Survival models with a cure fraction and mismeasured covariates

A thesis submitted in partial fulfillment of the requirements for the degree of Docteur en Sciences, Université catholique de Louvain, July 2017

Aurélie Bertrand

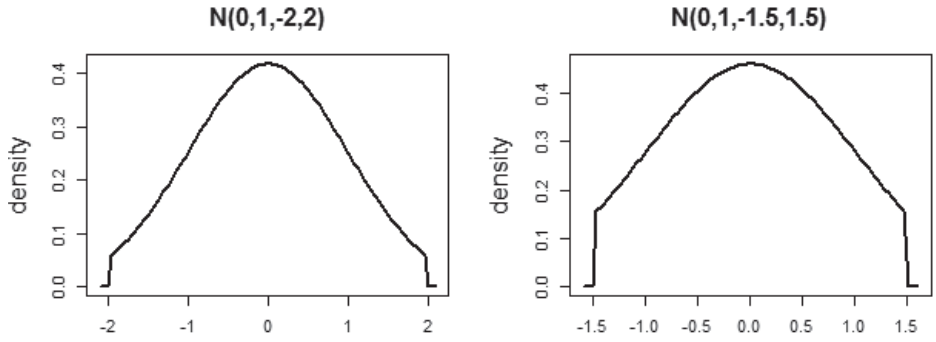
Thesis committee:

Catherine Legrand, Supervisor (Université catholique de Louvain)
Ingrid Van Keilegom, Supervisor (Katholieke Universiteit Leuven, Université catholique de Louvain)
Raymond J. Carroll (Texas A&M University)
Anouar El Ghouch, President (Université catholique de Louvain)
Paul Janssen (Hasselt University)
Philippe Lambert (Université de Liège, Université catholique de Louvain)

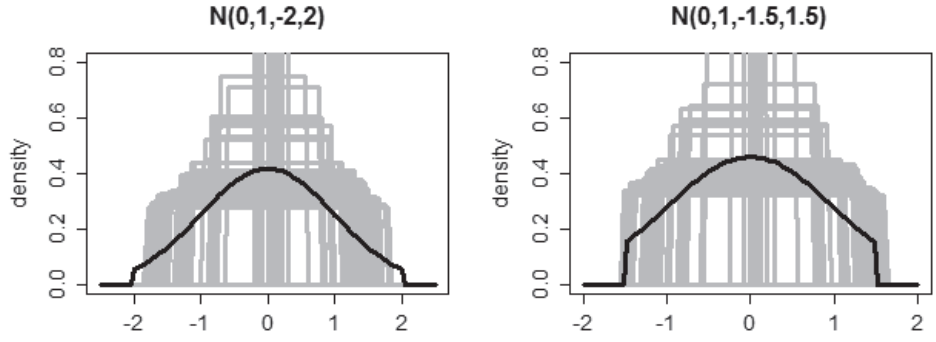
Erratum

In Sections 4.3 and 4.4, the Gaussian distributions actually used for data generation are not the ones reported in the manuscript. N(-1,4,-5,3) and N(-1,4,-4,2) should be read as N(0,1,-2,2) and N(0,1,-1.5,1.5), respectively. Below are the consequences in these sections.

- In Tables 4.4, 4.6 and 4.8, the correct theoretical values of a and b are 4 and -2 for the N$(0, 1, -2, 2)$ distribution, and 3 and -1.5 for the N$(0, 1, -1.5, 1.5)$ distribution.
- The values of NSR considered for the Gaussian distributions are 0, 0.5, 1 and 2 (instead of 0, 0.25, 0.5 and 0.75). Consequently, the results for these two distributions cannot be directly compared to the results obtained with the other distributions.
- In Figure 4.2, the two graphs related to the Gaussian distributions should be replaced by:



- In Figure 4.3, the two graphs related to the Gaussian distributions should be replaced by:



The results regarding the estimation of the density of X are hence actually better than reported.

Acknowledgments

I would like to thank all the people who supported me during this journey.

First of all, I wish to express my gratitude to my supervisors, Catherine Legrand and Ingrid Van Keilegom. Catherine, Ingrid, thank you for your guidance, your support, your availability, your ideas and, maybe the most important for me, your trust during these six years. It has been a very enriching experience to benefit from your complementary areas of expertise.

I would also like to thank the thesis committee members, Raymond J. Carroll, Anouar El Ghouch, Paul Janssen and Philippe Lambert: thank you for your interest, as well as your constructive comments during my private defense that helped improve this thesis. Paul, Philippe, thank you for your role in my PhD follow-up committee and for your interest in my work since the beginning. Ray, thank you for the fruitful discussions we had about SIMEX and measurement error during your stay in Belgium in 2012, as well as for your enthusiasm during my private defense.

Thanks to the whole ISBA and LSBA administrative staff, Nancy Guillaume, Maguy Hanon, Sophie Malali, Nadja Peiffer and Tatiana Regout, for their help in the daily life at the institute. Nancy, thank you for always taking the time to answer the (many) practical questions that can appear during a PhD thesis.

I also thank the whole SMCS team for giving me the opportunity to work on various consulting projects and to take part in several of their trainings. Alain, Catherine, Céline, Christian, Nathalie, thank you for trusting me and for sharing your experience with me. Special thanks to Alain, for his help with

R, Matlab and Fortran in the researchers' room!

This section wouldn't be complete without a special thank you to my colleagues who greatly contributed to making these six years (my thesis, but also the teaching activities, the conferences, the coffee breaks, the lunch breaks, the Friday drinks, the music quizzes, the ladies nights and even my weekends and holidays) such a great experience: Alain, Alexandre, Anna, Anne, Antoine, Baptiste, Benjamin, Catherine, Cédric, Céline, Cheikh, Edward, Federico, Florian, François, Gauthier, Hélène, Hugues, Joris, Josefine, Lieven, Maïlis, Majda, Manon, Marco, Mathieu, Mickaël, Nathalie Le., Nathalie Lu., Nathan, Nicolas, Rebecca, Samuel, Sylvie, Vincent.

I am also grateful to my friends, especially Anne-So, Christophe, Kristel and Sophie, for daring to mix with a statistician and, more seriously, for their encouragement.

Finalement, je remercie mes parents et mon frère d'avoir partagé leur passion des nombres avec moi, et de m'avoir soutenue de façon indéfectible pendant toute la durée de mes études.

Contents

CHAPTER 1

Introduction

In this thesis, we are interested in statistical methods to deal with **survival data** with a **cure fraction** and **mismeasured covariates**.

The characteristic of survival data is that the quantity of interest is the *length of time* until a well-defined event occurs. When some of the units of the population of interest will *never experience this event*, we call them the cure fraction. Finally, mismeasured covariates are variables that could influence the length of time of interest and that are *not necessarily measured very precisely*.

Data with such characteristics can appear in very different fields. For example, in medicine, if we are interested in aortic insufficiency, a cardiovascular disease which can be lethal, a cardiologist may wonder whether patients with a high initial value of ejection fraction (the fraction of blood that leaves the heart each time it contracts) tend to survive better to their heart disease. In an economic survey about people who lost their job after having worked for more than ten years, the researchers may want to identify the factors that influence the time spent in unemployment, studying for example the impact of the last income as reported by the surveyed individuals. The quality engineer in a factory may wish to predict the time until an electronic component of a machine breaks down before its planned renewal. The temperature and the pressure the component is subject to are known to influence that time. Common features of these three examples are the presence of a cure fraction and the fact that (some of) the covariates of interest are potentially measured with error (the ejection fraction computed based on an echocardiography, the self-reported income, and

the temperature and pressure inside a machine).

The remainder of this chapter is dedicated to a general presentation of these three topics (survival analysis without and with a cure fraction, measurement error in covariates), as well as to an outline of the following chapters.

1.1 Survival analysis

Survival analysis aims at describing and modeling time-to-event data, i.e. data about the time between a starting point and a predefined event of interest. This time is often called the *survival time*, the *event time* or the *failure time*. The event under consideration is not restricted to be the death of a patient: it can be, in a medical context, the recurrence of a disease or a tumor, or the recovery after a disease, among others. Moreover, survival analysis has applications in other fields, sometimes under different names: in sociology (e.g., the time until a former prisoner is rearrested), economics (*duration analysis*: e.g., the time until an unemployed person finds a new job), or engineering (*reliability analysis*: e.g., the time until a machine breaks down).

All the concepts and methods used in survival analysis are detailed in numerous textbooks, including Therneau and Grambsch (2000), Moore (2016) and Klein and Moeschberger (2003); in this section, we shortly review the notions that will be used in the forthcoming chapters.

1.1.1 Distinguishing features of survival data

Survival data have two main characteristics. First, a survival time, which we will denote by T, is a *positive* (continuous) random variable. Secondly, and more importantly, survival data typically exhibit some specific features that prevent the use of classical statistical techniques: *censoring* and *truncation*.

Censoring happens when the exact survival time of some individuals is not known. *Right censoring* occurs when only a lower bound for the survival time, called the *censoring time*, is known. That can happen, for example, if a subject is still alive at the end of the study. His/her survival time is known to be at least the time from the beginning of the follow up to the end of the study. More formally, right censoring can take one of the following three forms: type I, type II and random right censoring. *Type I right censoring* is found when all individuals enter the study at the same time point, and the study ends at a fixed date: all censored people have the same value of censoring time, c. In *type II right censoring*, all individuals are monitored from a common

starting point until a predefined number of them have experienced the event of interest. Here again, all censored individuals share a common censoring time c. Finally, *random right censoring* describes the case in which the censoring time is a random variable: this corresponds to the situation in which all people can enter the study at different points in time, as well as leave it: censoring can be due to a loss of follow up, a death not related to the event of interest, or the end of the study.

If we denote the random censoring time by C, the observed data consists in the *observed time*,

$$Y = \min(T, C),$$

and the *censoring indicator*,

$$\delta = I(T \leq C).$$

Truncation is found when the subjects have either to meet a condition or to have experienced the event in order to be included in the analysis. Compared with censoring, in which some information about the censored subjects is still known, truncation is more severe, as it implies that some individuals do not appear in the dataset at all.

In this work, we only consider random right censoring without truncation, which corresponds to many common situations in practice. We make the usual assumptions that T and C are independent (given the covariates, if present), and that the censoring is *uninformative*, i.e., the distribution of the censoring times does not depend on the parameters appearing in the survival distribution.

1.1.2 Survival quantities

While, in usual statistical analysis, the mathematical functions of interest are mainly the probability density function $f(t)$ and the cumulative distribution function $F(t) = P(T \leq t)$, the techniques used in survival analysis often focus on the survival function, the hazard function and the cumulative hazard function. The *survival function*,

$$S(t) = P(T > t) = 1 - F(t), \quad t \geq 0,$$

takes values in $[0, 1]$. The *hazard function* is

$$h(t) = \lim_{\epsilon \to 0} \frac{P(t \leq T < t + \epsilon | T \geq t)}{\epsilon}, \quad t \geq 0,$$

where $h(t) \geq 0$. For a small $\epsilon > 0$, $h(t)\epsilon$ is approximately the instantaneous risk of experiencing the event in $[t, t+\epsilon]$, if the individual has survived up to time t. Finally, the *cumulative hazard function* is

$$H(t) = \int_0^t h(u)du, \quad t \geq 0.$$

These functions are linked together, and once one of them is known, the other ones can be recovered uniquely: $S(t) = \exp\{-H(t)\}$, $f(t) = -dS(t)/dt$.

1.1.3 Survival likelihood

Censoring has consequences on the building of the likelihood. In the case of data with independent and uninformative right censoring and no truncation, the observed data is (Y_i, δ_i) for $i = 1, \ldots, n$. The likelihood is then

$$\begin{aligned} L &= \prod_{i=1}^{n} \{f(Y_i)\}^{\delta_i} \{S(Y_i)\}^{1-\delta_i} \\ &= \prod_{i=1}^{n} \{h(Y_i)\}^{\delta_i} S(Y_i), \end{aligned} \tag{1.1}$$

i.e., an individual who has experienced the event contributes $f(Y_i)$, while the contribution of a censored subject is $S(Y_i)$.

1.1.4 Estimation and modeling of survival quantities

In classical survival analysis, it is assumed that all individuals will experience the event of interest, so that $\lim_{t\to\infty} S(t) = 0$. Under this assumption, many statistical tools were developed for estimation and modeling.

A well-known nonparametric estimator of the survival function for right-censored data is the *Kaplan-Meier estimator* (Kaplan and Meier, 1958):

$$\widehat{S}(t) = \prod_{j:Y_{(j)} \leq t} \frac{R(Y_{(j)}) - d_{(j)}}{R(Y_{(j)})},$$

where $Y_{(j)}$ is the jth ordered survival time, $d_{(j)}$ is the corresponding number of events and $R(Y_{(j)})$ is the number of subjects at risk just before time $Y_{(j)}$.

As far as modeling is concerned, parametric, semiparametric and nonparametric methods exist in the literature. They allow one to incorporate a vector of covariates $\boldsymbol{X} = (X_1, \ldots, X_P)^T$. A first simple approach consists in assuming a specific distribution for T and letting its parameters depend on the

covariates. The distributions commonly used in this context include the exponential, Weibull, logistic, log-Normal and Gompertz distributions. However, other modeling strategies have been designed for survival data.

Proportional hazards models

The most widely used survival model is the *semiparametric proportional hazards model* (Cox, 1972), which assumes that

$$h(t|\boldsymbol{X}) = h_0(t)\exp(\boldsymbol{X}^T\boldsymbol{\beta}), \tag{1.2}$$

where $h_0(t)$, the baseline hazard function, is left unspecified. The parameters $\boldsymbol{\beta}$ can be estimated by maximizing the *partial likelihood* of the model, a profile likelihood in which the unknown $h_0(\cdot)$ does not appear. This model can also be used while assuming a parametric form for this hazard: this approach leads to *parametric proportional hazards models*.

All models of this family possess a proportional hazards structure: the ratio of the hazards of two subjects characterized by the covariates $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ is

$$\frac{h(t|\boldsymbol{X}_i)}{h(t|\boldsymbol{X}_j)} = \frac{\exp(\boldsymbol{X}_i^T\boldsymbol{\beta})}{\exp(\boldsymbol{X}_j^T\boldsymbol{\beta})},$$

which is independent of t. The effect of the covariates is a *multiplication* of the hazard at all times.

Other models

Other families of survival models include *proportional odds models*, as well as *accelerated failure time (AFT) models*. The latter can be expressed as

$$h(t|\boldsymbol{X}) = h_0\left\{\exp(\beta_0 + \boldsymbol{X}^T\boldsymbol{\beta})t\right\}\exp(\beta_0 + \boldsymbol{X}^T\boldsymbol{\beta}), \tag{1.3}$$

where the baseline hazard function $h_0(\cdot)$ can be modeled parametrically. In this model, the hazards are not proportional in general; the name of the model stems from the following property: if $S_0(t) = \exp\{-\int_0^t h_0(u)du\}$ is the survival function of a baseline individual (for which $\boldsymbol{X} = \boldsymbol{0}$), then

$$S(t|\boldsymbol{X}) = S_0\left\{t\exp(\beta_0 + \boldsymbol{X}^T\boldsymbol{\beta})\right\}.$$

The effect of the covariates is to change the time scale, to make the time pass more (or less) quickly. Model (1.3) can be rewritten as a log-linear model for

the survival time:

$$\log T = -\beta_0 - \boldsymbol{X}^T\boldsymbol{\beta} + \sigma\epsilon,$$

where σ is a scale parameter, ϵ is an error term and β_0 and $\boldsymbol{\beta}$ are the same as in Equation (1.3). Equivalently, we have that $T = \exp(-\beta_0 - \boldsymbol{X}^T\boldsymbol{\beta})\exp(\sigma\epsilon)$. The link between both expressions can be derived by defining $T_0 = \exp(\sigma\epsilon)$ and noticing that the survival function of $T = T_0\exp(-\beta_0 - \boldsymbol{X}^T\boldsymbol{\beta})$ is $S(t|\boldsymbol{X}) = P(T > t|\boldsymbol{X}) = S_0\left\{t\exp(\beta_0 + \boldsymbol{X}^T\boldsymbol{\beta})\right\}$, with $S_0(\cdot)$ the survival function of T_0.

1.2 Survival analysis with a cure fraction

In practice, there are many survival contexts in which it is known that some individuals will never experience the event of interest: this is the case, for example, when interest lies in the time until the recurrence of a cancer, the rearrest of a former prisoner or the start of a new job after a period of unemployment. In such situations, the population consists of two groups of subjects: the *susceptible* ones, who will one day experience the event of interest, and the *cured* ones, also called *long-term survivors*, for whom $T = \infty$. Mathematically speaking, the presence of cured individuals implies an improper survival function for the whole population, i.e.,

$$\lim_{t\to\infty} S(t) > 0,$$

and $1 - p = \lim_{t\to\infty} S(t)$ is the probability of being cured, called the *cure rate*. Since classical survival models assume that $\lim_{t\to\infty} S(t) = 0$ (as seen in the previous section), they should not be used when cure is a possibility. *Cure models*, also called *long-term survival models*, are survival models that were specifically developed for this context.

Compared to classical survival models, an additional difficulty appears here, due to the concomitant presence of a cure fraction and random right censoring. While all subjects having experienced the event of interest are known to be susceptible, the cured individuals cannot be identified: their survival times are censored, as is also the case for some susceptible ones. In other words, unlike the censoring indicator δ, the cure status $I(T < \infty)$ is unknown for some subjects. This will have consequences on model identifiability, as explained in section 1.2.3.

1.2.1 Likelihood

In the presence of cured individuals, the survival likelihood (1.1) presented in the previous section must be adapted in order to take into account the possibility of infinite event times. When such times can be identified, for example when the assumptions that the follow-up time is infinite and that some individuals will never experience right-censoring can be made (see Section 1.2.3 for a related discussion), three types of contributions can be identified: $f(Y_i)$ when $\delta_i = 1$ (in this case, it is known that $Y_i = T_i < \infty$), $S(Y_i)$ when $\delta_i = 0$ and $Y_i = C_i < \infty$, and $S(\infty)$ when $\delta_i = 0$ and $Y_i = C_i = T_i = \infty$. The likelihood then becomes

$$\prod_{i=1}^{n} \{f(Y_i)\}^{\delta_i} \{S(Y_i)\}^{(1-\delta_i)I(Y_i<\infty)} \{S(\infty)\}^{(1-\delta_i)I(Y_i=\infty)}. \tag{1.4}$$

1.2.2 Cure models

Two main classes of cure models have initially been developed: mixture cure models and promotion time cure models. In addition to that, there also exist some extensions of these models, as well as generalizations that encompass both families as special cases.

Mixture cure models

Mixture cure models assume that

$$S(t|\boldsymbol{X}_1, \boldsymbol{X}_2) = p(\boldsymbol{X}_1)S_u(t|\boldsymbol{X}_2) + 1 - p(\boldsymbol{X}_1), \tag{1.5}$$

where $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are two vectors of covariates (possibly sharing some components), $S_u(t|\boldsymbol{X}_2)$ is the proper survival function of the susceptible individuals, called the *latency* part of the model, while $p(\boldsymbol{X}_1)$ is the *incidence*, the probability of being susceptible. The first papers about these models assumed a *parametric* distribution (e.g., lognormal, exponential, piecewise exponential, Weibull) of the survival time for the susceptible subjects. Examples can be found in Boag (1949), Berkson and Gage (1952) and Farewell (1982).

Many papers about *semiparametric* mixture cure models then appeared, that assumed a semi- or a nonparametric form for the baseline survival function $S_{u0}(t)$ of the susceptible subjects. In Taylor (1995), the incidence part is modeled through a logistic regression, while the Kaplan-Meier estimator without any covariate is used for the latency. Kuk and Chen (1992), Peng and Dear (2000) and Sy and Taylor (2000) use a proportional hazards model with covariates for the survival of the non-cured individuals. Note that this model

possesses a proportional hazards structure for the susceptible individuals, but not for the whole population. For the same model, Corbière et al. (2009) propose a flexible estimation method based on splines for the baseline hazard of the susceptibles. Other examples of combinations of models for the incidence and the latency parts include a logit, probit or complementary log-log regression coupled with a proportional hazards model (Lam et al., 2005), a logistic/generalized F distribution model (Peng et al., 1998), a logistic/AFT model (Li and Taylor, 2002) or a logistic/semiparametric linear transformation model (Lu and Ying, 2004).

A fully *nonparametric* approach is possible too: in Lopez-Cheda et al. (2017), both the latency and the incidence parts are modeled nonparametrically and can include covariates. Xu and Peng (2014) propose a nonparametric estimator of the cure rate that allows for covariates.

Promotion time cure models

Promotion time cure models, also called *bounded cumulative hazard models*, or *proportional hazards cure models*, were introduced by Yakovlev et al. (1993) and Yakovlev and Tsodikov (1996). These models do not incorporate explicitly the division of the population into the susceptible and the cured groups. The existence of the cured individuals is taken into account by assuming an improper survival function for the whole population:

$$S(t|\boldsymbol{X}) = \exp\left\{-\theta(\boldsymbol{X})F(t)\right\}, \tag{1.6}$$

where $F(\cdot)$ is a proper baseline cumulative distribution function (cdf) and $\theta(\boldsymbol{X})$ is a known link function with an intercept. One often chooses $\theta(\boldsymbol{X}) = \exp(\boldsymbol{X}^T\boldsymbol{\beta})$, where the first component of the covariate $\boldsymbol{X}$ is supposed to be 1, in order to include an intercept in the model. Note that the classical Cox model (1.2) 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. The baseline cdf can be parametrically specified or left unknown: the latter solution yields the *semiparametric promotion time cure model*, which will be considered in the next two chapters.

With these models, the probability of being cured is

$$\lim_{t\to\infty} S(t|\boldsymbol{X}) = \exp\left\{-\theta(\boldsymbol{X})\right\}.$$

Contrary to mixture cure models, promotion time cure models (1.6) exhibit the proportional hazards property mentioned in the previous section:

$$\frac{h(t|\boldsymbol{X}_i)}{h(t|\boldsymbol{X}_j)} = \frac{\theta(\boldsymbol{X}_i)f(t)}{\theta(\boldsymbol{X}_j)f(t)} = \frac{\theta(\boldsymbol{X}_i)}{\theta(\boldsymbol{X}_j)},$$

where $f(t) = dF(t)/dt$ is the baseline density function.

These models have been studied mainly in the Bayesian literature, because the posterior distribution of the regression parameters is often proper when using noninformative improper priors, an attractive property in a Bayesian setting (Yin and Ibrahim, 2005). This is not the case with mixture cure models.

A mathematical and a biological derivations exist for this model. The former is due to Tsodikov (1998): the promotion time cure model is obtained from the proportional hazards model (1.2), by constraining the cumulative hazard to be bounded. This corresponds to constraining the survival function to be improper. The biological interpretation of the model can be found in Chen et al. (1999). When studying the relapse of some types of cancer, we can assume that individual i has N_i cells which can metastasize: N_i is an unobservable latent variable whose distribution is assumed to be $Poi(\theta(\boldsymbol{X}_i))$. The individuals with $N_i = 0$ are cured. The kth cell ($k = 1, \ldots, N_i$) takes a time Z_{ik} (the *promotion time*) to produce a detectable tumor mass. Conditionally on N_i, the variables Z_{ik}, $k = 1, \ldots, N_i$ are mutually independent, independent of N_i and have a common cdf $F(t)$. The time until the relapse of the cancer, which is the observed event time, is then $T_i = \min(Z_1, \ldots, Z_{N_i})$. In this case, the survival function of T is

$$\begin{aligned}
S(t|\boldsymbol{X}) &= P(N = 0|\boldsymbol{X}) + \sum_{j=1}^{\infty} P(T_1 > t, \ldots, T_j > t|N = j, \boldsymbol{X})P(N = j|\boldsymbol{X}) \\
&= \exp\{-\theta(\boldsymbol{X})\} + \sum_{j=1}^{\infty} \{1 - F(t)\}^j \frac{\theta(\boldsymbol{X})^j \exp\{-\theta(\boldsymbol{X})\}}{j!} \\
&= \exp\{-\theta(\boldsymbol{X})\} + \exp\left[\{1 - F(t)\}\theta(\boldsymbol{X}) - \theta(\boldsymbol{X})\right] - \exp\{-\theta(\boldsymbol{X})\} \\
&= \exp\{-\theta(\boldsymbol{X})F(t)\},
\end{aligned}$$

which is the survival function of the promotion time cure model.

Extensions Model (1.6) can be generalized in different ways. A first example, which generalizes other earlier approaches, can be found in Tsodikov (2002). The author introduces *extended hazard models*, cure models with a non-linear transformation model (including covariates) for the baseline cdf. Contrary to the covariates appearing in Equation (1.6), these new covariates

affect the short-term survival, but not the cure rate. Including two sets of covariates hence allows one to distinguish between the effect on the cure rate, called *long-term effect*, and the effect on the survival, called *short-term effect*. However, these models do not possess the proportional hazards property, and some restrictions on the covariates may be required for identifiability (see Bremhorst and Lambert (2016) for an identifiability result in one such model).

Zeng et al. (2006) study semiparametric transformation models that relax the assumption of mutual independence between the promotion times Z_{ik}, conditionally on the number of tumor cells N_i, by introducing a subject-specific frailty term ξ_i. The promotion times are assumed to be independent, conditionally on both N_i and ξ_i.

Unified approaches

Several general classes of cure models have been developed, which encompass both mixture and promotion time cure models as special cases. We briefly review some of them here.

Yin and Ibrahim (2005) propose a new class of cure models through a Box-Cox transformation of the survival function of the population. Two special cases are the mixture cure models and the promotion time cure models with covariates in the baseline cdf.

The extension of Cooner et al. (2007) is motivated through the biological interpretation of the promotion time cure model: in their paper, the distribution of the number of cells that can metastasize, N, is not necessarily Poisson-distributed, and the event is assumed to occur when r cells (and not necessarily one) have metastasized. r can be a constant, a function of N or a random variable. The mixture cure model is obtained by considering a Bernoulli-distributed N. Gu et al. (2011) study a proportional odds cure model, which can be obtained by using a geometric distribution for N.

Tournoud and Ecochard (2008) also derive their proposed class of models by using the biological interpretation of the promotion time cure model. They consider several distributions for N (all with a point mass at zero, to allow for cured individuals): not only Poisson, but also Bernoulli (yielding the mixture cure model), negative binomial and compound Poisson. Covariates can be included in the mean of N as well as in the distribution of the promotion time of each cell.

Ortega et al. (2014) study a flexible parametric class of models: they assume that N follows a negative binomial distribution and that the baseline cdf is the one of the generalized gamma distribution.

1.2.3 Identifiability issues

In a semiparametric cure model, (part of) the latency component $S_u(\cdot)$ of the mixture cure model (1.5) or the baseline cdf $F(\cdot)$ of the promotion time cure model (1.6) is left unspecified and estimated nonparametrically. In this case, some additional information is required for identifiability.

Zeng et al. (2006) explain that, in order for the semiparametric promotion time cure model to be identifiable in the regression parameters and in F, we need a threshold τ, called the *cure threshold*, such that all censored individuals with a censoring time larger than this threshold are treated as known to be cured, i.e., $T_i = C_i = Y_i = \infty$. However, since the estimation method introduced in the next chapter forces the estimated baseline cdf to be 1 beyond the largest observed failure time, this amounts to considering that no event can occur after this time: the cure threshold is then determined to be this largest event time.

In the mixture cure model, the equivalent approach is used by Taylor (1995) and Sy and Taylor (2000) when estimating nonparametrically (part of) the latency component: the survival function (the baseline survival function, respectively) is constrained to reach 0 at the largest observed event time. This is the *zero-tail constraint*. Taylor (1995) explains that such a constraint seems natural in contexts in which a cure model is justified, i.e., when cured subjects are known to exist and the follow-up is sufficiently long after the largest event time. This latter situation corresponds to the theoretical assumption (mentioned in Section 1.2.1) that $P(C = \infty|\boldsymbol{X}) > 0$ (Zeng et al., 2006).

An alternative to this approach is to model (part of) the baseline distribution parametrically: in such a case, a tail constraint is not needed any more. Peng (2003) introduces two *tail-completion methods* for the mixture cure model: the tail (after the largest event time) of the baseline survival function is modeled parametrically, so that it is allowed to decrease smoothly to 0.

As far as the promotion time cure model is concerned, an example of such an approach, in a Bayesian setting, is found in Ibrahim et al. (2001), where the baseline cdf is estimated by a piecewise constant hazard function. They allow some degree of parametricity (between full parametricity and full nonparametricity) in the right tail of the baseline cdf, by including a smoothing parameter in the model.

1.3 Model estimation with measurement error in the covariates

The topic of model estimation with measurement error in (some of) the covariates has been studied a lot in the literature: first in linear models (see, for example, Reiersøl (1950)), then in nonlinear models. Carroll et al. (2006) and Schennach (2016) provide a review of the existing methods in the latter case, while Yan (2014) focus on survival models.

We assume that $P + R$ covariates are available for inclusion in the model. The problem with measurement error is that the exact value of P covariates, $\boldsymbol{X} = (X_1, \ldots, X_P)^T$, cannot be observed; only a contaminated version of them, $\boldsymbol{W} = (W_1, \ldots, W_P)$, is available. The other R covariates, $\boldsymbol{Z} = (Z_1, \ldots, Z_R)^T$, are observed without error. The expression *measurement error* is usually associated with continuous variables, while the term *misclassification* is preferred for non-continuous variables. This latter case is not studied here.

This measurement error can have different causes, among which an inaccurate measurement device (for example, a thermometer), an imprecise recording (as can happen in self-report in surveys), or temporal variation in the covariate (as is the case with the blood pressure).

As explained by Carroll et al. (2006), ignoring this measurement error when estimating a model and making inference can have three consequences. First, the features of the data (i.e., the shape of the relationship between the covariate and the response, for nonlinear models) are less obvious. Second, the estimators become biased, since the relationship between the response and $(\boldsymbol{X}, \boldsymbol{Z})$ is unlikely to be the same as the one between the response and $(\boldsymbol{W}, \boldsymbol{Z})$. The sign and size of this bias depends on the model of interest and on the joint distribution of the error and of the other variables. In particular, the bias is not always an attenuation towards zero: this result is true in simple linear regression, but does not hold anymore in multivariate linear regression or in nonlinear models (Schennach, 2016). The bias in the proportional hazards model (Cox, 1972) is studied by Hughes (1993). Finally, if the variability in $\boldsymbol{W}$ is larger than the one in $\boldsymbol{X}$, the power for detecting relationships between variables is lower than expected.

As a consequence, taking the measurement error into account is essential to obtain valid estimates and inference. However, note that it is not necessary for prediction: in that case, the link between $\boldsymbol{W}$ and the response is the one of interest, since only $\boldsymbol{W}$ will be observed in the future. Most techniques in the literature about measurement error focus on a reduction in bias. In general, the variance of a corrected estimator is larger than that of the naive estimator, i.e., the estimator which does not take the measurement error into account.

As a result, except in some situations where the bias is the lone criterion of interest, a comparison of estimators in terms of the mean squared error (MSE) is useful.

1.3.1 Measurement error models

Error versus regression calibration models

There are two families of models for measurement error (Carroll et al., 2006): *error models*, and *regression calibration models*. The former consist in modeling the conditional distribution of $\boldsymbol{W}$ given $(\boldsymbol{X}, \boldsymbol{Z})$, while in the latter, the conditional distribution of $\boldsymbol{X}$ given $(\boldsymbol{W}, \boldsymbol{Z})$ is modeled.

Among the error models, the most famous one is the *classical measurement error model*:

$$\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U} \tag{1.7}$$

where $\boldsymbol{U} = (U_1, \ldots, U_P)^T$ is the (additive) error, with $E(\boldsymbol{U}|\boldsymbol{X}) = \boldsymbol{0}$ and a conditional variance-covariance matrix $V(\boldsymbol{U}|\boldsymbol{X})$ which may vary among individuals. This version of the model is what Schennach (2016) calls the *weakly classical model*; in the *strongly classical model*, $\boldsymbol{U}$ and $\boldsymbol{X}$ are assumed to be independent. This could be a reasonable model for the systolic blood pressure: W is the observed value, presumably different for each individual, X is the long-term value, and the variability of W is larger than the variability of X.

In constrast to this first model, the *Berkson model*, which is also widely used, is a regression calibration model, since it assumes that

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

where $E(\boldsymbol{U}|\boldsymbol{W}) = \boldsymbol{0}$, or more strongly, $\boldsymbol{U}$ and $\boldsymbol{W}$ are independent. This model suits covariates for which the variability in the true value is larger than the variability of the observed value. This can be used, for example, in an epidemiological study where past unmeasured radiation exposure is to be recovered from other available data: each individual is assigned a value W based on the value of other explanatory variables (for example, the geographical location at that time, age, diet habits), so that subjects sharing the same value of these variables are given exactly the same value of W. This is also a reasonable model for control variables which are set to a predefined value in an experiment: the actual value is likely to (slightly) differ from this predefined value (Schennach, 2016).

These models are only two special cases of possibly more general models for the measurement error: in practice, there can be a mixture of classical and Berkson

errors, more complex error structures (for example incorporating biases), or the error can be multiplicative. Examples can be found in Carroll et al. (2006).

In this work, we focus on the classical error model (1.7).

Differential versus nondifferential error

The error is often assumed to be *nondifferential*, meaning that $\boldsymbol{W}$ is conditionally independent of the response given $(\boldsymbol{X}, \boldsymbol{Z})$. $\boldsymbol{W}$ is then called a *surrogate*. In this case, $\boldsymbol{X}$ needs not be observable for the parameters of the model for the response given $\boldsymbol{X}$ and $\boldsymbol{Z}$ to be estimable. The reverse situation is *differential* error, when the distribution of the response given $(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{W})$ does depend on $\boldsymbol{W}$. This is equivalent to saying that $\boldsymbol{W}$ contains information about $\boldsymbol{Y}$ that is not contained in $\boldsymbol{X}$ and $\boldsymbol{Z}$. This can appear when $\boldsymbol{W}$ is linked to $\boldsymbol{X}$, but is actually not simply $\boldsymbol{X}$ observed with error.

From now on, we assume that the error is nondifferential.

Structural versus functional methods

Besides this first taxonomy of models, a further distinction is made in the statistical literature. On the one hand, *structural* methods assume a (parametric) model for the distribution of $\boldsymbol{X}$. *Functional*, or *semiparametric* methods, on the other hand, make no (or very weak) assumptions about the distribution of $\boldsymbol{X}$. The advantage of this approach is its robustness with respect to a misspecification of this distribution.

