Electronic Journal of Statistics
Vol. 16 (2022) 1831–1854
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS1996

Estimation of the variance matrix in bivariate classical measurement error models*

Elif Kekeç† **and Ingrid Van Keilegom**

ORSTAT, KU Leuven, Belgium
e-mail: elif.akca@kuleuven.be; ingrid.vankeilegom@kuleuven.be

Abstract: The presence of measurement errors is a ubiquitously faced problem and plenty of work has been done to overcome this when a single covariate is mismeasured under a variety of conditions. However, in practice, it is possible that more than one covariate is measured with error. When measurements are taken by the same device, the errors of these measurements are likely correlated.

In this paper, we present a novel approach to estimate the covariance matrix of classical additive errors in the absence of validation data or auxiliary variables when two covariates are subject to measurement error. Our method assumes these errors to be following a bivariate normal distribution. We show that the variance matrix is identifiable under certain conditions on the support of the error-free variables and propose an estimation method based on an expansion of Bernstein polynomials. To investigate the performance of the proposed estimation method, the asymptotic properties of the estimator are examined and a diverse set of simulation studies is conducted. The estimated matrix is then used by the simulation-extrapolation (SIMEX) algorithm to reduce the bias caused by measurement error in logistic regression models. Finally, the method is demonstrated using data from the Framingham Heart Study.

MSC2020 subject classifications: Primary 62F30, 62G07, 62H20; secondary 62J12, 62P99.
Keywords and phrases: Errors-in-variables, correlated measurement errors, identifiability, Bernstein polynomials, simulation-extrapolation, logistic regression.

Received December 2021.

1. Introduction

The term measurement error is commonly used to refer to situations whereby variables can only be measured with consequential error or be substituted by surrogate values because of not being accessible in principle. This might be caused by the nature of the measured quantity itself or by the measuring process. The latter could be exemplified by inaccuracies due to a measuring device,

*Financial support from the European Research Council (2016-2021, Horizon 2020 / ERC grant agreement No. 694409) is gratefully acknowledged.
†Corresponding author.
‡ORCID numbers: 0000-0002-8704-9282, 0000-0001-8827-7642.

a biased attitude during data collection, miscategorization, high expenses of a measuring process or incomplete information because of missing observations. If the presence of measurement error is not taken into account during the analysis, exceedingly biased estimates might be obtained. It has been of considerable interest to propose methods that correct measurement error in parametric, semiparametric or nonparametric models.

Various methods have been proposed to overcome measurement error such as the method of moments [22], regression calibration [14], score function based approaches [32], Bayesian methods [24], or simulation extrapolation [16]. [13], [9], [38] and [53] provide extensive collections of correction approaches for measurement errors.

In a multiple regression model, ignorance of measurement error and the application of naïve methods can bring about bias, and the consistency property of maximum likelihood estimators fails [33]. The direction in which the bias attenuates the estimates is not always the same. Even the coefficient estimates for precisely measured covariates can be biased because of the correlation with the error-prone covariates. This correlation determines the direction of the bias [11]. Besides, implementation of standard statistical techniques on data with errors-in-variables could result in concealment of meaningful characteristics of the data and loss of power in exploring the relations between covariates [13]. No matter what the nature of the constructed regression model is, neglecting measurement error could result in biased estimates. Independent of the type of model (longitudinal, survival, logistic, linear, etc.), measurement errors should therefore be taken into account when making statistical inferences [8].

If measurement error is not treated appropriately, it may cause serious problems. To avoid this, a proper specification of both the measurement error model and the distribution of the measurement error is essential. If any distributional assumption is made regarding the unobserved covariate(s), structural methods are implemented; otherwise, functional methods are used. Classical measurement error models are the most frequently considered models in the literature. They assume that

$$W = X + U, \tag{1.1}$$

where W and X refer to the surrogate and error-prone variables, respectively, while U represents the measurement error, which is presumed to be independent of X and to have zero mean. This model applies also to situations in which there are multiple covariates measured with error. In this case, all components of (1.1) are multi-dimensional.

While researchers have shown considerable interest in measurement error in a single covariate, dealing with measurement error in the case of multiple error-contaminated predictors has been studied much less. [36] offered a method to construct confidence intervals for the parameters of a logistic regression model with multiple covariates. It relies on a validation study and regression calibration. [52] constructed a bivariate measurement error model and built a logistic regression model by using replications and bio-markers. [41] noted that the regression calibration method is not capable of correcting the bias induced by

measurement errors if the individual measurement errors of the covariates of interest are associated and suggested to use a modified estimator which stands for generalized inverse-variance weighted average. In order to overcome the attenuation and problems preventing to obtain identifiable models, [45] introduced augmented validation study designs and presented semiparametric estimators based on instrumental variables. In a study by [19], measurement errors in three covariates are handled by multi-level multivariate regression calibration modeling. In order to estimate the measurement error variances, they combine the results of various research studies and perform meta-analysis to model the correlated measurement errors as well as the time effect. Besides, [1] obtain nonparametric maximum likelihood estimators for generalized linear models by the expectation-maximization (EM) algorithm under the assumption that the measurement errors of two mismeasured explanatory variables are independent. Although [31] recommend Bayesian hierarchical models via integrated nested Laplace approximations to jointly model measurement errors in two covariates, they assume no association among the covariates of interest as well as their corresponding measurement errors. [27] improve a procedure to allow for the implementation of regression calibration for the joint measurement errors in two covariates. In [6], unobserved covariates are treated as missing and analyzed by multiple overimputation based on the assumption that the observed data follow a multivariate normal distribution. Finally, [17] consider the presence of contaminated covariates in small area estimation problems. They assume structural measurement error models with uncorrelated errors and propose empirical best estimates for small area means.

In spite of the appealing properties of the above mentioned methods that constitute a vast literature to cope with a multivariate measurement error structure, the requirements for validation data, replications or auxiliary variables, the dependency on stringent assumptions or the difficulty of generalizing to the interrelated multivariate error structure, points out the need for a flexible method to estimate the covariance matrix of measurement errors. To the best of our knowledge, no practical method is offered at present that allows for the assessment of the covariance matrix that characterizes the measurement error if only one-record-at-a-time is available.

In this paper, we fulfill such a need to develop methods for bivariate correlated measurement errors in continuous covariates with an absence of auxiliary data, and we propose a methodology that is easy to implement and can be applied to error contaminated data from various fields. Our method extends the work of [5] and assumes the bivariate classical measurement error model

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} + \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \tag{1.2}$$

where $X = (X_1, X_2)^\top$ and $U = (U_1, U_2)^\top$ are independent, the density of X is unknown, and the vector U has a bivariate Gaussian distribution, i.e.

$$U = \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \right), \tag{1.3}$$

with Σ unknown. Note that not only we allow the errors U_1 and U_2 to be correlated, but we also allow the unobserved variables X_1 and X_2 to be correlated. The extension to multivariate measurement error models of dimension larger than two will be briefly discussed in Section 7. For the identifiability of measurement error models, we refer to the work of [39] as a pioneering paper in the univariate case. We will generalize their proof of the identifiability to the bivariate case, i.e. we will show that the density of X and the matrix Σ are identified even when both the errors and the unobserved variables of interest are correlated. Similarly as in the univariate case, we will only assume that X has a density on a compact (but unknown) support and that Σ is a symmetric, positive semi-definite matrix. To approximate the density of $X = (X_1, X_2)^\top$, we benefit from Bernstein polynomials [4]. The proposed methodology allows the use of the estimated error covariance matrix when applying any bias adjustment method to construct any type of regression model.

This paper is organized as follows. In the next section, we show the identifiability of the proposed model. In Section 3, we explain how to estimate the model, and we develop asymptotic properties of the model estimators. The finite-sample characteristics of the proposed estimators are reported in Section 4. In Section 5, the proposed estimators of the variance matrix are used to correct for measurement error in a logistic regression model by means of the SIMEX method. Section 6 contains an application of the proposed methodology to data from the Framingham Heart Study. Finally, conclusions and ideas for further research are given in Section 7. The online supplement [26] contains additional simulation results.

