UCL Université catholique de Louvain

INSTITUT DE STATISTIQUE
BIOSTATISTIQUE ET
SCIENCES ACTUARIELLES
(ISBA)

ISBA

DISCUSSION PAPER

2017/25

Flexible parametric approach to classical measurement error variance estimation without auxiliary data

Bertrand, A., Van Keilegom, I. and C. Legrand

Flexible parametric approach to classical measurement error variance estimation without auxiliary data

Aurélie Bertrand[1], Ingrid Van Keilegom[2,1], and Catherine Legrand[1]

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

[2]Operations Research and Business Statistics Research Center, KU Leuven, Leuven, Belgium (ingrid.vankeilegom@kuleuven.be)

August 4, 2017

Abstract

Measurement error in the continuous covariates of a model generally yields bias in the estimators. It is a frequent problem in practice, and many correction procedures have been developed for different classes of models. However, in most cases, some information about the measurement error distribution is required. When neither validation nor auxiliary data (e.g., replicated measurements) are available, this specification turns out to be tricky. In this paper, we develop a flexible likelihood-based procedure to estimate the variance of classical additive error of Gaussian distribution, without additional information. The performance of this estimator is investigated both in an asymptotic way and through finite sample simulations. The usefulness of the obtained estimator when using the simulation extrapolation (SIMEX) algorithm, a widely used correction method, is then analyzed in the Cox proportional hazards model through other simulations. Finally, the whole procedure is illustrated on real data.

Keywords: Error variance; measurement error; proportional hazards model; survival analysis.

1 Introduction

When analyzing real-life data, we often have to face measurement error in (some of) the variables. Measurement error can be the consequence of an inaccurate measurement device or an imprecise recording; a common example being the hemoglobin level. This error can also arise when the long-term average value of a variable is of interest while the instantaneous quantity fluctuates over time, as is the case, for example, with blood pressure. In the data analyzed in Section 5, we are interested in the potential impact of several variables on the survival of patients suffering from monoclonal gammopathy of undetermined significance (MGUS). However, three of the available covariates are likely to be measured with error: the hemoglobin level, the creatinine level, and the size of a spike on a serum protein electrophoresis.

If we want to use such mismeasured variables as covariates in a response model of interest, we should be aware that not taking this measurement error into account when estimating the model parameters may have serious consequences (Carroll et al., 2006). The estimates of the covariate effects are biased and their significance can be totally erroneous. In order to obtain valid results, methods aiming at correcting for this bias have been proposed.

The use of any correction method requires a model linking the error and the true covariate. One of the most commonly used ones is the classical measurement error model:

$$W = X + U, \tag{1}$$

where W is the observed (mismeasured) covariate, X is the true but unobservable covariate and U is the mean-zero error, where X and U are assumed to be independent.

Carroll et al. (2006) and Schennach (2016) provide a review of the existing correction methods in nonlinear models, while Yan (2014) focuses on survival analysis. A first distinction can be made based on the assumptions needed on X: structural methods assume a (parametric) model for the distribution of X, while functional methods make no (or very weak) assumptions about

its distribution. The correction methods can further be classified according to the information or assumptions required for the model to be identifiable. Most methods require some knowledge about the distribution of U, which can be recovered from validation or auxiliary data. Validation data contain the true values X (at least for some of the observations) in addition to the observed W, while an example of an auxiliary variable is a repeated measurement of W. However, it often happens that such additional information is actually not available from the data at hand, as is the case, for example, in the MGUS data we study in Section 5. In such a situation, the distribution of the error must usually be assumed to be completely known. There are a few exceptions in some specific contexts, when strong assumptions of independence are made to obtain identifiability (Reiersøl, 1950; Schennach, 2013). However, as pointed out by Carroll et al. (2006), such theoretical identifiability results in a measurement error context 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.