Both types of methods are considered in future sections.

1.3.2 Information required for identifiability

Whatever the model under consideration, some information about the measurement error distribution is most often needed for the response model parameters to be identifiable: about $\boldsymbol{W}$ given $\boldsymbol{X}$ and $\boldsymbol{Z}$ in an error model, or about $\boldsymbol{X}$ given $\boldsymbol{W}$ and $\boldsymbol{Z}$ in a regression calibration model. When this information is not available, making further assumptions can make the model identifiable.

Types of information

This information can be found either in the primary data collected for the study or in external data (or independent studies). The former is the ideal case (the precision is optimal), while associated with the latter are questions about the transportability of the information from one dataset to another.

This information can take different forms (Carroll et al., 2006; Schennach, 2016): *validation data* containing $\boldsymbol{X}$ (at least for some of the observations), or

auxiliary variables. An example of such variables is *replication data*, containing replicated measurements of $\boldsymbol{W}$, which are useful when the error is classical. *Factor data* containing a generalized form of replicates, $\boldsymbol{W} = \boldsymbol{\Lambda}\boldsymbol{X} + \boldsymbol{U}$, where Λ is a square matrix, can be used too. There are also *panel data*, as well as *instrumental data* containing other variables $\boldsymbol{T}$ related to $\boldsymbol{X}$ (called *instrumental variable* when internal) which must also satisfy some restrictions. In the context of nonlinear models, these conditions are the following: $\boldsymbol{T}$ is correlated with $\boldsymbol{X}$, $\boldsymbol{T}$ is independent of $\boldsymbol{U}$ and $(\boldsymbol{W}, \boldsymbol{T})$ is a surrogate for $\boldsymbol{X}$.

Use of the information

Schennach (2016) studies several nonlinear models, among which nonlinear (in the parameters or the covariates) regression models (either parametric or nonparametric), semi-, non- or parametric maximum likelihood models and parametric or semiparametric generalized method of moment models. In this context, the general use of the additional information which is available can be summarized as follows.

Validation data allow one to estimate the distribution of the error, or the distribution of $\boldsymbol{X}$ given the observed covariates, which is required by some correction methods. With replicated measurements, it is possible to recover the distribution of $\boldsymbol{X}$ (and $\boldsymbol{W}$), or at least its first moments (under weaker assumptions). With factor data, and under some conditions, the matrix $\boldsymbol{\Lambda}$ can be identified from the covariance matrix of $\boldsymbol{W}$; vectors of replicated measurements can then be built, which allow the estimation of the joint distribution of $\boldsymbol{X}$. In the absence of replicates or validation data, different approaches making use of instrumental variables have been developed; see Carroll et al. (2006) and Schennach (2016) for more details. With panel data, past or future observations can be used as instruments or replicated measurements, although the autocorrelation has to be taken into account.

When no auxiliary variables are available, Schennach (2016) explains that making *stronger assumptions of independence* can yield identifiability of the model. When $\boldsymbol{X}$, $\boldsymbol{U}$ and the model error are assumed to be mutually independent and $\boldsymbol{X}$ is not Gaussian, the regression parameters of a linear model are identified without auxiliary data (Reiersøl, 1950). Schennach (2013) present identification results for nonlinear (in the parameters or the covariates) regression models. However, Carroll et al. (2006) point out that such theoretical identifiability results do not always yield practical identifiability so that, in practice, additional information may still be required in order to be able to obtain stable estimates.

If independence assumptions cannot be made, only bounds on the coefficients (instead of point estimates) can be obtained.

1.3.3 Some functional methods

Several methods that correct for the bias due to measurement error without making parametric assumptions on $\boldsymbol{X}$ exist in the literature; the choice between them depends (among other criteria) on the response model of interest.

For linear regression, two possible approaches are described in Carroll et al. (2006). The method of moments consists in correcting a posteriori the naive estimate of the effect of the mismeasured covariate; in orthogonal regression, the classical least squares criterion is modified in order to take the measurement error into account. The former approach requires knowledge of the variance of the error, while the ratio of the variance of the regression error to the variance of the measurement error is needed for the latter.

For nonlinear models, three families of general methods that can be applied to a large class of models are detailed in Carroll et al. (2006) and reviewed here: regression calibration, simulation extrapolation and score function methods. Methods based on instrumental variables or designed to a specific response model are not considered here.

Regression calibration

Regression calibration is a first simple approximate functional method that works best in generalized linear models; the approach can be adapted to other highly nonlinear models (Carroll et al., 2006). It consists in replacing $\boldsymbol{X}$ by its regression on $\boldsymbol{Z}$ and $\boldsymbol{W}$, and then performing a classical analysis. Bootstrap can be used to estimate the standard errors. The regression of $\boldsymbol{X}$ on the covariates can be estimated with internal validation or replication data, an instrument, or external data (see Carroll et al. (2006) for more details).

Simulation extrapolation

Very similarily to the regression calibration approach, the simulation extrapolation (SIMEX) algorithm requires either that the error variance is known or sufficiently well estimated, or that there are replicate measurements of $\boldsymbol{X}$. It consists of two successive steps. In the *simulation step*, new contaminated data are created by adding artificial noise of known variance to $\boldsymbol{W}$, and the model parameters are estimated without taking the measurement error into account. This is done for different values of artificial error variance. In the *extrapolation step*, the naive estimates are modeled as a function of the measurement error variance. The SIMEX estimator is found by extrapolating this function to the case of no measurement error.

This method is computationally more intensive than regression calibration, but is easily implemented and very intuitive. Both methods can be applied to

many different models; however, they are approximate methods: in most cases, they yield what are sometimes called approximately consistent estimators.

SIMEX is described and studied in detail in the following two chapters.

Score function methods

Score function methods, like regression calibration and SIMEX, have a broad scope of application. However, they have the advantage of leading to fully consistent M-estimators[1] when the assumptions about the response model and the error model are met. They can be used when the variance of $\boldsymbol{U}$ is known (or estimated from external data or replicates). Carroll et al. (2006) discuss two methods that can be applied when the measurement error is classical, additive and normally distributed.

The *conditional score method* consists in defining a sufficient statistic for $\boldsymbol{X}$ which only depends on observed data and building unbiased score functions depending only on this statistic (and not on $\boldsymbol{X}$). This approach can still be used when the measurement error variance is not known, if an instrumental variable is available, or if the mismeasured covariate is measured longitudinally. This method was originally developed by Stefanski and Carroll (1987) in the context of canonical generalized linear models; it has been adapted to other models (see Carroll et al. (2006) for examples and references).

The *corrected score approach* is another score function method, which can be used when the model estimator in the absence of measurement error is an M-estimator. A corrected score function is a score function which depends only on the observed data and which is an unbiased estimator of the score function to be used without measurement error. This score function must be built and can then be used to construct estimating equations for the model parameters. The obtained estimator is fully consistent in some situations.

1.3.4 Some structural methods

Frequentist or Bayesian methods based on likelihoods are structural: they require stronger assumptions about $\boldsymbol{X}$ than functional approaches. Moreover, they can be computationally intensive. One advantage is that they can handle general models for the error: their scope of application is broader, and encompasses, for example, discrete mismeasured covariates and multiplicative error.

[1] An *M-estimator* is an estimator which is solution of an *estimating equation*, a sum (over all observations) of functions depending on the data and on the unknown parameters and taking values in a space with dimension equal to the number of parameters (Carroll et al., 2006). Least squares estimators and maximum likelihood estimators are M-estimators.

Frequentist likelihood methods

Frequentist likelihood methods consist in constructing the likelihood of $\boldsymbol{W}$ and the response given $\boldsymbol{Z}$, by integrating out the latent $\boldsymbol{X}$. This likelihood can then be maximized with respect to all unknown parameters. The specification of different models is required: for the response given $\boldsymbol{X}$, for the error and, if the measurement error includes classical error, for $\boldsymbol{X}$ given $\boldsymbol{Z}$. Since $\boldsymbol{X}$ is not observed, specifying the distribution for $\boldsymbol{X}$ given $\boldsymbol{Z}$ involves additional assumptions compared to the ones made in functional methods: this can yield robustness problems.

Bayesian methods

Most Bayesian methods are structural: because they rely on likelihoods, they require a model linking the response and $\boldsymbol{X}$, an error model, but also a specification of the distribution of $\boldsymbol{X}$ given $\boldsymbol{Z}$ if there are classical components in the error model.

These elements allow one to build the *likelihood* of $\boldsymbol{W}$, the response and $\boldsymbol{X}$ given $\boldsymbol{Z}$ and the parameters, as if $\boldsymbol{X}$ was observed. Moreover, a *prior distribution* is needed for each parameter to be estimated. The *posterior density* is then the conditional density of the parameters and the unobserved covariates given the observed data. When sampling from the posterior distribution, $\boldsymbol{X}$ is treated as missing data, and imputed from its conditional distribution given $\boldsymbol{Z}$, the response and the parameters.

Such an approach can incorporate validation data or replicates for an error model. Carroll et al. (2006) also explain that the prior information can help identify the parameters, in both identifiable and nonidentifiable models (Gustafson, 2005). However, as already mentioned, this comes at the price of stronger assumptions about the data.

Some flexible modeling strategies for $\boldsymbol{X}$

Some relatively flexible modeling solutions for $\boldsymbol{X}$ can be used in likelihood-based methods; with some of them, the distinction between structural and functional approaches is blurred.

In the frequentist literature, Carroll et al. (2006) mention *generalized linear models* for $\boldsymbol{X}$ given $\boldsymbol{Z}$, a *mixture of normal* densities, and the *semi-nonparametric (SNP)* density approximation as used by Zhang and Davidian (2001) in a context without measurement error (approximation of the random effects density in a linear mixed model). The latter approach relies on a class of densities of which the normal density is a particular case; however, the class encompasses deviations from the normality (skewed, multimodal, leptokurtic

or platykurtic distributions). These densities can be approximated, and hence estimated, by a truncated series expansion, leading to the SNP estimator. The number of terms in the series expansion can be chosen by using an information criterion. In Hu et al. (1998), $\boldsymbol{X}$ is modeled using three distributions: normal, SNP or none (in this latter case, the distribution of $\boldsymbol{X}$ is approximated by a *discrete distribution* with fixed support points). When analyzing clustered data, Li and Lin (2000) assume a *linear mixed model* (with a random effect for the cluster) for $\boldsymbol{X}$ given $\boldsymbol{Z}$.

Structural Bayesian techniques can also be made more robust by using flexible distributions for $\boldsymbol{X}$. Carroll et al. (1999b) and Richardson et al. (2002) consider a *mixture of normals* whose number of components can be either determined by a criterion (for example the BIC criterion) or chosen through a sensitivity analysis. In Müller and Roeder (1997), the joint distribution of $\boldsymbol{X}$, $\boldsymbol{W}$ and $\boldsymbol{Z}$ is modeled through a mixture of multivariate normal distributions, with a Dirichlet process prior on the mixing measure.

The distribution of $\boldsymbol{X}$ can also be modeled by a *discrete distribution*. This is the choice made for example by Aitkin and Rocci (2002) and Gustafson et al. (2002).

1.3.5 Deconvolution

The methods summarized in the previous two sections aim at estimating the parameters of the response model of interest, while taking into account the measurement error in the covariates. The objective of deconvolution is different: it allows one to estimate the density of the true variable from its mismeasured version. In this section, we follow Carroll et al. (2006) and consider a single variable X, with its associated observation $W = X + U$.

Deconvolution can be useful when the distribution of X is of direct interest, or when $E(X|W)$ is required by a method aiming at correcting for the measurement error (such as likelihood methods or regression calibration). In this context, the density of U in the classical error model is usually assumed to be known. *Parametric* deconvolution assumes a parametric (possibly flexible) model for X. In *nonparametric* deconvolution, no parametric assumptions are made on X, while the density of W must be estimated from the data. The case of a known error distribution has been studied a lot; see, for example, Taupin (2001) and Delaigle and Gijbels (2004), among others.

As will be studied in more details in Chapter 4, some authors relax the assumption of known density for U: for example, Matias (2002) and Schwarz and Van Bellegem (2010) leave the variance of the classical additive normal error unspecified, but restrict the class of possible distributions for X. In this

case, deconvolution is possible even without replicated or validation data. Delaigle and Hall (2016) consider the same classical additive error model without auxiliary data, but only assume that U has a symmetrical distribution (with unknown variance). The distribution of X cannot be symmetric or have a symmetric component.

1.3.6 In survival analysis

A review of some bias correction approaches available when dealing with survival data can be found in Carroll et al. (2006) and Yan (2014).

All functional approaches reviewed in the previous sections have been used in survival analysis. Clayton (1992) was one of the first authors to apply *regression calibration* to the survival Cox model. *SIMEX* has been adapted to the Cox model (Carroll et al., 2006), the Cox model for multiple events (Greene and Cai, 2004), the Cox model with nonlinear effect of mismeasured covariates (Crainiceanu et al., 2006), the frailty model (Li and Lin, 2003) and the AFT model (He et al., 2007). A *conditional score method* for the proportional hazards model with a longitudinal mismeasured covariate is studied by Tsiatis and Davidian (2001). The *corrected score approach* was adapted to the Cox model by Nakamura (1992), to a parametric promotion time cure model by Mizoi et al. (2007) and to the semiparametric promotion time cure model by Ma and Yin (2008). The approach of Huang and Wang (2000), which leaves the baseline hazard function and the distribution of $\boldsymbol{U}$ unspecified, is based on replicates. Zhou and Wang (2000) present an approach making no assumption on the baseline hazard function or on the covariate distribution, but requiring validation data.

As far as frequentist *likelihood methods* are concerned, Yi and Lawless (2007) and Hu et al. (1998) study the Cox model with Gaussian measurement error of known variance (or variance estimated through replicates) in the covariates. The latter reference considers three approaches (parametric, semi- and nonparametric), depending on the distribution assumed for $\boldsymbol{X}$: normal, SNP or none (the distribution of $\boldsymbol{X}$ is approximated by a *discrete distribution* with fixed support points). In the frailty model for clustered survival data, Li and Lin (2000) assume a normal distribution for the error, and a linear mixed model for $\boldsymbol{X}$.

With repeated measurements of the covariate with measurement error, joint modeling is available. This approach is considered, for example, in Crowther et al. (2013). They consider a longitudinal linear mixed effects submodel for the covariate with error, in which the true value of the explanatory variable is modeled using fixed and random effects. This true value appears in the survival submodel.

1.4 Outline

In this thesis, we first consider survival data with two features: measurement error in the covariates and a cure fraction in the population. We focus on the classical measurement error model (1.7) and on the promotion time cure model (1.6). Two estimation methods exist in the literature to estimate such a model, among which only one adapted to the semiparametric version of the promotion time cure model (Ma and Yin, 2008). However, despite its good performance, this method suffers from some drawbacks: it assumes a specific form for the link function of the promotion time cure model, and requires Gaussian measurement error. In Chapter 2, based on Bertrand et al. (2017a), we introduce an alternative approach which somewhat relaxes these assumptions: we extend a functional approach introduced in Section 1.3.3, the SIMEX algorithm, to the semiparametric promotion time cure model.

In Chapter 3, based on Bertrand et al. (2017b), we still consider the promotion time cure model with mismeasured covariates. We focus on three possible estimation methods: the naive one, which does not take the measurement error into account, Ma and Yin (2008)'s approach and the method that we introduced in Chapter 2. Both correction methods rely on some assumptions, among which the knowledge of the measurement error variance. As mentioned previously, Ma and Yin (2008)'s method also assumes the normality of the measurement error. We hence compare the robustness of these approaches with respect to these two assumptions, in order to derive recommendations about the use of these methods in practice.

Finally, in Chapter 4, based on Bertrand et al. (2017c), we focus on the general problem of measurement error, in a broader context. The methods described in Section 1.3 usually assume that the measurement error distribution is known. When no auxiliary or validation data are available, this assumption can be a real issue to use these methods in practice. We hence develop a flexible structural method to estimate the variance of classical Gaussian measurement error without validation or auxiliary data. This estimated variance can then be used with different correction methods, in various models. For illustration, we use it jointly with the SIMEX algorithm to estimate the parameters of a Cox proportional hazards model.

CHAPTER 2

The simulation extrapolation approach

The material presented in this chapter is published in the paper "Inference in a survival cure model with mismeasured covariates using a simulation-extrapolation approach", by Bertrand A., Legrand C., Carroll R.J., De Meester C. and Van Keilegom I. in Biometrika 2017, 104, 31–50.

2.1 Introduction

In this chapter, we consider survival data with a cure fraction and measurement error in some (or all) continuous covariates. The survival cure model that we will study is the promotion time cure model (1.6) introduced in Section 1.2.2, according to which the survival function of the population is

$$S(t|\boldsymbol{X}) = \exp\left\{-\theta(\boldsymbol{X})F(t)\right\}, \tag{2.1}$$

in which we leave F completely unspecified. The first component of the P-dimensional covariate $\boldsymbol{X}$ is supposed to be 1, in order to include an intercept in the model.

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 we do not observe whether the censored observations are cured or susceptible.

As far as the measurement error is concerned, we assume that the classical measurement error model (1.7) presented in Section 1.3.1 holds, i.e.,

$$\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}, \tag{2.2}$$

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 that it follows a continuous distribution with mean zero and known covariance matrix $\boldsymbol{V}$. In this chapter, contrary to what was done in Section 1.3, we denote by $\boldsymbol{X}$ the whole vector of covariates, whether or not they are measured with error. 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 error model studied, for example, by Cook and Stefanski (1994) and Ma and Yin (2008).

In order to correct for this measurement error, we choose to work with a functional approach, the SIMEX (simulation extrapolation) method introduced in Section 1.3.3. This algorithm has a number of important advantages that will be summarized in the Discussion (Section 2.7) and that make it a very popular method 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. SIMEX methodology has also been applied to nonparametric regression (Carroll et al., 1999a) and to general semiparametric problems (Apanasovich et al., 2009) where, if $\boldsymbol{X}$ denotes the mismeasured covariates and $\boldsymbol{Z}$, the covariates that are correctly measured, the log-likelihood is either of the form $\mathcal{L}\{Y, m(\boldsymbol{X}), \boldsymbol{Z}, \boldsymbol{\beta}\}$ or $\mathcal{L}\{Y, \boldsymbol{X}, m(\boldsymbol{Z}), \boldsymbol{\beta}\}$ where $m(\cdot)$ is an unknown nonparametric function.

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

(2007) study the problem of measurement error in the parametric promotion time cure model. They also consider $\theta(\boldsymbol{X}) = \exp(\boldsymbol{X}^T\boldsymbol{\beta})$ and Gaussian error.

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 consist 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, and can thus be considered as cured from the event of interest. Nowadays, the ejection fraction plays a major role in the choice of 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 chapter 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 2.3 contains the asymptotic properties of these estimators; the proofs can be found in Section 2.6. The finite sample properties of the estimator of $\boldsymbol{\beta}$, as well as its robustness with respect to some algorithm parameters and underlying assumptions, are investigated in Section 2.4 through a simulation study. Section 2.5 contains the analysis of the aortic insufficiency database. Finally, the obtained results are discussed in Section 2.7.

2.2 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 $Y_{(1)} < \cdots < Y_{(m)}$. We use model (2.1), 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 $\eta(\cdot) = \log\left[\exp(\cdot)/\left\{1 + \exp(\cdot)\right\}\right]$. We present the algorithm to be used when the error $\boldsymbol{U}$ is normally distributed. However, $\boldsymbol{U}$ does not have to have a normal distribution. In that case, the algorithm below remains valid without any modification, as long as the extrapolation function is correctly specified (Carroll et al., 1996).

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 $\lambda = \lambda_1, \ldots, \lambda_K$, $\lambda \geq 0$
   - 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 $\boldsymbol{W}_{i,\lambda,b} = \boldsymbol{W}_i + (\lambda \boldsymbol{V})^{1/2} \boldsymbol{Z}_{b,i}$ for each individual $i = 1, \ldots, n$, where $\boldsymbol{V}$ is the known covariance matrix of the error term, as defined in Section 2.1.
       The variance-covariance matrix of the contaminated $\boldsymbol{W}_{i,\lambda,b}$ is
       $$\mathrm{Var}\left(\boldsymbol{W}_{i,\lambda,b} | \boldsymbol{X}_i\right) = \mathrm{Var}(\boldsymbol{W}_i | \boldsymbol{X}_i) + \lambda \boldsymbol{V} = \boldsymbol{V} + \lambda \boldsymbol{V} = (1 + \lambda)\boldsymbol{V},$$
       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:
   $$\widehat{\beta}_{\lambda_k,p} = g_{\beta_p}(\boldsymbol{\gamma}_{\beta_p}, \lambda_k) + \pi_{p,k} = \gamma_{\beta_p,1} + \gamma_{\beta_p,2}\lambda_k + \gamma_{\beta_p,3}\lambda_k^2 + \pi_{p,k},$$
   $$p = 1, \ldots, P; \quad k = 1, \ldots, K.$$

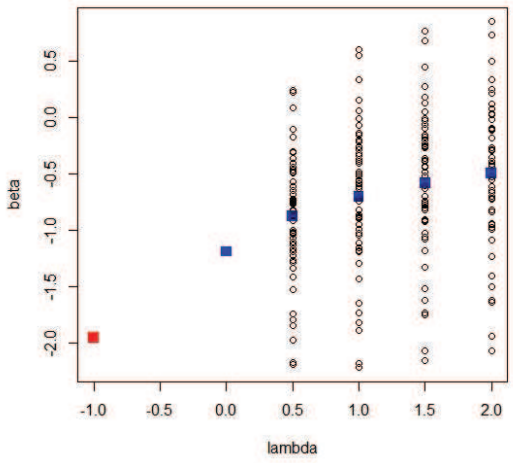


Figure 2.1: Illustration of the functioning of the SIMEX algorithm for one parameter. The circles are the estimated values (50 for each value of λ). The squares corresponding to $\lambda = 0.5, 1, 1.5, 2$ are the average estimated values. They are used together with the naive estimator corresponding 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.

where $\pi_{p,k}$ are the error terms in the extrapolant model, assumed to be independent and with mean 0. Fit these parametric models for each $p = 1, \ldots, P$ in order to obtain $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}} = (\widehat{\gamma}_{\beta_1}^T, \ldots, \widehat{\gamma}_{\beta_P}^T)^T$.

In practice, this function is often an approximation of the true extrapolation function, which will then yield an estimator that converges in probability to some constant being approximately equal to the true parameter (Cook and Stefanski, 1994). There are some cases in semiparametric models where the exact extrapolant is known: (a) Cox regression, with $\boldsymbol{X}$ and $\boldsymbol{W}$ normally distributed and homoscedastic (Prentice, 1982), and (b) the partially linear model $Y = m(\boldsymbol{Z}) + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \epsilon$, $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$, and $(\boldsymbol{X}, \boldsymbol{U})$ are independent (Liang et al., 1999).

3. Obtain the SIMEX estimated values

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

This algorithm is illustrated (for the coefficient of the mismeasured covariate in the cardiology database analyzed in Section 2.5) in Figure 2.1, with $B = 50$, $\lambda \in \{0, 0.5, 1, 1.5, 2\}$ and a quadratic extrapolation function.

As far as naive estimation is concerned, we use the results of Zeng et al. (2006) and Ma and Yin (2008) to estimate the model parameters $\boldsymbol{\beta}$ and F when there is no measurement error in the covariates. They show that the log-likelihood of the promotion time cure model without measurement error is (we use $\eta(\boldsymbol{X}^T\boldsymbol{\beta})$ instead of the particular case $\exp(\boldsymbol{X}^T\boldsymbol{\beta})$ considered by the authors)

$$\begin{aligned}\ell = \sum_{i=1}^{n} \big[&\delta_i I(Y_i < \infty) \left\{-F(Y_i)\eta(\mathbf{X}_i^T\boldsymbol{\beta}) + \log F\{Y_i\} + \log \eta(\mathbf{X}_i^T\boldsymbol{\beta})\right\} \\ &+(1-\delta_i) I(Y_i < \infty) \left\{-F(Y_i)\eta(\mathbf{X}_i^T\boldsymbol{\beta})\right\} - I(Y_i = \infty)\eta(\mathbf{X}_i^T\boldsymbol{\beta})\big], \end{aligned} \tag{2.3}$$

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 $F(Y_i) = \sum_{Y_{(j)} \leq Y_i} p_{(j)}$. The estimated baseline cumulative distribution function is forced to be 1 beyond the largest observed failure time, $\sum_{j=1}^{m} p_{(j)} = 1$. This implies that no event can occur at those times: the cure threshold needed for identifiability and introduced in Section 1.2.3 is then determined to be $\tau = Y_{(m)}$.

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 $\widehat{F}_{SIMEX}(t) = \sum_{Y_{(j)} \leq t} \widehat{p}^*_{SIMEX,(j)}$.

2.3 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, while Theorem 2 establishes their asymptotic normality. They are proved in Section 2.6. Both results rely on the assumption that the true extrapolation function is known, which is rarely the case in practice. As explained by Cook and Stefanski (1994) and mentioned in Section 2.2, since the extrapolation function used in the algorithm is often an approximation of the true one, we will obtain an estimator which converges in probability to some constant being approximately equal to the true parameter. This is sometimes called *approximate consistency*. In this case, in all the results that follow,

β_{TRUE} and F_{TRUE} are replaced in the results by β^* and F^*, the limiting values with this extrapolant. In the parametric case, when $\boldsymbol{X}$ is scalar, with v^2 being the measurement error variance, the bias of a polynomial extrapolant of order p is $O(v^{2+2p})$, see Cook and Stefanski (1994), page 1317, although they only consider $p = 2$ (quadratic).

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, which 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* $n^{1/2}(\widehat{\boldsymbol{\beta}}_{SIMEX} - \boldsymbol{\beta}_{TRUE}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$*, where* $\boldsymbol{\Sigma}$ *is given in Section 2.6. Moreover,* $n^{1/2}(\widehat{F}_{SIMEX} - F_{TRUE})$ *converges weakly to a zero-mean Gaussian process* $\mathcal{G}$ *whose covariance function is given in Section 2.6.*

The regularity conditions (C1)–(C4) of Zeng et al. (2006) applied to model (2.1) are the following:

(C1) The covariate $\boldsymbol{X}$ is bounded with probability 1, and if there exists a vector $\widetilde{\boldsymbol{\beta}}$ such that $\widetilde{\boldsymbol{\beta}}^T \boldsymbol{X} = 0$ with probability 1, then $\widetilde{\boldsymbol{\beta}} = \mathbf{0}$.

(C2) Conditional on $\boldsymbol{X}$, the right-censoring time C is independent of T, and $P(C = \infty | \boldsymbol{X}) > 0$.

(C3) The true value of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_0$, belongs to the interior of a known compact set $\mathcal{B}_0$, and the true promotion time cumulative distribution function F_0 is differentiable with $F_0'(x) > 0$ for all $x \in \mathbb{R}^+$.

(C4) The link function $\eta(\cdot)$ is strictly increasing and twice-continuously differentiable with $\eta(\cdot) > 0$.

As far as the asymptotic 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 $\lim_{\lambda\to-1}\left(\overline{\boldsymbol{\Sigma}}_\lambda - \widehat{\boldsymbol{\Sigma}}_\lambda\right)$, where

- $\overline{\boldsymbol{\Sigma}}_\lambda$ 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.

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

The intuition behind this approach is explained in Carroll et al. (2006). They show that $Var(\widehat{\boldsymbol{\beta}}_{SIMEX}) \approx Var(\widehat{\boldsymbol{\beta}}_{TRUE}) + Var(\widehat{\boldsymbol{\beta}}_{SIMEX} - \widehat{\boldsymbol{\beta}}_{TRUE})$, where $\widehat{\boldsymbol{\beta}}_{TRUE}$ is the estimator of $\boldsymbol{\beta}$ when $\boldsymbol{X}$ is observable. The variance of the SIMEX estimator can hence be decomposed into two components: the former is due to sampling variability, while the latter is due to measurement error variability. $Var(\widehat{\boldsymbol{\beta}}_{TRUE})$ can be naturally estimated by $\lim_{\lambda\to-1}\overline{\boldsymbol{\Sigma}}_\lambda$. Moreover, it can be shown that $Var(\widehat{\boldsymbol{\beta}}_{SIMEX} - \widehat{\boldsymbol{\beta}}_{TRUE}) = -\lim_{\lambda\to-1} Var(\widehat{\boldsymbol{\beta}}_{\lambda,b} - \widehat{\boldsymbol{\beta}}_\lambda)$ (Stefanski and Cook, 1995); the minus sign appears because, since $Var(\widehat{\boldsymbol{\beta}}_{\lambda,b} - \widehat{\boldsymbol{\beta}}_\lambda)$ is positive when $\lambda > 0$ and equals zero when $\lambda = 0$, the extrapolant is negative when $\lambda < 0$. A natural estimator is then $-\lim_{\lambda\to-1}\widehat{\boldsymbol{\Sigma}}_\lambda$.

2.4 Simulation studies

The objective of the first three simulation studies that we present here is to investigate the properties of the proposed estimator in samples of finite size and to compare it with both the naive method which was introduced at the end of Section 2.2 and is based on Equation (2.3), and the corrected score method (Ma and Yin, 2008).

In the last four subsections, we focus on our proposed estimator. In Subsections 2.4.4 and 2.4.5, we present the results of simulation studies aiming at studying the effect of the choice of the extrapolation function and of the grid of values of λ on the SIMEX estimator. Subsections 2.4.6 and 2.4.7 contain results pertaining to the robustness of the SIMEX estimator to a misspecification of the error distribution and variance.

An extensive simulation study investigating thoroughly the robustness of both the SIMEX algorithm and the corrected score approach (Ma and Yin, 2008)

with respect to the assumptions of normal distribution and known variance of the error can be found in Chapter 3.

In the following, for the SIMEX algorithm, we used $B = 50$, $\lambda \in \{0, 0.5, 1, 1.5, 2\}$ (except in Subsection 2.4.5) and a quadratic extrapolant (except in Subsection 2.4.4). For each setting, 500 simulated data sets were analyzed. Appendix A contains an example of R program that generates such data.

2.4.1 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 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 $W = X_1 + U_1$ is observed, where $U_1 \sim N(0, v^2)$, where v^2 is the only non-zero element of $\boldsymbol{V}$. 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 2.1, while those corresponding to $\mu = 1$ can be found in Table 2.2.

Table 2.1: 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	.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	.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	.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	.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

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

Table 2.2: Simulation results for the settings with one mismeasured covariate, when $\mu = 1$.

			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	.1	Bias	.002	.018	-.006	.054	-.102	-.002	.007	.004	-.005
		Emp. var.	.041	.102	.030	.036	.077	.030	.041	.098	.030
		Est. var.	.042	.100	.028	.037	.076	.028	.041	.096	.028
		95% cv	.948	.946	.942	.940	.928	.946	.948	.944	.942
		MSE	.041	.102	.030	.039	.087	.030	.041	.098	.030
200	.2	Bias	-.014	.059	-.008	.151	-.323	.003	.046	-.085	-.002
		Emp. var.	.052	.171	.032	.032	.060	.030	.042	.113	.031
		Est. var.	.056	.167	.030	.032	.057	.028	.041	.098	.028
		95% cv	.964	.960	.940	.870	.730	.948	.928	.916	.940
		MSE	.052	.174	.032	.055	.164	.030	.044	.120	.031
300	.1	Bias	-.004	.024	-.009	.048	-.094	-.006	.001	.013	-.009
		Emp. var.	.031	.071	.020	.028	.054	.020	.030	.069	.020
		Est. var.	.028	.066	.019	.024	.050	.018	.027	.064	.019
		95% cv	.954	.946	.956	.916	.904	.950	.944	.942	.954
		MSE	.031	.071	.021	.030	.063	.020	.030	.069	.021
300	.2	Bias	-.013	.049	-.012	.146	-.316	-.001	.039	-.076	-.009
		Emp. var.	.037	.104	.022	.024	.039	.020	.031	.075	.021
		Est. var.	.035	.104	.020	.021	.038	.018	.027	.065	.019
		95% cv	.964	.966	.948	.825	.638	.940	.922	.924	.947
		MSE	.037	.106	.022	.045	.139	.020	.033	.081	.021

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

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 corrected score method 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 a mean squared error (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 one.

2.4.2 Two mismeasured covariates

We now introduce, in the previous setting, one additional covariate with measurement error. In this case,

$$S(t|X_1, X_2, X_3) = \exp\left\{-\exp\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3\right) F(t)\right\}.$$

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 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 2.3, while those corresponding to $\mu = 1$ can be found in Table 2.4.

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 corrected score method is the best for β_0 and β_1, while SIMEX is preferred 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.

Table 2.3: Simulation results for the settings with two 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	.1	.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	.2	.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	.1	.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	.2	.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

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

Table 2.4: Simulation results for the settings with 2 mismeasured covariates, when $\mu = 1$

				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	.1	.1	Bias	.004	.007	.013	-.006	.051	-.117	-.009	-.001	.011	-.011	.010	-.005
			Emp. var.	.057	.118	.013	.032	.050	.088	.012	.031	.056	.113	.013	.031
			Est. var.	.049	.109	.012	.031	.042	.081	.011	.030	.047	.103	.012	.031
			95% cv	.930	.946	.948	.958	.918	.926	.948	.960	.924	.948	.942	.960
			MSE	.057	.118	.013	.032	.053	.101	.012	.031	.056	.113	.013	.031
200	.2	.2	Bias	-.009	.050	.024	-.011	.135	-.346	-.060	.010	.052	-.111	.003	-.003
			Emp. var.	.076	.205	.017	.035	.046	.068	.011	.031	.062	.134	.014	.033
			Est. var.	.066	.187	.015	.034	.037	.061	.010	.030	.048	.107	.013	.032
			95% cv	.944	.952	.950	.952	.864	.710	.902	.952	.910	.908	.952	.948
			MSE	.076	.208	.017	.036	.064	.187	.015	.031	.064	.146	.014	.033
300	.1	.1	Bias	.016	.012	.009	-.009	.062	-.110	-.013	-.003	.020	-.001	.007	-.008
			Emp. var.	.032	.076	.008	.020	.028	.058	.007	.019	.032	.075	.008	.020
			Est. var.	.032	.071	.008	.020	.028	.054	.007	.020	.031	.068	.008	.020
			95% cv	.958	.940	.960	.956	.932	.928	.952	.954	.948	.940	.960	.958
			MSE	.032	.076	.008	.020	.032	.070	.007	.019	.032	.075	.008	.020
300	.2	.2	Bias	.005	.045	.017	-.011	.146	-.338	-.063	.007	.062	-.097	.002	-.006
			Emp. var.	.042	.125	.010	.021	.025	.045	.007	.019	.035	.087	.009	.021
			Est. var.	.042	.117	.010	.022	.025	.040	.007	.020	.032	.071	.009	.021
			95% cv	.954	.948	.964	.954	.848	.608	.872	.954	.920	.906	.960	.950
			MSE	.042	.127	.010	.021	.046	.159	.011	.019	.039	.096	.009	.021

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

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 2.2) 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 $W = X_1 + U_1$ is observed, where $U_1 \sim N(0, v^2)$.