2. Identifiability

Identifiability of a model is a key property in statistics to ensure that accurate inferences can be made. In a likelihood-identifiable model, the observed data contain the essential information needed for the estimation of the model parameters ([28]; Chp. 5, p.124). In this section, we show that model (1.2)-(1.3) is likelihood-identifiable under certain additional specifications.

We first suppose that model (1.2)-(1.3) is satisfied, and that in addition, the support of $(X_1, X_2)^\top$ is compact, but unknown. For simplicity, we assume that it has a rectangular shape. Building on the work of [5], we write X_1 and X_2 as follows:

$$\begin{aligned} X_1 &= a_1 S_1 + b_1, \\ X_2 &= a_2 S_2 + b_2, \end{aligned} \tag{2.1}$$

where $S = (S_1, S_2)^\top$ is a bivariate continuous random vector defined on $[0,1] \times [0,1]$ and a_1, a_2, b_1 and b_2 are unknown parameters, with a_1 and a_2 positive. Therefore, the density of $W = (W_1, W_2)^\top$ can be written as:

$$\begin{aligned} f_{W_1,W_2}(w_1, w_2) &= \int\int f_{X_1,X_2}(x_1, x_2) f_{U_1,U_2}(w_1 - x_1, w_2 - x_2; \Sigma)\, dx_1\, dx_2 \\ &= \frac{1}{a_1 a_2} \int\int f_{S_1,S_2}\Big(\frac{x_1 - b_1}{a_1}, \frac{x_2 - b_2}{a_2}\Big) \end{aligned}$$

$$\times f_{U_1,U_2}(w_1 - x_1, w_2 - x_2; \Sigma)\, dx_1\, dx_2, \tag{2.2}$$

where $f_{S_1,S_2}(\cdot,\cdot)$ and $f_{X_1,X_2}(\cdot,\cdot)$ are the density of S and X respectively, and $f_{U_1,U_2}(\cdot,\cdot;\Sigma)$ is the bivariate Gaussian density with mean zero and variance Σ. This shows that

$$P^{W_1,W_2} = P^{X_1,X_2} * N_2(0,\Sigma),$$

where $*$ represents the convolution, and P^{W_1,W_2} and P^{X_1,X_2} are the laws of $(W_1,W_2)^\top$ and $(X_1,X_2)^\top$, respectively.

Let us now show that model (1.2)-(1.3)-(2.1) is identifiable. We will show this identifiability in a more general model which, instead of (2.1), assumes that the law P^{X_1,X_2} belongs to the set $\mathcal{P}_{2,0}$ defined as follows:

$$\mathcal{P}_{2,0} = \Big\{P \in \mathcal{P}_2 \,|\, \text{supp}(P) \subset [q_1,\infty) \times [q_2,\infty) \text{ for some } q_1, q_2 < \infty\Big\},$$

where $\mathcal{P}_2$ is the set of all bivariate probability laws, and $\text{supp}(P)$ is the support of the law P. It is worthy to note here that finite values of q_1 and q_2 are needed for the identifiability of both the diagonal and off-diagonal elements of the error covariance matrix. Many distributions fall into the class $\mathcal{P}_{2,0}$, like distributions corresponding to positive random variables (which is the case for many commonly studied variables in practice), variables defined on a compact interval (like proportions, percentages,etc.). On the other hand, it excludes the bivariate normal distribution, but this is not surprising, since it is impossible to identify a normal distribution from the convolution of two normals. So, this is in a sense, the price to pay for proving the identifiability of the error covariance matrix. Also note that in Section 4 we truncate the covariates to meet this condition. In addition, the matrix Σ needs to belong to

$$\Theta = \Big\{ \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \in \mathbb{R}^{2\times 2} \,|\, 0 < \sigma_1, \sigma_2 < \infty,\ \sigma_{12} \in \mathbb{R},\ \sigma_1^2\sigma_2^2 - \sigma_{12}^2 \geqslant 0\Big\},$$

i.e., Θ constitutes the space of 2×2 positive semi-definite covariance matrices. Note that if $(X_1,X_2)^\top$ satisfies (2.1), then P^{X_1,X_2} obviously belongs to the set $\mathcal{P}_{2,0}$, since the support of $(X_1,X_2)^\top$ is $[b_1, a_1+b_1] \times [b_2, a_2+b_2]$ in that case.

Theorem 2.1. *Suppose that* $P^{X_1,X_2} * N_2(0,\Sigma) = P^{\tilde{X}_1,\tilde{X}_2} * N_2(0,\tilde{\Sigma})$, *with* $P^{X_1,X_2}, P^{\tilde{X}_1,\tilde{X}_2} \in \mathcal{P}_{2,0}$ *and* $\Sigma, \tilde{\Sigma} \in \Theta$. *Then,* $P^{X_1,X_2} = P^{\tilde{X}_1,\tilde{X}_2}$ *and* $\Sigma = \tilde{\Sigma}$. *Hence, model (1.2)-(1.3) is identifiable in the parameter space* $\mathcal{P}_{2,0} \times \Theta$.

The proof relies on the following lemma.

Lemma 2.1. *Suppose that* $P^{X_1,X_2} \in \mathcal{P}_{2,0}$. *Then,* $P^{X_1}, P^{X_2} \in \mathcal{P}_{1,0}$, *where*

$$\mathcal{P}_{1,0} = \Big\{P \in \mathcal{P}_1 \,|\, \exists A \in \mathcal{B}(\mathbb{R}) : |A| > 0 \text{ and } P(A) = 0\Big\},$$

where $\mathcal{P}_1$ *is the set of all univariate laws,* $|A|$ *is the Lebesgue measure of a set* A *in* $\mathbb{R}$, *and* $\mathcal{B}(\mathbb{R})$ *contains all Borel sets in* $\mathbb{R}$.

Proof. Let $D = [d_1, d_2]$ such that $[d_1, d_2] \cap [q_1, \infty) = \varnothing$ and $d_1 < d_2$. Then,

$$P^{X_1}(D) = \int_{d_1}^{d_2} f_{X_1}(x_1)dx_1 = \int_{d_1}^{d_2} \int_{q_2}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2\, dx_1 = 0,$$

implying that $P^{X_1} \in \mathcal{P}_{1,0}$. A similar approach is applicable to show that $P^{X_2} \in \mathcal{P}_{1,0}$. □

Proof of Theorem 2.1. Suppose that $(P^{X_1,X_2}, \Sigma), (P^{\tilde{X}_1,\tilde{X}_2}, \tilde{\Sigma}) \in \mathcal{P}_{2,0} \times \Theta$ satisfy $P^{X_1,X_2} * N_2(0, \Sigma) = P^{\tilde{X}_1,\tilde{X}_2} * N_2(0, \tilde{\Sigma})$. Then,

$$\begin{aligned} &\int\int f_{X_1,X_2}(x_1, x_2) f_{U_1,U_2}(w_1 - x_1, w_2 - x_2; \Sigma)\, dx_1\, dx_2 \\ &= \int\int f_{\tilde{X}_1,\tilde{X}_2}(x_1, x_2) f_{\tilde{U}_1,\tilde{U}_2}(w_1 - x_1, w_2 - x_2; \tilde{\Sigma})\, dx_1\, dx_2 \end{aligned} \tag{2.3}$$

for all w_1, w_2. Substituting $f_{U_1,U_2}(w_1 - x_1, w_2 - x_2; \Sigma) = f_{U_2|U_1}(w_2 - x_2|w_1 - x_1) f_{U_1}(w_1 - x_1)$ and $f_{X_1,X_2}(x_1, x_2) = f_{X_2|X_1}(x_2|x_1) f_{X_1}(x_1)$ into (2.3) results in:

$$\begin{aligned} &\int\int f_{X_2|X_1}(x_2|x_1) f_{X_1}(x_1) f_{U_2|U_1}(w_2 - x_2|w_1 - x_1; \Sigma) f_{U_1}(w_1 - x_1; \sigma_1)\, dx_1\, dx_2 \\ &= \int\int f_{\tilde{X}_2|\tilde{X}_1}(x_2|x_1) f_{\tilde{X}_1}(x_1) f_{\tilde{U}_2|\tilde{U}_1}(w_2 - x_2|w_1 - x_1; \tilde{\Sigma}) \\ &\qquad \times f_{\tilde{U}_1}(w_1 - x_1; \tilde{\sigma}_1)\, dx_1\, dx_2. \end{aligned} \tag{2.4}$$

