case $\exp(\boldsymbol{X}^T\boldsymbol{\beta})$ considered by the authors):

$$\ell = \sum_{i=1}^{n} \big[\delta_i I(Y_i < \infty) \left\{-F(Y_i)\eta(\mathbf{X}_i^T\boldsymbol{\beta}) + \log F\{Y_i\} + \log \eta(\mathbf{X}_i^T\boldsymbol{\beta})\right\} \\ +(1-\delta_i) I(Y_i < \infty) \left\{-F(Y_i)\eta(\mathbf{X}_i^T\boldsymbol{\beta})\right\} - I(Y_i = \infty)\eta(\mathbf{X}_i^T\boldsymbol{\beta})\big], \qquad (3)$$

where $F\{Y_i\}$ is the jump size of F at Y_i. As Zeng et al. (2006) explain, it can be shown that the nonparametric maximum likelihood estimator for F is a function with point masses at the distinct observed failure times $Y_{(1)}, \ldots, Y_{(m)}$ only: if $p_{(j)}$ denotes the jump size of F at $Y_{(j)}$, then $F(Y_i) = \sum_{Y_{(j)} \leq Y_i} p_{(j)}$. Moreover, the authors also explain that, in order for this semiparametric model to be identifiable in $(\boldsymbol{\beta}, F)$, we need to choose a threshold τ (the *cure threshold*) such that all censored individuals with a censoring time larger than this threshold are treated as observed to be cured (with $T_i = C_i = Y_i = \infty$). The largest observed failure time is often used as the cure threshold in practice, so that $\tau = Y_{(m)}$. Thus, the estimated baseline cumulative

distribution function is forced to be 1 beyond the cure threshold, $\sum_{j=1}^{m} p_{(j)} = 1$, since no event can occur at those times.

The parameters can then be estimated by solving the score equations related to the likelihood in which the baseline cumulative distribution function is replaced by a step function.

If the interest also lies in estimating the baseline cumulative distribution function F, exactly the same SIMEX procedure can be applied to the $\widehat{p}_i$, yielding the $\widehat{p}_{SIMEX,i}$. In order to ensure that their sum is equal to 1, each of them is divided by their sum: $\widehat{p}^*_{SIMEX,i} = \widehat{p}_{SIMEX,i} / \sum_j \widehat{p}_{SIMEX,j}$. Finally, we obtain $\widehat{F}_{SIMEX}(t) = \sum_{Y_{(j)} \leq t} \widehat{p}^*_{SIMEX,(j)}$.

3 ASYMPTOTIC PROPERTIES

We present some theorems regarding the asymptotic properties (consistency and asymptotic normality) of the SIMEX estimators of the regression parameters $\boldsymbol{\beta}$ and the baseline cumulative distribution function F. Theorem 1 states their consistency; its proof can be found in the Supplementary Materials. Theorem 2 establishes their asymptotic normality and is proved in the Appendix. Here, we assume that the $\boldsymbol{Z}_{b,i}$ that are generated in the simulation step follow a truncated Gaussian distribution with large truncation limits (this will always be the case in practice). We also assume that the expectation of the log-likelihood has a unique maximizer, whether or not there is measurement error in the covariates.

Theorem 1. *Under the regularity conditions (C1)-(C4) of Zeng et al. (2006) (by replacing* $\mathbf{X}$ *that appears there by* $\mathbf{W}_{\lambda,b}$, $\forall b = 1, \ldots, B$ *and* $\forall \lambda$*), if the measurement error variance and the true extrapolant function are known, then, with probability 1,*

$$\|\widehat{\boldsymbol{\beta}}_{SIMEX} - \boldsymbol{\beta}_{TRUE}\| \to 0 \quad \textit{and} \quad \sup_{t \in \mathbb{R}^+} |\widehat{F}_{SIMEX}(t) - F_{TRUE}(t)| \to 0.$$

Theorem 2. *Under the regularity conditions (C1)-(C4) of Zeng et al. (2006) (by replacing* $\mathbf{X}$ *that appears there by* $\mathbf{W}_{\lambda,b}$, $\forall b = 1, \ldots, B$ *and* $\forall \lambda$*), if the measurement error variance and the true extrapolant function are known, then* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{SIMEX} - \boldsymbol{\beta}_{TRUE}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$*, where* Σ *is given in the Appendix. Moreover,* $\sqrt{n}(\widehat{F}_{SIMEX} - F_{TRUE})$ *converges weakly to a zero-mean Gaussian process* $\mathcal{G}$ *whose covariance function is given in the Appendix.*

As far as the variance of the SIMEX estimator is concerned, it can be estimated using the method introduced by Stefanski and Cook (1995) and summarized, for example, in Carroll et al. (2006). The variance estimator can be computed as $\widehat{\Sigma} = \lim_{\lambda \to -1} \left(\overline{\Sigma}_\lambda - \widehat{\Sigma}_\lambda \right)$, where

- $\overline{\Sigma}_\lambda$ is the extrapolation function corresponding to

$$\overline{Var}[\widehat{\boldsymbol{\beta}}_\lambda] = B^{-1} \sum_{b=1}^{B} \widetilde{Var}[\widehat{\boldsymbol{\beta}}_{\lambda,b}],$$

  where $\widetilde{Var}[\widehat{\boldsymbol{\beta}}_{\lambda,b}]$ is the estimated covariance matrix of $\widehat{\boldsymbol{\beta}}_{\lambda,b}$, when using the variance estimator corresponding to the naive estimation method.

- $\widehat{\Sigma}_\lambda$ is the extrapolation function corresponding to

$$\widehat{Var}[\widehat{\boldsymbol{\beta}}_\lambda] = B^{-1} \sum_{b=1}^{B} [\widehat{\boldsymbol{\beta}}_{\lambda,b} - \widehat{\boldsymbol{\beta}}_\lambda][\widehat{\boldsymbol{\beta}}_{\lambda,b} - \widehat{\boldsymbol{\beta}}_\lambda]^T,$$

  i.e. the empirical covariance matrix of $\{\widehat{\boldsymbol{\beta}}_{\lambda,b}\}_{b=1}^{B}$.

4 SIMULATION STUDIES

We perform here a simulation study to investigate the properties of the proposed estimator in samples of finite size and to compare it with both the naive method, i.e.

the one that does not take measurement error into account, which was introduced at the end of Section 2 and is based on Equation (3), and the corrected score method of Ma and Yin (2008). For the SIMEX algorithm, we used $B = 50$, $\lambda \in \{0, 0.5, 1, 1.5, 2\}$ and a quadratic extrapolant. For each setting, 500 simulated data sets were analyzed.