Therefore, in situations where there is no access to additional data, one is very often left with methods requiring strong assumptions about the distribution of the error. Some methods assume Gaussian error; however, this is not the case for the simulation extrapolation (SIMEX) algorithm (Cook and Stefanski, 1994), an intuitive functional correction method that has been applied to many different contexts. Due to its practical advantages, we are going to consider it in more detail in Section 4. Another common assumption, which is made, among many others, by SIMEX and appears to be more crucial in practice, is the knowledge of the measurement error variance. This assumption is often a real issue: this variance is rarely known, and there is a risk of obtaining (even more) incorrect results if it is incorrectly specified (see, for example, Bertrand et al. (2017) in a survival analysis context). This may restrain statisticians from attempting to take the error into account in the data analysis. This probably 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. This calls for a way of estimating the measurement error variance from the available data.

Some ideas towards this objective can be found in the field of nonparametric deconvolution. Deconvolution approaches aim at recovering the distribution of X (and, sometimes, of U) from the observed W. Some authors relax the assumption of completely known density for U. Matias (2002), Butucea and Matias (2005), Meister (2006), Butucea et al. (2008) and Schwarz and Van Bellegem (2010) assume an error of known distribution but 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 they lead to 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 paper, we address this problem of unknown variance in the case of measurement error in continuous covariates with not replicates and propose an easy-to-use solution to estimate it from the data at hand. We suppose the classical measurement error model (1) with a Gaussian error of unknown variance v^2, i.e., $U \sim N(0, v^2)$, in the case where no auxiliary or validation data are available. We build on the work of Schwarz and Van Bellegem (2010) mentioned above to derive a likelihood-based approach for estimating the measurement error variance (and the distribution of X, considered as a nuisance parameter in this context) while making only weak assumptions about the distribution of the true covariate X, i.e., it has compact support. A kernel density estimate could be used for X; however, this latter nonparametric estimate is expected to have a lower rate of convergence to the true density function and a flexible approximate parametric estimate is preferable. Some common flexible approaches to density approximation include using a mixture of Gaussian densities (Carroll et al., 1999) 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 paper, 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).

The proposed variance estimator should greatly help make existing bias correction methods much more useable in practice: this estimator can be used in combination with any correction method for various response models. In particular, we study its behaviour when used in combination with the SIMEX algorithm and, since our application involves survival data, we focus on the estimation of the Cox proportional hazards model (Cox, 1972). This solution is then applied to the MGUS database: the effect of several variables, among which some with measurement error, on the survival of these patients, is analyzed.

In Section 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 3. Section 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 Web Appendix A. In Section 5, we present the results of our application, while a discussion and some insight into further work can be found in Section 6.

2 Methodology

We consider the classical additive measurement error model (1) with a Gaussian error of unknown variance v^2, where X is only assumed to be continuous and 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

$$X = aS + b, \tag{2}$$

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

$$f_W(w) = \frac{1}{v} \int f_X(x) \phi \left(\frac{w - x}{v} \right) dx = \frac{1}{a\, v} \int f_S \left(\frac{x - b}{a} \right) \phi \left(\frac{w - x}{v} \right) dx, \tag{3}$$

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 1 states that this model is identifiable.

Theorem 1. *There exist unique a, b, v^2, and a unique density f_S such that (3) holds true. 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. A Bernstein polynomial of degree m (for some choice of $m \geq 0$, which will be discussed later) is

$$B_m(s) = \sum_{k=0}^{m} \alpha_{k,m} b_{k,m}(s), \qquad s \in [0,1], \tag{4}$$

where

$$b_{k,m}(s) = \binom{m}{k} s^k (1-s)^{m-k}, \qquad k = 0, \ldots, m. \tag{5}$$

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 $\alpha_{k,m} = f\left(\frac{k}{m}\right)$:

$$\lim_{m\to\infty} \sup_{0\leq s\leq 1} \left| \sum_{k=0}^{m} f\Big(\frac{k}{m}\Big) b_{k,m}(s) - f(s) \right| = 0. \tag{6}$$

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), (5) and (6) yield the following approximation of $f_S(s)$:

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

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 $\widetilde{f}_{S,m}(\cdot\ ; \overline{\boldsymbol{\theta}}_m)$ must be a density, $\sum_{k=0}^{m} \theta_{k,m} = 1$. This representation hence actually depends on the vector of parameters $\boldsymbol{\theta}_m = (\theta_{0,m}, \ldots, \theta_{m-1,m})^T$, and we denote it by $\widetilde{f}_{s,m}(s; \boldsymbol{\theta}_m)$. In the following, when $\theta_{m,m}$ appears, it should be understood as

$\theta_{m,m} = 1 - \sum_{k=0}^{m-1} \theta_{k,m}$. This representation shows that f_S is 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. Web Figure 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

$$\widetilde{f}_{X,m}(x; a, b, \boldsymbol{\theta}_m) = \frac{1}{a}\widetilde{f}_{S,m}\left(\frac{x-b}{a}; \boldsymbol{\theta}_m\right) = \frac{1}{a}\sum_{k=0}^{m} \theta_{k,m} Beta_{k+1,m-k+1}\left(\frac{x-b}{a}\right), \tag{7}$$

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.

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 (7), we have that

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

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

$$\begin{aligned}\mathcal{L}_n(v, a, b, \boldsymbol{\theta}_m) &= \sum_{i=1}^{n} \log \widetilde{f}_{W,m}(W_i; v, a, b, \boldsymbol{\theta}_m) \\ &= \sum_{i=1}^{n} \log\left\{\frac{1}{av}\sum_{k=0}^{m} \theta_{k,m} \int Beta_{k+1,m-k+1}\Big(\frac{x-b}{a}\Big)\phi\Big(\frac{W_i-x}{v}\Big)dx\right\}. \end{aligned} \tag{8}$$

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. Based on the results of our simulation study (see Section 3), we suggest using the BIC, which achieves a better performance than the AIC thanks to the choice of more parsimonious models: $BIC(m) = (m+3)\log(n) - 2\mathcal{L}_n(\widehat{v}_m, \widehat{a}_m, \widehat{b}_m, \widehat{\boldsymbol{\theta}}_m)$, $m \geq 0$. The choice of the maximal value of m to be considered is also of interest. In the simulation study of Section 3, where samples of sizes 200, 300 and 1000 are considered, a maximal value of 6 appeared to be enough in nearly all cases (only a few situations occurred where $m = 5$ or 6 were chosen, otherwise lower values were sufficient).

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, $E\left[\log\left\{f_W(W)/\widetilde{f}_{W,m}(W; v, a, b, \boldsymbol{\theta}_m)\right\}\right]$. We have the following results for fixed $m \geq 0$.

Theorem 2. *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 3. *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

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

with

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

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

$\boldsymbol{\gamma}_m = (v_m, a_m, b_m, \boldsymbol{\theta}_m)$ *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 3 is $\boldsymbol{A}(\boldsymbol{\gamma}_m^*)^{-1}$, the inverse Fisher matrix. The regularity conditions of White (1982) are assumptions regarding f_W, $\widetilde{f}_{W,m}$, $\boldsymbol{\gamma}_m^*$ and the differentiability of $\widetilde{f}_{W,m}$.

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.

3 Simulation Study

Data were simulated according to models (1) and (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$(0,1)$ truncated at $(t_L, t_U) = (-2, 2)$ or $(-1.5, 1.5)$ 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 black in Figure 1. 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 (8). The simulation results regarding the estimation of v for $n = 200$, $n = 300$ and $n = 1000$ can be found in Web Table 1, Table 1 and Web Table 2, respectively. As can be expected, the largest sample size is associated with lower standard errors; this trend also appears in the bias, except with the 2Beta$(3, 2) - 1$ and the Gaussian 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$(0, 1, -2, 2)$ distributions yield the worst results: v is always overestimated; however, this bias tends to decrease when v increases. These distributions are rather close to a Gaussian one (or to a mixture of two Gaussians), which could explain the poor performance of the method in these cases. Indeed, the observed W is the convolution of a Gaussian and a near-Gaussian variable in that case, 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.