Note that $\int f_{U_2|U_1}(w_2 - x_2|w_1 - x_1; \Sigma) dw_2 = 1$ and that $\int f_{X_2|X_1}(x_2|x_1) dx_2 = 1$. Hence, (2.4) reduces to:

$$\int f_{X_1}(x_1) f_{U_1}(w_1 - x_1)\, dx_1 = \int f_{\tilde{X}_1}(x_1) f_{\tilde{U}_1}(w_1 - x_1)\, dx_1,$$

which shows that $P^{X_1} * N_1(0, \sigma_1^2) = P^{\tilde{X}_1} * N_1(0, \tilde{\sigma}_1^2)$. From Lemma 2.1 we know that $(P^{X_1}, \sigma_1^2), (P^{\tilde{X}_1}, \tilde{\sigma}_1^2) \in \mathcal{P}_{1,0} \times (0, \infty)$. Therefore, it follows from [39] that

$$P^{X_1} = P^{\tilde{X}_1} \quad \text{and} \quad \sigma_1^2 = \tilde{\sigma}_1^2.$$

In a similar way we can show that $P^{X_2} = P^{\tilde{X}_2}$ and $\sigma_2^2 = \tilde{\sigma}_2^2$.

Next, note that the equality $P^{X_1,X_2} * N_2(0, \Sigma) = P^{\tilde{X}_1,\tilde{X}_2} * N_2(0, \tilde{\Sigma})$ implies that

$$\varphi_{X_1,X_2}(t_1, t_2) \exp\Big\{ -\frac{1}{2}(2\sigma_{12} t_1 t_2) \Big\} = \varphi_{\tilde{X}_1,\tilde{X}_2}(t_1, t_2) \exp\Big\{ -\frac{1}{2}(2\tilde{\sigma}_{12} t_1 t_2) \Big\}$$

for all t_1, t_2, where $\varphi_{X_1,X_2}(t_1, t_2)$ is the characteristic function of any vector $(X_1, X_2)^\top$. Now choose $t_1 = t_2 = t$ and let a be any value such that $2\sigma_{12} + a > 0$ and $2\tilde{\sigma}_{12} + a > 0$. Then,

$$\varphi_{X_1+X_2}(t)\, \varphi_{N(0,2\sigma_{12}+a)}(t) = \varphi_{\tilde{X}_1+\tilde{X}_2}(t)\, \varphi_{N(0,2\tilde{\sigma}_{12}+a)}(t),$$

which shows that $P^{X_1+X_2} * N_1(0, 2\sigma_{12} + a) = P^{\tilde{X}_1+\tilde{X}_2} * N_1(0, 2\tilde{\sigma}_{12} + a)$. It is easily verified that $(P^{X_1+X_2}, 2\sigma_{12} + a), (P^{\tilde{X}_1+\tilde{X}_2}, 2\tilde{\sigma}_{12} + a) \in \mathcal{P}_{1,0} \times (0, \infty)$. Hence, again by [39], it follows that

$$P^{X_1+X_2} = P^{\tilde{X}_1+\tilde{X}_2} \quad \text{and} \quad \sigma_{12} = \tilde{\sigma}_{12}.$$

Finally, since $P^{X_1,X_2} * N_2(0, \Sigma) = P^{\tilde{X}_1,\tilde{X}_2} * N_2(0, \tilde{\Sigma})$ and $\Sigma = \tilde{\Sigma}$, it follows from the convolution theorem that $\varphi_{X_1,X_2}(t_1, t_2) = \varphi_{\tilde{X}_1,\tilde{X}_2}(t_1, t_2)$ for all t_1, t_2, and hence by the inversion theorem, $P^{X_1,X_2} = P^{\tilde{X}_1,\tilde{X}_2}$. This completes the proof. □

3. Estimation

We are now ready to develop an estimation procedure for our model. This will be done by approximating the joint density of S_1 and S_2 by Bernstein polynomials with positive coefficients. A bivariate Bernstein polynomial of degree (m_1, m_2) defined on the unit square, is given by

$$\sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \alpha^m_{k_1,k_2}\, b^m_{k_1,k_2}(s_1, s_2),$$

for some real-valued coefficients $\alpha^m_{k_1,k_2}$, $k_1 = 0, 1, \ldots, m_1$, $k_2 = 0, 1, \ldots, m_2$, where $m = (m_1, m_2)^\top$, $m_1, m_2 \geq 0$,

$$b^m_{k_1,k_2}(s_1, s_2) = \binom{m_1}{k_1}\binom{m_2}{k_2} s_1^{k_1}(1 - s_1)^{m_1-k_1} s_2^{k_2}(1 - s_2)^{m_2-k_2},$$

and $(s_1, s_2) \in [0, 1] \times [0, 1]$.

Any continuous function f defined on $[0, 1] \times [0, 1]$, can be approximated by bivariate Bernstein polynomials in the sense that

$$\lim_{m_1,m_2\to\infty} \sup_{0\leq s_1,s_2\leq 1} \Big| B_{m_1,m_2}(f; s_1, s_2) - f(s_1, s_2) \Big| = 0,$$

where

$$B_{m_1,m_2}(f; s_1, s_2) = \sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} f\Big(\frac{k_1}{m_1}, \frac{k_2}{m_2}\Big)\, b^m_{k_1,k_2}(s_1, s_2).$$

See also [46] and [2]. Hence, if we assume that $S = (S_1, S_2)^\top$ has a continuous density $f_{S_1,S_2}(\cdot, \cdot)$, it can be approximated by

$$\sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \alpha^m_{k_1,k_2} \binom{m_1}{k_1}\binom{m_2}{k_2} s_1^{k_1}(1 - s_1)^{m_1-k_1} s_2^{k_2}(1 - s_2)^{m_2-k_2}, \tag{3.1}$$

for certain coefficients $\alpha^m_{0,0}, \alpha^m_{0,1}, ..., \alpha^m_{0,m_2}, \alpha^m_{1,0}, ..., \alpha^m_{m_1,m_2}$. Moreover, note that if we define $\theta^m_{k_1,k_2} = \alpha^m_{k_1,k_2}[(m_1 + 1)(m_2 + 1)]^{-1}$, then (3.1) can be written as

$$\tilde{f}^m_{S_1,S_2}(s_1, s_2; \theta_m) = \sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \theta^m_{k_1,k_2} Beta_{k_1+1,m_1-k_1+1}(s_1) Beta_{k_2+1,m_2-k_2+1}(s_2),$$

where $Beta_{\ell_1,\ell_2}(\cdot)$ is the density of a Beta variable with parameters ℓ_1 and ℓ_2, and $\theta_m = (\theta^m_{0,0}, \theta^m_{0,1}, ..., \theta^m_{0,m_2}, \theta^m_{1,0}, ..., \theta^m_{m_1,m_2})^\top$, which is a vector of size $(m_1+1)\times(m_2+1)$. This shows that we approximate f_{S_1,S_2} by a weighted sum of products of two beta densities.

In order for $\tilde{f}^m_{S_1,S_2}(s_1, s_2; \theta_m)$ to be a valid probability density function, some constraints on the coefficients are needed, namely the area under $\tilde{f}^m_{S_1,S_2}(s_1, s_2; \theta_m)$ needs to add up to 1 and the function needs to be non-negative on its support. These constraints are set by imposing that $\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2}\theta^m_{k_1,k_2} = 1$ and that $\theta^m_{k_1,k_2} \geq 0$.