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 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%. The results are summarized in Table 2.5.

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 corrected score method yields the largest MSE).

2.4.4 Impact of the choice of the extrapolation function

In order to compare the performance of SIMEX with different choices of extrapolant, we consider the same setting as in the previous subsection, and we estimate the model parameters using, in addition to the quadratic extrapolant, a linear and a cubic extrapolant. Table 2.6 reports the results.

Table 2.5: 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
.1	200	Bias	-.021	.013	-.022	.038	-.106	-.021	-.014	-.003	-.022
		Emp. var.	.126	.212	.067	.112	.162	.067	.124	.206	.067
		Est. var.	.114	.221	.063	.098	.167	.062	.107	.191	.063
		95% cv	.944	.972	.948	.942	.954	.95	.934	.962	.948
		MSE	.126	.212	.068	.114	.173	.067	.124	.206	.068
.25	200	Bias	-.056	.093	-.026	.199	-.429	-.019	.074	-.186	-.020
		Emp. var.	.185	.450	.071	.094	.100	.067	.121	.207	.068
		Est. var.	.218	.578	.069	.079	.105	.062	.097	.146	.063
		95% cv	.972	.974	.960	.886	.742	.954	.927	.886	.951
		MSE	.188	.459	.072	.133	.284	.067	.126	.242	.068
.1	300	Bias	-.031	.025	-.013	.025	-.094	-.012	-.025	.013	-.013
		Emp. var.	.100	.167	.043	.092	.128	.043	.098	.164	.043
		Est. var.	.079	.144	.041	.072	.110	.041	.076	.125	.041
		95% cv	.932	.948	.956	.930	.924	.958	.928	.938	.956
		MSE	.101	.168	.044	.093	.136	.043	.099	.164	.044
.25	300	Bias	-.068	.105	-.019	.187	-.412	-.011	.060	-.158	-.014
		Emp. var.	.153	.371	.047	.074	.082	.043	.099	.174	.044
		Est. var.	.144	.375	.046	.056	.069	.041	.069	.097	.042
		95% cv	.960	.968	.956	.850	.632	.958	.897	.839	.956
		MSE	.158	.382	.047	.109	.252	.043	.103	.199	.044

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

Table 2.6: Simulation results for the realistic settings, for SIMEX with three different extrapolation functions

			SIMEX (linear)			SIMEX (quadratic)			SIMEX (cubic)		
v	n	Estimate	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
.1	200	Bias	.000	-.030	-.022	-.014	-.003	-.022	-.014	.000	-.022
		Emp. var.	.120	.192	.067	.124	.206	.067	.125	.212	.067
		Est. var.	.102	.179	.063	.107	.191	.063	.110	.194	.063
		95% cv	.936	.956	.948	.934	.962	.948	.932	.950	.944
		MSE	.120	.193	.068	.124	.206	.068	.125	.212	.068
.25	200	Bias	.145	-.325	-.019	.074	-.186	-.020	.036	-.106	-.019
		Emp. var.	.104	.141	.067	.121	.207	.068	.138	.265	.068
		Est. var.	.088	.112	.062	.097	.146	.063	.160	.182	.063
		95% cv	.909	.808	.954	.927	.886	.951	.931	.899	.957
		MSE	.125	.247	.067	.126	.242	.068	.140	.276	.069
.1	300	Bias	-.011	-.016	-.013	-.025	.013	-.013	-.025	.019	-.012
		Emp. var.	.095	.152	.043	.098	.164	.043	.108	.171	.043
		Est. var.	.080	.118	.041	.076	.125	.041	.078	.128	.041
		95% cv	.928	.938	.956	.928	.938	.956	.928	.936	.954
		MSE	.095	.152	.043	.099	.164	.044	.108	.171	.044
.25	300	Bias	.131	-.300	-.012	.060	-.158	-.014	.019	-.068	-.016
		Emp. var.	.084	.117	.044	.099	.174	.044	.114	.224	.045
		Est. var.	.081	.074	.041	.069	.097	.042	.081	.121	.042
		95% cv	.872	.729	.956	.897	.839	.956	.918	.852	.952
		MSE	.101	.207	.044	.103	.199	.044	.114	.229	.045

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

In these simulations, varying the extrapolation function used in the SIMEX algorithm has no effect on the estimation of β_2. This is not the case for the other two parameters. When the measurement error variance is rather low, the smallest bias for β_0 is obtained by the linear extrapolant, while there is no clear conclusion regarding β_1. However, for the largest variance, the higher the extrapolation order, the smaller the bias. In terms of MSE, the lowest order of extrapolation yields the best results for β_0, but the differences among extrapolants are quite limited. For β_1, the MSE increases with the order of the extrapolation when the measurement error variance is low, while the quadratic extrapolant outperforms the other ones when the variance is larger.

These findings are consistent with the general behaviour of the SIMEX estimator: the lower the order of the extrapolant, the more conservative the correction (Cook and Stefanski, 1994). Consequently, the choice of the extrapolation function has to be considered in the context of the bias-variance tradeoff; the quadratic one has appeared to be a good compromise in many different cases (Cook and Stefanski, 1994; Carroll et al., 2006) and is widely used (He et al., 2007; Li and Lin, 2003).

2.4.5 Robustness with respect to the grid of λ

In this simulation study, we investigate the impact of the choice of the grid of values for λ in the SIMEX method. We still consider the same setting as in Subsection 2.4.3, and compare the estimates obtained by using $\lambda \in \{0, 0.5, 1, 1.5, 2\}$ (as in all previous simulation settings) with those obtained with $\lambda \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$. The results are reported in Table 2.7: both grids of λ values yield nearly identical results in this setting.

2.4.6 Robustness with respect to a misspecification of the error distribution

In this subsection, the effect of a misspecification of the measurement error distribution is investigated through another simulation study, again using the settings of Subsection 2.4.3 with $n = 200$. The measurement error is now generated using two distributions which are very different from the Gaussian one: a uniform and a Chi-squared distribution, with standard deviation $v = 0.1$ and $v = 0.25$ (assumed to be known). When $v = 0.1$, we see in Table 2.8 that, for this setting, the misspecification has no impact on the MSE; the impact on the bias is very limited (except, in some extent, for β_1 with the Chi-squared distribution). For the largest value of the measurement error variance (0.25^2), the estimation of β_2 is not influenced by the true distribution. The MSE of β_0 and β_1 increase slightly. With the uniform distribution, the bias of these two

Table 2.7: Simulation results investigating the robustness of SIMEX with respect to the grid of λ.

			$\lambda \in \{0, 0.5, 1, 1.5, 2\}$			$\lambda \in \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2\}$		
v	n		β_0	β_1	β_2	β_0	β_1	β_2
.1	200	Bias	-.014	-.003	-.022	-.014	-.003	-.022
		Emp. var.	.124	.206	.067	.124	.206	.067
		Est. var.	.108	.191	.063	.107	.191	.063
		95% cv	.934	.958	.948	.934	.962	.948
		MSE	.124	.206	.068	.124	.206	.068
.25	200	Bias	.072	-.182	-.020	.074	-.186	-.020
		Emp. var.	.121	.210	.068	.121	.207	.068
		Est. var.	.097	.148	.063	.097	.146	.063
		95% cv	.927	.888	.951	.927	.886	.951
		MSE	.127	.243	.068	.126	.242	.068
.1	300	Bias	-.025	.013	-.013	-.025	.013	-.013
		Emp. var.	.099	.165	.043	.098	.164	.043
		Est. var.	.076	.125	.041	.076	.125	.041
		95% cv	.928	.940	.956	.928	.938	.956
		MSE	.100	.165	.044	.099	.164	.044
.25	300	Bias	.058	-.153	-.014	.061	-.156	-.015
		Emp. var.	.100	.177	.044	.099	.174	.044
		Est. var.	.070	.098	.042	.069	.097	.042
		95% cv	.897	.843	.956	.897	.839	.956
		MSE	.103	.200	.045	.103	.199	.044

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

parameters stay nearly constant, while with a Chi-squared error, they increase quite markedly.

2.4.7 Robustness with respect to a misspecification of the error variance

We now investigate the behaviour of our estimator when the measurement error variance is misspecified. More precisely, we consider the same setting as in Section 2.4.3 with $n = 200$, where the error is simulated with $v_S = 0.1$ and $v_S = 0.25$. However, in the estimation process, the variance is misspecified as $v_E \in \{0.05, 0.15\}$ for $v_S = 0.1$ and as $v_E \in \{0.2, 0.3\}$ for $v_S = 0.25$.

The results, reported in Table 2.9, show that, in these simulations, a misspecification of the measurement error variance has no impact on the estimation of β_2. Both β_0 and β_1 are influenced, in terms of both the bias and the MSE. The latter increases with the value of the variance assumed in the estimation procedure, although this increase is less marked when switching from the underspecified variance to the true one (compared to when switching from the true variance to the overspecified one). As far as the bias is concerned, the conclusion depends on the true value of the measurement error variance. When it is low, the lowest bias is obtained when the correct variance is assumed, while, when the true variance is higher, the bias decreases when the specified variance increases.

2.5 Aortic insufficiency database

We illustrate the proposed methodology on data from patients suffering from aortic insufficiency (AI), a cardiovascular disease. Between 1995 et 2013, 393 patients underwent echocardiography for severe AI at the Brussels Saint-Luc University Hospital (Belgium). These data were collected by one of the authors (CdM) and include information from the diagnosis of the pathology, between 1981 and 2013. Although AI can be a lethal condition, it is known that a proportion of patients will never die from it. Since the patients considered in this study have no or limited other known morbidity, those who survive for a sufficiently long period after the diagnosis can be considered as *long-term survivors*. 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 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

Table 2.8: Simulation results investigating the robustness of SIMEX with respect to a misspecification of the error distribution.

		$v = 0.1$				$v = 0.25$		
True distribution	Estimate	β_0	β_1	β_2	Estimate	β_0	β_1	β_2
Gaussian	Bias	-.014	-.003	-.022	Bias	.072	-.182	-.020
	Emp. var.	.124	.206	.067	Emp. var.	.121	.210	.068
	Est. var.	.108	.191	.063	Est. Var.	.097	.148	.063
	95% cv	.934	.958	.948	95% cv	.927	.888	.951
	MSE	.124	.206	.068	MSE	.127	.243	.068
Chi-squared	Bias	-.004	-.020	-.021	Bias	.116	-.251	-.020
	Emp. var.	.128	.210	.068	Emp. var.	.124	.199	.070
	Est. var.	.108	.186	.063	Est. Var.	.092	.128	.063
	95% cv	.932	.938	.940	95% cv	.902	.833	.944
	MSE	.128	.210	.069	MSE	.138	.262	.070
Uniform	Bias	-.010	-.007	-.020	Bias	.071	-.177	-.019
	Emp. var.	.126	.202	.069	Emp. var.	.135	.219	.071
	Est. var.	.107	.191	.063	Est. var.	.826	.152	.063
	95% cv	.931	.950	.944	95% cv	.913	.861	.942
	MSE	.127	.202	.070	MSE	.140	.250	.071

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

Table 2.9: Simulation results investigating the robustness of SIMEX with respect to a misspecification of the error variance.

		$v_S = 0.1$				$v_S = 0.25$		
v_E		β_0	β_1	β_2	v_E	β_0	β_1	β_2
.05	Bias	.025	-.081	-.021	.20	.116	-.262	-.021
	Emp. var.	.115	.171	.067		.114	.172	.067
	Est. var.	.112	.173	.062		.092	.138	.063
	95% cv	.944	.958	.948		.923	.863	.954
	MSE	.115	.178	.067		.128	.240	.068
.1	Bias	-.014	-.003	-.022	.25	.072	-.182	-.020
	Emp. var.	.124	.206	.067		.121	.210	.068
	Est. var.	.108	.191	.063		.097	.148	.063
	95% cv	.934	.958	.948		.927	.888	.951
	MSE	.124	.206	.068		.127	.243	.068
.15	Bias	-.080	.130	-.022	.30	.029	-.093	-.021
	Emp. var.	.141	.266	.068		.133	.265	.070
	Est. var.	.117	.218	.063		.101	.159	.063
	95% cv	.932	.924	.946		.929	.888	.948
	MSE	.148	.283	.068		.134	.274	.070

Emp. var.: empirical variance; Est. var.: estimated variance; 95% cv: coverage probabilities of 95% confidence intervals computed based on the asymptotic normal distribution; MSE: mean squared error.

appearing in the guidelines used to decide whether a patient should be operated on (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 2.2). As explained in Section 2.2, the cure threshold is the largest observed event time: all patients surviving up to 17.21 years are considered as cured from their AI. 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 compare our results with those obtained by the corrected score method (Ma and Yin, 2008), as well as with those 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 v of the measurement error 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, standardized for the analysis) 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 2.10.

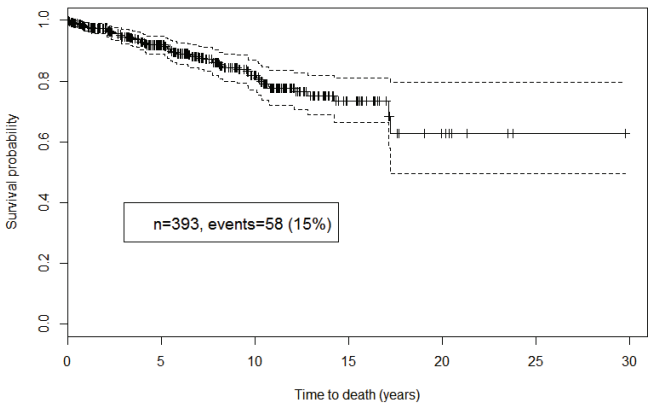


Figure 2.2: Kaplan-Meier estimate of the survival curve for the patients from the aortic insufficiency database.

Table 2.10: Regression coefficient estimates, estimated standard deviations and confidence intervals (based on the asymptotic Normal distribution)

Estimate	EF	Gender	Age (standardized)	Surgery < 3 months	Surgery > 3 months
Naive	-1.222	0.634	1.233	0.137	-0.730
(Estimated SD)	(1.376)	(0.284)	(0.195)	(0.374)	(0.396)
95% C.I. (lower bound)	-3.920	0.078	0.850	-0.597	-1.507
95% C.I. (upper bound)	1.475	1.190	1.616	0.871	0.046
Ma & Yin Method ($v = 0.05$)	-1.496	0.633	1.229	0.122	-0.732
(Estimated SD)	(1.685)	(0.285)	(0.195)	(0.377)	(0.396)
95% C.I. (lower bound)	-4.799	0.074	0.846	-0.616	-1.507
95% C.I. (upper bound)	1.807	1.193	1.612	0.860	0.043
Ma & Yin Method ($v = 0.10$)	-3.942	0.629	1.172	-0.034	-0.766
(Estimated SD)	(4.209)	(0.312)	(0.208)	(0.434)	(0.397)
95% C.I. (lower bound)	-12.193	0.017	0.763	-0.885	-1.545
95% C.I. (upper bound)	4.308	1.241	1.580	0.817	0.013
SIMEX ($v = 0.05$)	-1.454	0.631	1.228	0.121	-0.727
(Estimated SD)	(1.480)	(0.285)	(0.195)	(0.376)	(0.395)
95% C.I. (lower bound)	-4.354	0.072	0.845	-0.616	-1.500
95% C.I. (upper bound)	1.446	1.189	1.611	0.858	0.047
SIMEX ($v = 0.10$)	-2.091	0.621	1.221	0.104	-0.728
(Estimated SD)	(1.709)	(0.288)	(0.196)	(0.378)	(0.393)
95% C.I. (lower bound)	-5.441	0.057	0.838	-0.638	-1.498
95% C.I. (upper bound)	1.260	1.186	1.605	0.845	0.042

Table 2.11: Regression coefficient estimates when interaction terms are included in the model

Estimate	EF	Gender	Age (stand.)	Surgery < 3 months	Surgery > 3 months	EF * Surgery < 3 months	EF * Surgery > 3 months
Naive	-6.727	0.713	1.267	-4.332	-2.357	8.563	3.172
(Estimated SD)	(2.157)	(0.271)	(0.193)	(1.428)	(1.378)	(2.741)	(2.723)
95% C.I. (lower bound)	-11.040	0.170	0.881	-7.187	-5.112	3.082	-2.275
95% C.I. (upper bound)	-2.413	1.256	1.654	-1.477	0.398	14.045	8.619
SIMEX ($v = 0.05$)	-7.587	0.723	1.272	-5.095	-2.631	9.962	3.679
SIMEX ($v = 0.10$)	-10.556	0.772	1.269	-7.005	-3.098	13.619	4.653

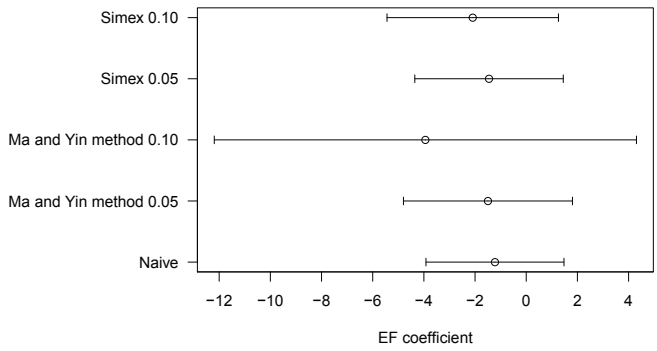


Figure 2.3: 95% confidence intervals for the EF parameter.

The parameter that is the most affected when taking the measurement error into account is the coefficient of the ejection fraction. Both methods correct in the same direction. However, the SIMEX approach (with a quadratic extrapolant) yields a more conservative correction, as already reported by Carroll et al. (2006). This smaller correction is associated with a smaller estimated standard deviation (which is consistent with what was observed in the simulation study), and hence a narrower confidence interval, as can be observed on Figure 2.3. It can be seen that correcting for the measurement error increases the size of the estimated effect of the ejection fraction. The correction also impacts the estimated effect of the group "surgery after 3 months". This can be explained by the fact that the mismeasured ejection fraction is linked to the surgery strategy, since existing guidelines advise to perform surgery when the EF is below a given threshold. However, the effects of the ejection fraction and of the surgery strategy are not significant.

We also considered introducing an interaction between the ejection fraction level and the surgery strategy. However, note that an interaction term between a mismeasured covariate and a correctly measured one is actually a mismeasured covariate whose variance depends on the latter covariate. The SIMEX algorithm can easily be tuned to accommodate such a case and yield parameter estimates; however, the asymptotic results of Section 2.3 do not hold anymore, and new asymptotic results are the subject of future research. Nevertheless, the bootstrap could be used to conduct inference on the estimated parameters. As far as the corrected score method is concerned, it is unclear how to modify it to take into account a dependence between an error term and a covariate. In Table 2.11, we report the results for this model. When the measurement error is not taken into account, one of the interaction terms is significant, and its

introduction in the model modifies the significance of other parameters. The estimated parameters obtained with SIMEX are also reported: as before, the correction leads to higher (in absolute value) estimated effects for the mismeasured covariates, but also for the covariates included in the interaction terms. We observe a negative estimated effect of the ejection fraction on the survival for patients with surgery after more than three months as well as for those without surgery. This effect is statistically significant (according to the naive estimates). In the promotion time cure model, a negative coefficient implies an increase in the cure probability and in the survival (at all times), when the value of the covariate increases. When this effect is significant, then, for patients without surgery, 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. There is no significant effect (according to the naive estimates) of the ejection fraction in the patients having undergone surgery within the first three months. This can probably be explained by the fact that these patients are operated on due to the presence of symptoms, independently of the value of their ejection fraction: the decision of whether to operate immediately was taken according to existing guidelines, based on the prognosis of the patients.

2.6 Proofs

2.6.1 Proof of Theorem 1

For a fixed λ and a fixed b, we have from Theorem 1 in Zeng et al. (2006) that

$$\|\widehat{\boldsymbol{\beta}}_{\lambda,b} - \boldsymbol{\beta}_\lambda\| \to 0 \quad \text{and} \quad \sup_{t\in\mathbb{R}^+} |\widehat{F}_{\lambda,b}(t) - F_\lambda(t)| \to 0,$$

with probability 1. As a result, for a fixed λ, since $\widehat{\boldsymbol{\beta}}_\lambda = \frac{1}{B}\sum_{b=1}^{B} \widehat{\boldsymbol{\beta}}_{\lambda,b}$ and $\widehat{F}_\lambda = \frac{1}{B}\sum_{b=1}^{B} \widehat{F}_{\lambda,b}$, by the continuous mapping theorem,

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

with probability 1.

An immediate result is that

$$\|\widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda}) - \boldsymbol{\beta}(\boldsymbol{\Lambda})\| \to 0 \quad \text{and} \quad \sup_{t\in\mathbb{R}^+} |\widehat{\boldsymbol{F}}(\boldsymbol{\Lambda}, t) - \boldsymbol{F}(\boldsymbol{\Lambda}, t)| \to 0,$$

with probability 1, where $\boldsymbol{\beta}(\boldsymbol{\Lambda}) = (\boldsymbol{\beta}_{\lambda_1}^T, \ldots, \boldsymbol{\beta}_{\lambda_K}^T)^T$ and $\boldsymbol{F}(\boldsymbol{\Lambda}, t) = \{F_{\lambda_1}(t), \ldots, F_{\lambda_K}(t)\}^T$, for a finite grid $\boldsymbol{\Lambda} = (\lambda_1, \ldots, \lambda_K)^T$.

We now take the extrapolation step of the SIMEX algorithm into account. We first show the consistency of $\widehat{\boldsymbol{\beta}}_{SIMEX}$. Similarly to what is done in Li and Lin (2003), 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 $\boldsymbol{\gamma}_{\boldsymbol{\beta}}$. Under the assumption that this 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

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

and hence, with probability 1,

$$\| \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}} - \boldsymbol{\gamma}_{\boldsymbol{\beta}} \| \to 0,$$

by the continuous mapping theorem, if $\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})$ is bounded and $\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, \boldsymbol{\Lambda})$ is invertible. Finally, by using once again the continuous mapping theorem, since $\widehat{\boldsymbol{\beta}}_{SIMEX} = \boldsymbol{g}_{\boldsymbol{\beta}}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}}, -1)$ and $\boldsymbol{\beta}_{TRUE} = \boldsymbol{g}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}}, -1)$, we find that, with probability 1,

$$\| \widehat{\boldsymbol{\beta}}_{SIMEX} - \boldsymbol{\beta}_{TRUE} \| \to 0.$$

We finally turn to the proof of the consistency of $\widehat{F}_{SIMEX}$. Similarly to what is done in Li and Lin (2003), suppose that, for a fixed t, $F_\lambda(t)$ can be specified using a parametric model $\boldsymbol{g}_t(\boldsymbol{\gamma}_t, \lambda)$ depending on a parameter vector $\boldsymbol{\gamma}_t$. If this is the true extrapolation function, we have $F_{TRUE}(t) = \boldsymbol{g}_t(\boldsymbol{\gamma}_t, -1)$ and $\widehat{F}_{SIMEX}(t) = \boldsymbol{g}_t(\widehat{\boldsymbol{\gamma}}_t, -1)$, where $\widehat{\boldsymbol{\gamma}}_t$ is a solution of

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

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

$$\widehat{\boldsymbol{\gamma}}_t - \boldsymbol{\gamma}_t = \left\{ \dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda}) \right\}^{-1} \dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \left\{ \widehat{\boldsymbol{F}}(\boldsymbol{\Lambda}, t) - \boldsymbol{F}(\boldsymbol{\Lambda}, t) \right\} + o_p(1)$$

and hence, with probability 1,

$$\sup_{t\in\mathbb{R}^+} \|\widehat{\gamma}_t - \gamma_t\| \to 0,$$

by the continuous mapping theorem, assuming that $\dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})$ is bounded uniformly in t and that $\dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})^T \dot{\boldsymbol{g}}_t(\boldsymbol{\gamma}_t, \boldsymbol{\Lambda})$ is invertible. Finally, by using once again the continuous mapping theorem, since $\widehat{F}_{SIMEX}(t) = \boldsymbol{g}_t(\widehat{\boldsymbol{\gamma}}_t, -1)$ and $F_{TRUE}(t) = \boldsymbol{g}_t(\boldsymbol{\gamma}_t, -1)$, we find that, with probability 1,

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

2.6.2 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

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

where f is the density function corresponding to F.

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

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

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

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

$$= n^{-1} \sum_{i=1}^{n} \psi_\lambda(T_i, \boldsymbol{W}_{i,\lambda,b}, \boldsymbol{h}_1, h_2) + o_p(n^{-1/2}),$$

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 $(\Omega^{-1}_{\lambda,\boldsymbol{\beta}}, \Omega^{-1}_{\lambda,F})$ 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,

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

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

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

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

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

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

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

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

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

$$n^{1/2}\begin{Bmatrix} (\widehat{\boldsymbol{\beta}}_{\lambda_1} - \boldsymbol{\beta}_{\lambda_1})^T\boldsymbol{h}_1 + \int_0^\infty h_2\ d(\widehat{F}_{\lambda_1} - F_{\lambda_1}) \\ \vdots \\ (\widehat{\boldsymbol{\beta}}_{\lambda_K} - \boldsymbol{\beta}_{\lambda_K})^T\boldsymbol{h}_1 + \int_0^\infty h_2\ d(\widehat{F}_{\lambda_K} - F_{\lambda_K}) \end{Bmatrix}$$

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

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

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 $n^{1/2}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda}) - \boldsymbol{\beta}(\boldsymbol{\Lambda}))$ 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 $n^{1/2}\left\{\widehat{\boldsymbol{F}}(\boldsymbol{\Lambda},t)-\boldsymbol{F}(\boldsymbol{\Lambda},t)\right\}$ 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

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\beta}}-\boldsymbol{\gamma}_{\boldsymbol{\beta}})=&\left\{\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})^T\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})\right\}^{-1}\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})^T n^{1/2}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\Lambda})-\boldsymbol{\beta}(\boldsymbol{\Lambda}))\\&+o_p(1)\end{aligned}$$

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

$$\begin{aligned}n^{1/2}(\widehat{\boldsymbol{\beta}}_{SIMEX}-\boldsymbol{\beta}_{TRUE})\longrightarrow&\,\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},-1)\left\{\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})^T\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})\right\}^{-1}\\&\times\dot{\boldsymbol{g}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}_{\boldsymbol{\beta}},\boldsymbol{\Lambda})^T N(\mathbf{0},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}).\end{aligned}$$

Finally, we will show that $n^{1/2}(\widehat{F}_{SIMEX}-F_{TRUE})$ 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

$$\begin{aligned}&n^{1/2}(\widehat{\boldsymbol{\gamma}}_t-\boldsymbol{\gamma}_t)\\&\quad=\left\{\dot{g}_t(\boldsymbol{\gamma}_t,\boldsymbol{\Lambda})^T\dot{g}_t(\boldsymbol{\gamma}_t,\boldsymbol{\Lambda})\right\}^{-1}\dot{g}_t(\boldsymbol{\gamma}_t,\boldsymbol{\Lambda})^T n^{1/2}\left\{\widehat{\boldsymbol{F}}(\boldsymbol{\Lambda},t)-\boldsymbol{F}(\boldsymbol{\Lambda},t)\right\}+o_p(1)\end{aligned}$$

for all t, and hence the process $n^{1/2}(\widehat{\gamma}_t - \gamma_t)$ (indexed by $t \in \mathbb{R}^+$) converges to the Gaussian process

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

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

$$n^{1/2}(\widehat{F}_{SIMEX} - F_{TRUE}) \longrightarrow \dot{g}_t(\gamma_t, -1)^T \left\{\dot{g}_t(\gamma_t, \mathbf{\Lambda})^T \dot{g}_t(\gamma_t, \mathbf{\Lambda})\right\}^{-1} \dot{g}_t(\gamma_t, \mathbf{\Lambda})^T \mathcal{G}.$$

2.7 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 corrected score approach of Ma and Yin (2008) cannot be clearly discriminated on the basis of the bias, while in the practical setting, SIMEX has a lower bias for the lower measurement error variance, but a bias larger than with the corrected score 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 SIMEX

and the corrected score approach could depend on the desired property of the obtained estimates: a pure reduction of the bias, or a decrease in the MSE.

Based on some further simulation studies, we concluded that the choice of the grid of λ has nearly no impact on the results, while the choice of the extrapolant must be made in view of the bias-variance tradeoff. The robustness of SIMEX with respect to two assumptions underlying it (as well as the method of Ma and Yin) was also investigated: our estimator seems quite robust in many cases, and a lack of robustness only appears in the most extreme cases (an asymmetrical distribution, a large variance). Finally, we analyzed, using our proposed estimator, the effect of an error-prone covariate on the survival time for patients suffering from a heart disease.

CHAPTER 3

Robustness of the estimation methods

The material presented in this chapter is published in the paper "Robustness of estimation methods in a survival cure model with mismeasured covariates", by Bertrand A., Legrand C., Léonard D. and Van Keilegom I. in Computational Statistics and Data Analysis 2017, 113, 3–18.

3.1 Introduction

In this chapter, as in the previous one, we study the semiparametric promotion time cure model (2.1) whose covariates are subject to classical measurement error, according to Equation (2.2), where $\boldsymbol{U}$ is independent of $\boldsymbol{X}$ and follows a continuous distribution with mean zero and known covariance matrix $\boldsymbol{V}$. The elements of $\boldsymbol{V}$ corresponding to covariates with no measurement error are set to 0. We still assume that, due to right censoring, we observe $(Y_i, \delta_i, \boldsymbol{X}_i)$ where $Y = \min(T, C)$ and $\delta = I(T \leq C)$, for $i = 1, \ldots, n$, and that the vectors $(T_i, C_i, \boldsymbol{X}_i)$ are independent and identically distributed, with the same distribution as a generic vector $(T, C, \boldsymbol{X})$.

When (some of) the covariates are mismeasured, Zeng et al. (2006)'s technique and Ma and Yin (2008)'s non-corrected approach, two estimation methods for the promotion time cure model introduced in Section 2.2, yield biased estimators. This is a well-known fact which holds for many statistical models; however, the form of this bias in the context of a promotion time cure model

has, to the best of our knowledge, never been presented. This chapter contains the first study of the form of the bias in this context.

Both Ma and Yin (2008)'s corrected score approach and our SIMEX method, introduced and compared via some simulations in the previous chapter, proved to be useful for correcting for bias due to mismeasured covariates in some simple settings. However, their utility could be limited in concrete applications, since they require precise information about the distribution of the measurement error, which is often not available in practice. In particular, these methods rely on two main assumptions. The first one is the normality of the measurement error. This is required for Ma and Yin (2008)'s method and, although this is not the case for SIMEX, its practical performance in the non-Gaussian situation has never been thoroughly verified for the promotion time cure model. The second strong assumption is the correct specification of the measurement error variance, which is necessary for both correction methods.

In practice, except in some specific contexts, the user rarely knows the exact form of the distribution of the error, or even its variance. In these cases, the superiority of a correction method over the naive one, which does not take the measurement error into account, is not guaranteed. As a consequence, it is important to study the robustness of these methods with respect to their assumptions. We therefore performed an extensive simulation study aiming at investigating these questions and providing practical recommendations as to whether a correction method is useful, and which one is likely to provide the best results in various settings. Throughout this simulation study, we denote by "Naive" the naive method (i.e., not correcting for the measurement error) based on the backfitting approach, by "MY" Ma and Yin (2008)'s corrected score approach and by "SIMEX" the SIMEX algorithm that we adapted to the promotion time cure model in Chapter 2. Our R (R Core Team, 2015) implementation of the different estimation methods used in these simulations is provided in the R package `miCoPTCM`, available on the CRAN website (cran.r-project.org). A usage example of this package can be found in Appendix B.

The structure of this chapter is as follows. In the next section, we derive an expression for the bias of the regression parameter estimators obtained with the naive method. The details of the calculations are presented in Section 3.5. Section 3.3 presents the simulation results and recommendations pertaining to the issues of non-Gaussian distribution and unknown variance of the measurement error. This allows us to elaborate a strategy for studying the time until recurrence of cancer for patients suffering from rectal cancer; the obtained results are presented in Section 3.4. Finally, the conclusions and limitations of

this study as well as some ideas for further research are discussed in Section 3.6.

3.2 The bias of the naive estimator

We first derive an expression for the asymptotic bias that appears in the naive estimator when we do not take the measurement error into account. Such an expression has never been obtained in the context of the promotion time cure model with mismeasured covariates, even though assessing this bias helps justify the need for correction methods, even in large samples.

We follow an approach similar to the ones used by Hughes (1993) and Li and Lin (2000), but allowing here for several covariates, in order to obtain an equation containing both the large-sample estimator of $\boldsymbol{\beta}$ and the true parameter.