4.1 One Mismeasured Covariate

The first set of simulation studies that we conduct is the first one used in Ma and Yin (2008). They assume that the follow-up is infinite, and that the censoring distribution and the failure distribution have an infinite support, so that each individual is known to be either cured (when $T_i = C_i = \infty$), dead (when $T_i < C_i \leq \infty$) or censored (when $C_i < T_i \leq \infty$; some of the censored individuals being actually cured, those for whom $C_i < T_i = \infty$). In such a case, a cure threshold is not needed for the estimation. Each subject has a probability of 60% of having an infinite censoring time. Because of the infinite follow-up, this setting clearly does not correspond to a realistic case, but is useful for assessing the proposed method in an "ideal" situation.

The model under study is:

$$S(t|X_1, X_2) = \exp\left\{-\exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2\right) F(t)\right\}$$

and we generate the data from the model with $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = -0.5$, $F(t) = 1 - \exp(-t)$ and $X_1 \sim Uniform[0, 1]$, $X_2 \sim Bernoulli(0.5)$, X_1 is subject to measurement error so that $W = X_1 + U$ is observed, where $U \sim N(0, v^2)$. Moreover, the censoring time C is independent of $\boldsymbol{X}$ and of T given $\boldsymbol{X}$, and the finite censoring times follow an exponential distribution with mean μ.

Eight different settings are obtained by considering two possible values for each of the following three parameters: sample size ($n = 200$ or $n = 300$), variance of the measurement error ($v^2 = 0.1^2$ or $v^2 = 0.2^2$) and mean of the finite censoring times ($\mu = 0.1$ or $\mu = 1$). The average cure rate ($T = \infty$) is 14%, the average proportion

of subjects who are considered cured for the estimation ($T = C = \infty$) is 8%, while the average censoring rate is 17% when $\mu = 1$ and 33% when $\mu = 0.1$.

The results for the four settings with $\mu = 0.1$ are summarized in Table 1, while those corresponding to $\mu = 1$ can be found in the Supplementary Materials.

The empirical and estimated variances are always quite close to each other, while both the corrected score and the SIMEX approaches yield coverage probabilities of the confidence intervals that are close to their nominal level of 95%. It also appears that, compared to the naive estimation method, both correction methods decrease the bias in the intercept and the parameter corresponding to the mismeasured covariate, but at the cost of a larger variance. Although the SIMEX algorithm and the method of Ma and Yin (2008) cannot really be discriminated on the basis of the bias, the former leads to a smaller variance for β_0 and β_1 when $v = 0.2$, and to similar variances when $v = 0.1$. This results in an MSE which is, when $v = 0.2$, the smallest for SIMEX (compared to the naive and corrected score methods). When the measurement error variance is smaller, the naive method yields a smaller MSE than both correction methods. This is to be expected since bias correction methods have a larger variance than the naive method.

4.2 Two Mismeasured Covariates

We now introduce, in the previous setting, one additional covariate with measurement error. In this case, $S(t|X_1, X_2, X_3) = \exp\{-\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3) F(t)\}$. We generate the data with $\beta_0 = 0.5$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = -0.5$, $F(t) = 1 - \exp(-t)$ and $X_1 \sim Uniform[0, 1]$, $X_2 \sim N(0, 1)$, $X_3 \sim Bernoulli(0.5)$, X_1 and X_2 are subject to measurement error so that $W_1 = X_1 + U_1$ and $W_2 = X_2 + U_2$ are observed, where $U_1 \sim N(0, v_1^2)$, $U_2 \sim N(0, v_2^2)$ and U_1 and U_2 are uncorrelated.

The average censoring rate is 17% when $\mu = 1$ and 32% when $\mu = 0.1$. The

Table 1: Simulation results for the settings with one mismeasured covariate, when $\mu = 0.1$.

			Ma & Yin method			Naive method			SIMEX method		
n	v	Estimate	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
200	0.1	Bias	-.010	.032	-.002	.043	-.092	.002	-.004	.015	-.001
		Emp. var.	.053	.132	.041	.047	.099	.040	.052	.128	.041
		Est. var.	.053	.130	.036	.046	.098	.036	.052	.124	.036
		95% cv	.946	.948	.936	.936	.934	.938	.942	.942	.938
		MSE	.053	.133	.041	.048	.108	.040	.052	.128	.041
200	0.2	Bias	-.030	.085	-.005	.142	-.315	.007	.036	-.076	.001
		Emp. var.	.069	.226	.044	.041	.075	.040	.054	.143	.042
		Est. var.	.074	.227	.039	.040	.074	.036	.052	.127	.037
		95% cv	.960	.968	.932	.884	.788	.934	.938	.924	.930
		MSE	.070	.234	.044	.061	.175	.040	.055	.148	.042
300	0.1	Bias	-.011	.037	-.014	.042	-.084	-.011	-.005	.023	-.014
		Emp. var.	.039	.094	.026	.034	.071	.025	.038	.090	.026
		Est. var.	.035	.087	.024	.031	.065	.024	.034	.083	.024
		95% cv	.956	.944	.966	.934	.932	.966	.954	.942	.966
		MSE	.039	.095	.026	.036	.078	.025	.038	.091	.026
300	0.2	Bias	-.021	.068	-.018	.142	-.310	-.006	.036	-.069	-.013
		Emp. var.	.049	.145	.027	.030	.051	.025	.039	.097	.026
		Est. var.	.046	.139	.026	.027	.049	.024	.034	.084	.025
		95% cv	.966	.958	.958	.852	.716	.954	.938	.934	.958
		MSE	.050	.149	.027	.050	.147	.025	.041	.102	.026

NOTE: The coverage probabilities of 95% confidence intervals (95% cv) are computed based on the asymptotic normal distribution.

average proportion of subjects considered cured for the estimation is 13%, while the average cure rate is 21%.

The results for the four settings with $\mu = 0.1$ are summarized in Table 2, while those corresponding to $\mu = 1$ can be found in the Supplementary Materials.

The three methods perform similarly as far as β_3 is concerned. None of the correction methods clearly outperforms the other one in terms of the bias. For the larger values of the measurement error variances, the method of Ma and Yin (2008) is the best for β_0 and β_1, while SIMEX is prefered for β_2. However, when also taking the variance of the estimators into account, the MSE indicates that the naive method is preferable for small values of v_1 and v_2, while SIMEX outperforms the corrected score approach and the naive method for larger values of v_1 and v_2.

4.3 A More Realistic Case

In practice, neither the failure times nor the censoring times can be infinite. Consequently, none of the cured subjects are observed to be cured. The use of the cure threshold (the largest observed event time, as mentioned in Section 2) is thus needed for the estimation of the model parameters. Moreover, depending on the context, the censoring and cure rates can be much larger than the values considered in the two previous settings.