Since we approximate $f_{S_1,S_2}(s_1, s_2)$ by $\tilde{f}^m_{S_1,S_2}(s_1, s_2; \theta_m)$, the joint density $f_{X_1,X_2}(x_1, x_2)$ of X_1 and X_2 can be approximated by

$$\begin{aligned}
&\tilde{f}^m_{X_1,X_2}(x_1, x_2; a, b, \theta_m)\\
&= \frac{1}{a_1 a_2}\tilde{f}^m_{S_1,S_2}\Big(\frac{x_1-b_1}{a_1}, \frac{x_2-b_2}{a_2}; \theta_m\Big)\\
&= \frac{1}{a_1 a_2}\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2}\theta^m_{k_1,k_2} Beta_{k_1+1,m_1-k_1+1}\Big(\frac{x_1-b_1}{a_1}\Big) Beta_{k_2+1,m_2-k_2+1}\Big(\frac{x_2-b_2}{a_2}\Big),
\end{aligned} \tag{3.2}$$

where $a = (a_1, a_2)^\top$, $b = (b_1, b_2)^\top$ and $x = (x_1, x_2)^\top \in [b_1, a_1+b_1]\times[b_2, a_2+b_2]$ due to (2.1). Note that in (2.1), we need to define X_1 and X_2 as such since bivariate Bernstein polynomials are defined only on the unit square.

As a result, the density of $(W_1, W_2)^\top$, given in (2.2), can be approximated by

$$\begin{aligned}
&\tilde{f}^m_{W_1,W_2}(w_1, w_2; a, b, \Sigma, \theta_m) \\
&= \int\int \tilde{f}^m_{X_1,X_2}(x_1, x_2; a, b, \theta_m) f_{U_1,U_2}(w_1-x_1, w_2-x_2; \Sigma)\, dx_1\, dx_2\\
&= \frac{1}{a_1 a_2}\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2}\theta^m_{k_1,k_2}\int\int \beta_{k_1,m_1;k_2,m_2}(x_1, x_2; a, b)\\
&\times f_{U_1,U_2}(w_1-x_1, w_2-x_2; \Sigma)\, dx_1\, dx_2,
\end{aligned} \tag{3.3}$$

with

$$\begin{aligned}
&\beta_{k_1,m_1;k_2,m_2}(x_1, x_2; a, b)\\
&= Beta_{k_1+1,m_1-k_1+1}\Big(\frac{x_1-b_1}{a_1}\Big) Beta_{k_2+1,m_2-k_2+1}\Big(\frac{x_2-b_2}{a_2}\Big).
\end{aligned}$$

If $f_{S_1,S_2}(\cdot,\cdot)$ is continuous, then,

$$\lim_{m_1,m_2\to\infty}\sup_{w_1,w_2}\Big|\tilde{f}^m_{W_1,W_2}(w_1, w_2; a, b, \Sigma, \theta_m) - f_{W_1,W_2}(w_1, w_2)\Big| = 0.$$

This equation shows that the density of the observed variables can be accurately approximated by means of bivariate Bernstein polynomials, provided the degree of these polynomials is sufficiently large.

This motivates us to use the following estimation procedure, based on an i.i.d. sample $W_i = (W_{i1}, W_{i2})^\top$, $i = 1, \ldots, n$, having the same distribution as W. The log-likelihood of the parameters a, b, Σ and θ_m is given by

$$\mathcal{L}_n(a, b, \Sigma, \theta_m) = \sum_{i=1}^{n} \log \left\{ \frac{1}{a_1 a_2} \sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \theta^m_{k_1,k_2} \int \int \beta_{k_1,m_1;k_2,m_2}(x_1, x_2; a, b) \times f_{U_1,U_2}(W_{i1} - x_1, W_{i2} - x_2; \Sigma)\, dx_1\, dx_2 \right\}. \tag{3.4}$$

Then, the maximum likelihood estimates of (a, b, Σ, θ_m), denoted by

$$(\hat{a}, \hat{b}, \hat{\Sigma}, \hat{\theta}_m) = \arg\max_{a,b,\Sigma,\theta_m} \mathcal{L}_n(a, b, \Sigma, \theta_m),$$

can be computed for fixed m_1 and m_2, where the maximization is done with respect to the parameter space

$$\mathcal{P} = \left\{ (a, b, \Sigma, \theta_m) : a \in [0, \infty)^2, b \in \mathbb{R}^2, \Sigma \in \Theta, \theta_m \in [0, \infty)^{(m_1+1)(m_2+1)} \right\}$$

under the constraint that $\sum_{k_1=0}^{m_1} \sum_{k_2=0}^{m_2} \theta^m_{k_1,k_2} = 1$. Let us now consider degree selection for Bernstein polynomials. Polynomials with larger degrees lead to better approximations. However, in this case, we need to estimate a larger number of parameters which increases the variance. Hence, we suggest choosing the degree $(m_1, m_2)^\top$ by means of the Bayesian Information Criterion (BIC):

$$BIC(m_1, m_2) = \log(n)[(m_1 + 1)(m_2 + 1) + 6] \; - 2\mathcal{L}_n(\hat{a}, \hat{b}, \hat{\Sigma}, \hat{\theta}_m)$$

for $m_1, m_2 \geq 0$, since the number of (free) parameters is $2 + 2 + 3 + (m_1 + 1)(m_2 + 1) - 1 = (m_1 + 1)(m_2 + 1) + 6$. In practice, we fit models for several values of m_1 and m_2 and select the one for which $BIC(m_1, m_2)$ is minimal.

Asymptotic properties of the proposed estimators can now be developed for fixed $m_1, m_2 \geq 0$. Note that the vector $(\hat{a}, \hat{b}, \hat{\Sigma}, \hat{\theta}_m)$ maximizes the likelihood of a potentially misspecified model. Therefore, we use White (1982), who developed sufficient conditions for consistency and asymptotic normality of maximum likelihood (ML) estimators under potential misspecification. In that case, the target vector of parameters is given by $(a^*, b^*, \Sigma^*, \theta^*_m)$, the minimizer (over (a, b, Σ, θ_m)) of the Kulback-Leibler information criterion, defined by

$$E\left[\log f_{W_1,W_2}(W_1, W_2) - \log \tilde{f}^m_{W_1,W_2}(W_1, W_2; a, b, \Sigma, \theta_m)\right].$$

Note here that $(a^*, b^*, \Sigma^*, \theta^*_m)$ equals to the true parameter vector (a, b, Σ, θ_m) if the model is correctly specified.

We have the following result.

Theorem 3.1. *Suppose that (1.2)-(1.3)-(2.1) hold true, and that assumptions* $(A1) - (A3)$ *in [51] are met. Then,*

$$(\hat{a}, \hat{b}, \hat{\Sigma}, \hat{\theta}_m) \xrightarrow{p} (a^*, b^*, \Sigma^*, \theta^*_m).$$

If in addition assumptions $(A4)-(A6)$ *in White (1982) are met, then,*

$$\sqrt{n}\left((\hat{a},\hat{b},\hat{\Sigma},\hat{\theta}_m)-(a^*,b^*,\Sigma^*,\theta_m^*)\right)\xrightarrow{d}N\Big(0,C(\gamma_m^*)\Big),$$

where

$$C(\gamma_m^*)=A(\gamma_m^*)^{-1}B(\gamma_m^*)A(\gamma_m^*)^{-1},$$
$$A(\gamma_m^*)=\left(E\left\{\frac{\partial^2}{\partial\gamma_i\partial\gamma_j}\log\tilde{f}^m_{W_1,W_2}(W_1,W_2;\gamma_m)\right\}\Big|_{\gamma_m=\gamma_m^*}\right)^M_{i,j=1},$$

and

$$B(\gamma_m^*)$$
$$=\left(E\left\{\frac{\partial}{\partial\gamma_i}\log\tilde{f}^m_{W_1,W_2}(W_1,W_2;\gamma_m).\frac{\partial}{\partial\gamma_j}\log\tilde{f}^m_{W_1,W_2}(W_1,W_2;\gamma_m)\right\}\Big|_{\gamma_m=\gamma_m^*}\right)^M_{i,j=1},$$

with $M=(m_1+1)(m_2+1)+6$ *and* $\gamma_m^*=(a^*,b^*,\Sigma^*,\theta_m^*)$.