When we assume that no measurement error is present in the data, the promotion time cure model can be fitted using Ma and Yin (2008)'s backfitting approach introduced in Section 2.2. This approach alternates between two steps. We first solve the score equations for the p_i's and λ_{MY} (a Lagrange multiplier) while holding $\boldsymbol{\beta}$ constant, and obtain $\widehat{\lambda}_{MY,\beta}$ and the $\widehat{p}_{i,\beta}$'s. Then, we solve the score equations for $\boldsymbol{\beta}$, holding the p_i's and λ_{MY} at the values $\widehat{\lambda}_{MY,\beta}$ and $\widehat{p}_{i,\beta}$. For convenience, in this section, we make explicit the fact that $\boldsymbol{X}$ contains a constant, so that the first element of $\boldsymbol{\beta}$ is β_0, an intercept. We let $\widetilde{\boldsymbol{X}}$ and $\widetilde{\boldsymbol{W}}$ be the vectors without constant, and $\widetilde{\boldsymbol{\beta}}$ the corresponding regression parameters, so that $\boldsymbol{\beta} = (\beta_0, \widetilde{\boldsymbol{\beta}}^T)^T$, $\boldsymbol{X} = (1, \widetilde{\boldsymbol{X}}^T)^T$ and $\boldsymbol{W} = (1, \widetilde{\boldsymbol{W}}^T)^T$. The equation to be solved for $\boldsymbol{\beta}$ is

$$n^{-1} \sum_{i=1}^{n} \left\{ \delta_i - \widehat{F}_{\beta}(Y_i) e^{\boldsymbol{X}_i^T \boldsymbol{\beta}} \right\} (1, \widetilde{\boldsymbol{X}}_i^T)^T = \boldsymbol{0}, \tag{3.1}$$

where $\widehat{F}_{\beta}(Y_i) = \sum_{Y_j \leq Y_i, \delta_j = 1} \widehat{p}_{j,\beta}$.

When there is measurement error in the covariates, such that $\boldsymbol{W}$ is observed instead of $\boldsymbol{X}$, then the naive estimator $\widehat{\boldsymbol{\beta}}^*$ solves (in $\boldsymbol{\beta}^*$)

$$n^{-1} \sum_{i=1}^{n} \delta_i (1, \widetilde{\boldsymbol{W}}_i^T)^T - n^{-1} \sum_{i=1}^{n} \widehat{F}_{\beta^*}(Y_i) e^{\boldsymbol{W}_i^T \boldsymbol{\beta}^*} (1, \widetilde{\boldsymbol{W}}_i^T)^T = \boldsymbol{0}. \tag{3.2}$$

Without loss of generality, $\widetilde{\boldsymbol{X}}$ is assumed to be standardized, so that $E(\widetilde{\boldsymbol{X}}) = \boldsymbol{0}$, $V(\widetilde{\boldsymbol{X}}) = \boldsymbol{I}_D$ ($D = P - 1$ being the dimension of $\widetilde{\boldsymbol{X}}$) and

$$V(\widetilde{\boldsymbol{W}} | \widetilde{\boldsymbol{X}}) = V(\widetilde{\boldsymbol{X}})^{-1/2} V(\boldsymbol{U}) V(\widetilde{\boldsymbol{X}})^{-1/2} = \boldsymbol{\Lambda},$$

assuming that the same transformation was applied to the error as to $\widetilde{\boldsymbol{X}}$, i.e. $V(\widetilde{\boldsymbol{X}})^{-1/2}(\cdot - E(\widetilde{\boldsymbol{X}}))$.

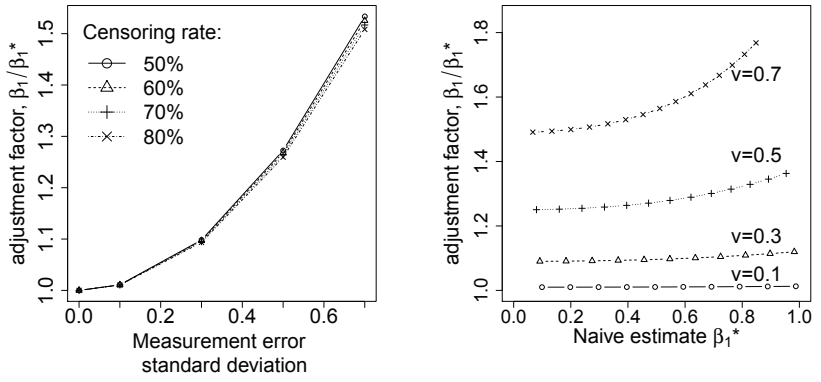
To derive the expression of the bias, as in Hughes (1993) and Li and Lin (2000), we assume type I censoring, such that the observed time for individual i is $Y_i = \min(T_i, t_c)$ for a fixed t_c. This is the case, for instance, when all patients in a study are monitored for a fixed length of time. When n is large, (3.2) tends in probability to

$$
\begin{aligned}
&E_{\widetilde{\boldsymbol{W}},T}(I(T \le t_c)(1, \widetilde{\boldsymbol{W}}^T)^T) - E_{\widetilde{\boldsymbol{W}},Y}\left[F_{\beta^*}(Y)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}(1, \widetilde{\boldsymbol{W}}^T)^T\right] \\
&= \begin{pmatrix} E_T(I(T \le t_c)) \\ E_{\widetilde{\boldsymbol{W}},T}(I(T \le t_c)\widetilde{\boldsymbol{W}}) \end{pmatrix} - \begin{pmatrix} E_{\widetilde{\boldsymbol{W}},Y}\left[F_{\beta^*}(Y)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\widetilde{\boldsymbol{W}},Y}\left[F_{\beta^*}(Y)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\widetilde{\boldsymbol{W}}\right] \end{pmatrix} \\
&= \begin{pmatrix} E_T(I(T \le t_c)) \\ E_{\widetilde{\boldsymbol{W}},T}(I(T \le t_c)\widetilde{\boldsymbol{W}}) \end{pmatrix} - \begin{pmatrix} E_{\widetilde{\boldsymbol{W}},T}\left[I(T \le t_c)F_{\beta^*}(T)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\widetilde{\boldsymbol{W}},T}\left[I(T \le t_c)F_{\beta^*}(T)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\widetilde{\boldsymbol{W}}\right] \end{pmatrix} \\
&\quad - \begin{pmatrix} E_{\widetilde{\boldsymbol{W}},T}\left[I(T > t_c)F_{\beta^*}(t_c)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\right] \\ E_{\widetilde{\boldsymbol{W}},T}\left[I(T > t_c)F_{\beta^*}(t_c)e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\widetilde{\boldsymbol{W}}\right] \end{pmatrix} \\
&= \begin{pmatrix} E_T(I(T \le t_c)) \\ E_T\left[I(T \le t_c)E_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{W}})\right] \end{pmatrix} \\
&\quad - \begin{pmatrix} E_T\left[I(T \le t_c)F_{\beta^*}(T)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\right)\right] \\ E_T\left[I(T \le t_c)F_{\beta^*}(T)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\widetilde{\boldsymbol{W}}\right)\right] \end{pmatrix} \\
&\quad - \begin{pmatrix} E_T\left[I(T > t_c)F_{\beta^*}(t_c)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\right)\right] \\ E_T\left[I(T > t_c)F_{\beta^*}(t_c)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\boldsymbol{W}^T\boldsymbol{\beta}^*}\widetilde{\boldsymbol{W}}\right)\right] \end{pmatrix} = \mathbf{0}. \qquad (3.3)
\end{aligned}
$$

Each of the three above expectations can be calculated in the case of the promotion time cure model (2.1) and the classical error model (2.2). Details are given in Section 3.5. If we assume that $(\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{x}}) \sim N(\widetilde{\boldsymbol{x}}, \boldsymbol{\Lambda})$, we obtain the following formulation of (3.3):

$$
\begin{pmatrix} 1 - \int e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}} \\ -\int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\widetilde{\boldsymbol{x}} \end{pmatrix}
$$

$$
-e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}
\times \begin{pmatrix} \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{\int_0^{t_c} F_{\beta^*}(t)\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\} d\widetilde{\boldsymbol{x}} \\ \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{\int_0^{t_c} F_{\beta^*}(t)\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt\right\} d\widetilde{\boldsymbol{x}} \end{pmatrix}
$$



(a) Adjustment factor as a function of the standard deviation of the measurement error, for different values of the censoring rate.

(b) Adjustment factor as a function of the naive estimate, for different values of the standard deviation v of the measurement error.

Figure 3.1: Two examples of plots that can be obtained from Equation (3.3).

$$-e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c) \times \begin{pmatrix} \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\, d\widetilde{\boldsymbol{x}} \\ \int (\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\, d\widetilde{\boldsymbol{x}} \end{pmatrix} = \mathbf{0}.$$

This equation can be solved numerically, for $\boldsymbol{\beta}^*$ given $\boldsymbol{\beta}$, in order to obtain the relationship between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$. The effect of each parameter of a particular setting on the asymptotic bias can then be investigated and displayed on a graph. As an illustration, we consider a model with $\beta_0 = -0.1$, $\widetilde{\boldsymbol{\beta}} = (0.5, -0.5)^T$, X_1 and X_2 are normally distributed with $V(X_1) = 1$, $V(X_2) = 1$ and $Cov(X_1, X_2) = 0$. We assume that X_1 is measured with some error, while X_2 is not. As mentioned by Li and Lin (2000) in the context of frailty models, the asymptotic bias is not easily computed when the baseline function is left unspecified. We hence proceed in a similar fashion to Li and Lin (2000) and assume (for this illustration) that the baseline cumulative distribution function is exponential with mean 8. We use four different values for the censoring rate, 50%, 60%, 70% and 80%. This can be achieved by appropriately choosing t_c, since the censoring rate equals $\int S(t_c|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\, d\widetilde{\boldsymbol{x}}$. Figure 3.1a represents the adjustment factor for the coefficient related to the mismeasured covariate, i.e. the ratio β_1/β_1^*, as a function of the measurement error variance. We see that, in this particular setting, the censoring rate does

not influence the adjustment factor very much. Moreover, as expected, bias increases with the measurement error variance.

However, as explained by Hughes (1993), a graphical representation of greater practical interest is one such as Figure 3.1b: it allows one to deduce the adjustment factor (and hence, the true value β_1) to be expected for a known value of the estimate β_1^* obtained using the naive method. Here, models with values of β_1 between 0.1 and 1.5 with a censoring rate of 70% were estimated. The ratio β_1/β_1^* is plotted as a function of the asymptotic naive estimate, for different values of the measurement error standard deviation. Here again, it appears that the adjustment factor increases with the measurement error standard deviation. Moreover, for the largest values of variance, the adjustment factor is not constant, but increases with the value of the naive estimate.

3.3 Simulation study

The previous section justifies the need for correction in order to avoid (or, at least, reduce) the bias of the estimated regression coefficients. However, as was discussed above, both the corrected score approach (Ma and Yin, 2008) and the SIMEX algorithm developed in Chapter 2 require two types of information about the error that are not always known in practice. In this section, we present the results of an extensive simulation study investigating the robustness of both approaches (as well as of the naive method) with respect to the assumptions of normality and known variance of the measurement error. These results provide practical recommendations as to whether a correction method should be used in a context, and if so, which one. These recommendations are based on two criteria, the mean squared error (MSE) and the bias. Techniques dealing with measurement error classically focus on the latter criterion; the former allows for a balance between the decrease in bias and the increase in variance.

The model used for simulating the data is

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

where the parameters were chosen to mimic a possible practical setting, i.e. $F(t)$ is the cumulative distribution function (cdf) of an exponential distribution with mean 6, truncated at $t = 20$, $X_2 \sim Bernoulli(0.5)$, $X_1 = (Y_1 - 0.5)(1/12)^{-1/2}$ where $Y_1 \sim Uni[0, 1]$, such that $E(X_1) = 0$ and $V(X_1) = 1$. X_2 can correspond to a treatment indicator, while X_1 can represent a continuous biomarker measured with error. The censoring times are generated (independently of the covariates and of the survival time) from an exponential

distribution with mean 5 truncated at $t = 30$. The sample size is 200. An example of data generation in R can be found in Appendix A.

For each setting, 500 replications were performed. We report the bias and MSE of both covariate coefficient estimators, $\widehat{\beta}_1$ and $\widehat{\beta}_2$. Another quantity of interest in practical applications is the estimated cure rate for given covariate values. We denote by CR_α the estimated cure rate at the quantile of order α of the continuous covariate X_1 (at a constant value of X_2). We computed the MSE and bias of $\widehat{CR}_{.25}$ and $\widehat{CR}_{.75}$ in all settings; these results are not presented here, but the most important conclusions are reported.

3.3.1 Non-normality of the measurement error

In this first section, we investigate the robustness of each method with respect to the assumption of normality of the measurement error, which can rarely be assessed in practice. This is required for MY; SIMEX can explicitly accommodate non-Gaussian distributions, but its practical performance in this context has never been verified.

For the sake of brevity, we restrict our attention to three representative settings (summarized in Table 3.1), which cover a large range of cure and censoring proportions and which include both continuous and binary covariates. The true cure rate reported in the table is the proportion of observations that are generated as cured, while the observed cure rate is the proportion of censored observations with a censoring time larger than the largest observed event time (the cure threshold).

Table 3.1: The three settings considered in the simulation study

	Setting 1	Setting 2	Setting 3
β_0	1	0.2	-0.5
β_1	0.5	1	0.75
β_2	-0.5	-0.5	-0.5
censoring rate	42%	56%	69%
true cure rate	16%	39%	58%
observed cure rate	3%	6%	9%

We assume that X_1 is measured with error, so that $W = X_1 + U$ is observed. Four different settings are used for the distribution of the error U. In the first one, $U \sim N(0, v^2)$ with variance $v^2 \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7\}$. This Gaussian case will be a standard of comparison, but it will also be used to investigate the effect of an increase in the measurement error variance on estimation. The second distribution under study is the Student-t: $U = aT$ where $T \sim t_\nu$, such that $V(U) = a^2 \frac{\nu}{\nu-2}$. We take $a = 0.5$ and $\nu \in \{4, 10\}$, hence

$V(U) \in \{0.5, 0.3125\}$. This case allows us to consider a distribution close to the Gaussian one, but with heavier tails. In the third setting, we use an asymmetrical distribution: $U = b(T - k)$ where $T \sim \chi^2_k$, such that $V(U) = 2b^2k$. We take $b = \sqrt{0.1}$ and $k \in \{1, 2, 3\}$, hence $V(U) \in \{0.2, 0.4, 0.6\}$. Finally, we consider the case where $U = T - \mu$ where $T \sim Laplace(\mu, b)$, with variance $2b^2$. This is another symmetrical distribution. We take $b \in \{\sqrt{0.05}, \sqrt{0.15}, \sqrt{0.25}\}$, so that $V(U) \in \{0.1, 0.3, 0.5\}$.

Results

The numerical results pertaining to the Gaussian distribution can be found in Tables 3.2, 3.4 and 3.6 for Settings 1, 2 and 3, respectively. The results for the non-Gaussian distributions can be found in Tables 3.3, 3.5 and 3.7.

As already observed in the previous chapter, SIMEX is usually the best method for decreasing the MSE, while MY often yields the smallest bias.

The sign of this bias clearly depends on the estimation method. The bias of $\widehat{\beta}_1$ is usually negative with Naive and SIMEX, and positive with MY (except with the Chi-squared distribution). This illustrates the fact that MY tends to make too much of a bias correction, a phenomenon that already appeared in the simulation results in Ma and Yin (2008) and in Chapter 2. Thus it corrects more than SIMEX, which is known to be more conservative, particularly with the quadratic extrapolant (Cook and Stefanski, 1994).

Moreover, the effect of an increasing measurement error variance also depends on the method under consideration. With Naive, both the bias and the MSE of $\widehat{\beta}_1$ increase with the measurement error variance, while this is not necessarily the case for $\widehat{\beta}_2$. This increase in MSE is due to the increase in bias, since variance tends to decrease (this decrease was already observed by Ma and Yin (2008), as a result of more variation in the covariate). The conclusion is the same for SIMEX, which could be expected, since this method depends heavily on the naive estimates. With MY, there is no clear trend for the bias, while the MSE of both $\widehat{\beta}_1$ and $\widehat{\beta}_2$ increases with the measurement error variance, driven by increases in the variance of the estimators.

The bias of $\widehat{\beta}_1$ has consequences on the estimated cure rates. $\widehat{CR}_{.25}$ is an increasing function of $\widehat{\beta}_1$, while $\widehat{CR}_{.75}$ is a decreasing function of this parameter. As a consequence, *for a fixed value of* $\widehat{\beta}_0$, an underestimation of β_1 yields a decrease in the estimation of $CR_{.25}$. This results in a decrease in the bias of $\widehat{CR}_{.25}$, and an increase in the bias of $\widehat{CR}_{.75}$. Furthermore, the conclusions for $\widehat{CR}_{.25}$ are less clear than those for $\widehat{CR}_{.75}$. When $\widehat{\beta}_1$ and X_1 do not have the same sign (in our settings, this is the case for $\widehat{CR}_{.25}$), $\widehat{\beta}_0 + \widehat{\beta}_1 X_1$ tends to be small, so that small changes in $\widehat{\beta}_1$ or in $\widehat{\beta}_0$ only have a small effect on $\exp(\widehat{\beta}_0 + \widehat{\beta}_1 X_1)$, and consequently on the estimated cure rate. On the other

Table 3.2: Assumption of normality of the measurement error in Setting 1: results for the Gaussian distribution.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Gaussian	.1	Bias	-.040	-.014	.015	-.020	.007	-.018	.007	-.018
		Var	.009	.038	.012	.040	.011	.040	.011	.040
		MSE	.011	.039	.012	.040	.011	.040	.011	.040
	.2	Bias	-.080	-.010	.022	-.022	-.001	-.018	-.001	-.018
		Var	.008	.038	.015	.041	.012	.040	.012	.040
		MSE	.015	.038	.015	.041	.012	.041	.012	.041
	.3	Bias	-.114	-.007	.030	-.024	-.014	-.017	-.014	-.017
		Var	.007	.038	.018	.042	.013	.041	.013	.041
		MSE	.020	.038	.019	.043	.013	.041	.013	.041
	.4	Bias	-.143	-.005	.039	-.027	-.030	-.016	-.030	-.016
		Var	.007	.038	.023	.044	.013	.041	.013	.041
		MSE	.027	.038	.024	.045	.014	.042	.014	.042
	.5	Bias	-.168	-.002	.048	-.030	-.048	-.015	-.048	-.015
		Var	.006	.038	.029	.046	.013	.042	.013	.042
		MSE	.034	.038	.031	.047	.015	.042	.015	.042
	.6	Bias	-.191	.001	.055	-.030	-.067	-.012	-.067	-.012
		Var	.006	.038	.034	.048	.012	.042	.012	.042
		MSE	.042	.038	.037	.049	.017	.042	.017	.042
	.7	Bias	-.212	.004	.061	-.029	-.087	-.008	-.087	-.008
		Var	.005	.038	.038	.050	.011	.041	.011	.041
		MSE	.050	.038	.042	.051	.019	.042	.019	.042

Table 3.3: Assumption of normality of the measurement error in Setting 1: results for the non-Gaussian distributions.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Student-t	.3125	Bias	-.125	-.002	.017	-.017	-.025	-.011	-.026	-.011
		Var	.007	.039	.019	.044	.013	.042	.013	.042
		MSE	.023	.039	.020	.044	.014	.042	.014	.042
	.5	Bias	-.170	.000	.043	-.026	-.051	-.012	-.050	-.012
		Var	.007	.038	.035	.048	.015	.041	.015	.042
		MSE	.036	.038	.037	.049	.018	.041	.018	.042
Chi-square	.2	Bias	-.098	-.006	-.011	-.016	-.017	-.014	-.028	-.013
		Var	.008	.040	.015	.044	.014	.043	.013	.043
		MSE	.018	.040	.015	.044	.015	.043	.014	.043
	.4	Bias	-.163	-.001	-.024	-.018	-.054	-.012	-.067	-.010
		Var	.007	.039	.021	.046	.016	.043	.014	.042
		MSE	.034	.039	.022	.046	.019	.043	.019	.042
	.6	Bias	-.213	.005	-.044	-.017	-.101	-.006	-.110	-.006
		Var	.006	.039	.027	.049	.014	.043	.013	.042
		MSE	.051	.039	.029	.049	.025	.043	.025	.042
Laplace	.1	Bias	-.043	-.010	.011	-.015	.004	-.014	.003	-.014
		Var	.009	.039	.012	.041	.012	.041	.012	.040
		MSE	.011	.039	.012	.041	.012	.041	.012	.041
	.3	Bias	-.119	-.001	.020	-.016	-.019	-.009	-.021	-.009
		Var	.008	.039	.018	.044	.014	.042	.014	.042
		MSE	.022	.039	.019	.044	.014	.042	.014	.042
	.5	Bias	-.172	.004	.033	-.018	-.053	-.005	-.055	-.005
		Var	.007	.039	.030	.049	.015	.043	.015	.043
		MSE	.037	.039	.031	.049	.018	.043	.018	.043

Table 3.4: Assumption of normality of the measurement error in Setting 2: results for the Gaussian distribution.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Gaussian	.1	Bias	-.104	-.006	.041	-.024	.010	-.018	.010	-.018
		Var	.017	.060	.030	.065	.025	.063	.025	.063
		MSE	.028	.060	.032	.066	.025	.064	.025	.064
	.2	Bias	-.196	.003	.073	-.033	-.020	-.016	-.020	-.016
		Var	.015	.060	.051	.074	.027	.067	.027	.067
		MSE	.053	.060	.056	.075	.028	.067	.028	.067
	.3	Bias	-.269	.011	.107	-.040	-.058	-.014	-.058	-.014
		Var	.013	.060	.074	.083	.028	.069	.028	.069
		MSE	.086	.060	.086	.085	.031	.069	.031	.069
	.4	Bias	-.348	.019	.099	-.042	-.130	-.005	-.130	-.005
		Var	.010	.060	.082	.097	.021	.070	.021	.070
		MSE	.131	.060	.092	.099	.038	.070	.038	.070
	.5	Bias	-.408	.021	.060	-.037	-.189	-.001	-.189	-.001
		Var	.008	.056	.154	.101	.017	.064	.017	.064
		MSE	.174	.057	.157	.103	.053	.064	.053	.064
	.6	Bias	-.461	.024	.019	-.048	-.247	.002	-.247	.002
		Var	.007	.058	.152	.111	.015	.067	.015	.067
		MSE	.219	.059	.152	.113	.076	.067	.076	.067
	.7	Bias	-.506	.034	-.056	-.045	-.298	.013	-.298	.013
		Var	.006	.057	.291	.129	.013	.065	.013	.065
		MSE	.262	.059	.294	.131	.102	.065	.102	.065

Table 3.5: Assumption of normality of the measurement error in Setting 2: results for the non-Gaussian distributions.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Student-t	.3125	Bias	-.293	.022	.066	-.020	-.086	.001	-.088	.001
		Var	.015	.064	.080	.089	.034	.075	.033	.074
		MSE	.101	.064	.084	.090	.041	.075	.041	.074
	.5	Bias	-.419	.042	-.005	-.004	-.208	.023	-.213	.022
		Var	.012	.057	.182	.099	.029	.068	.028	.069
		MSE	.187	.059	.182	.099	.073	.069	.073	.069
Chi-square	.2	Bias	-.266	.014	-.075	-.012	-.087	-.006	-.131	-.001
		Var	.021	.069	.063	.088	.050	.083	.040	.080
		MSE	.091	.069	.069	.088	.058	.084	.057	.080
	.4	Bias	-.410	.034	-.163	.001	-.207	.013	-.247	.017
		Var	.015	.066	.064	.103	.042	.085	.034	.081
		MSE	.183	.067	.090	.103	.085	.085	.095	.081
	.6	Bias	-.499	.037	-.231	-.004	-.297	.013	-.328	.015
		Var	.011	.066	.057	.109	.033	.088	.027	.084
		MSE	.260	.067	.111	.109	.121	.088	.135	.084
Laplace	.1	Bias	-.113	.001	.027	-.016	-.001	-.012	-.002	-.011
		Var	.018	.063	.030	.070	.026	.068	.025	.067
		MSE	.030	.063	.031	.070	.026	.068	.025	.068
	.3	Bias	-.284	.022	.055	-.018	-.081	.000	-.085	.001
		Var	.014	.064	.075	.090	.031	.076	.030	.075
		MSE	.095	.064	.078	.090	.037	.076	.037	.075
	.5	Bias	-.417	.042	-.004	-.009	-.205	.021	-.209	.021
		Var	.009	.062	.074	.104	.021	.075	.021	.075
		MSE	.183	.064	.074	.104	.063	.075	.064	.075

hand, when $\widehat{\beta}_1$ and X_1 have the same sign, the estimated cure rate is much more sensitive to variations in the estimated parameters.

Finally, the three settings under consideration can be compared. Settings 1 and 3 have the lowest true values for β_1 (0.5 and 0.75, instead of 1). This could explain some observations: a smaller impact of the correction (since this correction mainly impacts $\widehat{\beta}_1$) and of the increased measurement error variance and a better robustness to a misspecification of the error distribution in X_1. This could also justify the similarity of the MSEs of the cure rate across methods and measurement error variances in Settings 1 and 3. Moreover, in Setting 1 (with the smallest value of β_1), the MSE of $\widehat{\beta}_2$ is quite similar among methods. Finally, with Naive, the bias of $\widehat{\beta}_2$ is usually smaller in Setting 1 than in the other two settings (although the value of β_2 used in the simulations is the same, -0.5).

To assess the robustness of a method to a misspecification of the measurement error distribution, we compare, for a given measurement error variance, the value of one criterion (MSE or bias) in the Gaussian case with the value of this criterion obtained with the other distributions.

We first consider the two estimation methods which cannot explicitly accommodate non-Gaussian distributions: Naive and MY. The MSE of $\widehat{\beta}_2$ with Naive seems robust (the increases in MSE are not large); this method stays relatively robust for β_1, except when the distribution of the measurement error is asymmetrical. As far as bias is concerned, Naive seems quite robust for β_1, except when the measurement error distribution is asymmetrical (in which case, the increase in bias can be very large). Naive is quite robust for β_2 for Settings 1 and 3, but a bit less in Setting 2.

MY is not as robust as Naive in terms of the MSE of $\widehat{\beta}_1$: it can stay similar, decrease, or increase (differences can be dramatic when the measurement error variance is large). MY seems robust for β_2 (the increases or decreases in MSE are much smaller). When considering bias, there is a decrease (sometimes very large) in the bias of $\widehat{\beta}_2$ with MY when switching from a Gaussian to a non-Gaussian distribution. This counterintuitive result could be explained by MY's previously observed tendency to correct more than necessary in the Gaussian case. The conclusion for β_1 is the same as for β_2, except when the error distribution is asymmetrical: in that case, the bias can either increase or decrease.

Since SIMEX can be run with a non-Gaussian distribution, two questions are of interest here: is SIMEX robust with respect to the distribution of the measurement error? If so, is it worth using the true distribution when it is known? For β_1 and β_2, the increase in MSE when the true distribution is not Gaussian

Table 3.6: Assumption of normality of the measurement error in Setting 3: results for the Gaussian distribution.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Gaussian	.1	Bias	-.073	-.032	.024	-.038	.005	-.035	.005	-.035
		Var	.028	.108	.045	.112	.039	.110	.039	.110
		MSE	.034	.110	.045	.114	.039	.112	.039	.112
	.2	Bias	-.140	-.029	.036	-.040	-.017	-.035	-.017	-.035
		Var	.024	.108	.053	.114	.037	.111	.037	.111
		MSE	.043	.109	.054	.116	.038	.112	.038	.112
	.3	Bias	-.194	-.027	.053	-.043	-.044	-.035	-.044	-.035
		Var	.020	.108	.069	.119	.036	.112	.036	.112
		MSE	.058	.109	.071	.121	.038	.114	.038	.114
	.4	Bias	-.245	-.029	.037	-.043	-.084	-.037	-.084	-.037
		Var	.017	.107	.202	.147	.031	.112	.031	.112
		MSE	.077	.108	.203	.149	.038	.113	.038	.113
	.5	Bias	-.289	-.026	.031	-.042	-.123	-.034	-.123	-.034
		Var	.015	.106	.163	.145	.029	.112	.029	.112
		MSE	.098	.107	.164	.147	.044	.113	.044	.113
	.6	Bias	-.325	-.023	.001	-.040	-.156	-.031	-.156	-.031
		Var	.013	.107	.313	.156	.028	.113	.028	.113
		MSE	.119	.107	.313	.157	.052	.114	.052	.114
	.7	Bias	-.355	-.014	-.020	-.024	-.187	-.022	-.187	-.022
		Var	.012	.107	.385	.235	.026	.114	.026	.114
		MSE	.139	.107	.385	.235	.062	.114	.062	.114

Table 3.7: Assumption of normality of the measurement error in Setting 3: results for the non-Gaussian distributions.

			Naive		MY		SIMEX (correct dist.)		SIMEX (Gaussian dist.)	
Distribution	Var.		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
Student-t	.3125	Bias	-.203	-.024	.049	-.038	-.053	-.030	-.051	-.034
		Var	.021	.111	.074	.126	.039	.117	.039	.118
		MSE	.062	.111	.076	.127	.041	.118	.042	.119
	.5	Bias	-.293	-.021	.025	-.045	-.129	-.028	-.131	-.029
		Var	.015	.110	.084	.141	.032	.117	.031	.118
		MSE	.101	.110	.084	.143	.049	.118	.048	.118
Chi-square	.2	Bias	-.177	-.029	-.040	-.039	-.049	-.035	-.074	-.034
		Var	.023	.114	.050	.123	.044	.120	.038	.120
		MSE	.055	.114	.051	.125	.047	.122	.043	.121
	.4	Bias	-.283	-.021	-.087	-.036	-.131	-.027	-.152	-.027
		Var	.017	.109	.054	.129	.036	.119	.032	.119
		MSE	.097	.109	.061	.131	.053	.119	.055	.120
	.6	Bias	-.348	-.022	-.130	-.043	-.191	-.030	-.201	-.028
		Var	.014	.111	.107	.145	.032	.124	.031	.124
		MSE	.135	.112	.124	.147	.069	.124	.071	.125
Laplace	.1	Bias	-.075	-.030	.019	-.035	.003	-.033	.003	-.033
		Var	.027	.112	.042	.117	.037	.116	.037	.115
		MSE	.033	.113	.042	.118	.037	.117	.037	.116
	.3	Bias	-.199	-.024	.034	-.039	-.050	-.029	-.047	-.029
		Var	.020	.112	.064	.128	.036	.119	.038	.119
		MSE	.059	.113	.065	.130	.039	.120	.040	.120
	.5	Bias	-.291	-.018	.007	-.040	-.127	-.023	-.129	-.024
		Var	.015	.110	.154	.135	.030	.119	.029	.119
		MSE	.099	.110	.154	.137	.046	.120	.046	.119

(compared to when it is) is not large, except in Setting 2 (and, for β_1, when the measurement error distribution is asymmetrical). Taking the true distribution into account when applying SIMEX does not improve the MSE (except for β_1 in Setting 2, when the measurement error distribution is asymmetrical). The bias of $\widehat{\beta}_1$ does not increase much when passing from the Gaussian to the non-Gaussian case, except in Setting 2, and when the measurement error distribution is asymmetrical. There can be increases or decreases of relatively moderate size in the bias of $\widehat{\beta}_2$. Taking the true distribution into account in SIMEX improves the bias of $\widehat{\beta}_1$ when the measurement error distribution is asymmetrical; otherwise, it yields results very similar to those obtained without taking the true distribution into account.

Finally, the last piece of information that can be retrieved from the simulation results is the smallest value of the measurement error variance for which a correction is needed. In terms of the MSE of $\widehat{\beta}_1$, it is always useful to apply SIMEX (i.e., the decrease in MSE seems significant), except in some cases for the lowest value of the measurement error variance. The conclusions concerning MY are not so clear. For β_2, the coefficient of the variable measured without error, a correction is never needed. In terms of bias, a correction is always useful for β_1, even for a small measurement error variance. There is no clear conclusion for β_2.

Recommendations

Based on the observations from the previous subsection, we can make the following recommendations.

When the main objective is to decrease MSE, the best strategy depends on which covariate effect is of interest. For β_2, the coefficient of the covariate without measurement error, the two correction methods that are at our disposal are not useful. If we nevertheless want to use a correction method (because, for example, the focus of the study is on another parameter for which correcting is needed), then SIMEX should be chosen, since it does not increase the MSE much. For β_1, the parameter related to the mismeasured covariate, we should always correct with SIMEX (when MY decreases the MSE, the decrease is smaller than with SIMEX, and in some cases, MY actually increases the MSE).

The strategy to be used for decreasing bias is slightly different. For β_2, there is now no clear conclusion about the usefulness of the correction, although the best correction method is SIMEX, since it nearly always decreases bias. For β_1, using a correction method is always profitable; the best correction method is MY, except when the distribution is Gaussian and the variance is small (under 0.3 or 0.4), in which case SIMEX should be preferred.

3.3.2 Misspecification of the measurement error variance

The second strong assumption that is not often met in practice is the knowledge of the measurement error variance. As a result, we would like to know the consequences of its misspecification on each correction method. The objective is not only to obtain practical recommendations about which estimation method should be used in a given setting, but also to discover whether we could benefit, in some cases, from deliberately over- or underspecifying this variance.

We present the results for the same three settings as in Section 3.3.1, with the measurement error generated as $U \sim N(0, v^2)$ in the simulations but assumed to be $N(0, v_E^2)$ in the estimation procedure. We consider $v^2 = 0.1$ associated with $v_E^2 \in \{0.05, 0.10, 0.15\}$ and $v^2 = 0.3$ associated with $v_E^2 \in \{0.20, 0.25, 0.30, 0.35, 0.40\}$.

Results

The numerical results for Settings 1, 2 and 3 can be found in Tables 3.8, 3.9 and 3.10, respectively, and allow us to draw the following conclusions about the robustness of SIMEX and MY.

SIMEX is robust for the MSE of $\widehat{\beta}_2$; when considering the MSE of $\widehat{\beta}_2$ and $\widehat{\beta}_1$, SIMEX seems more robust than MY to variations in the assumed measurement error variance: changing the variance used in the estimation procedure leads to larger changes (increases or decreases) in the MSE with MY than with SIMEX. The changes are also larger for β_1 than for β_2.

As far as bias is concerned, for both methods, the results for β_1 greatly depend on the variance used in the estimation procedure. There is more change in the bias of $\widehat{\beta}_1$ than in the bias of $\widehat{\beta}_2$. SIMEX also seems a bit more robust than MY; it is quite robust for the bias of $\widehat{\beta}_2$.

Another interesting phenomenon appearing in the results is that MY works best (in terms of bias or MSE) when the measurement error variance is (to some extent) underspecified: this is because, as already noted in Section 3.3.1, MY tends to correct more than necessary. The opposite is true for β_1 with SIMEX: this is a conservative method, so we can benefit from overspecifying the error variance.

Recommendations

Here are the practical pieces of advice that can be derived from the simulation results.

We first consider the case in which we assume the same (incorrect) value of the measurement error variance for both approaches. The lowest MSE of $\widehat{\beta}_2$,

Table 3.8: Results for the case of misspecification of the measurement error variance in Setting 1.