We therefore consider the following model:

$$S(t|X_1, X_2) = \exp\left\{-\exp(-0.3 + X_1 - 0.5X_2)F(t)\right\},$$

where $X_1 \sim Uniform[0, 1]$, $X_2 \sim Bernoulli(0.5)$, X_1 is subject to measurement error so that $W = X_1 + U$ is observed, where $U \sim N(0, v^2)$.

For the baseline cumulative distribution function $F(t)$, we use an exponential distribution with mean 6 which is truncated at $t = 20$. Consequently, the maximum

Table 2: Simulation results for the settings with 2 mismeasured covariates, when $\mu = 0.1$.

				Ma & Yin method				Naive method				SIMEX method			
n	v_1	v_2	Estimate	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
200	0.1	0.1	Bias	.001	.020	.016	-.012	.049	-.107	-.007	-.006	.007	.001	.012	-.011
			Emp. var.	.070	.154	.017	.040	.061	.115	.016	.039	.068	.148	.017	.040
			Est. var.	.061	.138	.015	.039	.052	.102	.014	.038	.059	.131	.015	.039
			95% cv	.932	.938	.950	.948	.922	.934	.930	.950	.930	.946	.950	.950
			MSE	.070	.154	.017	.040	.064	.126	.016	.039	.068	.148	.017	.040
200	0.2	0.2	Bias	-.019	.081	.030	-.018	.133	-.334	-.056	.003	.047	-.094	.007	-.009
			Emp. var.	.098	.279	.022	.046	.056	.088	.015	.039	.076	.173	.019	.043
			Est. var.	.086	.252	.019	.043	.046	.077	.013	.038	.060	.136	.016	.041
			95% cv	.942	.962	.958	.942	.888	.758	.902	.940	.918	.924	.948	.942
			MSE	.098	.286	.023	.046	.074	.200	.018	.039	.079	.182	.019	.043
300	0.1	0.1	Bias	.020	.008	.012	-.013	.066	-.115	-.010	-.008	.024	-.007	.009	-.012
			Emp. var.	.041	.099	.009	.024	.036	.075	.009	.023	.040	.096	.009	.024
			Est. var.	.039	.090	.010	.026	.034	.068	.009	.025	.038	.086	.010	.026
			95% cv	.944	.942	.964	.952	.924	.936	.958	.956	.938	.942	.964	.952
			MSE	.041	.099	.010	.024	.040	.089	.009	.023	.041	.096	.009	.024
300	0.2	0.2	Bias	.008	.047	.022	-.016	.149	-.340	-.059	.002	.066	-.101	.005	-.010
			Emp. var.	.053	.163	.012	.026	.032	.059	.008	.024	.044	.112	.011	.025
			Est. var.	.052	.149	.012	.028	.030	.051	.009	.025	.039	.089	.011	.027
			95% cv	.950	.956	.962	.954	.868	.662	.906	.948	.906	.912	.958	.950
			MSE	.053	.166	.013	.027	.054	.174	.012	.024	.048	.122	.011	.025

NOTE: The coverage probabilities of 95% confidence intervals (95% cv) are computed based on the asymptotic normal distribution.

value for the event times is 20. The censoring times are independent of the covariates and are generated from an exponential distribution with mean $\mu = 5$, which is truncated at $t = 30$.

Four different settings are obtained by considering two possible values for these two parameters: sample size ($n = 200$ or $n = 300$) and variance of the measurement error ($v^2 = 0.1^2$ or $v^2 = 0.25^2$). The average censoring rate is 60% and the average proportion of cured subjects is 39%, while the average observed cure rate is 5%. For each setting, we performed analysis of 500 simulated data sets. The results are summarized in Table 3.

As can be expected, differences between the methods appear only for β_1 and, in some cases, for β_0. In terms of the bias, both correction methods are preferable to the naive one. When $v = 0.1$, SIMEX is the best for β_1, while the method of Ma and Yin (2008) is the best for this parameter when $v = 0.25$. When $v = 0.1$ the MSE of the naive estimator is the smallest. When $v = 0.25$, the MSE of the SIMEX estimator is the smallest (while the method of Ma and Yin yields the largest MSE).

5 AORTIC INSUFFICIENCY DATABASE

We illustrate the proposed methodology on data from patients suffering of aortic insufficiency (AI). Between 1995 et 2013, 393 patients underwent echocardiography for severe AI at the Brussels Saint-Luc University Hospital (Belgium). The (follow up) data were collected by one of the authors (CdM) and include information from the diagnosis of the pathology, between 1981 and 2013. The main event of interest in this study is death from AI. It is however known that a proportion of patients will not die from AI and will therefore be considered as cured from their heart disease. The main objective of this study is to investigate the link between the ejection fraction (measured at baseline) and the survival of the patients. The ejection fraction is the

Table 3: Simulation results for the realistic settings.

			Ma & Yin method			Naive method			SIMEX method		
v	n	Estimate	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
0.1	200	Bias	-0.021	0.013	-0.022	0.038	-0.106	-0.021	-0.014	-0.002	-0.022
		Emp. var.	0.126	0.212	0.067	0.112	0.162	0.067	0.124	0.206	0.067
		Est. var.	0.114	0.221	0.063	0.098	0.167	0.062	0.110	0.211	0.063
		95% cv	0.944	0.972	0.948	0.942	0.954	0.95	0.942	0.968	0.948
		MSE	0.126	0.212	0.068	0.114	0.173	0.067	0.124	0.206	0.068
0.25	200	Bias	-0.056	0.093	-0.026	0.200	-0.431	-0.018	0.076	-0.184	-0.020
		Emp. var.	0.185	0.450	0.071	0.093	0.100	0.066	0.123	0.213	0.068
		Est. var.	0.218	0.578	0.069	0.079	0.105	0.062	0.104	0.194	0.063
		95% cv	0.972	0.974	0.960	0.885	0.740	0.954	0.942	0.913	0.948
		MSE	0.188	0.459	0.072	0.133	0.286	0.067	0.129	0.247	0.068
0.1	300	Bias	-0.031	0.025	-0.013	0.028	-0.094	-0.012	-0.024	0.010	-0.013
		Emp. var.	0.100	0.167	0.043	0.088	0.128	0.043	0.098	0.160	0.043
		Est. var.	0.079	0.144	0.041	0.069	0.110	0.041	0.076	0.137	0.041
		95% cv	0.932	0.948	0.956	0.930	0.924	0.958	0.934	0.946	0.958
		MSE	0.101	0.168	0.044	0.089	0.136	0.043	0.099	0.160	0.044
0.25	300	Bias	-0.066	0.105	-0.019	0.187	-0.412	-0.011	0.059	-0.153	-0.014
		Emp. var.	0.153	0.371	0.047	0.074	0.082	0.043	0.100	0.171	0.044
		Est. var.	0.145	0.375	0.046	0.056	0.069	0.041	0.072	0.127	0.042
		95% cv	0.960	0.968	0.956	0.850	0.632	0.958	0.910	0.896	0.954
		MSE	0.158	0.382	0.047	0.109	0.252	0.043	0.103	0.195	0.044

NOTE: The coverage probabilities of 95% confidence intervals (95% cv) are computed based on the asymptotic normal distribution.

ratio of the difference between the end-diastolic and the end-systolic volumes over the end-diastolic volume and it therefore measures the fraction of blood which leaves the heart each time it contracts. It is typically high for healthy individuals and is one of the main indicators appearing in the guidelines used to decide whether a patient should be operated on or not (Bonow et al., 1998; Vahanian et al., 2007). However, the ejection fraction is known to be measured with error (Otterstad et al. 1997), and this should be taken into account when evaluating its impact on the survival of the patients.