4. Simulation study

In this section, in order to examine the numerical performance of the proposed method, we conduct simulation studies for various settings. Under model (1.2)-(1.3)-(2.1), we need to specify the bivariate distribution of (X_1,X_2) and the variance matrix of (U_1,U_2). For the latter, we will specify the values of σ_1,σ_2 and $\rho=Corr(U_1,U_2)=\sigma_{12}/(\sigma_1\sigma_2)$. The values of σ_1 and σ_2 will depend on the setting and can be found in Table 2, whereas for ρ we will work with $0.7, 0.1$ and -0.5 in each setting. For the bivariate distribution of (X_1,X_2) we consider four settings. In the first setting, S_1 and S_2 are simulated independently from a $Beta(1,1)$ distribution, and X_1 and X_2 equal $X_1=a_1S_1+b_1$ and $X_2=a_2S_2+b_2$ with $a_1=2$, $a_2=3$, $b_1=-1$ and $b_2=-2$. In the second setting, X_1 and X_2 are independent and distributed according to a $N(0,1;-1,1)$ and $N(0,1;0,2)$ distribution respectively, where the notation $N(0,1;t_L,t_U)$ stands for a standard normal distribution truncated to the interval $[t_L,t_U]$. The final two settings deal with the case where X_1 and X_2 are correlated. In the third setting, the vector (X_1,X_2) follows a zero-mean bivariate normal distribution with unit variance and correlation given by $\delta=0.1$, which we truncate to the rectangle determined by $t_L=(-1,-2)^\top$ and $t_U=(0,1)^\top$, whereas in the fourth setting $\delta=0.7$, $t_L=(-1,-2)^\top$ and $t_U=(2,0)^\top$.

Note that not only the measurement errors but also the unobserved covariates are correlated in the last two settings. To compute the bivariate integral in (3.4), we use Monte Carlo (MC) integration which allows to integrate any bivariate function defined on the unit square. Under this method, the integral is computed by evaluating the function of interest in a sufficiently large number of pseudo

uniform variables [25]. If the bounds of the integral are different from $[0,1]\times[0,1]$, they should be transformed. This is achieved as follows:

$$\int_{y_L}^{y_U}\int_{x_L}^{x_U} f(x,y)dxdy \cong (y_U-y_L)(x_U-x_L)\frac{1}{D} \times \sum_{i=1}^{D} f(R_{i1}(x_U-x_L)+x_L, R_{i2}(y_U-y_L)+y_L)$$

where $R_{11},\ldots,R_{D1}$ and $R_{12},\ldots,R_{D2}$ are two random samples of size D, obtained independently from a $U[0,1]$ distribution.

MC integral estimators are known to be unbiased and the accuracy increases with increasing values of D. For our case, we generated MC samples of size $D=10,000$. Therefore, the log-likelihood function in (3.4) can be approximated by

$$\mathcal{L}_n(a,b,\Sigma,\theta_m) \cong \sum_{i=1}^{n}\log\left\{\sum_{k_1=0}^{m_1}\sum_{k_2=0}^{m_2}\theta^m_{k_1,k_2}\frac{1}{D}\sum_{j=1}^{D}\beta^m_{k_1;k_2}(R_{j1},R_{j2})\gamma(W_{i1},R_{j1},W_{i2},R_{j2})\right\}, \tag{4.1}$$

where

$$\gamma(W_{i1},R_{j1},W_{i2},R_{j2}) = f_{U_1,U_2}(W_{i1}-(a_1R_{j1}+b_1),W_{i2}-(a_2R_{j2}+b_2);\Sigma),$$
$$\beta^m_{k_1,;k_2}(R_{j1},R_{j2}) = Beta_{k_1+1,m_1-k_1+1}(R_{j1})Beta_{k_2+1,m_2-k_2+1}(R_{j2}).$$

The use of MC integration significantly reduces the computation time to calculate the likelihood function. Next, we use constrained optimization algorithms for the numerical maximization of the log-likelihood in (4.1).

For each setting, $N=200$ samples of size $n=\{300,500,4000\}$ are drawn. To determine the necessary number of Bernstein polynomials, seventeen different degrees (m_1,m_2) are fitted, namely (m_1,m_2) for $m_1,m_2=0,1,2$, and $(m_1,0)$ and $(0,m_2)$ for $m_1,m_2=3,4,5,6$. Then, these degrees are evaluated by BIC. Following degree selection and parameter estimation, we evaluate the performance of the proposed method for approximating the joint density of (X_1,X_2) via the mean integrated absolute error (MIAE):

$$MIAE = \frac{1}{N}\sum_{r=1}^{N}\int\int\left|f_{X_1,X_2}(x_1,x_2)-\tilde{f}^{m,r}_{X_1,X_2}(x_1,x_2;\hat{a},\hat{b},\hat{\theta}_m)\right|dx_1dx_2, \tag{4.2}$$

where $\tilde{f}^{m,r}_{X_1,X_2}(x_1,x_2;\hat{a},\hat{b},\hat{\theta}_m)$ represents the estimated density using the data on replication r, while $f_{X_1,X_2}(x_1,x_2)$ is the unknown density of (X_1,X_2). The results of the simulations are presented in Tables 1-3. Table 1 and the first two columns of Table 2 summarize the simulation outcomes for the marginal error distributions. Tables 1 and 2 show that the estimation of the unknown parameters is in general accurate. Compared with the results for $n=\{300,4000\}$ (not

TABLE 1

Simulation results for the estimation of $f_{X_1,X_2}(x_1,x_2)$ when $n = 500$. For each distribution, the three lines correspond to three values for ρ, namely $\rho = 0.7, 0.1$ and -0.5 (B: Bias, SE: Standard error).

Covariate Distribution	Estimation of a_1			Estimation of b_1			Estimation of a_2			Estimation of b_2			
	a_1	$B(\hat{a}_1)$	SE($\hat{a}_1$)	b_1	$B(\hat{b}_1)$	SE($\hat{b}_1$)	a_2	$B(\hat{a}_2)$	SE($\hat{a}_2$)	b_2	$B(\hat{b}_2)$	SE($\hat{b}_2$)	MIAE
$2\text{Beta}(1,1)-1$		0.002	0.067		-0.001	0.040		0.009	0.119		-0.006	0.069	1.068
$3\text{Beta}(1,1)-2$	2	0.006	0.074	-1	-0.003	0.042	3	0.003	0.112	-2	-0.004	0.072	1.095
		0.006	0.068		-0.003	0.041		0.007	0.143		-0.006	0.086	1.087
$N(0,1;-1,1)$		-0.160	0.088		0.081	0.050		0.019	0.140		0.000	0.059	0.822
$N(0,1;0,2)$	2	-0.168	0.081	-1	0.084	0.046	2	0.011	0.149	0	0.004	0.066	0.830
		-0.166	0.079		0.084	0.046		0.025	0.138		-0.001	0.067	0.826
		-0.014	0.041		0.018	0.030		-0.390	0.368		0.214	0.283	1.017
$N_2, \delta = 0.1$	1	-0.015	0.039	-1	0.020	0.030	3	-0.420	0.392	-2	0.228	0.291	1.034
		-0.015	0.036		0.019	0.029		-0.404	0.305		0.217	0.258	1.039
		-1.575	0.808		0.295	0.277		-0.193	0.179		0.158	0.163	0.970
$N_2, \delta = 0.7$	3	-1.462	0.814	-1	0.275	0.261	2	-0.194	0.229	-2	0.161	0.190	0.889
		-1.484	0.876		0.285	0.294		-0.159	0.144		0.141	0.139	0.852

TABLE 2

Simulation results for the estimation of the covariance matrix of the measurement errors when $n = 500$.