			Naive		MY		SIMEX	
v^2	v_E^2		β_1	β_2	β_1	β_2	β_1	β_2
.1	.05	Bias	-.040	-.014	-.014	-.017	-.017	-.016
		Var	.009	.038	.010	.039	.010	.039
		MSE	.011	.039	.010	.039	.010	.039
	.1	Bias	-.040	-.014	.015	-.020	.007	-.018
		Var	.009	.038	.012	.040	.011	.040
		MSE	.011	.039	.012	.040	.011	.040
	.15	Bias	-.040	-.014	.049	-.024	.032	-.021
		Var	.009	.038	.014	.040	.013	.040
		MSE	.011	.039	.017	.041	.014	.041
	.2	Bias	-.040	-.014	.088	-.028	.056	-.023
		Var	.009	.038	.018	.042	.014	.041
		MSE	.011	.039	.025	.042	.017	.041
.3	.2	Bias	-.114	-.007	-.030	-.017	-.047	-.014
		Var	.007	.038	.013	.040	.011	.040
		MSE	.020	.038	.013	.041	.013	.040
	.25	Bias	-.114	-.007	-.002	-.020	-.031	-.016
		Var	.007	.038	.015	.041	.012	.040
		MSE	.020	.038	.015	.042	.013	.041
	.3	Bias	-.114	-.007	.030	-.024	-.014	-.017
		Var	.007	.038	.018	.042	.013	.041
		MSE	.020	.038	.019	.043	.013	.041
	.35	Bias	-.114	-.007	.067	-.029	.002	-.019
		Var	.007	.038	.023	.044	.014	.041
		MSE	.020	.038	.027	.045	.014	.042
	.4	Bias	-.114	-.007	.113	-.035	.017	-.020
		Var	.007	.038	.031	.046	.015	.042
		MSE	.020	.038	.044	.047	.015	.042

Table 3.9: Results for the case of misspecification of the measurement error variance in Setting 2.

v^2	v_E^2		Naive		MY		SIMEX	
			β_1	β_2	β_1	β_2	β_1	β_2
.1	.05	Bias	-.104	-.006	-.039	-.014	-.048	-.012
		Var	.017	.060	.022	.062	.021	.061
		MSE	.028	.060	.023	.062	.023	.061
	.1	Bias	-.104	-.006	.041	-.024	.010	-.018
		Var	.017	.060	.030	.065	.025	.063
		MSE	.028	.060	.032	.066	.025	.064
	.15	Bias	-.104	-.006	.145	-.038	.066	-.024
		Var	.017	.060	.046	.071	.029	.066
		MSE	.028	.060	.067	.072	.034	.066
	.2	Bias	-.104	-.006	.284	-.055	.120	-.030
		Var	.017	.060	.080	.078	.034	.069
		MSE	.028	.060	.161	.081	.048	.069
.3	.2	Bias	-.269	.011	-.068	-.016	-.126	-.006
		Var	.013	.060	.034	.070	.023	.065
		MSE	.086	.060	.039	.070	.039	.066
	.25	Bias	-.269	.011	.008	-.027	-.092	-.010
		Var	.013	.060	.047	.075	.026	.067
		MSE	.086	.060	.047	.076	.034	.067
	.3	Bias	-.269	.011	.107	-.040	-.058	-.014
		Var	.013	.060	.074	.083	.028	.069
		MSE	.086	.060	.086	.085	.031	.069
	.35	Bias	-.290	.012	.180	-.052	-.061	-.013
		Var	.010	.059	.085	.096	.022	.070
		MSE	.095	.059	.117	.099	.026	.070
	.4	Bias	-.310	.011	.211	-.053	-.068	-.014
		Var	.009	.057	.191	.104	.018	.066
		MSE	.105	.057	.236	.107	.023	.066

Table 3.10: Results for the case of misspecification of the measurement error variance in Setting 3.

v^2	v_E^2		Naive β_1	Naive β_2	MY β_1	MY β_2	SIMEX β_1	SIMEX β_2
.1	.05	Bias	-.073	-.032	-.029	-.035	-.035	-.034
		Var	.028	.108	.034	.110	.033	.109
		MSE	.034	.110	.035	.111	.034	.110
	.1	Bias	-.073	-.032	.024	-.038	.005	-.035
		Var	.028	.108	.045	.112	.039	.110
		MSE	.034	.110	.045	.114	.039	.112
	.15	Bias	-.075	-.032	.084	-.041	.041	-.037
		Var	.027	.108	.055	.114	.042	.111
		MSE	.033	.109	.062	.115	.043	.113
	.2	Bias	-.079	-.032	.160	-.046	.074	-.039
		Var	.026	.108	.075	.117	.044	.112
		MSE	.032	.109	.101	.119	.050	.114
.3	.2	Bias	-.191	-.028	-.054	-.037	-.089	-.033
		Var	.021	.108	.043	.113	.033	.110
		MSE	.058	.108	.046	.114	.041	.111
	.25	Bias	-.192	-.028	-.006	-.040	-.066	-.034
		Var	.021	.108	.052	.115	.034	.111
		MSE	.058	.108	.053	.116	.039	.112
	.3	Bias	-.194	-.027	.053	-.043	-.044	-.035
		Var	.020	.108	.069	.119	.036	.112
		MSE	.058	.109	.071	.121	.038	.114
	.35	Bias	-.204	-.035	.101	-.057	-.036	-.043
		Var	.018	.107	.073	.125	.033	.113
		MSE	.060	.108	.083	.128	.035	.115
	.4	Bias	-.213	-.024	.158	-.046	-.029	-.032
		Var	.017	.109	.085	.132	.033	.116
		MSE	.063	.110	.110	.134	.034	.117

the estimated parameter related to the covariate without measurement error, is obtained with Naive. If the analyst still wants to use a correction method, SIMEX should be chosen, as it yields lower MSE than MY. For β_1, the coefficient of the mismeasured covariate, the conclusions about the MSE are a bit more subtle. When the measurement error variance is small, especially if this small variance is overspecified, a correction is not needed. For a larger variance, we should always correct with SIMEX. MY can also be useful (although not as much as SIMEX), but only if the variance is underspecified.

If we are mainly interested in decreasing bias, then the recommendations are as follows. For β_2, a correction is not needed, but the correction method which decreases the bias the most is SIMEX. For β_1, when the true variance is low and is overspecified, we should not use MY; we should apply SIMEX if the variance is not too overspecified. When the variance is low and underspecified, we should always use a correction method, the best one being MY. When the true variance is larger, we always benefit from correcting for the measurement error. The best correction method is SIMEX when the variance is overspecified, and MY when the variance is underspecified.

However, in practice, we do not necessarily know whether the measurement error variance is over- or underspecified. In order to take this into account, we can express the recommendations as follows. The simulation results showed that we can lower MSE or bias by deliberately assuming a "low" or "high" value for the measurement error variance (even if we know its true value). We first focus on MSE. For β_1 and β_2, with MY, we should underspecify the variance. With SIMEX, there is no clear conclusion for β_1. For β_2, we could underspecify the variance, but the effect is very small (since SIMEX is robust for the MSE of $\widehat{\beta}_2$).

Similar recommendations can be made for the bias. With MY, we should always underspecify the variance for β_1 and β_2 (but preferably slightly for β_1), even if we know its true value. This, again, can be linked to the tendency of this method to over-correct for the measurement error. With SIMEX, for β_1, we should correctly estimate the variance when it is low (if it is not well estimated, it should be underspecified); when the variance is large, we should generally overspecify the variance. For β_2 with SIMEX, we should underspecify the variance when it is low; when the variance is high, we should also underspecify it, since the effect of overspecifying is not clear (but the bias does not change much in function of the assumed variance).

3.3.3 Implementation of the different methods in the R package miCoPTCM

We implemented both Ma and Yin (2008)'s corrected score approach and the SIMEX method adapted to the promotion time cure model in Chapter 2 using the open-source statistical language R (R Core Team, 2015). These implementations can be found in the package `miCoPTCM` (for **mi**smeasured **Co**variates in the **P**romotion **T**ime **C**ure **M**odel), available on the CRAN website.

This package consists of two functions. The first one, `PTCMestimBF`, is the R implementation of Ma and Yin (2008)'s corrected score approach. In addition to the data and the formula of the model, the function only requires the variance-covariance matrix of the measurement error, and a set of initial values for the regression parameters $(\beta_0, \boldsymbol{\beta}^T)^T$. This function also allows one to estimate a model without measurement error (or without taking the measurement error into account), through the backfitting procedure presented in Ma and Yin (2008), by simply setting all the elements of the variance-covariance matrix to 0.

The second function is `PTCMestimSIMEX`. It makes it possible to estimate a promotion time cure model with mismeasured covariates by using the SIMEX approach introduced in Chapter 2. The user can tune the parameters related to the SIMEX algorithm: these parameters include not only the variance-covariance matrix of the measurement error and initial values, but also the grid of levels for the variance of the artificial noise (λ), the number of replicates for each of these values (B) and the order of the extrapolation function. Moreover, the error distribution to be used can also be chosen, among the four distributions that are considered in this chapter: Gaussian, Student-t, chi-square and Laplace.

Both functions return the estimated coefficients, their standard errors and the estimated baseline cumulative distribution function F. Their use is illustrated in Appendix B.

3.4 Analysis of recurrence in rectal cancer patients

We would like to study the impact of several risk factors on the time until recurrence for patients suffering from rectal cancer and having been operated on. The data we are going to analyze consist of 224 patients with complete data, without confirmed metastasis at diagnosis and without synchronous metastasis indicated at follow-up. They were monitored between 1998 and 2014 at the Cliniques universitaires Saint-Luc (Brussels, Belgium). In such a context, we

know that some patients will never experience a recurrence: they can be considered as cured. In addition, one of the covariates is known to be measured with some error (the hemoglobin level). We want to take this into account, in case the error is not negligible. However, the distribution of this error is not known for our data. In addition to the hemoglobin level (median 13.45, range 5.5-17.7, mean 13.30, standard deviation 1.92), we want to include three other characteristics of the patients in this analysis: the ratio of infiltrated versus examined nodes (median 0, range $[0, 1]$), the presence of lymphatic permation (20% yes) and of peri-nervous permation (11% yes).

85% of the patients considered for the analysis were still alive at the end of the follow-up (median follow-up time: 3.9 years). We see a plateau in the Kaplan-Meier estimated survival curve represented on Figure 3.2, which confirms the medical assumption of the presence of cured patients.

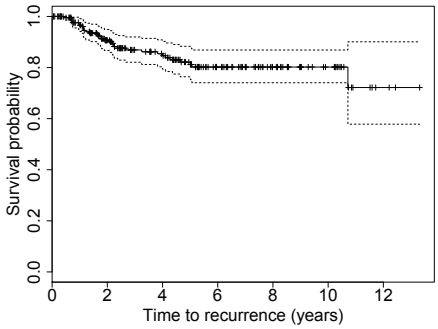


Figure 3.2: Kaplan-Meier estimate of the distribution of the time to recurrence for the patients from the rectal cancer database.

As we know neither the distribution nor the variance of the measurement error in the hemoglobin level, we estimated several models, assuming different values for the standard error v of the noise: 0.25, 0.5 and 1. Results are presented in Table 3.11.

We observe that correcting for the measurement error yields larger estimated (negative) effects of the hemoglobin level. Moreover, the larger the assumed error variance, the larger (in absolute value) the estimated coefficient. MY and SIMEX yield very similar results for the two lowest values of the error variance (0.25^2 and 0.5^2); for the third one (unit variance), the correction is larger with MY than with SIMEX.

Using SIMEX, if we assume Gaussian errors and a value of 0.25 for the measurement error standard deviation, we obtain the following results. The coefficient of the hemoglobin level is negative: a larger value of this covariate

is associated with a better survival, and a larger probability of being cured. The situation is reversed for the ratio of infiltrated versus examined nodes: an increase in this variable yields worse survival and cure probability. However, the effects of the hemoglobin level and of the presence of lympathic permation are not significant (at a level 5%). For a patient with the median hemoglobin level and ratio of infiltrated versus examined nodes and with no lymphatic permation nor peri-nervous permation, the estimated probability of being cured is 83%. This probability drops to 23% when both permations are present.

Table 3.11: Regression coefficient estimates and estimated standard deviations in the case of Gaussian errors.

Estimate	Intercept	Hemoglobin	Nodes ratio	Lymphatic permation	Peri-nervous permation
Naive	.258	-.142	2.619	.739	1.313
(Estim. s.d.)	(1.188)	(.087)	(.618)	(.435)	(.471)
MY ($v = 1$)	.906	-.190	2.613	.732	1.310
(Estim. s.d.)	(1.471)	(.111)	(.613)	(.440)	(.479)
MY ($v = .5$)	.392	-.152	2.618	.738	1.312
(Estim. s.d.)	(1.249)	(.092)	(.617)	(.436)	(.473)
MY ($v = .25$)	.290	-.144	2.619	.739	1.313
(Estim. s.d.)	(1.203)	(.089)	(.618)	(.435)	(.471)
SIMEX ($v = 1$)	.644	-.170	2.612	.734	1.313
(Estim. s.d.)	(1.274)	(.095)	(.615)	(.435)	(.472)
SIMEX ($v = .5$)	.402	-.152	2.622	.735	1.310
(Estim. s.d.)	(1.192)	(.088)	(.617)	(.435)	(.472)
SIMEX ($v = .25$)	.299	-.145	2.621	.738	1.312
(Estim. s.d.)	(1.188)	(.088)	(.618)	(.435)	(.472)

However, the measurement error could be non-Gaussian, and its standard deviation could be different from 0.25. From our simulation results, we know that, if we are interested in the effect of hemoglobin and if we want to decrease the bias in its estimation, we should use a correction method. The estimate obtained with Naive, -0.142, is probably somewhat too small (in absolute value). If we assume that the measurement error is Gaussian and that its variance is not too large (less than 0.6 or 0.8, according to the simulation results), then SIMEX is expected to yield the best results, even if the variance is (slightly) overspecified. In that case, the estimated effect of the hemoglobin level is -0.145 or -0.152. However, if we suspect that the variance could be underspecified, MY should be preferred. In this example, both strategies lead to very similar results. If the variance is larger (in this example, 1) and we tend to underspecify it, then MY should decrease the bias the most. The estimated

effect of -0.190 is hence expected to be larger (in absolute value) than the true value, unless the measurement error standard deviation is larger than 1.

On the other hand, if we prefer decreasing the MSE in the estimated coefficient of the level of hemoglobin, we should use SIMEX, whatever the error distribution, unless the error variance is really too low and overspecified (in which case, Naive should be used). For a Gaussian error, MY with an underspecified error variance should also yield useful results. If we suspect a measurement error standard deviation around 0.5, then we would expect the effect to be between -0.144 (MY with $v = 0.25$) and -0.152 (SIMEX with $v = 0.5$).

The estimated cure rate is robust to both the estimation method and the assumed error variance: for example, for a patient with the median hemoglobin level and ratio of infiltrated versus examined nodes and with no lymphatic permation nor peri-nervous permation, the cure rate takes values between 0.8245 and 0.8251.

In this particular study, it appears that the conclusions are very similar, no matter the method and parameters used for estimation: the sign and significance of each coefficient are not affected.

3.5 Calculations for the derivation of the bias of the naive estimator

We proceed similarly to Hughes (1993). We first compute $f_T(t)$ and $f_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{w}}|t)$. We easily find

$$f_T(t) = \int f_{T,\widetilde{\boldsymbol{X}}}(t, \widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} = \int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}$$

and, consequently,

$$\begin{aligned} f_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{w}}|t) &= \int f_{\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{X}}|T}(\widetilde{\boldsymbol{w}}, \widetilde{\boldsymbol{x}}|t) d\widetilde{\boldsymbol{x}} \\ &= \int \frac{f_{T|\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{w}}, \widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{W}},\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}, \widetilde{\boldsymbol{x}})}{f_T(t)} d\widetilde{\boldsymbol{x}} \\ &= \frac{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}. \end{aligned}$$

In these expressions, $f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ has to be chosen according to the context, while $f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})$ depends on the model assumed for the error. Here, we assume that $(\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{x}}) \sim N(\widetilde{\boldsymbol{x}}, \boldsymbol{\Lambda})$. $f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})$ is found using the promotion time

cure model, i.e.

$$\begin{aligned} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) &= -\frac{d}{dt}S(t|\widetilde{\boldsymbol{x}}) = -\frac{d}{dt}\exp\left\{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)\right\} \\ &= \eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)\exp\left\{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)\right\}. \end{aligned}$$

We can now obtain the conditional expectations in (3.3). The first one is

$$\begin{aligned} E_{\widetilde{\boldsymbol{W}}|T=t}(\widetilde{\boldsymbol{W}}) &= \int \widetilde{\boldsymbol{w}} f_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{w}}|t)d\widetilde{\boldsymbol{w}} \\ &= \frac{\int\int \widetilde{\boldsymbol{w}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}d\widetilde{\boldsymbol{w}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= \frac{\int \left(\int \widetilde{\boldsymbol{w}} f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{w}}\right) f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= \frac{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= \frac{\int \widetilde{\boldsymbol{x}}\ \eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int \eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= \frac{\int \widetilde{\boldsymbol{x}}\ \eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int \eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}, \end{aligned} \tag{3.4}$$

where the fourth equality is due to the fact that $E(\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}} = \widetilde{\boldsymbol{x}}) = \widetilde{\boldsymbol{x}}$. The second conditional expectation in (3.3) is

$$\begin{aligned} E_{\widetilde{\boldsymbol{W}}|T=t}\left(e^{\beta_0^*+\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*}\right) &= \int e^{\beta_0^*+\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{w}}|t)d\widetilde{\boldsymbol{w}} \\ &= \frac{\int\int e^{\beta_0^*+\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}d\widetilde{\boldsymbol{w}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= e^{\beta_0^*}\frac{\int \left(\int e^{\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{w}}\right) f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= e^{\beta_0^*}\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int \eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\ &= e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}\eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int \eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}, \end{aligned} \tag{3.5}$$

where the fourth equality follows directly from the definition of the moment generating function of $(\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}=\widetilde{\boldsymbol{x}}) \sim N(\widetilde{\boldsymbol{x}}, \boldsymbol{\Lambda})$.

The third conditional expectation in (3.3) is

$$
\begin{aligned}
E_{\widetilde{\boldsymbol{W}}|T=t}\left(e^{\beta_0^*+\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{W}}\right) &= \int e^{\beta_0^*+\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{w}} f_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{w}}|t)d\widetilde{\boldsymbol{w}} \\
&= e^{\beta_0^*}\frac{\int\int e^{\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{w}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}d\widetilde{\boldsymbol{w}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\
&= e^{\beta_0^*}\frac{\int\left(\int e^{\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{w}} f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{w}}\right) f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\
&= e^{\beta_0^*}\frac{\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\
&= e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\quad\times\frac{\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \\
&= e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\quad\times\frac{\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}\eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int\eta(\boldsymbol{x}^T\boldsymbol{\beta})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}},
\end{aligned}
\tag{3.6}
$$

where the fourth equality can be found by using the moment generating function of $(\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}=\widetilde{\boldsymbol{x}}) \sim N(\widetilde{\boldsymbol{x}}, \boldsymbol{\Lambda})$, since

$$
\begin{aligned}
&\int e^{\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{w}} f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{w}} = \frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^*}\int e^{\widetilde{\boldsymbol{w}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{w}}|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{w}} = \frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^*}m_{\widetilde{\boldsymbol{W}}|\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{\beta}}^*) \\
&= \frac{\partial}{\partial\widetilde{\boldsymbol{\beta}}^*}e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} = (\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}.
\end{aligned}
$$

The first element of the first term of (3.3) can now be written as

$$
\begin{aligned}
E_T(I(T\le t_c)) &= \int I(t\le t_c)f_T(t)dt = \int\int I(t\le t_c)f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}dt \\
&= \int_0^{t_c}\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}dt \\
&= \int\left(\int_0^{t_c} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})dt\right)f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}} \\
&= \int(1-S(t_c|\widetilde{\boldsymbol{x}}))f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}
\end{aligned}
$$

$$= \int f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}} - \int S(t_c|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}$$
$$= 1 - \int e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}.$$

Using (3.4), the second element of the first term of (3.3) becomes

$$E_T\left[I(T \le t_c)E_{\widetilde{\boldsymbol{W}}|T}(\widetilde{\boldsymbol{W}})\right] = E_T\left[I(T \le t_c)\frac{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\right]$$
$$= \int I(t \le t_c)\frac{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} f_T(t)dt$$
$$= \int I(t \le t_c)\frac{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}} \int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\ dt$$
$$= \int I(t \le t_c)\left\{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\right\}dt$$
$$= \int_0^{t_c}\left\{\int \widetilde{\boldsymbol{x}} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\right\}dt$$
$$= \int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\left\{\int_0^{t_c} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})dt\right\}d\widetilde{\boldsymbol{x}}$$
$$= \int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\left\{1 - S(t_c|\widetilde{\boldsymbol{x}})\right\}d\widetilde{\boldsymbol{x}}$$
$$= \int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}} - \int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})S(t_c|\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}$$
$$= E(\widetilde{\boldsymbol{X}}) - \int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\widetilde{\boldsymbol{x}}$$
$$= -\int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}d\widetilde{\boldsymbol{x}}.$$

Thanks to (3.5), the first element of the second term in (3.3) can be rewritten as

$$E_T\left[I(T \le t_c)F_{\beta^*}(T)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\beta_0^* + \widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*}\right)\right]$$
$$= E_T\left[I(T \le t_c)F_{\beta^*}(T)e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\right]$$
$$= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}E_T\left[I(T \le t_c)F_{\beta^*}(T)\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\right]$$

$$
\begin{aligned}
&= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int_0^{t_c} F_{\beta^*}(t) \frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}} f_T(t) dt \\
&= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\quad \times \int_0^{t_c} F_{\beta^*}(t) \frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}} \left(\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} \right) dt \\
&= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int_0^{t_c} F_{\beta^*}(t) \left(\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} \right) dt \\
&= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{ \int_0^{t_c} F_{\beta^*}(t) f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) dt \right\} d\widetilde{\boldsymbol{x}} \\
&= e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\quad \times \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{ \int_0^{t_c} F_{\beta^*}(t) \eta(\boldsymbol{x}^T\boldsymbol{\beta}) F'_\beta(t) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta}) F_\beta(t)} dt \right\} d\widetilde{\boldsymbol{x}}.
\end{aligned}
$$

Equation (3.6) allows us to express the second element of the second term of (3.3) as

$$
\begin{aligned}
&E_T \left[I(T \le t_c) F_{\beta^*}(T) E_{\widetilde{\boldsymbol{W}}|T} \left(e^{\beta_0^* + \widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*} \widetilde{\boldsymbol{W}} \right) \right] \\
&\quad = E_T \left[I(T \le t_c) F_{\beta^*}(T) e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \right. \\
&\qquad \left. \times \frac{\int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}} \right] \\
&\quad = e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\qquad \times E_T \left[I(T \le t_c) F_{\beta^*}(T) \frac{\int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}} \right] \\
&\quad = e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \\
&\qquad \times \int_0^{t_c} F_{\beta^*}(t) \frac{\int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}}} \\
&\qquad\quad \left\{ \int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} \right\} dt \\
&\quad = e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int_0^{t_c} F_{\beta^*}(t) \left\{ \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} \right\} dt \\
&\quad = e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{ \int_0^{t_c} F_{\beta^*}(t) f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}}) dt \right\} d\widetilde{\boldsymbol{x}} \\
&\quad = e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \\
&\qquad \times \left\{ \int_0^{t_c} F_{\beta^*}(t) \eta(\boldsymbol{x}^T\boldsymbol{\beta}) F'_\beta(t) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta}) F_\beta(t)} dt \right\} d\widetilde{\boldsymbol{x}}
\end{aligned}
$$

and the elements of the third term of (3.3) as

$$
\begin{aligned}
&E_T\left[I(T>t_c)F_{\beta^*}(t_c)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\beta_0^*+\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*}\right)\right]\\
&\quad=E_T\left[I(T>t_c)F_{\beta^*}(t_c)e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\right]\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\\
&\qquad\times\int_{t_c}^{\infty}F_{\beta^*}(t_c)\frac{\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\left\{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\right\}\,dt\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int_{t_c}^{\infty}\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\,dt\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\left\{\int_{t_c}^{\infty}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})dt\right\}\,d\widetilde{\boldsymbol{x}}\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})S(t_c|\widetilde{\boldsymbol{x}})\,d\widetilde{\boldsymbol{x}}\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\,d\widetilde{\boldsymbol{x}}
\end{aligned}
$$

and

$$
\begin{aligned}
&E_T\left[I(T>t_c)F_{\beta^*}(t_c)E_{\widetilde{\boldsymbol{W}}|T}\left(e^{\beta_0^*+\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{\beta}}^*}\widetilde{\boldsymbol{W}}\right)\right]\\
&\quad=E_T\left[I(T>t_c)F_{\beta^*}(t_c)e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\frac{\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(T|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\right]\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}\\
&\qquad\times\int_{t_c}^{\infty}F_{\beta^*}(t_c)\frac{\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}}\\
&\qquad\qquad\left\{\int f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\right\}\,dt\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int_{t_c}^{\infty}\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})d\widetilde{\boldsymbol{x}}\,dt\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\left\{\int_{t_c}^{\infty}f_{T|\widetilde{\boldsymbol{X}}}(t|\widetilde{\boldsymbol{x}})dt\right\}\,d\widetilde{\boldsymbol{x}}\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})S(t_c|\widetilde{\boldsymbol{x}})\,d\widetilde{\boldsymbol{x}}\\
&\quad=e^{\beta_0^*+\widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2}F_{\beta^*}(t_c)\int(\widetilde{\boldsymbol{x}}+\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*)e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*}f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)}\,d\widetilde{\boldsymbol{x}}.
\end{aligned}
$$

The final expression of (3.3) is then

$$\begin{pmatrix} 1 - \int e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) d\widetilde{\boldsymbol{x}} \\ -\int \widetilde{\boldsymbol{x}} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} d\widetilde{\boldsymbol{x}} \end{pmatrix}$$

$$-e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} \times \begin{pmatrix} \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{ \int_0^{t_c} F_{\beta^*}(t)\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt \right\} d\widetilde{\boldsymbol{x}} \\ \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) \left\{ \int_0^{t_c} F_{\beta^*}(t)\eta(\boldsymbol{x}^T\boldsymbol{\beta})F'_\beta(t)e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t)}dt \right\} d\widetilde{\boldsymbol{x}} \end{pmatrix}$$

$$-e^{\beta_0^* + \widetilde{\boldsymbol{\beta}}^{*T}\boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*/2} F_{\beta^*}(t_c) \times \begin{pmatrix} \int e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} d\widetilde{\boldsymbol{x}} \\ \int (\widetilde{\boldsymbol{x}} + \boldsymbol{\Lambda}\widetilde{\boldsymbol{\beta}}^*) e^{\widetilde{\boldsymbol{x}}^T\widetilde{\boldsymbol{\beta}}^*} f_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) e^{-\eta(\boldsymbol{x}^T\boldsymbol{\beta})F_\beta(t_c)} d\widetilde{\boldsymbol{x}} \end{pmatrix} = \mathbf{0}.$$

3.6 Discussion

Time-to-event data often exhibit features that prevent one from using classical statistical tools of survival analysis. The first of the two characteristics that we have considered here is the presence of cured subjects in the data, i.e. subjects who will never experience the event of interest. The other data property which is often ignored is the presence of measurement error in some of the continuous covariates. Our motivating example consists of the analysis of the impact of hemoglobin level on the time until recurrence for rectal cancer patients having undergone surgery.

In this chapter, we focused on one of the survival models taking cured subjects into account, the semiparametric promotion time cure model. We first derived an expression of the asymptotic bias of the traditional estimator when the measurement error is not taken into account. We illustrated the use of this expression by creating plots allowing us to visually investigate the effects of the different parameters (including the censoring rate and measurement error variance) on the bias. We then presented the results of an extensive simulation study investigating the performance of several estimation methods in real settings in which two of their underlying assumptions are not met: the correct specification of the measurement error variance, and the Gaussian distribution of this error. These results allowed us to provide practical recommendations that can help investigators build a strategy to estimate a promotion time cure model. It appeared that the conclusions depend on the estimation method, on the criterion to be used (bias or MSE) and on the covariate of interest. For

instance, the coefficient related to the covariate without measurement error is typically not very influenced by a deviation from the normality assumption with the naive method, meaning that a correction method is rarely needed when the objective of the study is to estimate this effect. We also discovered that, with Ma and Yin (2008)'s corrected score method, both the bias and the MSE can be decreased by assuming a low value of the measurement error variance (even if we actually know it). We also briefly described our R package `miCoPTCM`, which is freely available on the CRAN website, and which encompasses the implementation of all estimation methods used in this chapter.

These recommendations allowed us to analyze the impact of the hemoglobin level on the time until recurrence of cancer in a rectal cancer database. However, in this application, it appeared that the conclusions do not change much in function of the method and the parameters used for estimating the coefficients: the estimated effect of the hemoglobin is always negative (a higher level of hemoglobin is associated with a longer time before recurrence, and with a higher cure probability), but never significant.

Of course, the conclusions and recommendations given in this chapter should always be considered with care as, for practical reasons, only a limited number of simulation settings were considered. The results may not always be generalized; however, an explanation was provided for some observed phenomena, thus allowing the reader to understand the mechanism at work, and, as a result, to infer the possible consequences in a different setting.

An interesting further step in the analysis of time-to-event data with a cure fraction and mismeasured covariates would be the direct estimation of the measurement error variance in the model estimation procedure. This is currently not possible, except when repeated measurements of the covariates are available (Carroll et al., 2006), in which case approaches such as joint modeling can be used. For example, Crowther et al. (2013) use a longitudinal linear mixed effects submodel for the true covariate and a survival submodel depending on this true covariate. However, in many studies, such as the one considered in this work, repeated measures are not available. The estimation of the measurement error variance would hence make the use of promotion time cure models with mismeasured covariates much easier in practice.

CHAPTER 4

Estimation of the measurement error variance

The material presented in this chapter is to be submitted as "Flexible parametric approach to classical measurement error variance estimation without auxiliary data", by Bertrand A., Van Keilegom I. and Legrand C.

4.1 Introduction

In this chapter, we consider the general context of measurement error in variables, without focusing on one particular model for the response variable. We study the classical measurement error model in the case of a single covariate,

$$W = X + U, \tag{4.1}$$

where $U \sim N(0, v^2)$ and X and U are assumed to be independent.

As explained in Section 1.3, many different methods have been developed to correct the bias due to this error in several contexts. Most of them require some knowledge about the error distribution, which can be recovered from validation or auxiliary data. However, such data are not always available; in this case, the distribution of the error must usually be assumed to be completely known. These assumptions are very often a real issue in practice. In many cases, the (justified) fear of obtaining incorrect results because of inaccurate assumptions restrains the analyst from tempting to take the error into account in the data analysis. This explains why, despite a strong interest for measurement error in

the statistical literature and the plethora of methods which have been developed to deal with it, bias correction methods are still too rarely used in practical applications.

A first step towards more pragmatic solutions can be found in the field of nonparametric deconvolution, where some authors relax the assumption of known density for U. Deconvolution approaches aim at recovering the distribution of X (and, sometimes, of U) from the observed W. Matias (2002), Butucea and Matias (2005), Meister (2006), Butucea et al. (2008) and Schwarz and Van Bellegem (2010) assume a known distribution with unknown variance, while the approaches of Meister (2007) and Delaigle and Hall (2016) make still weaker assumptions about the error distribution. However, all these methods suffer from one or more of the following problems: the focus is on estimating the distribution of X, there are many tuning parameters for which no clear guidelines are given, no (or limited) practical implementation is discussed, and there are theoretical and practical identifiability issues. There is currently, to the best of our knowledge, no method which can easily be used in practice to recover the variance or distribution of the error when W is measured only once.

In this chapter, we address this problem of unknown variance in the case of measurement error in continuous covariates and propose an easy-to-use solution to estimate it from the data at hand. We suppose that no auxiliary or validation data are available. Schwarz and Van Bellegem (2010) show that model (4.1) is identifiable if we assume that X has compact support. We build on this result and present a likelihood-based approach for estimating the measurement error variance while making only weak assumptions about the distribution of the true covariate X. A kernel density estimate could be used. However, this latter nonparametric estimate is expected to have a lower rate of convergence to the true density function than an approximate parametric estimate. Some common flexible approaches to density approximation include using a mixture of Normal densities (Carroll et al., 1999b) or the semi-nonparametric density (Zhang and Davidian, 2001). However, they cannot be used in the current context, since they do not suit a variable with compact support. An approximation through a discrete distribution could be considered, as in Hu et al. (1998), but, as mentioned in their paper, it may not well approximate a smooth distribution, and can require a large number of parameters. In this chapter, we propose to use Bernstein polynomials (Bernstein, 1912). Compared to the aforementioned approaches, this robust method has the advantage of relying on only one regularization parameter, since it amounts to using a mixture of Beta density functions with known parameters. Moreover, such polynomials have been shown to approximate any continuous function with compact support. They have been used to estimate, among others, densities (Guan, 2016), distribution

functions (Leblanc, 2012), copulas (Janssen et al., 2012) and copula densities (Janssen et al., 2014).

Our proposed estimator can be used with any correction method and response model. As an illustration, we study its behaviour when used in combination with the simulation extrapolation (SIMEX) algorithm (Cook and Stefanski, 1994) introduced in Section 1.3.3, in the context of the survival Cox proportional hazards model (Cox, 1972).

In Section 4.2, the details of the proposed method are described. The results of a simulation study investigating the finite-sample properties of this technique are then presented in Section 4.3. Section 4.4 illustrates the use of the obtained variance estimator in the Cox proportional hazards model (Cox, 1972) estimated with the SIMEX approach. The asymptotic normality of the SIMEX estimator in this case is proved in Section 4.6. In Section 4.5, the proposed solution is applied on survival data of patients with monoclonal gammopathy of undetermined significance (MGUS): the effect of several variables, among which some with measurement error, on the survival of these patients, is analyzed. Finally, a discussion and some insight into further work can be found in Section 4.7.