After a median follow-up of 7.2 years, only 58 patients (15%) had died, and the Kaplan-Meier estimate of the survival curve for these patients shows a clear plateau after about 17 years (Figure 2). To take into account the presence of cured patients and the measurement error in the covariate of interest, we apply the promotion time cure model estimated with the SIMEX algorithm (with the quadratic extrapolant). We also compare our results with those obtained from a "naive" promotion time cure model (ignoring measurement error). In our data the ejection fraction (EF) takes value between 0.19 and 0.84 (median 0.56) and based on previous work (Otterstad et al. 1997), we consider a standard deviation of the measurement error (v) of 0.05 and 0.10. Our model is adjusted for other patient characteristics, measured without error, namely: gender (79% male), age at diagnosis (median 52, range 17-88) and surgery strategy chosen by the cardiologist for this patient (no surgery, 15% - surgery within the first 3 months, 39% - surgery after the first 3 months, 46%). Results are presented in Table 4.

The parameters that are the most affected when taking the measurement error into account are the intercept and the coefficient of the EF. It can be seen that correcting for the measurement error increases the size of the estimated effect of the ejection fraction. In the promotion time cure model, a negative coefficient implies an

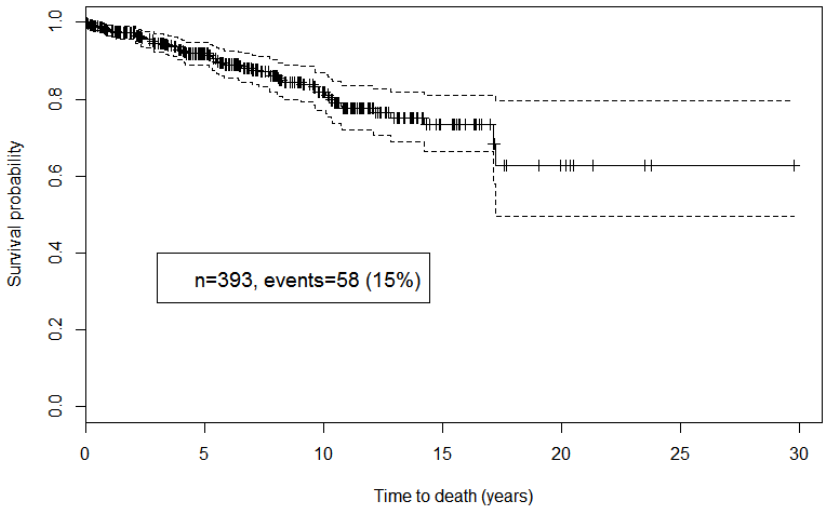

Figure 2: Kaplan-Meier estimate of the survival curve for the patients from the aortic insufficiency database.

increase in the cure probability and in the survival (at all times), when the value of the covariate increases. The results hence indicate that, all other things being equal, the higher the ejection fraction, the higher the cure probability and the better the survival for the susceptible subjects. This is consistent with the expectations and with the existing guidelines, which advise to perform surgery when the EF is below a given threshold. As far as the surgery strategy is concerned, our results indicate a better survival for the patients having undergone surgery more than three months after the discovery of the disease, and the worst, for those with surgery within the first three months (although the effect is reduced when taking measurement error into account). These results should however be interpreted carefully. First, the patients in the group "surgery after 3 months" have, by definition, lived at least three months after the discovery of their disease. Second, and probably more importantly, the two groups are not comparable at baseline as the decision of whether to operate immediately was taken according to existing guidelines, based on the prognosis of

Table 4: Regression coefficient estimates, estimated variances and confidence intervals (based on the asymptotic Normal distribution)

Estimate	Intercept	EF	Age	Gender	Surgery < 3 months	Surgery > 3 months
Naive	-4.1230	-1.1905	0.0759	0.6545	0.1429	-0.7332
(Estimated var.)	(1.2238)	(1.9144)	(0.0001)	(0.0825)	(0.1426)	(0.1481)
95% C.I. (lower bound)	-6.2913	-3.9024	0.0563	0.0915	-0.5972	-1.4875
95% C.I. (upper bound)	-1.9547	1.5214	0.0955	1.2175	0.8830	0.0211
SIMEX ($v = 0.05$)	-3.9752	-1.4072	0.0757	0.6483	0.1272	-0.7357
(Estimated var.)	(1.4598)	(2.6817)	(0.0001)	(0.0829)	(0.1436)	(0.1469)
95% C.I. (lower bound)	-6.3433	-4.6169	0.0561	0.0840	-0.6155	-1.4869
95% C.I. (upper bound)	-1.6071	1.8025	0.0953	1.2126	0.8699	0.0155
SIMEX ($v = 0.10$)	-3.6157	-1.9533	0.0750	0.6339	0.1038	-0.7335
(Estimated var.)	(1.9068)	(4.1702)	(0.0001)	(0.0840)	(0.1463)	(0.1451)
95% C.I. (lower bound)	-6.3222	-5.9558	0.0554	0.0658	-0.6459	-1.4801
95% C.I. (upper bound)	-0.9092	2.0492	0.0946	1.2020	0.8535	0.0131

the patients. Therefore, the worse survival for patients with surgery within the first three months can be explained by the fact that 80% of these patients had at least one of the guideline criteria for surgery, including the presence of symptoms in 62% of them. It is indeed known that the survival of severe AI patients with symptoms is worse than for those without symptoms (Dujardin et al. 1999), as also observed in post operative survival (Klodas et al. 1997).

6 DISCUSSION

We have proposed a method for estimating the parameters of a promotion time cure model with mismeasured covariates. The SIMEX algorithm has several advantages that make it a very appealing method, especially in applied problems. First, since

it allows one to graphically represent the effect of the measurement error and of the correction on the bias, it helps justify the need for a correction. Secondly, its intuitive nature makes it appealing in applied problems, with users not necessarily familiar with the issue of measurement error. Finally, the scope of the correction can be tuned, making a conservative correction possible. Compared to the alternative approach introduced by Ma and Yin (2008), SIMEX can be applied to a broader class of models, since $\theta(\boldsymbol{x})$ can take any parametric form, including non-penalized fixed knot B-splines. Also, when using the SIMEX approach, the additive error can have any distribution, whereas Ma and Yin (2008) only study the normal case in detail. Moreover, the practical implementation of the SIMEX method is easier, since it only requires software to estimate the parameters of the model without measurement error.