Covariate Distribution	Estimation of σ_1				Estimation of σ_2				Estimation of ρ			
	σ_1	Bias	SE	MSE	σ_2	Bias	SE	MSE	ρ	Bias	SE	MSE
$2\text{Beta}(1,1)-1$		-0.001	0.024	0.001		-0.006	0.040	0.002	0.7	0.009	0.264	0.070
$3\text{Beta}(1,1)-2$	0.14	-0.002	0.026	0.001	0.28	-0.008	0.043	0.002	0.1	0.000	0.410	0.168
		-0.002	0.024	0.001		-0.008	0.045	0.002	-0.5	-0.025	0.346	0.120
$N(0,1;-1,1)$		0.032	0.031	0.002		0.001	0.029	0.001	0.7	-0.095	0.280	0.088
$N(0,1;0,2)$	0.18	0.033	0.032	0.002	0.16	0.002	0.032	0.001	0.1	-0.006	0.324	0.105
		0.032	0.032	0.002		0.001	0.031	0.001	-0.5	0.046	0.288	0.085
		0.002	0.015	$2e^{-4}$		0.092	0.077	0.014	0.7	-0.140	0.330	0.129
$N_2, \delta = 0.1$	0.07	0.003	0.015	$2e^{-4}$	0.21	0.097	0.086	0.017	0.1	0.036	0.409	0.169
		0.003	0.013	$2e^{-4}$		0.097	0.076	0.015	-0.5	0.193	0.382	0.183
		0.223	0.134	0.068		0.041	0.038	0.003	0.7	0.058	0.161	0.029
$N_2, \delta = 0.7$	0.166	0.204	0.139	0.061	0.125	0.036	0.044	0.003	0.1	0.498	0.175	0.279
		0.196	0.151	0.061		0.023	0.030	0.001	-0.5	0.912	0.202	0.872

shown here, but provided in the online supplement), we see that the performance of the proposed method improves as the sample size increases.

The tables also show that our method has some difficulties in differentiating the contributions of the measurement error and the covariates when there is a curvature in the true covariate density. If the generated covariates have densities with a relatively flat shape, the method performs more satisfactorily. The relatively poor performance in the bivariate normal settings could be explained in this way. When the length of the truncation interval is shorter, the problem vanishes.

The last column of Table 2 shows the performance of the estimator of the error correlation. Overall, regardless of the sign and the magnitude of the error correlation, a good estimation performance is observed in most settings. Note however that the accuracy is less good than for the estimation of the marginal parameters $a_1, a_2, b_1, b_2, \sigma_1$ and σ_2, which can be explained by the fact that the correlation between two completely unobserved variables is a hard quantity to identify and estimate. The table shows nevertheless that the method works.

Table 3 shows the distribution of the selected pairs (m_1, m_2). In the first setting with two Beta densities the mostly chosen pair is $(0, 0)$, which is as

TABLE 3

Simulation results for the selection of (m_1, m_2) *when* $n = 500$*. For each distribution, the three lines correspond to three values for* ρ*, namely* $\rho = 0.7, 0.1$ *and* -0.5*.*

Covariate	Distribution of the selected (m_1, m_2) (in proportion)																
Distribution	(0,0)	(0,1)	(1,0)	(1,1)	(0,2)	(2,0)	(1,2)	(2,1)	(2,2)	(0,3)	(3,0)	(0,4)	(4,0)	(0,5)	(5,0)	(0,6)	(6,0)
$2\text{Beta}(1,1)-1$	0.82	0.06	0.065	0.005	0.02	0.03	0	0	0	0	0	0	0	0	0	0	0
$3\text{Beta}(1,1)-2$	0.935	0.02	0.02	0	0.01	0.01	0	0	0	0	0.005	0	0	0	0	0	0
	0.91	0.02	0.015	0	0.025	0.02	0	0	0	0	0.005	0.005	0	0	0	0	0
$N(0,1;-1,0)$	0	0.67	0	0.03	0.185	0	0.005	0	0	0.065	0	0.025	0.005	0.01	0	0.005	0
$N(0,1;-2,1)$	0.005	0.72	0	0.04	0.15	0	0	0	0	0.04	0	0.04	0	0.005	0	0	0
	0	0.68	0	0.055	0.14	0	0.005	0.005	0	0.09	0	0.02	0	0.005	0	0	0
$N_2, \delta = 0.1$	0.08	0.54	0.105	0.05	0.115	0.02	0.005	0.005	0	0.06	0	0.015	0	0.005	0	0	0
	0.09	0.625	0.095	0.025	0.08	0.015	0.005	0.005	0	0.03	0.01	0.015	0	0.005	0	0	0
	0.095	0.595	0.125	0.04	0.09	0.015	0	0	0	0.03	0	0.01	0	0	0	0	0
$N_2, \delta = 0.7$	0	0.255	0.005	0.245	0.18	0.015	0.155	0.1	0	0.035	0	0.005	0	0.005	0	0	0
	0	0.235	0.01	0.275	0.185	0.01	0.095	0.165	0.015	0.005	0	0.005	0	0	0	0	0
	0	0.24	0.005	0.315	0.14	0.005	0.11	0.145	0.015	0.025	0	0	0	0	0	0	0

expected. In the other settings, the distribution of the selected degrees is more diverse. The final column of Table 1 shows the performance of the estimated density of the unobserved covariates (X_1, X_2). Note that the last setting leads to the highest values for the MIAE, defined in (4.2) above. This can be due to the poor performance of the estimator of the error correlation in this case.

Finally, we ran a simulation experiment for the scenario where no measurement error is present. The results can be found in the online supplement, and show that our method also performs well when the data are not contaminated.

Although the proposed method offers overall a promising estimation performance, it might sometimes suffer from practical identification problems. Making inference about the density of the unobserved covariates and differentiating the contributions of the measurement error and the covariates in the absence of replication data is a complicated problem. [13] state that technical and theoretical identifiability may not necessarily imply practical identifiability when data are observed one-at-a-time and no validation data are available. If the covariates are highly correlated as well as the measurement errors, the problem becomes even more complex. It is a common issue in measurement error problems without additional data.

5. Logistic regression based on SIMEX and the estimated variance matrix

The methodology developed in Section 3 can be used not only for estimating the error variance-covariance matrix and the bivariate density of (X_1, X_2), but also for estimating a regression model with error-prone covariates, of which the errors are correlated. We will demonstrate this for the special case of a logistic regression model in which we use the SIMEX method to correct for the measurement errors, but the estimated error variance matrix can also be used in any other regression model and with any other method that corrects for measurement errors.

In a logistic regression model, the conditional mean of a binary response Y is given by

$$E(Y|X) = \frac{\exp(X^\top \beta)}{1 + \exp(X^\top \beta)} \tag{5.1}$$

where X is a vector of covariates and β is a vector of corresponding regression coefficients. If the covariates are exposed to measurement errors, the naïve methods cannot adequately estimate the conditional mean in (5.1), and thus are expected to lead to biased estimates [47]. To address this issue, a large variety of measurement error correction methods for logistic regression have been proposed in the literature. For instance, [47] proposed three correction methods that all reduce the bias of the naïve estimator, one of them being a corrected score approach. [48] generalized the latter approach to more general models. [37] introduced a method for constructing confidence intervals for coefficients in logistic regression when one of the covariates is error-prone, while [36] extended this method to the case where multiple covariates are subject to measurement error. In addition, [35] presented a nonparametric methodology to take the covariate measurement error into account while constructing a logistic regression model. We also refer to [15] for the use of multiple imputation in the same context. These methods rely on distributional assumptions on the covariate(s) or the presence of validation, replication or cohort data to specify the error distribution. In contrast to the aforementioned methods, the simulation-extrapolation algorithm (SIMEX) is applicable to a large class of regression models (logistic, linear, survival, etc.) as a remedy to correct for the consequences of measurement errors. For these reasons, we prefer to use SIMEX to get error-corrected coefficient estimates of a logistic model.

SIMEX, introduced by [16], attenuates the bias due to measurement errors when an additive measurement error model is present. The principle of this method is to mimic the influence of measurement errors on the coefficient estimates with Monte Carlo simulations. SIMEX is based on two steps. In the simulation step, the data are contaminated with noise and parameter estimates are obtained. This step is repeated for varying noise levels. Then, in the extrapolation step, a (parametric) curve is fit through the parameter estimates corresponding to these different noise levels and the curve is extrapolated to the case without measurement error.

It is still common practice to use the naïve method that neglects the measurement errors in the covariates. In order to be fully aware of the merits introduced by SIMEX, we compare the performance of both methods. On the other hand, applying SIMEX without taking the measurement error correlation into account is also a common practice in such situations. Therefore, we will also compare our method with this simplified method that ignores the correlation. Finally, we will compare these three estimators with the SIMEX estimators based on the true Σ and based on the true variances but by neglecting the correlation. To obtain the SIMEX estimates, we use a quadratic extrapolation function.