4.2 Methodology

We consider the classical additive measurement error model (4.1) with a Gaussian error of unknown variance, where X is only assumed to have compact support. The objective is to obtain an estimate of v, the measurement error standard deviation, which will then be available for use in a bias-correction method. The key idea is to build the probability density function (pdf) of W (and then the corresponding log-likelihood) with only weak assumptions on the distribution of X. This is possible if we assume that

$$X = aS + b, \tag{4.2}$$

with a and b two (unknown) constants with $a > 0$ for identifiability reasons, and S a continuous random variable taking values in $[0, 1]$. Consequently, the density of W is

$$\begin{aligned} f_W(w) &= \frac{1}{v} \int f_X(x) \phi\left(\frac{w-x}{v}\right) dx \\ &= \frac{1}{a\,v} \int f_S\left(\frac{x-b}{a}\right) \phi\left(\frac{w-x}{v}\right) dx, \end{aligned} \tag{4.3}$$

where f_S is the pdf of S and $\phi(\cdot)$ is the pdf of a normally distributed random variable with zero mean and unit variance. Theorem 3 states that this model is identifiable.

Theorem 3. *If, for all w,*

$$\frac{1}{a_1\, v_1}\int f_{S1}\left(\frac{x-b_1}{a_1}\right)\phi\left(\frac{w-x}{v_1}\right)\,dx = \frac{1}{a_2\, v_2}\int f_{S2}\left(\frac{x-b_2}{a_2}\right)\phi\left(\frac{w-x}{v_2}\right)\,dx,$$

then $a_1 = a_2$, $b_1 = b_2$, $v_1 = v_2$ and $f_{S1} = f_{S2}$. Hence, the model is identifiable.

The proof follows from Schwarz and Van Bellegem (2010), who prove the identifiability when U is Gaussian and when the law of X belongs to

$$\{P \in \mathcal{P} | \exists A \in \mathcal{B}(\mathbb{R}) : |A| > 0 \text{ and } P(A) = 0\}$$

where $\mathcal{B}(\mathbb{R})$ is the set of Borel sets in $\mathbb{R}$, $\mathcal{P}$ is the set of all probability distributions on $\mathbb{R}$ and $|A|$ is the Lebesgue measure of A. Note that in our situation, the set A equals the complementary of $[b, a+b]$.

The objective is now to specify the distribution of S with minimal assumptions. In this work, we propose to use Bernstein polynomials, as explained in Section 4.1. A Bernstein polynomial of degree m (for some choice of $m \geq 0$, which will be discussed later) is

$$B_m(s) = \sum_{k=0}^{m} \alpha_{k,m} b_{k,m}(s), \qquad s \in [0,1], \tag{4.4}$$

where

$$b_{k,m}(s) = \left(\begin{array}{c} m \\ k \end{array}\right) s^k (1-s)^{m-k}, \qquad k = 0, \ldots, m. \tag{4.5}$$

A key result shown in Bernstein (1912) is that any continuous function $f(s)$ defined on $[0,1]$ can be uniformly approximated by such a polynomial, by taking $\alpha_{k,m} = f\left(\frac{k}{m}\right)$:

$$\lim_{m\to\infty} \sup_{0\leq s\leq 1} \left| \sum_{k=0}^{m} f\Big(\frac{k}{m}\Big) b_{k,m}(s) - f(s) \right| = 0. \tag{4.6}$$

This property allows us to develop a parametric yet flexible likelihood approach to the estimation of v.

Since a pdf is always positive, we restrict our attention to Bernstein polynomials with positive coefficients. When applied to the approximation of f_S, Equations (4.4), (4.5) and (4.6) yield the following approximation of $f_S(s)$:

$$\widetilde{f}_{S,m}(s;\overline{\boldsymbol{\theta}}_m) = \sum_{k=0}^{m} f_S\left(\frac{k}{m}\right) b_{k,m}(s) = \sum_{k=0}^{m} \theta_{k,m} \frac{(m+1)!}{k!\ (m-k)!} s^k (1-s)^{m-k}, \quad s \in [0,1]$$

with $\overline{\boldsymbol{\theta}}_m = (\theta_{0,m}, \ldots, \theta_{m,m})^T$,

$$\theta_{k,m} = f_S\left(\frac{k}{m}\right) \binom{m}{k} \frac{k!\ (m-k)!}{(m+1)!} = \frac{1}{m+1} f_S\left(\frac{k}{m}\right),$$

and $\theta_{k,m} \geq 0$. Moreover, since $\widetilde{f}_{S,m}(\cdot\ ;\boldsymbol{\theta}_m)$ must be a density, $\sum_{k=0}^{m} \theta_{k,m} = 1$. This representation hence depends on the vector of parameters $\boldsymbol{\theta}_m = (\theta_{0,m}, \ldots, \theta_{m-1,m})^T$, and shows that f_S is actually approximated by a mixture of Beta$(k+1, m-k+1)$ densities, i.e. densities with known parameters. This result is used by Guan (2016) in the context of density estimation. Figure 4.1 shows the densities involved in the mixture, for $m = 0, 1, 2, 3$.

Once the density of S is specified, the likelihood of W can be built easily. If f_S is approximated by $\widetilde{f}_{S,m}(s;\boldsymbol{\theta}_m)$, then the probability density function of $X = aS + b$, $f_X(\cdot\ ; a, b)$, can be approximated by

$$\begin{aligned}
\widetilde{f}_{X,m}(x;a,b,\boldsymbol{\theta}_m) &= \frac{1}{a}\widetilde{f}_{S,m}\left(\frac{x-b}{a};\boldsymbol{\theta}_m\right) \\
&= \frac{1}{a}\sum_{k=0}^{m-1} \theta_{k,m} \frac{(m+1)!}{k!\ (m-k)!} \left(\frac{x-b}{a}\right)^k \left(1-\frac{x-b}{a}\right)^{m-k} \\
&\quad + \frac{1}{a}\left(1 - \sum_{k=0}^{m-1} \theta_{k,m}\right)(m+1)\left(\frac{x-b}{a}\right)^m \\
&= \frac{1}{a}\sum_{k=0}^{m-1} \theta_{k,m} Beta_{k+1,m-k+1}\left(\frac{x-b}{a}\right) \\
&\quad + \frac{1}{a}\left(1 - \sum_{k=0}^{m-1} \theta_{k,m}\right) Beta_{m+1,1}\left(\frac{x-b}{a}\right),
\end{aligned} \tag{4.7}$$

for $x \in [b, a+b]$, where $Beta_{\alpha,\beta}(\cdot)$ denotes the pdf of a Beta-distributed random variable with parameters α and β.

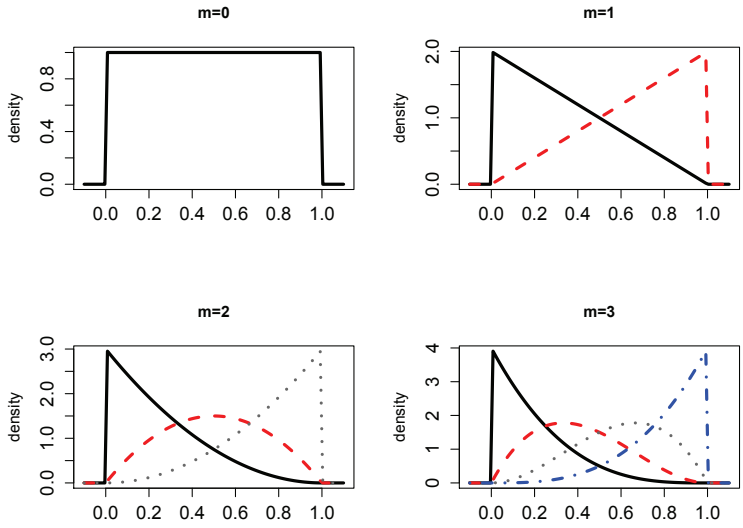


Figure 4.1: Representation of the Beta densities appearing in the Bernstein polynomials of degree $m = 0$, 1, 2 and 3. For $m = 0$, Beta$(1, 1)$. For $m = 1$, Beta$(1, 2)$ (solid, black) and Beta$(2, 1)$ (dashed, red). For $m = 2$, Beta$(1, 3)$ (solid, black), Beta$(2, 2)$ (dashed, red) and Beta$(3, 1)$ (dotted, gray). For $m = 3$, Beta$(1, 4)$ (solid, black), Beta$(2, 3)$ (dashed, red), Beta$(3, 2)$ (dotted, gray) and Beta$(4, 1)$ (dotdash, blue).

As a consequence, since X and U are assumed to be independent, the pdf of $W = X + U$ is approximately equal to

$$\widetilde{f}_{W,m}(w; v, a, b, \boldsymbol{\theta}_m) = \frac{1}{v} \int \widetilde{f}_{X,m}(x; a, b, \boldsymbol{\theta}_m) \phi\left(\frac{w-x}{v}\right) dx.$$

By (4.7), we have that

$$\widetilde{f}_{W,m}(w; v, a, b, \boldsymbol{\theta}_m) = \frac{1}{av} \sum_{k=0}^{m-1} \theta_{k,m} \int Beta_{k+1,m-k+1}\left(\frac{x-b}{a}\right) \phi\left(\frac{w-x}{v}\right) dx$$
$$+ \frac{1}{av}\left(1 - \sum_{k=0}^{m-1} \theta_{k,m}\right) \int Beta_{m+1,1}\left(\frac{x-b}{a}\right) \phi\left(\frac{w-x}{v}\right) dx$$

for $w \in \mathbb{R}$. This is a flexible $m+3$-dimensional parametric family of densities. Moreover, note that

$$\lim_{m\to\infty} \sup_w \left| \widetilde{f}_{W,m}(w; v, a, b, \boldsymbol{\theta}_m) - f_W(w) \right| = 0,$$

as long as f_S is continuous.

When a sample of W, $W_1, \ldots, W_n \sim^{iid} W$, is available, the log-likelihood function of the set of parameters $(v, a, b, \boldsymbol{\theta}_m)$ given the observed data is then

$$\begin{aligned}
\mathcal{L}_n(v, a, b, \boldsymbol{\theta}_m) &= \sum_{i=1}^{n} \log \widetilde{f}_{W,m}(W_i; v, a, b, \boldsymbol{\theta}_m) \\
&= \sum_{i=1}^{n} \log \left\{ \frac{1}{av} \sum_{k=0}^{m-1} \theta_{k,m} \int Beta_{k+1,m-k+1}\Big(\frac{x-b}{a}\Big) \phi\Big(\frac{W_i - x}{v}\Big) dx \right. \\
&\quad \left. + \frac{1}{av}\left(1 - \sum_{k=0}^{m-1} \theta_{k,m}\right) \int Beta_{m+1,1}\left(\frac{x-b}{a}\right) \phi\left(\frac{W_i - x}{v}\right) dx \right\} \quad (4.8)
\end{aligned}$$

This function can be maximized numerically with respect to $(v, a, b, \boldsymbol{\theta}_m)$ in order to obtain $(\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m)$, the estimators of the model parameters v, a, b and $\boldsymbol{\theta}_m$ for a given value of m, the degree of the Bernstein polynomial.

Theoretically, the quality of the approximation improves when m increases: Wang and Ghosh (2012) show that a model with $m = m_1$ is nested in a model with $m = m_2$, for $m_1 < m_2$. However, in practice, a large value of m implies a large number of parameters $\theta_{k,m}$ to be estimated, which could impair the quality of the estimated model. This is why we suggest choosing m by comparing the fit of the models estimated with several values of m, using a model selection criterion (see Sin and White (1996) for a theoretical justification of the use of such criteria). Based on the results of our simulation study (see Section 4.3),

we suggest using the BIC, which achieves a better performance thanks to the choice of more parsimonious models:

$$BIC(m) = (m+3)\log(n) - 2\mathcal{L}_n(\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m), \qquad m \geq 0.$$

The choice of the maximal value of m to be considered is also of interest. In the simulation study of Section 4.3, where samples of sizes 200 and 300 are considered, a maximal value of 6 appeared to be enough in nearly all cases (only few situations occured where $m = 5$ or 6 were chosen).

Let us now consider the asymptotic theory of our estimator. For a given value of m, this estimator maximizes the likelihood of a potentially misspecified model. In this context, the results of White (1982) on misspecified parametric models can be used to derive asymptotic properties. Let $(v_m^*, a_m^*, b_m^*, \boldsymbol{\theta}_m^*)$ be the parameter vector that minimizes the Kullback-Leibler Information Criterion:

$$E\left[\log\left\{\frac{f_W(W)}{\widetilde{f}_{W,m}(W; v, a, b, \boldsymbol{\theta}_m)}\right\}\right].$$

We then have the following results regarding consistency and asymptotic normality of our estimator.

Theorem 4. *Under regularity conditions (A1) to (A3) in White (1982),*

$$(\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m) \xrightarrow{P} (v_m^*, a_m^*, b_m^*, \boldsymbol{\theta}_m^*).$$

Theorem 5. *Under regularity conditions (A1) to (A6) in White (1982),*

$$n^{1/2}\Big((\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m) - (v_m^*, a_m^*, b_m^*, \boldsymbol{\theta}_m^*)\Big) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{C}),$$

where

$$\boldsymbol{C} = A(\boldsymbol{\gamma}_m^*)^{-1} B(\boldsymbol{\gamma}_m^*) A(\boldsymbol{\gamma}_m^*)^{-1},$$

with

$$A(\boldsymbol{\gamma}_m) = \left(E\left\{\frac{\partial^2}{\partial\gamma_i \partial\gamma_j}\log\widetilde{f}_{W,m}(W;\boldsymbol{\gamma}_m)\right\}\right)_{i,j=1}^{m+3},$$

$$B(\boldsymbol{\gamma}_m) = \left(E\left\{\frac{\partial}{\partial\gamma_i}\log\widetilde{f}_{W,m}(W;\boldsymbol{\gamma}_m)\cdot\frac{\partial}{\partial\gamma_j}\log\widetilde{f}_{W,m}(W;\boldsymbol{\gamma}_m)\right\}\right)_{i,j=1}^{m+3},$$

$\boldsymbol{\gamma}_m = (v_m, a_m, b_m, \boldsymbol{\theta}_m)$ *and* $\boldsymbol{\gamma}_m^* = (v_m^*, a_m^*, b_m^*, \boldsymbol{\theta}_m^*)$.

Note that, if the model is correctly specified, $\boldsymbol{C}$ appearing in Theorem 5 is $A(\boldsymbol{\gamma}_m^*)^{-1}$, the inverse Fisher matrix.

The regularity conditions of White (1982) as applied to our context are the following:

(A1) The independent random variables W_i, $i = 1, \ldots, n$, have common distribution function F_W on Ω, a measurable Euclidean space, with measurable Radon-Nikodým density $f_W = dF_W/dw$.

(A2) The family of distribution functions $\widetilde{F}_{W,m}(w; \boldsymbol{\gamma}_m)$ has Radon-Nikodým densities $\widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m) = d\widetilde{F}_{W,m}(w; \boldsymbol{\gamma}_m)/dw$ which are measurable in w for every $\boldsymbol{\gamma}_m$ in Γ_m, a compact subset of a $(m+3)$– dimensional Euclidean space, and continuous in $\boldsymbol{\gamma}_m$ for every w in Ω.

(A3) (a) $E\left[\log\left\{f_W(W_i)\right\}\right]$ exists and $|\log \widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m)| \leq g(w)$ for all γ_m in Γ_m, where g is integrable with respect to F_W;
(b) $E\left[\log\left\{f_W(W_i)/\widetilde{f}_{W,m}(W_i; \boldsymbol{\gamma}_m)\right\}\right]$ has a unique minimum at $\boldsymbol{\gamma}_m^*$ in Γ_m.

(A4) $\partial \log \widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m)/\partial\gamma_{m,j}$, $j = 1, \ldots, m+3$, are measurable functions of w for each $\boldsymbol{\gamma}_m$ in Γ_m and continuously differentiable functions of $\boldsymbol{\gamma}_m$ for each w in Ω.

(A5) $|\partial^2 \log \widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m)/\partial\gamma_{m,j}\gamma_{m,k}|$ and
$|\partial \log \widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m)/\partial\gamma_{m,j} \cdot \partial \log \widetilde{f}_{W,m}(w; \boldsymbol{\gamma}_m)/\partial\gamma_{m,k}|$, $j, k = 1, \ldots, m+3$ are dominated by functions integrable with respect to F_W for all w in Ω and $\boldsymbol{\gamma}_m$ in Γ_m.

(A6) (a) $\boldsymbol{\gamma}_m^*$ is interior to Γ_m; (b) $B(\boldsymbol{\gamma}_m^*)$ is nonsingular; (c) $A(\boldsymbol{\gamma}_m)$ has constant rank in some open neighborhood of $\boldsymbol{\gamma}_m^*$.

Assumptions (A1) and (A2) are general assumptions regarding the true structure that generates the observations and the family of distribution functions chosen to approximate the true structure. White (1982) explains that Assumption (A3)(a) makes sure that the Kullback-Leibler Information Criterion is well-defined, while (A3)(b) is the fundamental identification condition. Assumptions (A4) to (A6) are technical conditions required for proving the asymptotic normality of the estimator.

The methodology and asymptotic results presented here assume a single mismeasured covariate. However, in the case of multiple mutually independent covariates $X_1, \ldots, X_P$ subject to mutually independent errors $U_1, \ldots, U_P$, the methodology can be applied separately to each mismeasured covariate.

4.3 Simulation study

Data were simulated according to models (4.1) and (4.2) using four values for the noise-to-signal ratio (NSR), $v/\sigma_X = \{0, 0.25, 0.50, 0.75\}$, where σ_X is the standard deviation of X, and eight different distributions for X. In the first four cases, S is assumed to follow a Beta(α, β) distribution with $(\alpha, \beta) = \{(1,1), (1,2), (0.7, 0.5), (3,2)\}$, and $X = aS + b$ with $a = 2$ and $b = -1$. The other four distributions used for X are Normal$(-1, 4)$ truncated at $(t_L, t_U) = (-5, 3)$ or $(-4, 2)$ and Exponential$(\mu, t_U) - 1$ with $(\mu, t_U) = (0.5, 4)$ or $(10, 20)$ of mean μ and truncated at t_U. The density functions corresponding to these eight cases are represented in Figure 4.2. Only three of these densities can be written as a Bernstein polynomial: Beta$(1,1)$, Beta$(1,2)$ and Beta$(3,2)$. Two sample sizes ($n = 200$ and $n = 300$) were considered for all distributions; for three distributions, an additional sample size ($n = 1000$) was also investigated.

For each setting, 500 replicated datasets were created. For each of them, v was estimated using $m = \{0, \ldots, 6\}$ in Equation (4.8). The fmincon function of the Matlab software was used to maximize the log-likelihood function. The simulation results regarding the estimation of v for $n = 200$, $n = 300$ and $n = 1000$ can be found in Tables 4.1, 4.2 and 4.3, respectively.

As can be expected, a larger the sample size is usually associated with lower standard errors; this trend also appears in the bias, except with the 2Beta$(3,2)-1$ and Normal distributions.

When there is actually no measurement error in the data, the estimated measurement error variance is generally close to 0. For the other values of NSR, the smallest relative biases (smaller than 0.15) are found for the 2Beta$(1,1)-1$, 2Beta$(1,2) - 1$ and both Exponential distributions. The 2Beta$(3,2) - 1$ and N$(-1, 4, -5, 3)$ distributions yield the worst results: v is always overestimated; however, this bias decreases when v increases. These distributions are rather close to a Gaussian one, which could explain the poor performance of the method in theses cases: the Gaussian error could be more difficult to separate from a near-Gaussian signal. The observed W is the convolution of a Gaussian and a near-Gaussian variable, and hence looks very much like a Gaussian variable. Since a Gaussian variable can be decomposed as the sum of two independent Gaussian variables in an infinite number of ways, we are close to a non-identifiable situation. For the other distributions, the relative bias also generally decreases with the NSR; however, there is no systematic pattern in the evolution of the MSE.

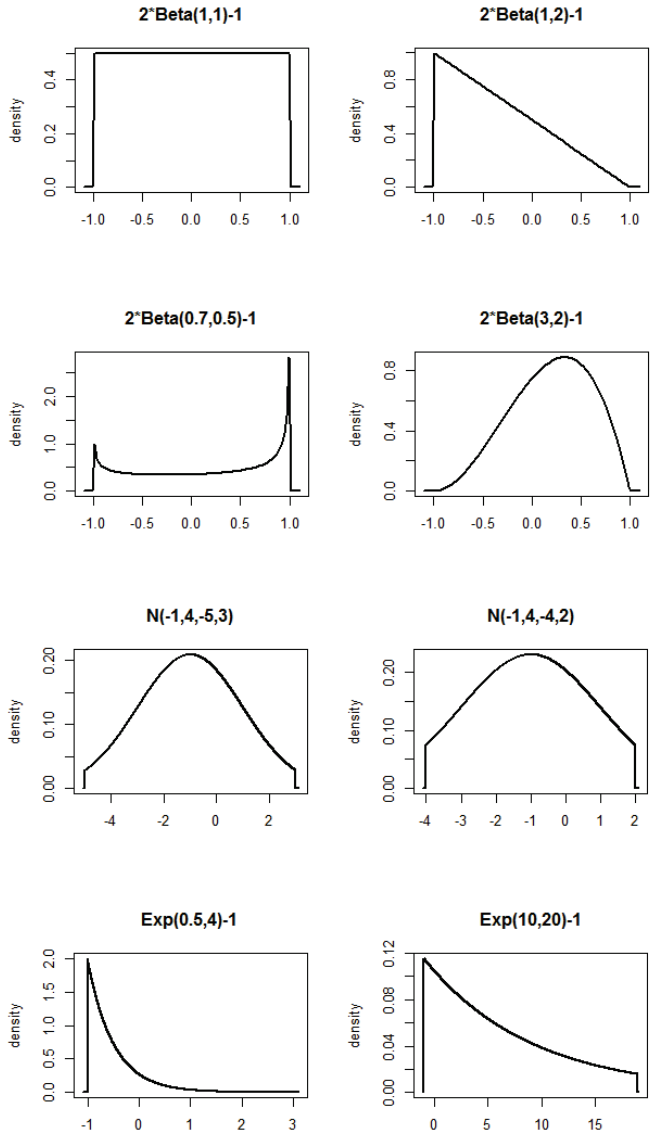


Figure 4.2: Representation of the eight densities assumed for X in the simulation study.

Table 4.1: Simulation results regarding the estimation of v, for samples of size $n = 200$

Distribution	v	Bias	RB	SE	MSE
$2{\cdot}\text{Beta}(1,1)-1$	0	.022		.106	.012
	.144	-.016	-.112	.113	.013
	.289	-.030	-.105	.124	.016
	.433	-.019	-.045	.124	.016
$2{\cdot}\text{Beta}(1,2)-1$	0	.021		.092	.009
	.118	-.016	-.134	.057	.004
	.236	-.014	-.058	.089	.008
	.354	.020	.056	.121	.015
$2{\cdot}\text{Beta}(.7,.5)-1$	0	.026		.122	.016
	.166	-.048	-.287	.078	.008
	.332	-.080	-.240	.097	.016
	.499	-.056	-.113	.131	.020
$2{\cdot}\text{Beta}(3,2)-1$	0	.076		.091	.014
	.100	.052	.524	.101	.013
	.200	.062	.308	.087	.011
	.300	.053	.176	.096	.012
$\text{N}(-1,4,-5,3)$	0	.254		.268	.136
	.440	.195	.444	.184	.072
	.880	.092	.104	.219	.057
	1.319	.001	.001	.244	.059
$\text{N}(-1,4,-4,2)$	0	.025		.112	.013
	.371	.070	.188	.164	.032
	.743	.041	.055	.182	.035
	1.114	-.021	-.019	.220	.049
$\text{Exp}(.5,4)-1$	0	.014		.078	.006
	.124	-.015	-.117	.061	.004
	.247	-.026	-.105	.068	.005
	.371	-.030	-.080	.103	.012
$\text{Exp}(10,20)-1$	0	.169		.889	.820
	1.313	-.067	-.051	.988	.980
	2.627	-.137	-.052	1.109	1.248
	3.940	.224	.057	1.341	1.849

RB: relative bias; SE: empirical standard error; MSE: empirical mean squared error.

Table 4.2: Simulation results regarding the estimation of v, for samples of size $n = 300$

Distribution	v	Bias	RB	SE	MSE
$2{\cdot}\text{Beta}(1,1)-1$	0	.020		.099	.010
	.144	-.010	-.069	.064	.004
	.289	-.011	-.037	.076	.006
	.433	-.003	-.007	.092	.009
$2{\cdot}\text{Beta}(1,2)-1$	0	.021		.092	.009
	.118	-.008	-.069	.053	.003
	.236	-.006	-.024	.072	.005
	.354	.014	.040	.101	.010
$2{\cdot}\text{Beta}(.7,.5)-1$	0	.026		.123	.016
	.166	-.037	-.221	.063	.005
	.332	-.067	-.202	.077	.010
	.499	-.052	-.104	.102	.013
$2{\cdot}\text{Beta}(3,2)-1$	0	.071		.095	.014
	.100	.073	.727	.083	.012
	.200	.065	.326	.072	.009
	.300	.059	.196	.078	.010
$\text{N}(-1,4,-5,3)$	0	.205		.255	.107
	.440	.209	.475	.137	.063
	.880	.116	.132	.177	.045
	1.319	.032	.024	.229	.053
$\text{N}(-1,4,-4,2)$	0	.016		.095	.009
	.371	.088	.238	.127	.024
	.743	.053	.071	.154	.027
	1.114	.006	.005	.189	.036
$\text{Exp}(.5,4)-1$	0	.015		.080	.007
	.124	-.008	-.066	.058	.003
	.247	-.026	-.104	.037	.002
	.371	-.029	-.077	.064	.005
$\text{Exp}(10,20)-1$	0	.150		.844	.734
	1.313	.001	.001	.947	.897
	2.627	-.049	-.019	1.019	1.042
	3.940	.121	.031	1.142	1.318

RB: relative bias; SE: empirical standard error; MSE: empirical mean squared error.

Table 4.3: Simulation results regarding the estimation of v, for samples of size $n = 1000$

Distribution	v	Bias	RB	SE	MSE
$2{\cdot}\text{Beta}(1,1) - 1$	0	.018		.097	.010
	.144	.008	.058	.067	.005
	.289	.011	.038	.066	.004
	.433	.011	.024	.064	.004
$2{\cdot}\text{Beta}(.7,.5) - 1$	0	.014		.089	.008
	.166	-.008	-.048	.093	.009
	.332	-.043	-.130	.073	.007
	.499	-.046	-.093	.060	.006
$\text{N}(-1,4,-5,3)$	0	.038		.155	.025
	.440	.223	.507	.078	.056
	.880	.139	.157	.126	.035
	1.319	.075	.057	.162	.032

RB: relative bias; SE: empirical standard error; MSE: empirical mean squared error.

Although the estimation of the pdf of X is not the main focus of this chapter, the results regarding $\widehat{f}_X$ could also be of interest. Tables 4.4, 4.6 and 4.8 contain the results pertaining to the estimation of a and b. The best results are found for the $2\text{Beta}(1,1)-1$, $2\text{Beta}(1,2)-1$ and $\text{Exp}(10,20)-1$ distributions, while the Gaussian distributions yield the worst results. The quality of the estimation of the support of X is moderate in the $2\text{Beta}(3,2)-1$ and $\text{Exp}(0.5,4)-1$ cases, although b is rather well estimated in the latter case. In general, an increase in the NSR is associated with a deterioration in the estimation of a and b, while increasing the sample size does not seem to improve the results.

There is a clear link between the estimation of v and the estimated support of X: the latter is too small when v is overestimated (in the $2\text{Beta}(3,2)-1$ and Gaussian cases), and too large when v is underestimated (with the $2\text{Beta}(0.7,0.5)-1$ distribution).

The results regarding the selection of m using the BIC can be found in Tables 4.5, 4.7 and 4.9. The true value of m for the $2\text{Beta}(1,1)-1$ and the $2\text{Beta}(1,2)-1$ distributions (0 and 1, respectively) are well recovered by the BIC criterion (and the true value is more often chosen when n increases). It is not the case for the $2\text{Beta}(3,2)-1$ (where m should be 3): this could, again, be a consequence of the closeness of this distribution to the Gaussian one. In all cases, the selected m tends to decrease with the NSR, but to increase with the sample size (except with the Gaussian distributions).

The graphical representation of the results regarding the estimation of f_X can be found in Figure 4.3. For each distribution of X, the estimated pdf of X for a random sample of 50 datasets is represented, for $n = 300$ and NSR$= 0.5$. As could be expected from the previous results, the quality of the estimation is good for the $2\text{Beta}(1,1)-1$, $2\text{Beta}(1,2)-1$ and exponential distributions, and very poor for the Gaussian ones.

Although Model (4.1) was shown to be identifiable, there appears to be some practical identifiability problems in some of the settings under study: with the Beta distributions (although the problem is smaller with the $2\text{Beta}(0.7,0.5)$ distribution) and the Gaussian ones. In some cases, the estimated value of v is either very low (close to 0) or very high (close to the standard deviation of W). The frequency of this problem decreases with the sample size and the NSR and increases with m.

This phenomenon disappears (for all m) if we set a and b to their true value, and estimate only v and $\boldsymbol{\theta}_m$. Such a lack of practical identifiability despite theoretical identifiability is not uncommon in the context of model estimation with mismeasured covariates (Carroll et al., 2006).