We have proved that, under some conditions, our estimator is consistent and asymptotically normally distributed. Three simulation studies have shown the good properties of this estimator in samples of finite size, both in theoretical and in pratical settings. In the theoretical settings (when the cured subjects are observed), the SIMEX method and the method of Ma and Yin (2008) cannot be clearly discriminated on the basis of the bias, while in the practical setting, SIMEX is preferred for the lower measurement error variance, and the Ma and Yin method for larger variances. As far as the MSE is concerned, in all settings, SIMEX yields the lowest MSE when the measurement error variance is relatively large. For smaller values of the variance, the naive method outperforms both correction methods in terms of the MSE. Therefore, when the measurement error variance is small, it is better not to perform any correction. Moreover, the choice between the SIMEX and the Ma and Yin approaches could depend on the desired property of the obtained estimates: a pure reduction of the bias, or a decrease in the MSE.

ACKNOWLEDGEMENTS

The authors thank Professor Jean-Louis Vanoverschelde (Cardiovascular Research group, Institute of Experimental and Clinical Research, Université catholique de Louvain, Brussels, Belgium) for providing the data.

The authors also thank Professor Yanyuan Ma for providing her programs, which were used in our simulations.

Aurélie Bertrand, Catherine Legrand and Ingrid Van Keilegom were supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy) and by the contract "Projet d'Actions de Recherche Concertées" (ARC) 11/16-039 of the "Communauté française de Belgique" (granted by the "Académie universitaire Louvain"). Ingrid Van Keilegom was also supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 203650. Raymond J. Carroll was supported by a grant from the U.S. National Cancer Institute (R37-CA05730).

APPENDIX: PROOF OF THEOREM 2

For showing the asymptotics under this model, we follow the approach proposed by Zeng et al. (2006), using $G(\cdot) = \exp(-\cdot)$. In the case of no measurement error, the log-likelihood function is

$$\tilde{\ell}(\boldsymbol{\beta}, F) = I(Y < \infty) \left[\delta \log f + \delta \log \left\{-G'(\eta(\boldsymbol{X}^T\boldsymbol{\beta})F(Y))\eta(\boldsymbol{X}^T\boldsymbol{\beta})\right\} \right.$$
$$\left. +(1-\delta) \log G(\eta(\boldsymbol{X}^T\boldsymbol{\beta})F(Y))\right] + I(Y = \infty) \log G(\eta(\boldsymbol{X}^T\boldsymbol{\beta})).$$

Then, the true $(\boldsymbol{\beta}_{TRUE}, F_{TRUE})$ maximizes the expected log-likelihood $E\left\{\tilde{\ell}(\boldsymbol{\beta}, F)\right\}$ over the class $\mathcal{H} = \{(\boldsymbol{\beta}, F) : \boldsymbol{\beta} \in B, F \; cdf\}$, for some compact set B.

With measurement error, we define

$$\ell_\lambda(\boldsymbol{\beta}, F) = I(Y < \infty)\left[\delta \log f + \delta \log\left\{-G'(\eta(\boldsymbol{W}_\lambda^T\boldsymbol{\beta})F(Y))\eta(\boldsymbol{W}_\lambda^T\boldsymbol{\beta})\right\}\right.$$
$$\left.+(1-\delta)\log G(\eta(\boldsymbol{W}_\lambda^T\boldsymbol{\beta})F(Y))\right] + I(Y=\infty)\log G(\eta(\boldsymbol{W}_\lambda^T\boldsymbol{\beta})),$$

where $\boldsymbol{W}_\lambda = \boldsymbol{W} + \lambda^{1/2}\boldsymbol{U}^*$ and $\boldsymbol{U}^* \sim N(\boldsymbol{0}, V)$, and we suppose that $E\{\ell_\lambda(\boldsymbol{\beta}, F)\}$ has a unique maximizer $(\boldsymbol{\beta}_\lambda, F_\lambda)$. Therefore, we can follow exactly the same reasoning as in Zeng et al. (2006), replacing $\boldsymbol{X}$ by $\boldsymbol{W}_\lambda$ in all their calculations.

For a fixed λ and a fixed b, it follows from Equation (A.7) in Zeng et al. (2006) that

$$\begin{aligned}
&(\widehat{\boldsymbol{\beta}}_{\lambda,b} - \boldsymbol{\beta}_\lambda)^T\boldsymbol{h}_1 + \int_0^\infty h_2 \; d(\widehat{F}_{\lambda,b} - F_\lambda) \\
&= -(\boldsymbol{P}_n - \boldsymbol{P})\left\{\ell_{\lambda,\boldsymbol{\beta}}(\boldsymbol{\beta}_\lambda, F_\lambda)^T\Omega_{\lambda,\boldsymbol{\beta}}^{-1}(\boldsymbol{h}_1, h_2) + \ell_{\lambda,F}(\boldsymbol{\beta}_\lambda, F_\lambda)\left[\int \Omega_{\lambda,F}^{-1}(\boldsymbol{h}_1, h_2)dF_\lambda\right]\right\} \\
&\quad + o_p(n^{-1/2}) \\
&= n^{-1}\sum_{i=1}^n \psi_\lambda(T_i, \boldsymbol{W}_{i,\lambda,b}, \boldsymbol{h}_1, h_2) + o_p(n^{-1/2}),
\end{aligned}$$

uniformly over all $(\boldsymbol{h}_1, h_2) \in S_0$. Here, $\boldsymbol{P}_n[g(\delta, Y, \boldsymbol{X})] = n^{-1}\sum_{i=1}^n g(\delta_i, Y_i, \boldsymbol{X}_i)$ is the empirical measure of n iid observations, $\boldsymbol{P}[g(\delta, Y, \boldsymbol{X})] = E[g(\delta_i, Y_i, \boldsymbol{X}_i)]$ is the expectation, $\ell_{\lambda,\boldsymbol{\beta}}(\boldsymbol{\beta}, F)$ is the derivative of $\ell_\lambda(\boldsymbol{\beta}, F)$ with respect to $\boldsymbol{\beta}$, $\ell_{\lambda,F}(\boldsymbol{\beta}, F)[\int h_2 dF_\lambda]$ is the derivative of $\ell_\lambda(\boldsymbol{\beta}, F)$ along the path $(\boldsymbol{\beta}, F_{\epsilon,\lambda}(t) = F_\lambda(t) + \epsilon\int_0^t h_2(u)dF_\lambda(u))$, $\epsilon \in (-\epsilon_0, \epsilon_0)$ for a small constant ϵ_0, and $(\Omega_{\lambda,\boldsymbol{\beta}}^{-1}, \Omega_{\lambda,F}^{-1})$ is the inverse of the linear operator $(\Omega_{\lambda,\boldsymbol{\beta}}(\boldsymbol{h}_1, h_2), \Omega_{\lambda,F}(\boldsymbol{h}_1, h_2))$ defined in Appendix A.2 in Zeng et al. (2006). Finally,