Although the estimation of the pdf of X is not the main focus of this work, the results regarding $\widehat{f}_X$ could also be of interest. Web Table 1, Table 1 and Web Table 2 contain the results pertaining to the estimation of a and b. The best results are found for the 2Beta$(1, 1) - 1$,

2Beta$(1,2)-1$, 2Beta$(0.7,0.5)$ and Exp$(10,20)-1$ distributions. The quality of the estimation of the support of X is moderate in the other cases, although b is rather well estimated for the Exp$(0.5,4)-1$ distribution. In general, an increase in the NSR is associated with a deterioration in the estimation of a and b.

The results regarding the selection of m using the BIC can be found in Web Table 3, Table 2 and Web Table 4. The true value of m for the 2Beta$(1,1)-1$ and the 2Beta$(1,2)-1$ distributions (0 and 1, respectively) is well recovered by the BIC criterion (and the true value is more often chosen when n increases). It is not the case for the 2Beta$(3,2)-1$ (where m should be 3): this could, again, be a consequence of the closeness of this distribution to a mixture of Gaussian distributions. In all cases, the selected m tends to decrease with the NSR, but to increase with the sample size (except with the Gaussian distributions).

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 2Beta$(3,2)-1$ and Gaussian cases), and too large when v is underestimated (with the 2Beta$(0.7,0.5)-1$ distribution), which is to be expected.

The graphical representation of the results regarding the estimation of f_X can be found in Figure 1. 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 2Beta$(1,1)-1$, 2Beta$(1,2)-1$ and exponential distributions, and rather poor for the Gaussian ones.

Although Model (1) was shown to be identifiable, there appears to be some practical identifiability problems in some of the settings under study: with all Beta distributions (although the problem is smaller with the 2Beta$(0.7,0.5)$ 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).

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 Web Figure 2. 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 at $m = 5$. However, the problems become obvious when looking at the evolution of $\widehat{v}$.

4 Use of our Estimator to Fit a 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 (PH) model (Cox, 1972):

$$h(t|\boldsymbol{x}) = h_0(t) \exp(\boldsymbol{x}^T \boldsymbol{\beta}), \quad t \geq 0, \tag{9}$$

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). Some correction methods have been developed for or adapted to this model (see Yan (2014) for a review) or to its extensions (for example, Gorfine et al. (2004) in the stratified Cox PH model), among which the simulation extrapolation (SIMEX) approach (Crainiceanu et al., 2006). When no validation or replication data is available, all these approaches assume a known measurement error variance.

The SIMEX algorithm developed by Cook and Stefanski (1994) is a functional method aiming at correcting for the bias due to measurement error. The method consists in estimating the effect of the error on the estimated coefficients through the generation of new contaminated datasets (the simulation step) and then extrapolating to the case of no measurement error (the extrapolation step). Since the extrapolation function used in this latter step is often approximated, SIMEX yields what are sometimes called approximately consistent estimators. This method has enjoyed great popularity in many different statistical fields since its creation, notably because of its intuitive working, the ease of its implementation, and the fact that it is associated with a graphical representation of the effect of both the error and the correction on the estimated values. However, as previously mentioned, it requires knowledge of the measurement error variance.

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 2. In Web Appendix A, we prove that this estimator is normally distributed. We assume that the components of $\boldsymbol{X}$ are mutually independent, that the components 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 Web Appendix A.

In the following, we present the results of a simulation study comparing the quality of the parameter estimates for model (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 and X_2, 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 3: those with sample size 300 for the 2Beta$(1, 1) - 1$, 2Beta$(0.7, 0.5) - 1$ and N$(0, 1, -2, 2)$ 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 $i = 1, \ldots, n$. 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. The corresponding 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 our estimation approach as well as two 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 Table 3. The parameter related to the covariate without mea-

surement 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.