Table 4.4: Simulation results regarding the estimation of a and b, for samples of size $n = 200$

				Average		SE	
Distribution	a	b	v	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
$2{\cdot}\text{Beta}(1,1)-1$	2	-1	0	1.91	-.95	.38	.21
			.144	1.98	-.99	.45	.24
			.289	2.02	-1.01	.48	.25
			.433	1.99	-.99	.49	.26
$2{\cdot}\text{Beta}(1,2)-1$	2	-1	0	1.82	-.96	.39	.16
			.118	1.91	-1.01	.21	.09
			.236	1.85	-1.02	.37	.15
			.354	1.52	-.96	.71	.28
$2{\cdot}\text{Beta}(.7,.5)-1$	2	-1	0	1.92	-.96	.40	.23
			.166	2.23	-1.09	.29	.18
			.332	2.46	-1.19	.35	.24
			.499	2.42	-1.12	.52	.37
$2{\cdot}\text{Beta}(3,2)-1$	2	-1	0	1.55	-.70	.34	.21
			.100	1.41	-.60	.38	.25
			.200	1.24	-.48	.39	.26
			.300	1.11	-.38	.52	.30
$\text{N}(-1,4,-5,3)$	8	-5	0	3.08	-1.54	.96	.48
			.440	2.45	-1.23	.83	.42
			.880	2.24	-1.11	1.34	.70
			1.319	2.31	-1.16	1.75	.90
$\text{N}(-1,4,-4,2)$	6	-4	0	2.88	-1.44	.45	.22
			.371	2.38	-1.19	.63	.34
			.743	2.13	-1.06	1.04	.53
			1.114	2.12	-1.06	1.47	.75
$\text{Exp}(.5,4)-1$	4	-1	0	2.86	-.98	.68	.08
			.124	2.84	-1.03	.63	.09
			.247	2.73	-1.05	.70	.13
			.371	2.45	-1.06	.99	.22
$\text{Exp}(10,20)-1$	20	-1	0	19.50	-.76	3.56	1.13
			1.313	20.09	-1.07	4.21	1.29
			2.627	20.39	-1.23	5.77	1.91
			3.940	17.63	-1.02	7.96	2.82

Table 4.5: Simulation results regarding the selection of m, for samples of size $n = 200$

		Distribution (in %) of the selected m						
Distribution	v	0	1	2	3	4	5	6
$2{\cdot}\text{Beta}(1,1)-1$	0	91.4	4.4	1.6	.4	1.4	.4	.4
	.144	78.2	1.6	15.4	1.8	1.6	.8	.6
	.289	82.6	2.6	10.6	1.6	1.2	.8	.6
	.433	92.6	1.4	3.2	.8	1.2	.4	.4
$2{\cdot}\text{Beta}(1,2)-1$	0	.4	86.2	9.2	2.4	.6	.6	.6
	.118	1.2	86.0	3.6	6.6	2.4	.2	.0
	.236	17.2	72.4	2.8	6.4	.6	.4	.2
	.354	60.4	35.0	1.0	1.2	.4	1.2	.8
$2{\cdot}\text{Beta}(.7,.5)-1$	0	.6	1.0	27.8	40.0	10.6	11.2	8.8
	.166	28.4	33.2	35.6	2.0	.4	.2	.2
	.332	54.6	37.0	4.4	3.0	.4	.2	.4
	.499	79.0	16.8	1.2	1.6	.6	.2	.6
$2{\cdot}\text{Beta}(3,2)-1$	0	19.8	25.6	11.4	37.6	4.4	1.2	.0
	.100	42.4	32.0	4.8	16.8	2.6	.4	1.0
	.200	73.0	19.2	1.0	3.8	1.6	.8	.6
	.300	88.0	7.6	1.2	1.4	.6	.4	.8
$\text{N}(-1,4,-5,3)$	0	46.4	1.0	36.8	6.2	8.0	1.0	.6
	.440	92.0	1.4	3.0	1.2	1.2	.8	.4
	.880	94.8	2.2	.6	.6	1.4	.2	.2
	1.319	94.4	2.0	1.2	.2	1.2	.4	.6
$\text{N}(-1,4,-4,2)$	.0	18.2	.4	71.8	7.2	1.2	.4	.8
	.371	87.8	2.4	5.4	2.0	1.4	.4	.6
	.743	95.8	2.8	.6	.4	.2	.0	.2
	1.114	96.8	2.0	.2	.2	.2	.6	.0
$\text{Exp}(.5,4)-1$	0	.4	.0	1.2	24	38.2	27.8	8.4
	.124	.0	.0	6.2	42.6	33.8	13.4	4
	.247	.4	.0	19.8	49.8	22.6	6.2	1.2
	.371	9.2	7.2	35.4	32.2	13.6	2.0	.4
$\text{Exp}(10,20)-1$	0	.2	56.2	34.0	6.4	2.2	1.0	.0
	1.313	1.0	57.0	35.4	3.8	2.2	.6	.0
	2.627	5.2	78.8	10.0	2.8	1.0	1.0	1.2
	3.940	45.4	47.0	4.2	2.0	.4	.6	.4

Table 4.6: Simulation results regarding the estimation of a and b, for samples of size $n = 300$

				Average		SE	
Distribution	a	b	v	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
$2{\cdot}\text{Beta}(1,1)-1$	2	-1	0	1.92	-.96	.36	.19
			.144	2.01	-1.00	.23	.13
			.289	2.01	-1.00	.27	.14
			.433	1.95	-.98	.44	.22
$2{\cdot}\text{Beta}(1,2)-1$	2	-1	0	1.84	-.97	.39	.14
			.118	1.90	-1.00	.24	.10
			.236	1.86	-1.00	.35	.14
			.354	1.65	-.99	.59	.22
$2{\cdot}\text{Beta}(.7,.5)-1$	2	-1	0	1.92	-.95	.40	.24
			.166	2.19	-1.08	.26	.16
			.332	2.42	-1.17	.32	.20
			.499	2.43	-1.14	.44	.30
$2{\cdot}\text{Beta}(3,2)-1$	2	-1	0	1.57	-.72	.38	.22
			.100	1.37	-.59	.34	.22
			.200	1.26	-.50	.37	.24
			.300	1.11	-.39	.49	.29
$\text{N}(-1,4,-5,3)$	8	-5	0	3.27	-1.64	.92	.46
			.440	2.45	-1.23	.70	.36
			.880	2.16	-1.08	1.28	.67
			1.319	2.12	-1.08	1.74	.93
$\text{N}(-1,4,-4,2)$	6	-4	0	2.93	-1.46	.38	.19
			.371	2.34	-1.17	.55	.28
			.743	2.12	-1.06	.98	.51
			1.114	2.00	-1.01	1.43	.75
$\text{Exp}(.5,4)-1$	4	-1	0	3.00	-.98	.69	.09
			.124	2.97	-1.02	.66	.10
			.247	2.90	-1.06	.55	.06
			.371	2.66	-1.07	.77	.14
$\text{Exp}(10,20)-1$	20	-1	0	19.57	-.79	3.27	1.11
			1.313	19.50	-.87	5.36	1.71
			2.627	19.24	-.79	7.02	2.32
			3.940	17.43	-.73	8.07	2.62

Table 4.7: Simulation results regarding the selection of m, for samples of size $n = 300$

		Distribution (in %) of the selected m						
Distribution	v	0	1	2	3	4	5	6
$2{\cdot}\text{Beta}(1,1)-1$	0	92.6	3.6	1.2	.8	.4	1.0	.4
	.144	90.0	2.4	6.2	.8	.0	.2	.4
	.289	92.4	3.6	3.0	.6	.2	.0	.2
	.433	93.2	3.0	1.0	1.2	1.2	.2	.2
$2{\cdot}\text{Beta}(1,2)-1$	0	1.0	88.0	7.4	1.6	1.0	.0	1.0
	.118	.2	90.8	2.0	4.6	1.8	.6	.0
	.236	7.8	85.2	2.2	3.2	.4	.6	.6
	.354	44.8	50.0	1.6	.8	1.2	1.0	.6
$2{\cdot}\text{Beta}(.7,.5)-1$	0	.4	1.2	12.0	27.8	14.2	25.4	19.0
	.166	10.4	28.4	55.6	4.8	.6	.0	.2
	.332	44.2	49.2	4.2	2.2	.0	.0	.2
	.499	74.2	22.8	1.0	.8	1.0	.2	.0
$2{\cdot}\text{Beta}(3,2)-1$	0	11.8	28.4	5.4	45.6	7.0	1.6	.2
	.100	35.4	49.2	1.0	10.8	2.0	1.2	.4
	.200	65.2	29.2	1.6	1.2	1.4	1.4	.0
	.300	84.2	12.0	.8	.6	1.2	.4	.8
$\text{N}(-1,4,-5,3)$	.000	38.2	.8	43.0	3.4	12.2	1.6	.8
	.440	93.6	1.6	1.8	.6	1.0	1.0	.4
	.880	94.2	3.2	.4	.8	.6	.0	.8
	1.319	93.0	2.6	.4	1.0	1.4	.6	1.0
$\text{N}(-1,4,-4,2)$	0	6.2	.2	85.0	5.6	2.0	.2	.8
	.371	92.4	1.4	2.4	1.6	1.2	.2	.8
	.743	96.0	2.6	.2	.2	.4	.2	.4
	1.114	97.6	2.0	.2	.0	.2	.0	.0
$\text{Exp}(.5,4)-1$	0	.0	.4	.2	12.2	34.8	36.6	15.8
	.124	.4	.2	1.2	30.8	41.8	19.4	6.2
	.247	.2	.4	8.6	52.2	27.8	10.0	.8
	.371	3.2	2.0	27.6	46.6	18.4	1.4	.8
$\text{Exp}(10,20)-1$	0	.4	52.8	35.8	8.0	2.0	.8	.2
	1.313	1.6	61.0	29.4	5.6	1.6	.6	.2
	2.627	6.0	68.6	19.6	3.2	1.6	.4	.6
	3.940	42.8	37.2	15.6	2.6	1.2	.0	.6

Table 4.8: Simulation results regarding the estimation of a and b, for samples of size $n = 1000$

				Average		SE	
Distribution	a	b	v	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
$2{\cdot}\text{Beta}(1,1)-1$	2	-1	0	1.94	-.97	.34	.16
			0.144	1.96	-.98	.30	.15
			0.289	1.93	-.96	.38	.19
			0.433	1.91	-.96	.42	.22
$2{\cdot}\text{Beta}(.7,.5)-1$	2	-1	0	1.96	-.97	.29	.18
			0.166	2.09	-1.02	.38	.21
			0.332	2.32	-1.14	.39	.21
			0.499	2.42	-1.18	.38	.23
$\text{N}(-1,4,-5,3)$	8	-5	0	3.84	-1.92	.69	.34
			0.440	2.44	-1.22	.54	.28
			0.880	2.22	-1.11	1.01	.50
			1.319	2.06	-1.03	1.51	.75

Table 4.9: Simulation results regarding the selection of m, for samples of size $n = 1000$

		Distribution (in %) of the selected m						
Distribution	v	0	1	2	3	4	5	6
$2{\cdot}\text{Beta}(1,1)-1$	0	96.4	1.8	.8	.6	.0	.2	.2
	0.144	96.4	1.4	.6	.2	.6	.4	.4
	0.289	95.6	1.2	.0	.6	.6	1.2	.8
	0.433	94.6	1.6	.6	.2	1.0	.8	1.2
$2{\cdot}\text{Beta}(.7,.5)-1$	0	.4	.6	.6	.0	.2	8.6	89.6
	0.166	0.4	1.0	86.4	10.0	.4	1.0	.8
	0.332	2.6	70.2	24.8	1.6	.4	.2	.2
	0.499	32.8	65.2	1.0	.4	.2	.4	.0
$\text{N}(-1,4,-5,3)$	0	3.0	.4	22.2	1.0	70.4	2.4	.6
	0.440	95.6	1.8	.2	.6	.8	.2	.8
	0.880	96.4	1.6	.2	.8	.2	.2	.6
	1.319	95.2	2.0	.8	.4	.6	.0	1.0

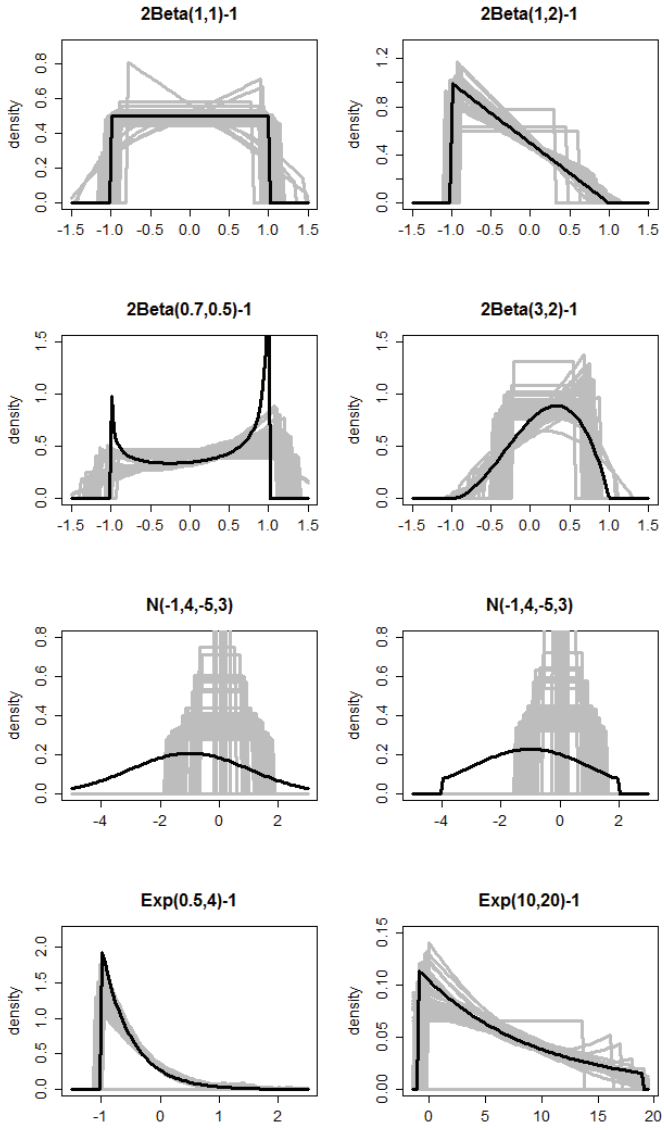


Figure 4.3: For each distribution of X, $\widehat{f}_X$ obtained for 50 datasets (in gray) and f_X (in black), in the case $n = 300$ and NSR$= 0.5$.

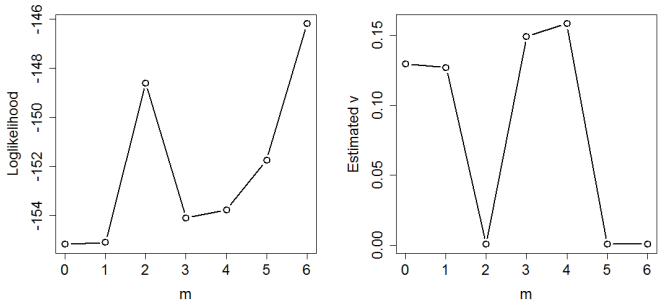


Figure 4.4: Example of a simulated dataset for which the estimation of v failed for $m = 2$, 5 and 6.

The extreme values of $\widehat{v}$ are associated with either a low or a high value of the maximum log-likelihood. In practice, these problems can be detected by examination of the value of the maximum log-likelihood as a function of m: any non-smooth pattern (a sudden large drop or increase) should be investigated, bearing in mind that a model with $m = m_1$ is nested within a model with $m = m_2 > m_1$ (Wang and Ghosh, 2012). An example of a problematic dataset is illustrated in Figure 4.4. The peak in the log-likelihood at $m = 2$ is easily detected. The issues with $m = 5$ and $m = 6$ are a bit more subtle: the growth rate of the log-likelihood increases from $m = 5$. However, the problems become obvious when looking at the evolution of $\widehat{v}$.

4.4 Illustration on the Cox proportional hazards model estimated with SIMEX

The measurement error standard deviation estimated by the proposed method could sometimes have intrinsic interest; however, in many practical situations, the final objective is not the estimation of the measurement error distribution, but rather the estimation of the parameters of a model of interest, taking this error into account. We hence illustrate the use of our proposed estimator of v in the context of survival analysis with random right censoring: the response, T, is censored by C, and T and C are assumed to be independent, conditionally on the covariates $\boldsymbol{X}$. We observe $Y = \min(T, C)$ and the censoring indicator, $\delta = I(T \leq C)$. We consider one of the most well-known survival models, the Cox proportional hazards model (Cox, 1972) introduced in Section 1.1.4 and

recalled here for convenience:

$$h(t|\boldsymbol{X}) = h_0(t)\exp(\boldsymbol{X}^T\boldsymbol{\beta}), \quad t \geq 0, \tag{4.9}$$

where $\boldsymbol{X}$ is a vector of covariates (with no intercept), $h(t|\boldsymbol{X})$ is the hazard rate at time t, i.e. the instantaneous risk of failure of T at time t given that one has survived up to this time (conditional on the covariates $\boldsymbol{X}$), and h_0 is an unspecified baseline hazard. When there is no measurement error in the covariates, the regression parameters $\boldsymbol{\beta}$ can be estimated by maximizing a partial likelihood which does not involve the baseline function h_0. However, these estimators become biased when measurement error is present (Prentice, 1982). As explained in Section 1.3, some correction methods have been developed for or adapted to this model, among which the simulation extrapolation (SIMEX) approach introduced in Section 1.3.3 and studied in a context with a cure fraction in Chapters 2 and 3. When no validation or replication data is available, all these approaches assume that the measurement error variance is known.

In this section, we study the behaviour of the SIMEX estimator of the regression parameters in the Cox model, when the measurement error variance is estimated by the approach introduced in Section 4.2.

Crainiceanu et al. (2006) adapted the SIMEX algorithm to a Cox model with a nonlinear effect of the covariates. They present the asymptotic distribution of their estimator when the error variance is assumed to be known, as well as when it is estimated. However, the result they obtain in the latter case assumes that the error variance is estimated through an estimating equation which only depends on this variance, and not on other parameters, as is the case in the approach we suggest in Section 4.2. In Section 4.6, we prove that the SIMEX estimator used with the error variance estimated through our approach is still normally distributed. We assume that the components of $\boldsymbol{X}$ and of $\boldsymbol{U}$ are mutually independent, and that $\boldsymbol{X}$ and $\boldsymbol{U}$ are independent. Under some conditions, if the correct extrapolation function is used in SIMEX and the parametric model assumed for X is correct, then

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{SIMEX}),$$

where $\boldsymbol{\Sigma}_{SIMEX}$ is given in Theorem 1 in Crainiceanu et al. (2006), where their $\psi(\cdot)$-function is replaced by the function η defined in Equation (4.11).

In the following, we present the results of a simulation study comparing the quality of the parameter estimates for model (4.9) when one covariate is measured with an error of unknown variance. More specifically, we use $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (1, -0.5)^T$ and assume that X_1 is subject to measurement error, so that $W_1 = X_1 + U$ is observed, with $U \sim N(0, v^2)$ independent of X_1,

while X_2 is measured precisely. X_1 and X_2 are independent. For X_1 and W_1, we use some of the data that were generated in the simulation study of Section 4.3: those with sample size 300 for the $2\text{Beta}(1,1)-1$, $2\text{Beta}(0.7,0.5)-1$ and $\text{N}(-1,4,-5,3)$ distributions (which correspond to situations where the bias of $\widehat{v}$ was low, moderate and high, respectively). X_2 is drawn from a Bernoulli(0.5) distribution. The baseline hazard is the hazard of an exponential distribution with mean 2, so $h_0(t) = 0.5$. The right-censoring time C is generated, independently from T, from an exponential distribution with mean 3. This results in a censoring rate of 43%. The iid sample which is available is $(Y_i, \delta_i, W_{i1}, X_{i2})$ for $i = 1, \ldots, n$. Different estimation methods are compared: the naive one (which does not take measurement error into account) and SIMEX (in order to take the error in X_1 into account) using $\widehat{v}$ from the simulation study presented in Section 4.3 as well as two arbitrarily misspecified values of v ($0.75v$ and $1.25v$). In addition, for the sake of comparison, we also consider the unrealistic case in which v is known and used in SIMEX. A quadratic extrapolation function is used in the second step of the SIMEX algorithm.

The results are reported in Tables 4.10 and 4.11. The parameter related to the covariate without measurement error, β_2, is not affected a lot by the measurement error in X_1 nor by the correction. As expected, the situation is very different for β_1.

In a practical situation, three approaches are possible: ignore the measurement error, correct with SIMEX using a given (likely misspecified) value of the error variance, or correct with SIMEX with the error variance estimated through our approach. When the error is not taken into account, the estimator of β_1 is biased, and this bias increases with the NSR. Using SIMEX with the estimated v always decreases this bias. Due to the tendency of SIMEX with a quadratic extrapolant to yield conservative corrections, as reported, for example, by Carroll et al. (2006), this bias does not completely disappear. This decrease in bias comes at the cost of a higher variance; however, the former counterbalances the latter, since the MSE is smaller with SIMEX than with the Naive method for the largest two values of NSR. SIMEX with the estimated error variance should hence clearly be preferred over no correction. One might also wonder whether we actually gain from estimating v rather than using a (likely incorrectly) specified value for it. The two values that we chose, $0.75v$ and $1.25v$, are not too far from the true value, and allow one to investigate the effect of both under- and overspecifying this value. Using SIMEX with these misspecified values also decreases the bias and the MSE, compared to the naive method. Since the SIMEX correction is conservative, overspecifying the value of v sometimes yields better results (in terms of both bias and MSE) than when v is estimated. However, unless the NSR is low, an underspecification of

v always deteriorates the results. As a consequence, in situations where it is difficult to consciously (and slightly!) overspecify the error variance, it should rather be estimated. Note that the two misspecified values of v considered in this simulation are actually not so badly specified; the superiority of the approach using the estimated v should become more apparent when considering misspecified values that are further away from the true value.

The impact of the estimation of v on the estimation of β_1 can be of theoretical interest, in order to quantify the loss in the quality of the estimation of β_1. Compared to the (hypothetical) situation in which v is known, having to estimate it has nearly no impact on the bias with the $2\text{Beta}(1,1)-1$ distribution, for which v was well estimated in Section 4.3. The bias and MSE when using the estimated error variance are nearly as good as the ones obtained when correcting with SIMEX, using the true value of v. In the $2\text{Beta}(0.7,0.5)-1$ setting, the estimation of v deteriorates the performance of the SIMEX estimation (in terms of both bias and MSE). Finally, with the $N(-1,4,-5,3)$ distribution, the estimation of v yields larger (for a low NSR), equivalent (for a medium NSR) and smaller (for a large NSR) bias and MSE. The good performance of SIMEX with a large NSR is due to the positive bias in the estimation of v.

Finally, some insight into the behaviour of SIMEX in this model can be gained by comparing the results obtained in the non-realistic settings in which v is assumed to be known when it is either correctly specified or misspecified. The results regarding the bias in the estimator of β_1 are similar for both Beta distributions: for a low or medium NSR, the lowest bias is obtained by using the true value of v. When the NSR is large, overspecifying this variance yields better results than using the true v, while the worst results appear with an underspecification of v. This result holds for the Normal distribution with a medium or high NSR.

4.5 Application

In this section, we are interested in patients with monoclonal gammopathy of undetermined significance (MGUS), a blood condition without overt malignancy, characterized by the dominance of a monoclonal protein, visible as a spike on a serum protein electrophoresis. The objective is to investigate the effect of some covariates on the survival of these individuals. The database contains information about 1341 patients monitored at the Mayo Clinic and is available in the R package Survival (R Core Team, 2015).

Table 4.10: Simulation results for the illustration on the Cox proportional hazards model (for the first two distributions)

				Naive		SIMEX (estimated v)		SIMEX (true v)		SIMEX ($.75v$)		SIMEX ($1.25v$)	
Distribution	NSR	v		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$2\cdot\text{Beta}(1,1)-1$	.25	.144	Bias	-.054	-.007	.013	-.013	.012	-.012	-.017	-.010	.050	-.015
			SE	.136	.164	.163	.165	.150	.165	.144	.165	.156	.166
			MSE	.021	.027	.027	.027	.023	.027	.021	.027	.027	.028
	.50	.289	Bias	-.217	.001	-.038	-.012	-.036	-.012	-.114	-.007	.059	-.020
			SE	.139	.168	.197	.173	.182	.172	.163	.170	.199	.175
			MSE	.066	.028	.040	.030	.034	.030	.040	.029	.043	.031
	.75	.433	Bias	-.387	.026	-.146	.011	-.147	.010	-.246	.016	-.038	.003
			SE	.112	.162	.186	.173	.168	.173	.147	.169	.190	.179
			MSE	.162	.027	.056	.030	.050	.030	.082	.029	.038	.032
$2\cdot\text{Beta}(0.7,0.5)-1$	.25	.166	Bias	-.057	-.004	-.007	-.008	.017	-.010	-.017	-.007	.059	-.014
			SE	.131	.164	.160	.166	.147	.167	.140	.165	.155	.169
			MSE	.021	.027	.026	.028	.022	.028	.020	.027	.028	.029
	.50	.332	Bias	-.234	.008	-.109	-.003	-.045	-.008	-.128	-.001	.050	-.017
			SE	.124	.161	.166	.165	.165	.168	.148	.164	.185	.172
			MSE	.070	.026	.040	.027	.029	.028	.038	.027	.037	.030
	.75	.499	Bias	-.399	.028	-.196	.009	-.156	.004	-.258	.014	-.050	-.006
			SE	.101	.157	.160	.167	.155	.169	.133	.163	.176	.175
			MSE	.169	.025	.064	.028	.048	.029	.084	.027	.033	.031

SE: empirical standard error; MSE: empirical mean squared error.

Table 4.11: Simulation results for the illustration on the Cox proportional hazards model (for the last distribution)

				Naive		SIMEX (estimated v)		SIMEX (true v)		SIMEX ($.75v$)		SIMEX ($1.25v$)	
Distribution	NSR	v		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
N(-1,4,-5,3)	.25	.440	Bias	-.238	.023	.157	-.024	-.040	-.001	-.127	.009	.060	-.013
			SE	.088	.173	.186	.202	.123	.187	.108	.180	.139	.195
			MSE	.064	.030	.060	.042	.017	.035	.028	.032	.023	.038
	.50	.880	Bias	-.557	.056	-.273	.020	-.320	.025	-.413	.037	-.227	.012
			SE	.071	.160	.144	.189	.121	.184	.103	.173	.137	.194
			MSE	.316	.029	.095	.036	.117	.034	.181	.031	.071	.038
	.75	1.319	Bias	-.739	.081	-.560	.061	-.564	.061	-.629	.068	-.505	.054
			SE	.054	.175	.106	.192	.097	.194	.082	.185	.110	.202
			MSE	.549	.037	.324	.041	.327	.041	.402	.039	.267	.044

SE: empirical standard error; MSE: empirical mean squared error.

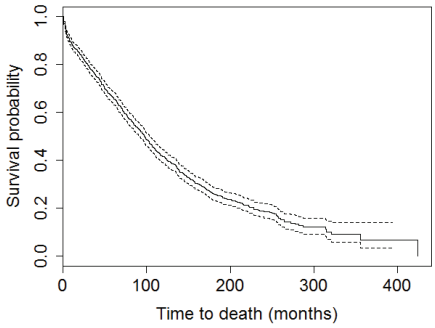


Figure 4.5: Kaplan-Meier estimate of the survival curve for the patients from the MGUS database.

The Kaplan-Meier estimate of the survival function of these patients can be found in Figure 4.5. There is no medical or empirical evidence of cure: we will use a Cox proportional hazards model to analyze these data. The median follow-up time is 84 months, while the censoring rate is 30%.

The available covariates are age (between 24 and 96, median: 72, standard deviation: 12.17), gender (54% male), hemoglobin level (between 5.7 and 18.9, median: 13.5, standard deviation: 2.02), creatinine log-level (between -0.9 and 3.1, median: 0.1, standard deviation: 0.39) and size of the monoclonal serum spike (between 0.0 and 3.0, median: 1.2, standard deviation: 0.57). The objective is to study the effect of gender, hemoglobin, log-creatinine and the size of the spike on survival while adjusting for age, in a Cox proportional hazards model. The potential problem is that, except for age and gender, these variables could be measured with some error. As previously mentioned, trying to assess their impact on the survival without taking their error into account is not helpful: we therefore analyze these data correcting for these possible measurement errors and will discuss the impact on the size and significance of the effects after correction.

The results pertaining to the estimation of the measurement error standard deviation can be found in Tables 4.12, 4.13 and 4.14 for the hemoglobin level, the creatinine log-level and the size of the monoclonal spike, respectively. The procedure introduced in Section 4.2 was applied, for each covariate, using $m = 0, \ldots, 14$, but the BIC and $\widehat{v}$ are only reported for selected values of m, around the value that was chosen by the BIC.

Table 4.15 contains the results regarding the estimation of the Cox proportional hazards model with two methods: the naive one, as well as SIMEX using,

Table 4.12: Results of the estimation of the measurement error standard deviation for the hemoglobin level.

	$m=1$	$m=2$	$\boldsymbol{m=3}$	$m=4$	$m=5$
BIC$\cdot 10^{-3}$	5.7986	5.8053	**5.7770**	5.7834	5.7904
$\widehat{v}$	1.2351	1.4911	**1.3616**	1.2798	1.3385

Table 4.13: Results of the estimation of the measurement error standard deviation for the creatinine log-level.

	$m=9$	$m=10$	$\boldsymbol{m=11}$	$m=12$	$m=13$
BIC$\cdot 10^{-3}$	0.7345	0.7284	**0.7265**	0.7271	0.7294
$\widehat{v}$	0.1836	0.1871	**0.1907**	0.1944	0.1975

for each covariate, the measurement error standard deviation as estimated previously. The covariances between the measurement errors were assumed to be zero. In an analysis that ignores the measurement error, all covariates have a significant effect on the survival, except the size of the monoclonal spike. When the measurement error is taken into account, the significance of the covariates is not modified. However, the estimated effect of the hemoglobin level is larger (in absolute value): a one-unit increase in this covariate is expected to be related to an even higher survival probability. This is also the case for gender; taking the errors in other covariates into account slightly modifies its estimated effect. This illustrates the fact that measurement error in one covariate can yield bias in the estimator of the effect of another, correctly measured covariate, when both variables are dependent (Carroll et al., 2006).

Table 4.14: Results of the estimation of the measurement error standard deviation for the size of the monoclonal spike.

	$m=7$	$m=8$	$\boldsymbol{m=9}$	$m=10$	$m=11$
BIC$\cdot 10^{-3}$	2.2785	2.2810	**2.2749**	2.2755	2.2792
$\widehat{v}$	0.1743	0.1731	**0.1780**	0.1685	0.1706

Table 4.15: Regression coefficient estimates and estimated standard deviations without and with correction for the measurement error.

		Age	Gender	Hemoglobin	Log-creatinine	Spike
No correction	Estimate	.055	-.389	-.121	.367	.037
	SE	.003	.070	.018	.079	.060
SIMEX	Estimate	.053	-.449	-.183	.349	.030
	SE	.004	.075	.026	.094	.066

4.6 Asymptotic theory for the Cox proportional hazards model

The proof of the asymptotic normality of the SIMEX estimator in the Cox proportional hazards model follows the two steps of the estimation procedure: first, v is estimated by the method introduced in Section 4.2. Then, $\boldsymbol{\beta}$ is estimated with the SIMEX algorithm.

When estimating v, we know that $\widehat{\boldsymbol{\gamma}}_m := (\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m)$ maximizes the log-likelihood

$$\mathcal{L}_n(\boldsymbol{\gamma}_m) = \mathcal{L}_n(v, a, b, \boldsymbol{\theta}_m) = n^{-1} \sum_{i=1}^{n} \log \widetilde{f}_{W,m}(W_i; \boldsymbol{\gamma}_m).$$

Hence, $\widehat{\boldsymbol{\gamma}}_m$ solves the system of equations

$$S_n(\boldsymbol{\gamma}_m) = \nabla_{\boldsymbol{\gamma}} \mathcal{L}_n(\boldsymbol{\gamma}_m) = \mathbf{0},$$

where $\nabla_{\boldsymbol{\gamma}} \mathcal{L}_n(\boldsymbol{\gamma}_m)$ is the vector of partial derivatives of $\mathcal{L}_n(\boldsymbol{\gamma}_m)$ with respect to $\boldsymbol{\gamma}_m$.

It follows from the proof of Theorem 3.2 in White (1982) that

$$\widehat{\boldsymbol{\gamma}}_m - \boldsymbol{\gamma}_m^* = -A(\boldsymbol{\gamma}_m^*)^{-1} S_n(\boldsymbol{\gamma}_m^*) + o_p(n^{-1/2}),$$

and hence

$$\widehat{v}_m - v_m^* = n^{-1} \sum_{i=1}^{n} \xi(W_i, \boldsymbol{\gamma}_m^*) + o_p(n^{-1/2}),$$

for some function ξ that is such that $E\{\xi(W, \boldsymbol{\gamma}_m^*)\} = 0$. We suppose now that the parametric model is correct, i.e. $f_W(\cdot) \equiv \widetilde{f}_{W,m}(\cdot\,; \boldsymbol{\gamma}_m)$ and so $\boldsymbol{\gamma}_m^* = \boldsymbol{\gamma}$ and

$v_m^* = v$.

The second step of the procedure consists in estimating $\boldsymbol{\beta}$ in the proportional hazards model (4.9), using the estimated value $\widehat{v}_m$ of the measurement error variance. Let d denote the dimension of the covariate vector $\boldsymbol{X}$. We allow (some components of) $\boldsymbol{X}$ to be subject to measurement error, i.e. we observe $\boldsymbol{W} = \boldsymbol{X} + \boldsymbol{U}$ where $\boldsymbol{U}$ is independent of $\boldsymbol{X}$, $\boldsymbol{U} \sim N_d(\boldsymbol{0}, \boldsymbol{V})$ and

$$\boldsymbol{V} = \begin{pmatrix} v_1^2 & & 0 \\ & \ddots & \\ 0 & & v_d^2 \end{pmatrix}.$$

For ease of notation, we assume that all $v_k^2 > 0$ (i.e., all components are subject to measurement error). We suppose that $v_1, \ldots, v_d$ are estimated using the procedure of Section 4.2:

$$\widehat{\boldsymbol{V}} = \begin{pmatrix} \widehat{v}_1^2 & & 0 \\ & \ddots & \\ 0 & & \widehat{v}_d^2 \end{pmatrix},$$

such that, for $k = 1, \ldots, d$,

$$\widehat{v}_k - v_k = n^{-1} \sum_{i=1}^{n} \xi_k(W_{i,k}, \boldsymbol{\gamma}_k) + o_p(n^{-1/2}).$$

Let $\xi(\boldsymbol{W}, \boldsymbol{\Gamma}) = (\xi_1(W_1, \boldsymbol{\gamma}_1), \ldots, \xi_d(W_d, \boldsymbol{\gamma}_d))^T$, where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)^T$.

First, consider the case where there is no measurement error. We know that $\boldsymbol{\beta}$ is estimated by solving the score equations corresponding to the partial likelihood:

$$S_n(\boldsymbol{\beta}) := \left(n^{-1} \sum_{i=1}^{n} \delta_i \left\{ X_{ik} - \frac{\sum_{j \in R_i} X_{jk} \exp(\boldsymbol{X}_j^T \boldsymbol{\beta})}{\sum_{j \in R_i} \exp(\boldsymbol{X}_j^T \boldsymbol{\beta})} \right\} \right)_{k=1}^{d} = \boldsymbol{0},$$

where R_i is the risk set of observation i (we assume for simplicity that there are no ties). Note that

$$S_n(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \psi(Y_i, \delta_i, \boldsymbol{X}_i; \boldsymbol{\beta}) + o_p(n^{-1/2}) = \boldsymbol{0},$$

for some function ψ, which is asymptotically equivalent to solving

$$n^{-1}\sum_{i=1}^{n}\psi(Y_i,\delta_i,\boldsymbol{X}_i;\boldsymbol{\beta})=\mathbf{0},$$

(note that the latter is a sum of iid terms, whereas $S_n(\boldsymbol{\beta})$ is not). The form of $n^{-1}\sum_{i=1}^{n}\psi(\cdot)$ can be found in Tsiatis (1981), who first defines some quantities:

$$\begin{aligned}
Q(t) &= P(Y\geq t,\delta=1),\\
Q(t|\boldsymbol{X}) &= P(Y\geq t,\delta=1|\boldsymbol{X}),\\
E(t) &= E\left\{\exp(\boldsymbol{X}^T\boldsymbol{\beta})I(Y\geq t)\right\}=E[\exp(\boldsymbol{X}^T\boldsymbol{\beta})P(Y\geq t|\boldsymbol{X})],\\
E_x(t) &= E\left\{\boldsymbol{X}\exp(\boldsymbol{X}^T\boldsymbol{\beta})I(Y\geq t)\right\}=E[\boldsymbol{X}\exp(\boldsymbol{X}^T\boldsymbol{\beta})P(Y\geq t|\boldsymbol{X})],\\
E_1(g(\cdot),t) &= E\left\{g(\boldsymbol{X})I(Y\geq t,\delta=1)\right\}=E[g(\boldsymbol{X})Q(t|\boldsymbol{X})].
\end{aligned}$$

Let $\widehat{Q}(t)$, $\widehat{E}(t)$, $\widehat{E}_x(t)$ and $\widehat{E}_1(g(\cdot),t)$ denote the empirical estimates of the corresponding quantities, and T_0, the bound on the censoring times: $C\leq T_0<\infty$. According to Tsiatis (1981), we have

$$n^{-1}\sum_{i=1}^{n}\psi(Y_i,\delta_i,\boldsymbol{X}_i;\boldsymbol{\beta})=C_{1n}+C_{2n}+C_{3n}+C_{4n}+C_{5n}+C_{6n},$$

where

$$\begin{aligned}
C_{1n} &= n^{1/2}[\widehat{E}_1(\boldsymbol{x},0)-E_1(\boldsymbol{x},0)],\\
C_{2n} &= -n^{1/2}[\widehat{Q}(0)-Q(0)]E_x(0)/E(0),\\
C_{3n} &= -\int_0^{T_0}[n^{1/2}(\widehat{Q}-Q)/E]dE_x,\\
C_{4n} &= \int_0^{T_0}[n^{1/2}(\widehat{Q}-Q)E_x/E^2]dE,\\
C_{5n} &= \int_0^{T_0}[n^{1/2}(\widehat{E}_x-E_x)/E]dQ,\\
C_{6n} &= -\int_0^{T_0}[n^{1/2}(\widehat{E}-E)E_x/E^2]dQ.
\end{aligned}$$

Consider now the case with measurement error. Under the SIMEX method, we generate new errors $\lambda^{1/2}\widehat{\boldsymbol{V}}^{1/2}\boldsymbol{Z}_i^b$ ($\lambda=\lambda_1,\ldots,\lambda_K$; $b=1,\ldots,B$), where $\boldsymbol{Z}_i^b\sim N(\mathbf{0},I_d)$, $i=1,\ldots,n$, independent of the data. According to the SIMEX method, we need to solve

$$n^{-1}\sum_{i=1}^{n}\psi(Y_i,\delta_i,\boldsymbol{W}_i+\lambda^{1/2}\widehat{\boldsymbol{V}}^{1/2}\boldsymbol{Z}_i^b;\boldsymbol{\beta})=\mathbf{0}, \tag{4.10}$$

which gives an estimator $\widehat{\boldsymbol{\beta}}_\lambda^b$. Note that (4.10) is equal to

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\psi(Y_i,\delta_i,\boldsymbol{W}_i+\lambda^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{Z}_i^b;\boldsymbol{\beta}) \\
& \quad + n^{-1}\sum_{i=1}^{n}\nabla_w^T\psi(Y_i,\delta_i,\boldsymbol{w};\boldsymbol{\beta})|_{\boldsymbol{w}=\boldsymbol{W}_i+\lambda^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{Z}_i^b}\lambda^{1/2}(\widehat{\boldsymbol{V}}^{1/2}-\boldsymbol{V}^{1/2})\boldsymbol{Z}_i^b \\
& \quad +o_p(n^{-1/2})=\mathbf{0},
\end{aligned}
$$

which is asymptotically equivalent to solving

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\psi(Y_i,\delta_i,\boldsymbol{W}_i+\lambda^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{Z}_i^b;\boldsymbol{\beta}) \\
& \quad + \lambda^{1/2}E\left\{\nabla_w^T\psi(Y,\delta,\boldsymbol{w};\boldsymbol{\beta})|_{\boldsymbol{w}=\boldsymbol{W}+\lambda^{1/2}\boldsymbol{V}^{1/2}\boldsymbol{Z}}\boldsymbol{Z}\right\}n^{-1}\sum_{i=1}^{n}\xi(\boldsymbol{W}_i,\boldsymbol{\Gamma}) \\
& := n^{-1}\sum_{i=1}^{n}\eta_\lambda^b(Y_i,\delta_i,\boldsymbol{W}_i;\boldsymbol{\beta},\boldsymbol{\gamma})=\mathbf{0}. \qquad (4.11)
\end{aligned}
$$

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

The rest is now similar to the usual SIMEX method. Hence, the asymptotic theory for $\widehat{\boldsymbol{\beta}}$ is as in the case where $\boldsymbol{V}$ is known (see Theorem 1 in Crainiceanu et al. (2006)), except that the ψ-function appearing in their development is replaced by the function η defined in (4.11).