In order to carry out this comparison, we will study the five estimation methods in four different settings. In each setting there will be five covariates in the logistic model, so $X = (X_1, \ldots, X_5)^\top$. The covariates X_1 and X_2 are generated in exactly the same way as in the four settings studied in Section 4. We will even work with the same sample size ($n = \{300, 500, 4000\}$) and the same samples, so that the error covariance matrix does not need to be re-estimated, since it has already been computed in Section 4. On the other hand, the vari-

TABLE 4
Simulation results for the logistic regression model when $n = 500$ and $\rho = 0.7$.

Covariate Distribution	Estimation Method	$\beta_0 = 0$ Bias	$\beta_0 = 0$ MSE	$\beta_1 = 4.2$ Bias	$\beta_1 = 4.2$ MSE	$\beta_2 = -2.1$ Bias	$\beta_2 = -2.1$ MSE	$\beta_3 = 3.9$ Bias	$\beta_3 = 3.9$ MSE	$\beta_4 = 0.42$ Bias	$\beta_4 = 0.42$ MSE	$\beta_5 = 0.21$ Bias	$\beta_5 = 0.21$ MSE
$2\text{Beta}(1,1) - 1$ $3\text{Beta}(1,1) - 2$	Naïve	-0.718	0.614	-0.462	0.371	0.234	0.103	-1.892	3.757	-0.062	0.087	-0.020	0.024
	SIMEX ($\hat{\Sigma}$)	-0.519	0.399	-0.213	0.258	0.091	0.074	-1.268	1.937	-0.041	0.096	-0.010	0.026
	SIMEX ($\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$)	-0.464	0.407	0.245	0.392	-0.202	0.146	-0.723	1.124	-0.008	0.120	0.007	0.032
	SIMEX (Σ)	-0.514	0.390	-0.214	0.252	0.088	0.073	-1.255	1.901	-0.042	0.096	-0.010	0.026
	SIMEX (σ_1, σ_2, σ_3)	-0.471	0.388	0.240	0.374	-0.204	0.140	-0.737	1.053	-0.009	0.117	0.008	0.032
$N(0,1;-1,0)$ $N(0,1;-2,1)$	Naïve	-0.701	0.679	-0.573	0.504	0.063	0.165	-1.908	3.874	-0.022	0.127	-0.024	0.021
	SIMEX ($\hat{\Sigma}$)	-0.551	0.573	0.002	0.330	-0.057	0.249	-1.267	2.074	0.009	0.149	-0.009	0.024
	SIMEX ($\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$)	-0.244	0.446	0.575	0.953	-0.537	0.669	-0.718	1.432	0.042	0.182	0.010	0.030
	SIMEX (Σ)	-0.537	0.544	-0.158	0.295	-0.001	0.215	-1.276	2.057	0.001	0.143	-0.012	0.023
	SIMEX (σ_1, σ_2, σ_3)	-0.228	0.385	0.309	0.525	-0.460	0.540	-0.789	1.304	0.027	0.171	0.003	0.028
$N_2, \delta = 0.1$	Naïve	-0.813	0.834	-0.282	0.518	0.220	0.127	-1.945	3.973	-0.001	0.124	-0.010	0.029
	SIMEX ($\hat{\Sigma}$)	-0.626	0.620	-0.036	0.598	-0.044	0.128	-1.337	2.145	0.017	0.138	0.001	0.032
	SIMEX ($\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$)	-0.388	0.440	0.454	1.034	-0.268	0.262	-0.853	1.391	0.039	0.157	0.012	0.037
	SIMEX (Σ)	-0.588	0.560	-0.139	0.560	0.096	0.108	-1.345	2.144	0.011	0.134	-0.003	0.031
	SIMEX (σ_1, σ_2, σ_3)	-0.355	0.373	0.286	0.781	-0.055	0.130	-0.922	1.347	0.027	0.148	0.006	0.035
$N_2, \delta = 0.7$	Naïve	-0.833	0.800	-0.639	0.517	0.215	0.150	-1.952	3.935	-0.010	0.073	-0.035	0.020
	SIMEX ($\hat{\Sigma}$)	-0.628	0.544	0.349	0.438	-0.053	0.186	-1.282	1.879	0.037	0.097	-0.014	0.024
	SIMEX ($\hat{\sigma}_1$, $\hat{\sigma}_2$, $\hat{\sigma}_3$)	-0.917	1.094	1.856	4.936	-1.212	1.955	-0.544	0.823	0.115	0.159	0.020	0.035
	SIMEX (Σ)	-0.582	0.470	-0.274	0.226	0.113	0.143	-1.329	1.990	0.012	0.084	-0.025	0.021
	SIMEX (σ_1, σ_2, σ_3)	-0.607	0.533	0.103	0.224	-0.298	0.269	-0.868	1.093	0.034	0.099	-0.014	0.023

ables X_3, X_4 and X_5 are new, and are generated in the same way in each of the four settings. They are generated independently of each other and of (X_1, X_2). The variable X_3 is contaminated, so we observe $W_3 = X_3 + U_3$ with $U_3 \sim N(0, \sigma_3^2)$. Note that U_1, U_2 and U_3 are independent of all other variables. The variable X_3 is drawn from a Beta(1,1)-1 distribution and the variance of U_3 equals $\sigma_3^2 = 0.07$. The variables X_4 and X_5 are not error-prone and are generated from a Bernoulli(0.7) and $N(0,1)$ distribution, respectively. For an i.i.d. sample of size n, the i-th data point is composed of $(Y_i, W_{i1}, W_{i2}, W_{i3}, X_{i4}, X_{i5})$, $i = 1, \ldots, n$. Finally, the vector of regression coefficients is given by $\beta = (\beta_0, \ldots, \beta_5)^\top = (0, 4.2, -2.1, 3.9, 0.42, 0.21)^\top$.

The comparative simulation results for these five methods when $\rho = 0.7$ are presented in Table 4. For almost all cases, the naïve method leads to highly biased estimates. When the measurement errors in the covariates are taken into account, less biased estimates are obtained. Overall, SIMEX based on $\hat{\Sigma}$ outperforms the naïve estimator and the SIMEX estimator based on $(\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3)$ in terms of bias reduction. On the other hand, the SIMEX estimates have a higher variance than the naïve ones. This phenomenon is referred to as 'variance versus bias tradeoff' in the measurement error literature. The complexity of correction procedures often leads to higher variability than when using the simple naïve procedure that ignores the measurement error. See e.g. Stefanski and Carroll (1985), who observed that the naïve method is not always worse than their proposed correction methods when the sample size is small, i.e. when variance dominates bias. Since small bias is often considered as being more important than small variance, we believe that the proposed method should be preferred in practice, despite its larger variance. Finally, we note that when the true Σ and/or $(\sigma_1, \sigma_2, \sigma_3)$ are used in the SIMEX procedure, the results are better than when their corresponding estimators are used, but the differences are not very significant, which suggests that the estimation of the matrix Σ does not have an important impact on the performance of the SIMEX estimator.

It is also interesting to note that the SIMEX method that ignores the correlation between the measurement errors does not perform very well. This shows

that it is essential to take the correlation between measurement errors into account in order to do correct inference. The simulation results for ρ equal to 0.1 and -0.5 are available in the online supplement. The same conclusions hold. However, the outperformance of SIMEX($\hat{\Sigma}$) with respect to SIMEX($\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$) is more significant when the magnitude of the error correlation is larger (i.e. when $\rho = 0.7$).

6. Application

In this section, we apply the proposed method on data from the Framingham Heart Study. Our main interest is to model the 10-year risk of coronary heart disease (CHD) using various covariates on the demographical and medical characteristics of the patients. There are 3,827 individuals with complete cases and 592 of them have the disease.

The covariates in this data set are gender (male (1) 44.24 %), age (min: 32, max: 70, median: 49, standard deviation: 8.6), stroke (yes (1) 0.63 %), hypertansive (yes (1) 0.31 %), cigarettes per day (CPD) (min: 0, max: 70, median: 0, standard deviation: 8.6), diastolic blood pressure (DBP) (min: 48, max: 142.5, median: 82, standard deviation: 11.97), systolic blood pressure (SBP) (min: 83.5, max: 295, median: 128, standard deviation: 22.1) and glucose (min: 40, max: 394, median: 78, standard deviation: 24). Note that in this application, SBP, DBP and glucose are the error-prone variables. Although self-reported variables are known to be potentially mismeasured in epidemiological studies, CPD is not considered as a contaminated covariate in this application. [7] conducted an analysis for a comparison of self-reported CPD counts with the returned cigarette butts and blood sample measurements regarding the nicotine level and concluded that self-reported counts are reliable sources of smoking behavior information. Therefore, we assume that CPD is an error-free covariate.