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, except in one case (the $\mathrm{N}(0, 1, -2, 2)$ distribution with a NSR of 0.25, for which the relative bias in the estimation of v was the largest). 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 $2\mathrm{Beta}(1, 1)$ distribution, for which v was well estimated in Section 3. In the $2\mathrm{Beta}(0.7, 0.5) - 1$

setting, the estimation of v deteriorates the performance of the SIMEX estimation (in terms of both bias and MSE). Finally, with the $N(0, 1, -2, 2)$ distribution, the estimation of v yields larger (for a low or medium NSR) or 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.

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).

The Kaplan-Meier estimate of the survival function of these patients can be found in Web Figure 3. The median follow-up time is 84 months, while the censoring rate is 30%. The estimated median survival time is 98 months, while the estimated survival probability is 0.87 at one year, and 0.66 at five years.

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. 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 may be misleading: we therefore analyze these data with a Cox PH model 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 for the hemoglobin level, the creatinine log-level and the size of the monoclonal spike can be found in Table 4. The procedure introduced in Section 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 5 contains the results regarding the estimation of the Cox PH model with two methods: the naive one, as well as SIMEX using, 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.

6 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, correction methods developed to deal with it 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. In most cases, the methods developed to take that error into account require some knowledge about its distribution. Many of them, among which the SIMEX algorithm, one of the most widely used correction methods, require knowledge of the measurement error variance. When no validation or auxiliary data are available, it is usually impossible to estimate this variance, but misspecifying it can yield worse results than no correction.

In this paper, 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 relies in its easiness of implementation with a single tuning parameter, which can be selected by using an information criterion, such as the BIC. Moreover, once the variance estimate is obtained, it can be used in combination with any correction method, for any model of interest.

The results of a simulation study are very encouraging, and show that our variance estimator performs quite well in many situations. Some practical identifiability problems appeared in some settings, but they can easily be dealt with in practice.

In Section 4, we ascertained the use of our approach in combination with the SIMEX algorithm, in the context of a Cox PH model for survival data. We showed in a simulation study that correcting with SIMEX always yielded better results than no correction at all, even in the cases where the measurement error variance estimation was of limited quality. Nevertheless, it must be emphasized that the measurement error variance procedure described in Section 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 could be performed for any model and estimation procedure of interest.

We also 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 method introduced in this paper 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.

Acknowledgements

The authors thank Professor Raymond J. Carroll (Department of Statistics, Texas A&M University, USA) and Professor Paul Janssen (Center for Statistics, Hasselt University, Belgium) for their useful ideas and comments that helped improve this manuscript.

Aurélie Bertrand, Catherine Legrand and Ingrid Van Keilegom were supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). Ingrid Van Keilegom was also supported by the European Research Council (2016–2021, Horizon 2020, grant agreement nr. 694409).

Supplementary Materials

Web Appendix, Tables, and Figures referenced in Sections 1, 2, 3, 4 and 5 are available with this paper at the Biometrics website on Wiley Online Library.

References

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

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., Roeder, K., and Wasserman, L. (1999). 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.

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

Cox, D. 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,** 1–22.

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.

Gorfine, M., Hsu, L., and Prentice, R. L. (2004). Nonparametric correction for covariate measurement error in a stratified cox model. *Biostatistics* **5,** 75–87.

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

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.

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.

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

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.

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.

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.

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

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

Yan, Y. (2014). *Statistical Methods on Survival Data with Measurement Error*. Doctoral thesis, University of Waterloo, Waterloo, Canada.