4.7 Discussion

Continuous variables measured with error are frequent in practice. When they are incorporated as covariates in a response model, the measurement error must be taken into account in order to obtain valid inference. However, in most cases, the procedures designed to take that error into account require some knowledge about its distribution. When no validation or auxiliary data are available, specifying this distribution can be very tricky, especially as incorrect assumptions can yield worse results than without correction, as was seen in Chapter 3.

In this chapter, we proposed a flexible parametric approach to the estimation of the measurement error variance of Gaussian classical error. The flexibility is obtained by approximating the density of the unobserved covariate by a Bernstein polynomial. The advantage of this approach is its dependence on a single tuning parameter, which can be selected by using an information criterion, such as the BIC. We investigated the finite-sample properties of our estimator in a

simulation study. Despite some current practical identifiability problems, the results are very encouraging. We also ascertained the use of our approach in a widely-used method that corrects for the bias due to measurement error, the SIMEX algorithm. We considered the context of a Cox proportional hazards model for survival data. It appeared in a simulation study that correcting with SIMEX always yielded better results than no correction at all, even in the cases when the measurement error variance estimation was of limited quality. Finally, we analyzed the impact of several covariates on the survival of patients with monoclonal gammopathy of undetermined significance. Three covariates were thought to be subject to measurement error: the proposed approach allowed us to take that error into account despite the lack of replicated measurements or auxiliary information.

The very first challenge when it comes to measurement error is to widen the use of methods that are suited to deal with it. They are still too rarely used in practice: either because of a lack of knowledge of the consequences of the measurement error, or because they are (thought to be) difficult to use in concrete situations, sometimes with good reason.

The method introduced in this chapter is a first step towards an easier use of any method that corrects for the bias due to measurement error. However, some challenges remain unsolved: the error must be Gaussian, and it is assumed to be independent of the true value of the covariate. In the case of multiple mismeasured covariates, the true variables and the errors are also assumed to be mutually independent. These assumptions might not hold in practice. Further work is needed in order to obtain an even more flexible approach, which could suit all situations encountered in practice.

The identifiability issues that appeared in Section 4.3 would certainly deserve more investigation; however, their importance in practice should not be overstated, as such problems can often easily be dealt with, as explained at the end of that section.

Finally, in Section 4.4, the use of the proposed approach was illustrated in the context of a Cox proportional hazards model estimated by the SIMEX algorithm. Nevertheless, it must be emphasized that the measurement error variance procedure described in Section 4.2 is very general, and does not depend on a specific response model or correction method. Consequently, the theoretical and numerical analyses of Section 4.4 could be performed for any model and estimation procedure of interest.

CHAPTER 5

Discussion

5.1 Conclusion

In survival analysis, there are many situations in which a fraction of the population is known to be immune to the event of interest: these individuals are said to be *cured*. Cure models are survival models that take into account the existence of such subjects. There are two main classes of cure models: mixture cure models and promotion time cure models. Extensions and generalizations of these two classes of models have been developed too. In mixture cure models, the probability of being cured and the survival for those who are not are modeled separately. In promotion time cure models, the existence of the cured fraction is assumed by directly considering an improper survival function for the whole population. In Chapters 2 and 3, we studied the semiparametric version of the promotion time cure model: the baseline cumulative distribution function (cdf) appearing in the model was left unspecified. Compared to a fully parametric approach, it implies a computationally more involved estimation procedure and the need for a constraint on the estimated baseline cdf (the so-called *cure threshold*) to ensure identifiability, but these points are outweighed by the more flexible and robust results that are obtained.

Another phenomenon which appears quite often in statistical practice is the presence of measurement error in (some of) the covariates of interest. When these covariates are categorical, this error is called misclassification. In this

work, we studied measurement error in continuous covariates. We focused on the most famous error model, the so-called classical error model. One of the consequences of this error is a bias in the estimators of the parameters in the model of interest. This consequence is not widely known, or its importance is underestimated, and hence the measurement error is still too often ignored in practice. In Section 3.2, we derived an expression for the asymptotic bias of the regression parameter estimator in the promotion time cure model. From a practical point of view, this mathematical expression can be used to graphically represent the effect of different parameters (e.g., censoring and cure rate, measurement error variance) on the large-sample bias. This is a useful tool that can help justify the need for a correction when working with statisticians or practitioners who are not aware of the consequences of including mismeasured covariates in their model.

Many different techniques have been developed to reduce or remove the bias due to measurement error in several classes of models, among which linear and nonlinear models, as well as in survival analysis. One of these methods is SIMEX, the simulation extrapolation algorithm, which has been applied to many different contexts. SIMEX is a method based on simulations which consists in estimating the effect of the measurement error on the bias and in extrapolating to the case of no measurement error. Its drawbacks (the obtained estimators are generally only approximately consistent, the procedure can be computationally quite demanding) are largely compensated by its advantages: it has been shown to perform well in practice and its working is very intuitive. Moreover, it allows a graphical representation of the effect of both the measurement error and the correction on the bias. This is a very interesting feature when collaborating with non-statisticians.

The issue of mismeasured covariates in the promotion time cure model has not been studied a lot in the literature. To the best of our knowledge, only Ma and Yin (2008) and Mizoi et al. (2007) focused on this topic, in the semiparametric and parametric versions of the model, respectively. They both suggested a corrected score approach. In Ma and Yin (2008), the link function is restricted to be the exponential function, and the error must be Gaussian. The same assumptions are required for Mizoi et al. (2007)'s method; moreover, only one covariate can be subject to measurement error. In addition to these restrictions, the practical implementation of both approaches is not widely available.

In Chapter 2, we proposed an approach that overcomes these limitations: we adapted the SIMEX algorithm to the semiparametric promotion time cure model with mismeasured covariates. In addition to inheriting all the aforementioned advantages of the SIMEX method, our estimation procedure allows for

a link function that is more general than the one considered in Ma and Yin (2008) and Mizoi et al. (2007). Moreover, we only assumed a continuous distribution (not necessarily Gaussian) for the error. We proved the asymptotic consistency and normality of our estimator. We investigated its finite-sample behavior in a simulation study, using some of the settings considered in Ma and Yin (2008). However, in these settings, both the event and the censoring times were allowed to be infinite, so that some of the cured individuals were observable. As this is not the case in practice, we also considered a realistic setting with finite failure and censoring times, in which the cure threshold had to be used for identifiability purposes. In this study, we also compared SIMEX with two competitors, Ma and Yin (2008)'s corrected score approach and the naive method, in terms of performance. There is a tendency for SIMEX to yield a smaller bias than the naive method, but possibly a larger bias than the corrected score approach when the measurement error variance is large. However, SIMEX tends to outperform the corrected score method in terms of variance (and also of MSE).

We implemented in the R language our proposed approach, as well as the corrected score and the naive methods, and made this program available in the R package `miCoPTCM`, which is briefly described in Section 3.3.3, and whose use is illustrated in Appendix B.

In Section 2.5, we illustrated the use of our SIMEX estimator by analyzing the effect of the ejection fraction, measured with error, on the survival of patients suffering from aortic insufficiency, a heart disease. We observed that taking the measurement error into account yielded a larger (in absolute value) estimate of this effect.

One feature of SIMEX that sometimes triggers some criticism is the extrapolation function that is used in the second step of the algorithm. Carroll et al. (2006) explain that the choice of this function should be faced like any modeling decision. In this thesis, we mainly used the quadratic extrapolant: although it yields somewhat conservative corrections, it has been shown to perform quite well in many other models. Since, in practice, this extrapolant could be misspecified, we performed a simulation study investigating the impact of the choice of this extrapolant (linear, quadratic or cubic) on the results in the promotion time cure model. The results, reported in Section 2.4.4, should reassure practitioners. As can be expected from the previous literature about SIMEX, when the measurement error variance is not too small, an extrapolation function of higher order yields a smaller bias, but a larger variance. In practice, we believe that this choice should be based on a visual inspection of the graph of the estimates as a function of the measurement error variance, keeping in

mind that the function degree can be seen as a tuning parameter allowing us to control the degree of the bias correction.

In practice, a more problematic assumption of SIMEX is that the error variance is assumed to be known. Nearly all correction methods actually require some information about the error distribution. Sometimes, this information can be recovered from validation or replicate data. Many techniques were also developed to deal with instrumental variables. When such data are not available, external studies or educated guesses can provide some insights, although they must be used with care: their transportability is not always guaranteed. Because we are interested in contexts in which no auxiliary data are available, the variance used in our SIMEX estimator is likely to be misspecified. We hence investigated, in a simulation study in Section 2.4.7, the robustness of our approach with respect to a misspecification of this variance. This study provided us with valuable insights: since SIMEX with a quadratic extrapolant tends to yield conservative corrections, it appears that assuming a larger value for the error variance can actually improve the bias, at least when the true measurement error variance is large enough.

Knowledge of the error variance is also required by Ma and Yin (2008)'s corrected score method. Another common assumption is that the error distribution is known: normality is required for the corrected score approach, and often assumed for SIMEX. These constraints can hinder the use of correction methods in a practical setting: one may fear the consequences of a misspecification of this variance or distribution on the obtained results.

As a first step toward addressing this problem in the context of the promotion time cure model, we wanted to understand the behaviour of the estimation methods in different realistic settings, and hence the consequences of a possible misspecification of the variance or distribution of the error. Such a knowledge should allow practitioners to make informed decisions about whether to use a correction method in a given setting and, if so, about the specific method that is likely to perform best. We hence performed an extensive simulation study, whose results are reported and analyzed in Chapter 3. The objective was to investigate and compare the robustness with respect to both assumptions (known variance and normality) of three possible estimation methods for the semiparametric promotion time cure model: SIMEX, Ma and Yin (2008)'s corrected score approach, and the naive method. Although generalizations of these results to other settings should be considered with care, some trends appeared, that allowed us to make some recommendations regarding the questions that could be faced by a practitioner. The conclusions appeared to depend on

the estimation method, on the criterion to be used (bias or MSE) and on the covariate of interest. A key result that influences the recommendations, and hence that should be kept in mind, is that the corrected score approach tends to correct too much, while the opposite is true for SIMEX.

We illustrated the use of these recommendations in Section 3.4, where we analyzed the effect of the hemoglobin level on the time until recurrence in rectal cancer patients. It appeared that, in this specific example, the conclusions from the different estimation methods were very similar to each others: the sign and significance of each coefficient remained unchanged. Correcting for the error yielded larger estimated values for the intercept and the coefficient of the mismeasured hemoglobin level.

The difficulties encountered in practice when trying to use a correction method for the promotion time cure model motivated us to address one of these issues in a general context, outside any specific response model. In Chapter 4, we proposed a way of recovering one of the key elements required by SIMEX and many other correction methods: the measurement error variance.

We studied the case of a covariate with Gaussian classical measurement error of unknown variance and considered the situation where this variable is observed only once per observation. Schwarz and Van Bellegem (2010) showed that this error model is identifiable when making only weak assumptions about the distribution of the true covariate. However, they did not give any detail regarding the practical implementation of their proposed estimator. In a closely related context, Delaigle and Hall (2016) developed an estimation method for the distribution of the true covariate. This approach relies on several tuning parameters whose choice is rather tough in practice. Other methods that appeared in theoretical papers are subject to the same criticism and are hence of very limited use for practitioners.

In Chapter 4, we developed an estimation procedure for the error variance which makes minimal assumptions about the distribution of the true covariate, which is not too computationally demanding, and which depends on only one tuning parameter. These advantages are the consequences of the main two features of our method. First, the estimator is obtained by maximizing a likelihood with a finite number of parameters. Second, the distribution of the true covariate is not assumed to be known, but is approximated by a Bernstein polynomial, which can be seen as a mixture of Beta probability density functions. The only tuning parameter is the degree of the polynomial. We suggested using an information criterion, e.g., the BIC, to select the optimal value of this parameter.

We performed a simulation study to assess the quality of the obtained estimates in different settings: the results, presented in Section 4.3, are very encouraging. They are very good in some cases, a bit less in other situations, but even an imprecise estimation of the variance can be an improvement over a total guess, especially when there is no available information at all about this variance.

There seems to be currently some practical identifiability issues with some distributions of the true covariate, especially for models with many parameters: a quite large fraction of estimates were either close to 0 or to the variance of the observed covariate. The BIC, by selecting more sparse models, reduced the occurence of these extreme values in the simulation study. Fortunately, we discovered that such problems can most often be detected and solved in practice. This can easily be done by inspecting the evolution of the likelihood evaluated at the estimated value: since it should always increase with the polynomial degree, any decrease should be investigated.

The final objective is rarely to estimate this error variance, but rather to estimate the parameters of a response model of interest, while taking this error into account. In Section 4.4, we thus investigated the usefulness of our variance estimator, i.e., the quality of the estimators in the response model obtained by a given correction method. Due to its practical advantages, we still used SIMEX. However, since the issue addressed in this chapter is very general, and our approach, quite innovative, we wanted to study a very general model: we chose the survival proportional hazards model. Here again, the simulation results are encouraging: even when the error variance is rather badly estimated, SIMEX using this estimated variance outperforms the naive method.

Finally, we analyzed a real database about the survival of patients with monoclonal gammopathy of undetermined significance.

5.2 Further work

Despite the growing interest in cure models in the statistical literature, these models are still not widely used by practitioners, as was reported by Othus et al. (2012) in the clinical research field. Some approaches have been implemented in R and SAS; however, non-programming software like JMP and SPSS, widely used in practice, do not seem to contain such features. Making the estimation of cure models broadly available could clearly help propagate their use.

Statisticians themselves are not always fully aware of the relevance of cure models. More communication, e.g. in basic survival analysis courses, could help make them realize the advantages of these approaches, such as the evaluation of the impact of covariates on the probability of being cured, in situations

where cure is a possibility.

Exactly the same can be said of methods correcting for the bias due to mismeasured covariates: their use is still not widely spread, partly because of a lack of knowledge of the possible consequences of measurement error, but also because they are only implemented (when implemented at all) in programming software. Here again, awareness should be raised among statisticians and practitioners.

The main focus should however be on *developing* correction methods that are easy to use and meet the constraints faced in practice regarding the available information about the error. The method we developed in Chapter 4 is an important first step towards this objective, since it allows one to estimate the variance of Gaussian classical measurement error without validation or auxiliary information. As mentioned in the previous section, this method currently seems to suffer from identifiability problems in some situations. These problems should be further analyzed, with the objective of either finding an automated way of dealing with them in practice, or even suppressing them. It would be interesting to investigate whether a Bayesian approach to the estimation problem could be a solution.

A useful extension of our approach would consist in relaxing the assumption of Gaussian error: Schwarz and Van Bellegem (2010) explain that their identification result also holds for Cauchy distributions[1] with a zero location parameter and for stable distributions with a fixed exponent in the interval $(0, 2]$ and null skewness and location parameters. It could also be worth investigating under which assumptions the error model with a flexible modeling of the error distribution (e.g., through Bernstein polynomials) is identifiable.

Such generalizations call for practical tools to assess the plausibility of the assumptions about the error (e.g., Normal distribution) in a dataset. A way of checking the classical error model (independence of the error and the true covariate, normality) is suggested in Carroll et al. (2006), but this method requires replicated measures. Ma and Yin (2008) mention and discuss several transformations to normality, which also rely on replicates.

Another feature that can happen in practice, as was the case in the database that we analyzed in Chapter 4, is that several independent covariates are measured with error. When they are measured independently, their measurement errors should be independent, so that our procedure can be applied separately to all covariates to obtain the estimated variances. However, in some cases, the

[1] The pdf of a Cauchy-distributed random variable with location parameter $\mu \in \mathbb{R}$ and scale parameter $\gamma > 0$ is $\left[\pi\gamma\left\{1 + \left(\frac{x-\mu}{\gamma}\right)^2\right\}\right]^{-1}$, $x \in \mathbb{R}$.

measurement errors may be linked to each other. It could be worth investigating whether our approach can be extended to such a situation.

As far as cure models with mismeasured covariates are concerned, the computational aspects of the SIMEX procedure that we introduced in Chapter 2 could be improved by using an alternative estimation method for the semiparametric promotion time cure model. Portier et al. (2016) suggest using a profiling approach which reduces the size of the optimization problem, but the asymptotic variance of the estimator must be estimated by bootstrap.

In Section 2.5, we briefly considered the case of an interaction between a mismeasured covariate and another one, observed without error. This results in a new variable with an error depending on the value of the correctly measured covariate. Such a situation, as well as one in which the error variance depends on the covariate subject to the error, could happen quite frequently in practice. Consequently, it would be interesting to study the asymptotic and finite-sample properties of SIMEX in this case.

The issue of measurement error could also be studied in other versions of the promotion time cure model. A parametric or semiparametric distribution could be assumed for the baseline cdf, which would avoid the use of the so-called cure threshold for identifiability reasons. Letting this baseline distribution depend on covariates, which yields models that do not possess a proportional hazards structure, could also have some advantages: it would allow one to distinguish the effect of covariates on the probability of being cured and on the short-term survival. This feature is shared by the mixture cure model, for which, to the best of our knowledge, the issue of measurement error has not been studied: this large gap should be filled by future research.

Bibliography

Aitkin, M. and Rocci, R. (2002). A general maximum likelihood analysis of measurement error in generalized linear models. *Statistics and Computing*, 12:163–174.

Apanasovich, T. V., Carroll, R. J., and Maity, A. (2009). Simex and standard error estimation in semiparametric measurement error models. *Electronic Journal of Statistics*, 3:318.

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

Bernstein, S. (1912). Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. *Communications de la Société mathématique de Kharkow*, 13:1–2.

Bertrand, A., Legrand, C., Carroll, R. J., De Meester, C., and Van Keilegom, I. (2017a). Inference in a survival cure model with mismeasured covariates using a SIMEX approach. *Biometrika*, 104:31–50.

Bertrand, A., Legrand, C., Léonard, D., and Van Keilegom, I. (2017b). Robustness of estimation methods in a survival cure model with mismeasured covariates. *Computational Statistics and Data Analysis*, 113:3–18.

Bertrand, A., Van Keilegom, I., and Legrand, C. (2017c). Flexible parametric approach to classical measurement error variance estimation without auxiliary data. (in preparation).

Boag, J. W. (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., O'Gara, P. T., O'Rourke, 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.

Bremhorst, V. and Lambert, P. (2016). Flexible estimation in cure survival models using bayesian p-splines. *Computational Statistics and Data Analysis*, 93:270–284.

Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11:309–340.

Butucea, C., Matias, C., and Pouet, C. (2008). Adaptivity in convolution models with partially known noise distribution. *Electronic Journal of Statistics*, 2:897–915.

Carroll, R. J., Küchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Asymptotics for the simex estimator in nonlinear measurement error models. *Journal of the American Statistical Association*, 91:242–250.

Carroll, R. J., Maca, J. D., and Ruppert, D. (1999a). Nonparametric regression in the presence of measurement error. *Biometrika*, 86:541–554.

Carroll, R. J., Roeder, K., and Wasserman, L. (1999b). Flexible parametric measurement error models. *Biometrics*, 55:44–54.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective. 2nd edition.* Chapman and Hall/CRC, Boca Raton.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (1999). A new bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94:909–919.

Clayton, D. G. (1992). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In Dwyer, J. H., Feinleib, M., Lipsert, P., and Hoffmeister, H., editors, *Statistical Models for Longitudinal Studies of Health*, chapter 12, pages 301–331. Oxford University Press, New York.

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

Cooner, F., Banerjee, S., Carlin, B., and Sinha, D. (2007). Flexible cure rate modeling under latent activation scheme. *Journal of the American Statistical Association*, 102:560–572.

Corbière, F., Commenges, D., Taylor, J. M. G., and Joly, P. (2009). A penalized likelihood approach for mixture cure models. *Statistics in Medicine*, 28:510–524.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Crainiceanu, C. M., 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.

Crowther, M. J., Lambert, P. C., and Abrams, K. R. (2013). Adjusting for measurement error in baseline prognostic biomarkers included in a time-to-event analysis: a joint modelling approach. *BMC Medical Research Methodology*, 13:1–8.

Delaigle, A. and Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis*, 45:249–267.

Delaigle, A. and Hall, P. (2016). Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society, Series B*, 78:231–252.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38:1041–1046.

Greene, W. F. and Cai, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, 60:987–996.

Gu, Y., Sinha, D., and Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis*, 17:123–134.

Guan, Z. (2016). Efficient and robust density estimation using Bernstein type polynomials. *Journal of Nonparametric Statistics*, 28:250–271.

Gustafson, P. (2005). On model expansion, model contraction, identifiability and prior information: two illustrative scenarios involving mismeasured variables. *Statistical Science*, 20:111–140.

Gustafson, P., Le, N. D., and Vallée, M. (2002). A Bayesian approach to case-control studies with errors in covariables. *Biostatistics*, 3:229–243.

He, W., Yi, G. Y., and Xiong, J. (2007). Accelerated failure time models with covariates subject to measurement error. *Statistics in Medicine*, 26:4817–4832.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics*, 54:1407–1419.

Huang, Y. and Wang, C. Y. (2000). Cox regression with accurate covariates unascertainable: A nonparametric-correction approach. *Journal of the American Statistical Association*, 45:1209–1219.

Hughes, M. D. (1993). Regression dilution in the proportional hazards model. *Biometrics*, 49:1056–1066.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). Bayesian semiparametric models for survival data with a cure fraction. *Biometrics*, 57:383–388.

Janssen, P., Swanepoel, J. W. H., and Veraverbeke, N. (2012). Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, 142:1189–1197.

Janssen, P., Swanepoel, J. W. H., and Veraverbeke, N. (2014). A note on the asymptotic behavior of the Bernstein estimator of a copula density. *Journal of Multivariate Analysis*, 124:480–487.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

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

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

Lam, K. F., Fong, D. Y. T., and Tang, O. Y. (2005). Estimating the proportion of cured patients in a censored sample. *Statistics in Medicine*, 24:1865–1879.

Leblanc, A. (2012). On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics*, 64:919–943.

Li, C.-S. and Taylor, J. M. G. (2002). A semi-parametric accelerated failure time cure model. *Statistics in Medicine*, 21:3235–3247.

Li, Y. and Lin, X. (2000). Covariate measurement errors in frailty models for clustered survival data. *Biometrika*, 87:849–866.

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.

Liang, H., Härdle, W., and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *Annals of Statistics*, 27(5):1519–1535.

Lopez-Cheda, A., Cao, R., Jacome, M. A., and Van Keilegom, I. (2017). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics and Data Analysis*, 105:144–165.

Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika*, 91:331–343.

Ma, Y. and Yin, G. (2008). Cure rate model with mismeasured covariates under transformation. *Journal of the American Statistical Association*, 103:743–756.

Matias, C. (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM: Probability and Statistics*, 6:271–282.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, 16:195–211.

Meister, A. (2007). Deconvolving compactly supported densities. *Mathematical Methods of Statistics*, 16:63–76.

Mizoi, M. F., Bolfarine, H., and Pedroso-De-Lima, A. C. (2007). Cure rate model with measurement error. *Communications in Statistics - Simulation and Computation*, 36:185–196.

Moore, D. F. (2016). *Applied Survival Analysis Using R*. Springer, New York.

Müller, P. and Roeder, K. (1997). A bayesian semiparametric model for case-control studies with errors in variables. *Biometrika*, 84:523–537.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48:829–838.

Ortega, E. M. M., Barriga, G. D. C., Hashimoto, E. M., Cancho, V. G., and Cordeiro, G. M. (2014). A new class of survival regression models with cure fraction. *Journal of Data Science*, 12:107–136.

Othus, M., Barlogie, B., LeBlanc, M. L., and Crowley, J. J. (2012). Cure models as a useful statistical tool for analyzing survival. *Statistics in Clinical Cancer Research*, 18:3731–3736.

Otterstad, J. E., 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. B. G. (2000). A nonparametric mixture model for cure rate estimation. *Biometrics*, 56:237–243.

Peng, Y., Dear, K. B. G., and Denham, J. W. (1998). A generalized f mixture model for cure rate estimation. *Statistics in Medicine*, 17:813–830.

Portier, F., El Ghouch, A., and Van Keilegom, I. (2016). Efficiency and bootstrap in the promotion time cure model. *Bernoulli*, (to appear).

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69:331–342.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrika*, 18:375–389.

Richardson, S., Leblond, L., Jaussent, I., and Green, P. J. (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *Journal of the Royal Statistical Society, Series A*, 165:549–566.

Schennach, S. M. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association*, 108:177–186.

Schennach, S. M. (2016). Recent advances in the measurement error literature. *Annual Review of Economics*, 8:341–377.

Schwarz, M. and Van Bellegem, S. (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, 80:236–241.

Sin, C.-Y. and White, H. (1996). Information criteria for selecting possibly misspecified parametric models. *Journal of Econometrics*, 71:207–225.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores in generalized linear measurement error models. *Biometrika*, 74:703–716.

Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, 90:1247–1256.

Sy, J. P. and Taylor, J. M. G. (2000). Estimation in a Cox promotional hazards cure model. *Biometrics*, 56:227–236.

Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *The Annals of Statistics*, 29:66–93.

Taylor, J. M. G. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, 51:899–907.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Tournoud, M. and Ecochard, R. (2008). Promotion time models with time-changing exposure and heterogeneity: Application to infectious diseases. *Biometrical Journal*, 50:395–407.

Tsiatis, A. A. (1981). A large sample study of cox's regression model. *The Annals of Statistics*, 9:93–108.

Tsiatis, A. A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudian lcovariates measured with error. *Biometrika*, 88:447–458.

Tsodikov, A. (1998). A proportional hazards model taking account of long-term survivors. *Biometrics*, 54:1508–1516.

Tsodikov, A. (2002). Semi-parametric models of long- and short-term survival: an application to the analysis of breast cancer survival in utah by age and stage. *Statistics in Medicine*, 21:895–920.

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. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

Wang, J. and Ghosh, S. K. (2012). Shape restricted nonparametric regression with bernstein polynomial. *Computational Statistics and Data Analysis*, 56:2729–2741.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50:1–25.

Xu, K. and Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canadian Journal of Statistics*, 42:1–17.

Yakovlev, A. Y., Asselai, B., Bardou, V.-J., Fourquet, A., Hoang, T., Rochefedière, A., and Tsodikov, A. D. (1993). A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. In Asselain, B., Boniface, M., Duby, C., Lopez, C., Masson, J.-P., and Tranchefort, J., editors, *Biométrie et Analyse de Données Spatio-Temporelles*, pages 66–82. Société française de biométrie, ENSA Rennes.

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

Yan, Y. (2014). Statistical methods on survival data with measurement error.

Yi, G. Y. and Lawless, J. F. (2007). A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. *Journal of Statistical Planning and Inference*, 137:1816–1828.

Yin, G. and Ibrahim, J. G. (2005). Cure rate models: a unified approach. *Canadian Journal of Statistics*, 33:559–570.

Zeng, D., Yin, G., and Ibrahim, J. G. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, 101:670–684.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57:795–802.

Zhou, H. and Wang, C. Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, 62:657–665.

APPENDIX A

Example of data generation in R

```
set.seed(123)

# Set some parameter values

n <- 200 # sample size
meanC <- 5 # mean of censoring distribution
meanT <- 6 # mean of baseline distribution
maxC <- 30 # truncation limit for censoring
maxT <- 20 # truncation limit for baseline cdf
varCov <- matrix(nrow=3,ncol=3,0) # meas. error var-cov matrix
varCov[2,2] <- 0.1^1


# Generate covariates

   # with measurement error (standardized Uni[0,1])
X1 <- (runif(n)-.5)/sqrt(1/12)
V <- round(X1 + rnorm(n,0,varCov[2,2]),7)

   # without measurement error (Bernoulli(0.5))
Xc <- round(as.numeric(runif(n)<0.5),7)
```

```
# Generate censoring times

   # unrealistic setting: exponential distribution
C_un <- round(rexp(n,1/meanC),5)
C_un[runif(n)<0.6] <- -1 # infinite censoring times

   # realistic setting: truncated exponential distribution,
   # using inverse transform sampling
Uc <- runif(n)
C <- round((-meanC)*log(1-Uc*(1-exp(-maxC/meanC))),5)


# Compute cure probabilities

expb <- exp(0.5+X1-0.5*Xc)
cure <- exp(-expb)


# Use inverse transform sampling for generating event times
# with baseline cdf of an exponential distribution

U <- runif(n)

   # unrealistic setting: exponential distribution
T_un <- round(-meanT*log( 1+ log(1-(1-cure)*U)/expb),5)
T_un[(runif(n)<cure)] <- -1 # cured subjects

   # realistic setting: truncated exponential distribution
T <- round(-meanT*log(1+(1-exp(-maxT/meanT))
                      *log(1-(1-cure)*U)/expb),5)
T[(runif(n)<cure)] <- 99999 # cured subjects


# Compute observed times and status indicators

   # unrealistic setting
Tobs_un <- rep(NA,n) # observed times
dT <- (T_un<0)
dC <- (C_un<0)
```

```
         # infinite observed times:
Tobs_un[dT & dC] <- 99999
         # observed events:
Tobs_un[!dT & dC] <- T_un[!dT & dC]
         # censored observations:
Tobs_un[dT & !dC] <- C_un[dT & !dC]
         # censored observations or observed events:
Tobs_un[!dT & !dC] <- pmin(C_un[!dT & !dC], T_un[!dT & !dC])

d_un <- rep(NA,n) # status indicator
d_un[dT & dC] <- 2 # infinite observed time
d_un[!dT & dC] <- 1 # observed event
d_un[dT & !dC] <- 0 # censored observation
d_un[!dT & !dC] <- (Tobs_un[!dT & !dC]==T[!dT & !dC])

   # realistic setting
Tobs <- rep(NA,n)
Tobs <- pmin(C,T) # observed times
d <- rep(NA,n)
d <- (Tobs==T) # censoring indicator
```

APPENDIX B

Example of use of the R package miCoPTCM

```
library(survival)
library(miCoPTCM)

fm <- formula(Surv(Tobs,d) ~ V + Xc)

# Naive method

varCov0 <- matrix(nrow=3,ncol=3,0)
resNaive <- PTCMestimBF(fm, Dat, varCov = varCov0,
                        init = rnorm(3))
summary(resNaive)

   # Output:
Call:
PTCMestimBF.formula(formula = fm, data = Dat, varCov =  varCov0,
    init = rnorm(3))

           Estimate   StdErr z.value   p.value
Intercept  0.45590  0.19165  2.3788 0.0173677 *
V          0.92943  0.11652  7.9764 1.507e-15 ***
Xc        -0.76436  0.21994 -3.4753 0.0005104 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Ma and Yin (2008) method

resMY <- PTCMestimBF(fm, Dat, varCov = varCov, init = rnorm(3))
summary(resMY)

   # Output:
Call:
PTCMestimBF.formula(formula = fm, data = Dat, varCov = varCov,
    init = rnorm(3))

          Estimate   StdErr z.value   p.value
Intercept  0.49265  0.20525  2.4002 0.0163843 *
V          1.09160  0.15745  6.9330 4.121e-12 ***
Xc        -0.80008  0.23258 -3.4400 0.0005817 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


# SIMEX method

resSimex <- PTCMestimSIMEX(fm, Dat, errorDistEstim = "normal",
            varCov = varCov,  init = rnorm(3), orderExtrap = 1:3)
summary(resSimex)

   # Output:
Call:
PTCMestimSIMEX.formula(formula = fm, data = Dat,
    errorDistEstim = "normal", varCov = varCov, init = rnorm(3),
    orderExtrap = 1:3)

Results with extrapolant of order  1 :
          Estimate   StdErr z.value   p.value
Intercept  0.47460  0.20165  2.3536 0.0185950 *
V          1.01835  0.12686  8.0273 9.966e-16 ***
Xc        -0.78780  0.22553 -3.4931 0.0004774 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Results with extrapolant of order  2 :
          Estimate   StdErr z.value  p.value
Intercept  0.48418  0.21140  2.2903 0.022002 *
V          1.05075  0.14203  7.3982 1.38e-13 ***
Xc        -0.81181  0.23351 -3.4765 0.000508 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Results with extrapolant of order  3 :
          Estimate   StdErr z.value   p.value
Intercept  0.48653  0.21285  2.2858 0.0222686 *
V          1.04823  0.14827  7.0696 1.554e-12 ***
Xc        -0.83482  0.23765 -3.5128 0.0004434 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```