$$S_0 = \left\{\boldsymbol{h}_1 \in \mathbb{R}^P : \|\boldsymbol{h}_1\| \le 1\right\} \times \left\{h_2 : \mathbb{R}^+ \to \mathbb{R} : \|h_2\|_V \le 1, \int_0^\infty h_2(y)dF_\lambda(y) = 0\right\}$$

with the total variation of h_2 defined as the supremum over all finite partitions $0 = t_1 < t_2 < \cdots < t_{m+1} = \infty$:

$$\|h_2\|_V = \sup_{0=t_1<t_2<\cdots<t_{m+1}=\infty} \sum_{i=1}^m |h_2(t_{i+1}) - h_2(t_i)|.$$

Of course, $E_\lambda\left\{\psi_\lambda(T, \boldsymbol{W}_\lambda, \boldsymbol{h}_1, h_2)\right\} = 0$, $\forall(\boldsymbol{h}_1, h_2) \in S_0$.

Next, for fixed λ, the class $\{(t, \boldsymbol{w}) \to \psi_\lambda(t, \boldsymbol{w}, \boldsymbol{h}_1, h_2) : (\boldsymbol{h}_1, h_2) \in S_0\}$ is Donsker (see Zeng et al. (2006)), and hence the class

$$\left\{(t, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_B) \to B^{-1}\textstyle\sum_{b=1}^{B}\psi_\lambda(t, \boldsymbol{w}_b, \boldsymbol{h}_1, h_2) : (\boldsymbol{h}_1, h_2) \in S_0\right\}$$

is also Donsker, since sums of Donsker classes are Donsker, see Van der Vaart and Wellner (1996), Lemma 2.10.6. It now follows that the process

$$\begin{aligned}
&\sqrt{n}\left\{(\widehat{\boldsymbol{\beta}}_\lambda - \boldsymbol{\beta}_\lambda)^T\boldsymbol{h}_1 + \int_0^\infty h_2\ d(\widehat{F}_\lambda - F_\lambda)\right\} \\
&= \sqrt{n}\left\{B^{-1}\sum_{b=1}^{B}(\widehat{\boldsymbol{\beta}}_{\lambda,b} - \boldsymbol{\beta}_\lambda)^T\boldsymbol{h}_1 + \int_0^\infty h_2\ d\left(\frac{1}{B}\sum_{b=1}^{B}(\widehat{F}_{\lambda,b} - F_\lambda)\right)\right\} \\
&= n^{-1/2}\sum_{i=1}^{n}B^{-1}\sum_{b=1}^{B}\psi_\lambda(T_i, \boldsymbol{W}_{i,\lambda,b}, \boldsymbol{h}_1, h_2) + o_p(1)
\end{aligned}$$

converges weakly to a zero-mean Gaussian process GP indexed by $(\boldsymbol{h}_1, h_2) \in S_0$ (see Zeng et al. (2006), under Equation (A.7)).

The covariance between $GP(\boldsymbol{h}_1, h_2)$ and $GP(\boldsymbol{h}_1^*, h_2^*)$ is

$$\begin{aligned}
&E\left[\left\{\ell_{\lambda,\boldsymbol{\beta}}(\boldsymbol{\beta}_\lambda, F_\lambda)^T\Omega_{\lambda,\boldsymbol{\beta}}^{-1}(\boldsymbol{h}_1, h_2) + \ell_{\lambda,F}(\boldsymbol{\beta}_\lambda, F_\lambda)\left[\int\Omega_{\lambda,F}^{-1}(\boldsymbol{h}_1, Q_{F_\lambda}(h_2))dF_\lambda\right]\right\}\right. \\
&\left.\times\left\{\ell_{\lambda,\boldsymbol{\beta}}(\boldsymbol{\beta}_\lambda, F_\lambda)^T\Omega_{\lambda,\boldsymbol{\beta}}^{-1}(\boldsymbol{h}_1^*, h_2^*) + \ell_{\lambda,F}(\boldsymbol{\beta}_\lambda, F_\lambda)\left[\int\Omega_{\lambda,F}^{-1}(\boldsymbol{h}_1^*, Q_{F_\lambda}(h_2^*))dF_\lambda\right]\right\}\right].
\end{aligned}$$

However, by noting that, for any h_2 in the class

$$S = \left\{\boldsymbol{h}_1 \in \mathbb{R}^P : \|\boldsymbol{h}_1\| \le 1\right\} \times \left\{h_2 : \mathbb{R}^+ \to \mathbb{R} : \|h_2\|_V \le 1\right\},$$

we have $\int_0^\infty h_2\ d(\widehat{F}_\lambda - F_\lambda) = \int_0^\infty g_2\ d(\widehat{F}_\lambda - F_\lambda)$, where $g_2 = h_2 - \int_0^\infty h_2\ dF_\lambda$, we can also consider this process as a process indexed by $(\boldsymbol{h}_1, h_2) \in S$.

Finally, we take a finite grid $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_K)^T$. The foregoing reasoning which was based on a single value of λ can be redone in exactly the same way for the vector

$(\lambda_1, \ldots, \lambda_K)$. At the end we have that

$$\sqrt{n}\left\{\begin{array}{c} (\widehat{\boldsymbol{\beta}}_{\lambda_1} - \boldsymbol{\beta}_{\lambda_1})^T \boldsymbol{h}_1 + \int_0^\infty h_2 \; d(\widehat{F}_{\lambda_1} - F_{\lambda_1}) \\ \vdots \\ (\widehat{\boldsymbol{\beta}}_{\lambda_K} - \boldsymbol{\beta}_{\lambda_K})^T \boldsymbol{h}_1 + \int_0^\infty h_2 \; d(\widehat{F}_{\lambda_K} - F_{\lambda_K}) \end{array}\right\}$$

converges to a K-dimensional Gaussian process of mean zero. The covariance function between the ith and jth components $(i, j = 1, \ldots, K)$ is