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

Table 1: Simulation results regarding the estimation of v, a and b, for samples of size $n = 300$

			Estimation of v					Average		SE	
Distribution	a	b	v	Bias	RB	SE	MSE	$\widehat{a}$	$\widehat{b}$	$\widehat{a}$	$\widehat{b}$
$2{\cdot}\text{Beta}(1,1)-1$	2	-1	0	.020		.099	.010	1.92	-.96	.36	.19
			.144	-.010	-.069	.064	.004	2.01	-1.00	.23	.13
			.289	-.011	-.037	.076	.006	2.01	-1.00	.27	.14
			.433	-.003	-.007	.092	.009	1.95	-.98	.44	.22
$2{\cdot}\text{Beta}(1,2)-1$	2	-1	0	.021		.092	.009	1.84	-.97	.39	.14
			.118	-.008	-.069	.053	.003	1.90	-1.00	.24	.10
			.236	-.006	-.024	.072	.005	1.86	-1.00	.35	.14
			.354	.014	.040	.101	.010	1.65	-.99	.59	.22
$2{\cdot}\text{Beta}(.7,.5)-1$	2	-1	0	.026		.123	.016	1.92	-.95	.40	.24
			.166	-.037	-.221	.063	.005	2.19	-1.08	.26	.16
			.332	-.067	-.202	.077	.010	2.42	-1.17	.32	.20
			.499	-.052	-.104	.102	.013	2.43	-1.14	.44	.30
$2{\cdot}\text{Beta}(3,2)-1$	2	-1	0	.071		.095	.014	1.57	-.72	.38	.22
			.100	.073	.727	.083	.012	1.37	-.59	.34	.22
			.200	.065	.326	.072	.009	1.26	-.50	.37	.24
			.300	.059	.196	.078	.010	1.11	-.39	.49	.29
$\text{N}(0,1,-2,2)$	4	-2	0	.205		.255	.107	3.27	-1.64	.92	.46
			.220	.254	1.156	.190	.101	2.64	-1.31	.74	.38
			.440	.209	.475	.137	.063	2.45	-1.23	.70	.36
			.660	.152	.230	.151	.046	2.34	-1.17	.92	.48
$\text{N}(0,1,-1.5,1.5)$	3	-1.5	0	.016		.095	.009	2.93	-1.46	.38	.19
			.186	.036	.194	.173	.031	2.67	-1.33	.59	.30
			.371	.088	.238	.127	.024	2.34	-1.17	.55	.28
			.557	.068	.123	.126	.021	2.26	-1.13	.64	.34
$\text{Exp}(.5,4)-1$	4	-1	0	.015		.080	.007	3.00	-.98	.69	.09
			.124	-.008	-.066	.058	.003	2.97	-1.02	.66	.10
			.247	-.026	-.104	.037	.002	2.90	-1.06	.55	.06
			.371	-.029	-.077	.064	.005	2.66	-1.07	.77	.14
$\text{Exp}(10,20)-1$	20	-1	0	.150		.844	.734	19.57	-.79	3.27	1.11
			1.313	.001	.001	.947	.897	19.50	-.87	5.36	1.71
			2.627	-.049	-.019	1.019	1.042	19.24	-.79	7.02	2.32
			3.940	.121	.031	1.142	1.318	17.43	-.73	8.07	2.62

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

Table 2: 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
$\text{N}(0,1,-2,2)$	.000	38.2	.8	43.0	3.4	12.2	1.6	.8
	.220	85.0	1.0	8.0	1.6	2.8	1.2	.4
	.440	93.6	1.6	1.8	.6	1.0	1.0	.4
	.660	96.4	2.0	.0	.4	.4	.6	.2
$\text{N}(0,1,-1.5,1.5)$	0	6.2	.2	85.0	5.6	2.0	.2	.8
	.186	63.4	1.8	29.0	3.6	1.4	.2	.6
	.371	92.4	1.4	2.4	1.6	1.2	.2	.8
	.557	95.2	1.8	.6	.8	.8	.4	.4
$\text{Exp}(.5,4)-1$	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