$$E\left[\left\{\ell_{\lambda_i,\boldsymbol{\beta}}(\boldsymbol{\beta}_{\lambda_i}, F_{\lambda_i})^T \Omega^{-1}_{\lambda_i,\boldsymbol{\beta}}(\boldsymbol{h}_1, h_2) + \ell_{\lambda_i,F}(\boldsymbol{\beta}_{\lambda_i}, F_{\lambda_i}) \left[\int \Omega^{-1}_{\lambda_i,F}(\boldsymbol{h}_1, Q_{F_{\lambda_i}}(h_2)) dF_{\lambda_i}\right]\right\}\right.$$
$$\left.\times \left\{\ell_{\lambda_j,\boldsymbol{\beta}}(\boldsymbol{\beta}_{\lambda_j}, F_{\lambda_j})^T \Omega^{-1}_{\lambda_j,\boldsymbol{\beta}}(\boldsymbol{h}_1^*, h_2^*) + \ell_{\lambda_j,F}(\boldsymbol{\beta}_{\lambda_j}, F_{\lambda_j}) \left[\int \Omega^{-1}_{\lambda_j,F}(\boldsymbol{h}_1^*, Q_{F_{\lambda_j}}(h_2^*)) dF_{\lambda_j}\right]\right\}\right]$$

We consider two particular cases.

1. Consider the class

$$\{(\boldsymbol{h}_1, h_2) \in S : \boldsymbol{h}_1 = (0, \ldots, 0, 1, 0, \ldots, 0) \text{ and } h_2 \equiv 0\}$$

   where $\boldsymbol{h}_1$ is a vector containing 1 at the jth position $(j = 1, \ldots, P)$ and 0 elsewhere. Then, we get the weak convergence of $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda}) - \boldsymbol{\beta}(\boldsymbol{\Lambda}))$ to a multivariate normal random variable of dimension PK, $N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$, where $\boldsymbol{\beta}(\boldsymbol{\Lambda}) = (\boldsymbol{\beta}_{\lambda_1}^T, \ldots, \boldsymbol{\beta}_{\lambda_K}^T)^T$.

2. Consider the class

$$\left\{(\boldsymbol{h}_1, h_2) \in S : \boldsymbol{h}_1 = \mathbf{0} \text{ and } h_2(\cdot) = I(\cdot \leq t), t \in \mathbb{R}^+\right\}.$$

   Then, we get the weak convergence of $\sqrt{n}\left\{\widehat{\boldsymbol{F}}(\boldsymbol{\Lambda}, t) - \boldsymbol{F}(\boldsymbol{\Lambda}, t)\right\}$ to a Gaussian process $\mathcal{G}$ indexed by $t \in \mathbb{R}^+$, where $\boldsymbol{F}(\boldsymbol{\Lambda}, t) = (F_{\lambda_1}(t), \ldots, F_{\lambda_K}(t))^T$.

We will now prove the asymptotic normality of $\widehat{\boldsymbol{\beta}}_{SIMEX}$. To this end, suppose that $\boldsymbol{\beta}_\lambda$ can be specified using a parametric model $\boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \lambda)$ depending on a vector of parameters $\boldsymbol{\gamma}_{\boldsymbol{\beta}}$. Assuming that $\boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \lambda)$ is the true extrapolation function, we have

that $\boldsymbol{\beta}_{TRUE} = \boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, -1)$ and $\widehat{\boldsymbol{\beta}}_{SIMEX} = \boldsymbol{g}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}}, -1)$, where $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}}$ solves (by the least squares estimation method)

$$\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \left\{ \boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda}) - \widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda}) \right\} = \mathbf{0}$$

and $\dot{\boldsymbol{g}}_{\beta}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})$ is the $PK \times \dim(\boldsymbol{\gamma}_{\boldsymbol{\beta}})$ matrix of partial derivatives of the elements of $\boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})$ with respect to the elements of $\boldsymbol{\gamma}_{\boldsymbol{\beta}}$. We then have that

$$\sqrt{n}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}} - \boldsymbol{\gamma}_{\boldsymbol{\beta}}) = \left\{ \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda}) \right\}^{-1} \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \sqrt{n}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda}) - \boldsymbol{\beta}(\boldsymbol{\Lambda})) + o_p(1)$$

converges to

$$\left\{ \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda}) \right\}^{-1} \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}).$$

Because $\widehat{\boldsymbol{\beta}}_{SIMEX} = \boldsymbol{g}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}}, -1)$ and $\boldsymbol{\beta}_{-1} = \boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, -1) = \boldsymbol{\beta}_{TRUE}$, using the Delta method, we have that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{SIMEX} - \boldsymbol{\beta}_{TRUE}) \longrightarrow \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, -1) \left\{ \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda}) \right\}^{-1} \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}).$$

Finally, we will show that $\sqrt{n}(\widehat{F}_{SIMEX} - F_{TRUE})$ converges weakly to a Gaussian process. For a fixed t, suppose that $F_\lambda(t)$ is determined by a parametric model $g_t(\boldsymbol{\gamma}_t, \lambda)$ depending on a parameter vector $\boldsymbol{\gamma}_t$. Under the assumption that this is the true extrapolation function, we have that $F_{TRUE}(t) = g_t(\boldsymbol{\gamma}_t, -1)$ and $\widehat{F}_{SIMEX}(t) = g_t(\widehat{\boldsymbol{\gamma}}_t, -1)$, where $\widehat{\boldsymbol{\gamma}}_t$ is a solution of

$$\dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \left\{ g_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda}) - \widehat{\boldsymbol{F}}(\boldsymbol{\Lambda}, t) \right\} = 0$$

and $\dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda}) = \partial g_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda}) / \partial \boldsymbol{\gamma}_t^T$. It now follows that

$$\begin{aligned} &\sqrt{n}(\widehat{\boldsymbol{\gamma}}_t - \boldsymbol{\gamma}_t) \\ &\quad = \left\{ \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda}) \right\}^{-1} \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \sqrt{n} \left\{ \widehat{\boldsymbol{F}}(\boldsymbol{\Lambda}, t) - \boldsymbol{F}(\boldsymbol{\Lambda}, t) \right\} + o_p(1) \end{aligned}$$

for all t, and hence the process $\sqrt{n}(\widehat{\boldsymbol{\gamma}}_t - \boldsymbol{\gamma}_t)$ (indexed by $t \in \mathbb{R}^+$) converges to the Gaussian process

$$\left\{\dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})\right\}^{-1} \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \mathcal{G}.$$

Since by definition $\widehat{F}_{SIMEX} = g_t(\widehat{\boldsymbol{\gamma}}_t, -1)$ and $F_{-1} = g_t(\boldsymbol{\gamma}_t, -1) = F_{TRUE}$, using the Delta method, we obtain that

$$\sqrt{n}(\widehat{F}_{SIMEX} - F_{TRUE}) \longrightarrow \dot{g}_t(\boldsymbol{\gamma}_t, -1)^T \left\{\dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})\right\}^{-1} \dot{g}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \mathcal{G}.$$

SUPPLEMENTARY MATERIALS

Proof of Theorem 1: Proof of the Theorem stating the consistency of the proposed estimator. (pdf file)

Simulation results: Simulation results for the settings of Sections 4.1 and 4.2, for a lower censoring rate (when $\mu = 1$). (pdf file)

References

Berkson, J. and Gage, R. (1952). "Survival curve for cancer patients following treatment." *Journal of the American Statistical Association*, 47:501–515.

Boag, J. (1949). "Maximum Likelihood Estimates of the Proportion of Patients Cured by Cancer Therapy." *Journal of the Royal Statistical Society, Series B*, 11:15–44.

Bonow, R. O., Carabello, B., de Leon, A. C., Edmunds, L. H., Fedderly, B. J., Freed, M. D., Gaasch, W. H., McKay, C. R., Nishimura, R. A., OGara, P. T., ORourke, R. A., Rahimtoola, S. H., Ritchie, J. L., Cheitlin, M. D., Eagle, K. A., Gardner, T. J., Garson, A., Gibbons, R. J., Russell, R. O., Ryan, T. J., and Smith, S. C. (1998). "Guidelines for the Management of Patients With Valvular Heart Disease: Executive Summary A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Patients With Valvular Heart Disease)." *Circulation*, 98:1949–1984.

Carroll, R., Kuchenhoff, H., Lombard, F., and Stefanski, L. (1996). "Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models." *Journal of the American Statistical Association*, 91:242–250.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. *Measurement error in nonlinear models: A modern perspective. 2nd edition*. Chapman and Hall/CRC (2006).

Carvalho Lopes, C. and Bolfarine, H. (2012). "Random effects in promotion time cure rate models." *Computational Statistics and Data Analysis*, 56:75–87.

Chen, M.-H., Ibrahim, J., and Sinha, D. (1999). "A New Bayesian Model For Survival Data With a Surviving Fraction." *Journal of the American Statistical Association*, 94:909–919.

Cook, J. R. and Stefanski, L. A. (1994). "Simulation-extrapolation in parametric measurement error models." *Journal of the American Statistical Association*, 89:1314–1328.

Cox, D. (1972). "Regression Models and Life-Tables." *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Crainiceanu, C., Ruppert, D., and Coresh, J. (2006). "Cox models with nonlinear effect of covariates measured with error: A case study of chronic kidney disease incidence." *Johns Hopkins University, Dept. of Biostatistics Working Papers*, 116.

Dujardin, K., Enriquez-Sarano, M., Schaff, H., Bailey, K., Seward, J., and Tajik, A. (1999). "Mortality and Morbidity of Aortic Regurgitation in Clinical Practice: A Long-Term Follow-Up Study." *Circulation*, 99:1851–1857.

Farewell, V. (1982). "The Use of Mixture Models for the Analysis of Survival Data with Long-Term Survivors." *Biometrics*, 38:1041–1046.

Greene, W. and Cai, J. (2004). "Measurement Error in Covariates in the Marginal Hazards Model for Multivariate Failure Time Data." *Biometrics*, 60:987–996.

Ibrahim, J., Chen, M.-H., and Sinha, D. (2001). "Bayesian Semiparametric Models for Survival Data With a Cure Fraction." *Biometrics*, 57:383–388.

Klodas, E., Enriquez-Sarano, M., Tajik, A., Mullany, C., Bailey, K., and Seward, J. (1997). "Optimizing Timing of Surgical Correction in Patients With Severe Aortic Regurgitation: Role of Symptoms." *Journal of the American College of Cardiology*, 30:746–752.

Kuk, A. and Chen, C.-H. (1992). "A mixture model combining logistic regression with proportional hazards regression." *Biometrika*, 79:531–541.

Li, Y. and Lin, X. (2003). "Functional Inference in Frailty Measurement Error Models for Clustered Survival Data Using the SIMEX Approach." *Journal of the American Statistical Association*, 98:191–203.

Lu, W. (2008). "Maximum likelihood estimation in the proportional hazards cure model." *Annals of the Institute of Statistical Mathematics*, 60:545–574.

Ma, Y. and Yin, G. (2008). "Cure Rate Model With Mismeasured Covariates Under Transformation." *Journal of the American Statistical Association*, 103:743–756.

Otterstad, J., Froeland, G., St John Sutton, M., and Holme, I. (1997). "Accuracy and reproducibility of biplane two-dimensional echocardiographic measurements of left ventricular dimensions and function." *European Heart Journal*, 18:507–513.

Peng, Y. (2003). "Fitting semiparametric cure models." *Computational Statistics and Data Analysis*, 41:481–490.

Peng, Y. and Dear, K. (2000). "A Nonparametric Mixture Model for Cure Rate Estimation." *Biometrics*, 56:237–243.

Stefanski, L. and Cook, J. (1995). "Simulation-Extrapolation: The Measurement Error Jackknife." *Journal of the American Statistical Association*, 90:1247–1256.

Sy, J. and Taylor, J. (2000). "Estimation in a Cox Promotional Hazards Cure Model." *Biometrics*, 56:227–236.

Taylor, J. (1995). "Semi-parametric Estimation in Failure Time Mixture Models." *Biometrics*, 51:899–907.

Tsodikov, A. (1998a). "Asymptotic efficiency of a proportional hazards model with cure." *Statistics and Probability Letters*, 39:237–244.

— (1998b). "A Proportional Hazards Model Taking Account of Long-Term Survivors." *Biometrics*, 54:1508–1516.

— (2001). "Estimation of Survival Based on Proportional Hazards When Cure is a Possibility." *Mathematical and Computer Modelling*, 33:1227–1236.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003). "Estimating Cure Rates From Survival Data: An Alternative to Two-Component Mixture Models." *Journal of the American Statistical Association*, 98:1063–1078.

Vahanian, A., Baumgartner, H., Bax, J., Butchart, E., Dion, R., Filippatos, G., Flachskampf, F., Hall, R., Iung, B., Kasprzak, J., Nataf, P., Tornos, P., Torracca, L., and Wenink, A. (2007). "Guidelines on the management of valvular heart disease: The Task Force on the Management of Valvular Heart Disease of the European Society of Cardiology." *European Heart Journal*, 28:230–268.

Van der Vaart, A. and Wellner, J. *Weak Convergence and Empirical Processes*. New-York: Springer-Verlag (1996).

Yakovlev, A. and Tsodikov, A. *Stochastic Models of Tumor Latency and Their Biostatistical Applications*. World Scientific (1996).

Zeng, D., Yin, G., and Ibrahim, J. (2006). "Semiparametric Transformation Models for Survival Data with a Cure Fraction." *Journal of the American Statistical Association*, 101:670–684.