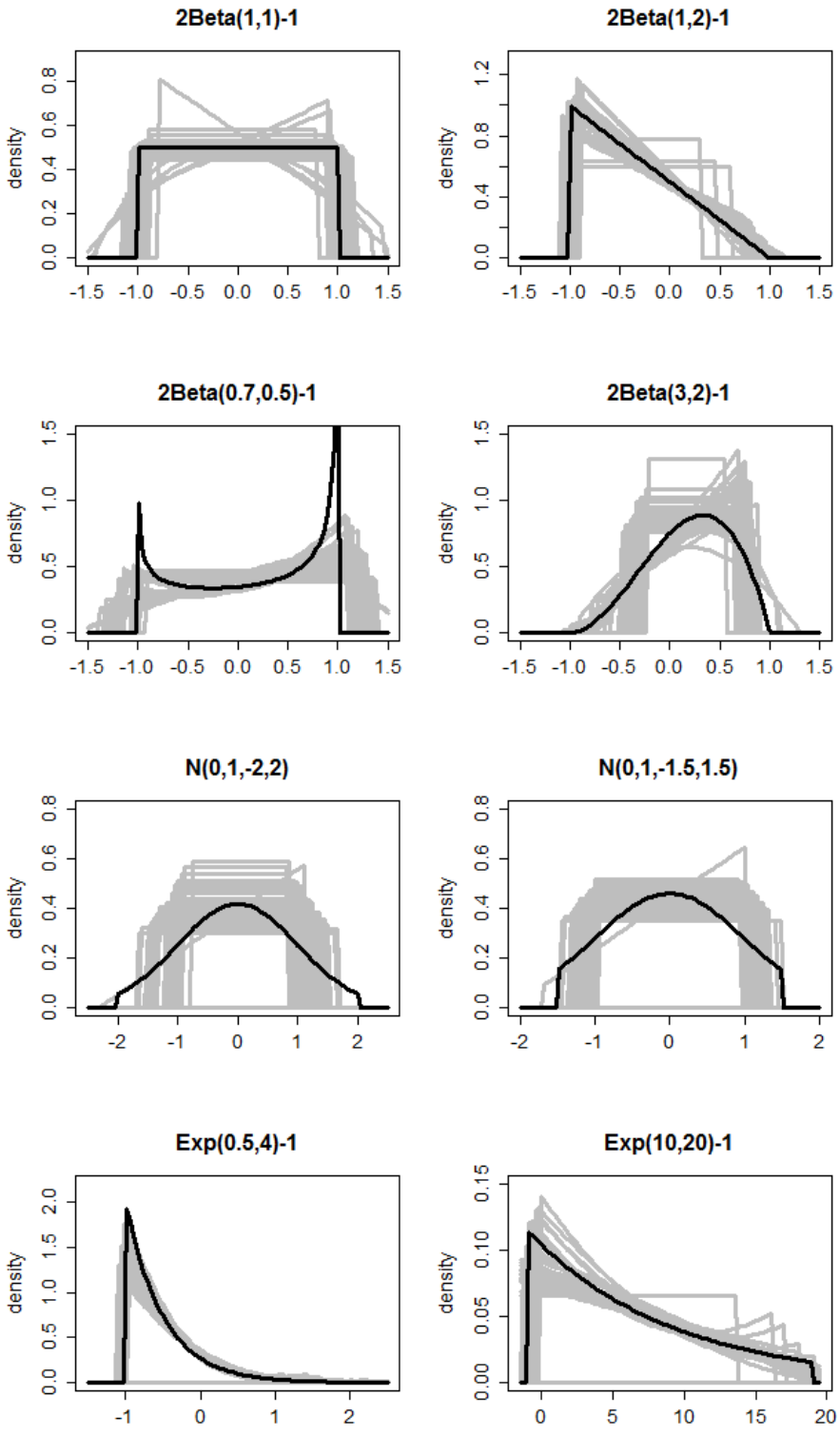

Figure 1: For each distribution of X, $\widehat{f}_X$ obtained for 50 datasets (in gray) and f_X (in black), in the case $n = 300$ and NSR$= 0.5$.

Table 3: Simulation results for the illustration on the Cox proportional hazards model

Distribution	v		Naive		Simex ($\widehat{v}$)		Simex (true v)		Simex ($.75v$)		Simex ($1.25v$)	
			β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$2\cdot\text{Beta}(1,1)-1$	.144	Bias	-.05	-.01	.01	-.01	.01	-.01	-.02	-.01	.05	-.02
		SE	.14	.16	.16	.17	.15	.17	.14	.17	.16	.17
		MSE	.02	.03	.03	.03	.02	.03	.02	.03	.03	.03
	.289	Bias	-.22	.00	-.04	-.01	-.04	-.01	-.11	-.01	.06	-.02
		SE	.14	.17	.20	.17	.18	.17	.16	.17	.20	.18
		MSE	.07	.03	.04	.03	.03	.03	.04	.03	.04	.03
	.433	Bias	-.39	.03	-.15	.01	-.15	.01	-.25	.02	-.04	.00
		SE	.11	.16	.19	.17	.17	.17	.15	.17	.19	.18
		MSE	.16	.03	.06	.03	.05	.03	.08	.03	.04	.03
$2\cdot\text{Beta}(.7,.5)-1$	.166	Bias	-.06	-.00	-.01	-.01	.02	-.01	-.02	-.01	.06	-.01
		SE	.13	.16	.16	.17	.15	.17	.14	.17	.16	.17
		MSE	.02	.03	.03	.03	.02	.03	.02	.03	.03	.03
	.332	Bias	-.23	.01	-.11	-.00	-.05	-.01	-.13	-.00	.05	-.02
		SE	.12	.16	.17	.17	.17	.17	.15	.16	.19	.17
		MSE	.07	.03	.04	.03	.03	.03	.04	.03	.04	.03
	.499	Bias	-.40	.03	-.20	.01	-.16	.00	-.26	.01	-.05	-.01
		SE	.10	.16	.16	.17	.16	.17	.13	.16	.18	.18
		MSE	.17	.03	.06	.03	.05	.03	.08	.03	.03	.03
N(0,1,-2,2)	.220	Bias	-.07	-.01	.30	-.06	.01	-.02	-.03	-.01	.05	-.03
		SE	.10	.17	.24	.20	.11	.18	.11	.17	.12	.18
		MSE	.02	.03	.15	.05	.01	.03	.01	.03	.02	.03
	.440	Bias	-.24	.02	.16	-.02	-.04	.00	-.13	.01	.06	-.01
		SE	.09	.17	.19	.20	.12	.19	.11	.18	.14	.20
		MSE	.06	.03	.06	.04	.02	.04	.03	.03	.02	.04
	.660	Bias	-.42	.04	-.08	.00	-.17	.01	-.28	.02	-.07	.00
		SE	.08	.16	.16	.19	.13	.18	.11	.17	.15	.19
		MSE	.18	.03	.03	.04	.05	.03	.09	.03	.03	.04

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

Table 4: Results of the estimation of the measurement error standard deviation for the three mismeasured covariates.

Hemoglobin level		$m=1$	$m=2$	$\boldsymbol{m=3}$	$m=4$	$m=5$
	$\text{BIC}\cdot 10^{-3}$	5.7986	5.8053	**5.7770**	5.7834	5.7904
	$\widehat{v}$	1.2351	1.4911	**1.3616**	1.2798	1.3385
Creatinine log-level		$m=9$	$m=10$	$\boldsymbol{m=11}$	$m=12$	$m=13$
	$\text{BIC}\cdot 10^{-3}$	0.7345	0.7284	**0.7265**	0.7271	0.7294
	$\widehat{v}$	0.1836	0.1871	**0.1907**	0.1944	0.1975
Monoclonal spike		$m=7$	$m=8$	$\boldsymbol{m=9}$	$m=10$	$m=11$
	$\text{BIC}\cdot 10^{-3}$	2.2785	2.2810	**2.2749**	2.2755	2.2792
	$\widehat{v}$	0.1743	0.1731	**0.1780**	0.1685	0.1706

Table 5: 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