Neglecting the presence of measurement errors and applying a naïve logistic regression analysis is expected to lead to incorrect inferences. Since both blood pressures, SBP and DBP, are measured by the same device, it is likely for these covariates to have correlated measurement errors. However, glucose level in the blood is measured independently by using a separate device. For this reason, we assume that the measurement error of glucose level is independent of the measurement errors in the two blood pressures. Hence, we take these issues into account and assess the covariate effects after remedying the measurement errors. We use the following transformed variables:

$$\begin{aligned} SBP^* &= \log(SBP - 50) \\ DBP^* &= \log(DBP - 30) \\ Glucose^* &= \log(Glucose - 35) \end{aligned}$$

This transformation has e.g. been used for SBP by [13] (page 118) in order to homogenize the error variance of SBP. We followed a similar approach for DBP and Glucose, and replaced the value 50 which was used for SBP by respectively

TABLE 5
Estimates for each error-prone variable separately.

Covariate	$\hat{a}$	$\hat{b}$	$\hat{\sigma}$	$\hat{m}$
Glucose*	0.378	3.588	0.322	0
SBP*	0.528	4.141	0.1897	-
DBP*	0.328	3.83	0.225	-

TABLE 6
Estimates for the correlated measurement errors in SBP and DBP*.*

Covariate	$\hat{\rho}$	$\hat{\sigma}_{12}$	$(\hat{m}_1, \hat{m}_2)$	$\hat{\theta}_m$
SBP* & DBP*	0.038	0.887	(0,0)	1

TABLE 7
Coefficient estimates and standard errors without and with correction for the measurement errors.

	Naïve Method	SIMEX
Intercept	-12.30 (1.285)	-14.43 (1.841)
Gender	0.513 (0.103)	0.541 (0.107)
Age	0.065 (0.006)	0.061 (0.006)
Stroke	1.010 (0.443)	1.009 (0.415)
Hypertensive	0.267 (0.135)	0.109 (0.166)
CPD	0.020 (0.004)	0.021 (0.004)
Glucose*	0.441 (0.127)	0.586 (0.188)
SBP*	1.576 (0.345)	2.247 (0.515)
DBP*	-0.503 (0.331)	-0.789 (0.505)

30 and 35, taking the different support of these variables into account. Based on these transformations, the parameter estimates related to the measurement errors in Glucose*, SBP* and DBP* are obtained by the method explained in Section 4. The selection of the degree of the Bernstein polynomials is as before based on BIC.

Tables 5 and 6 show the results of the estimation. Table 5 indicates that the density of Glucose* can be estimated using Bernstein polynomials of degree 0. On the other hand, Table 6 suggests that the bivariate density of (SBP*,DBP*) can be estimated by bivariate Bernstein polynomials of degree $(0,0)^\top$. The estimated error variance-covariance matrix of $(Glucose^*, SBP^*, DBP^*)$ is found to be:

$$\hat{\Sigma} = \begin{pmatrix} 0.104 & 0 & 0 \\ 0 & 0.036 & 0.038 \\ 0 & 0.038 & 0.050 \end{pmatrix}$$

The correlation between the measurement errors in SBP* and DBP* is hence estimated to be around 0.887, which points to a strong positive linear relationship.

Let us now use this estimated variance matrix to estimate the coefficients of the logistic regression model by means of the SIMEX method. SIMEX plots for each covariate are provided in Figure 1. A comparison between the coefficient estimates obtained from the naïve method and from the SIMEX method is given in Table 7. Standard errors for the SIMEX estimates are also provided.

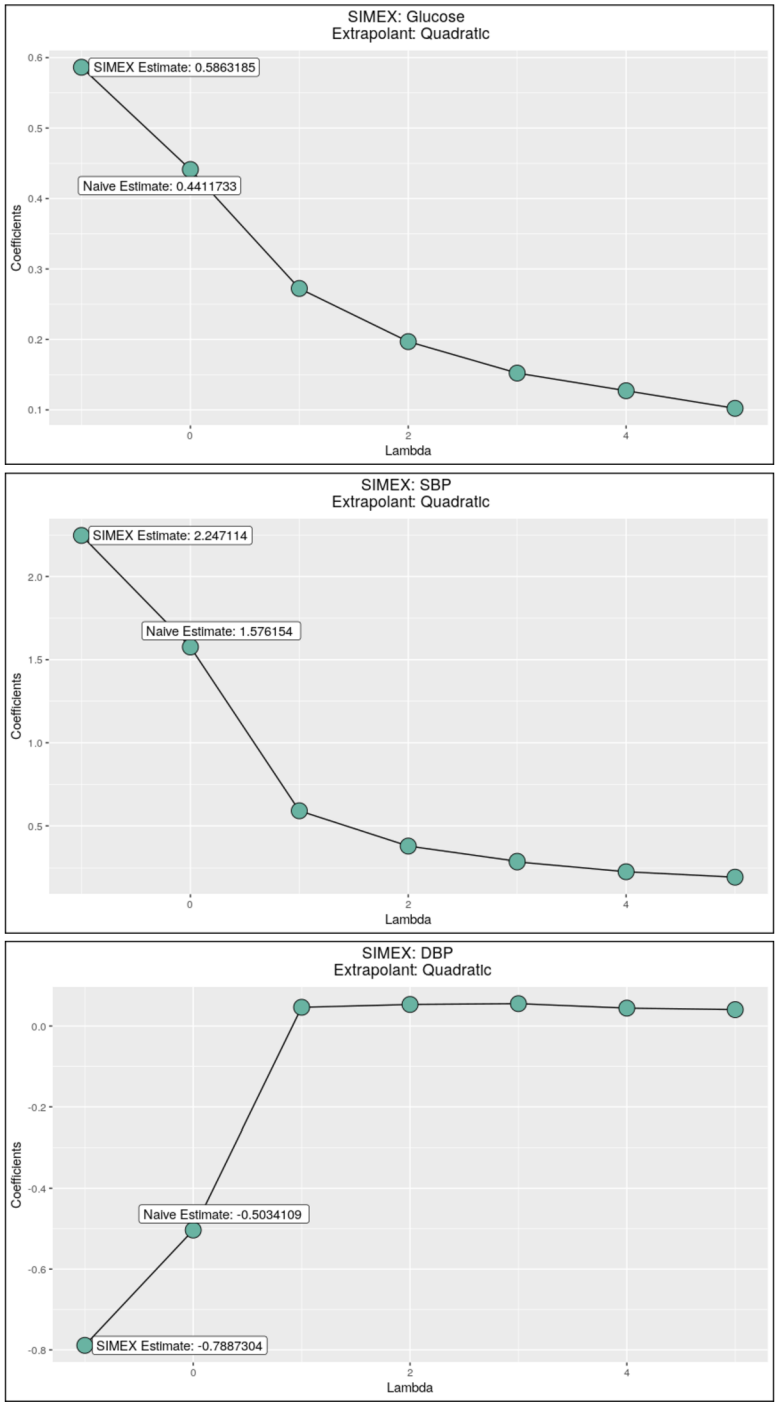

Fig 1: SIMEX plots for the covariates Glucose*, SBP* and DBP*, respectively.

They are obtained from the sandwich estimators of the variance given in Carroll et al. (2006) (pp. 367-374), see also Carroll et al. (1996) (page 245). After adjusting for the measurement errors, the estimates for Glucose*, SBP* and DBP* have considerably changed, which suggests that correcting for correlated measurement errors is essential.

7. Discussion and future research

When dealing with measurement error models, it is common to assume that the measurement errors in multiple covariates are independent of each other. When this assumption is violated, the correlation between the errors must be taken into account in order to make correct inferences. On the other hand, in the literature on correlated measurement errors many existing methods suppose that external information sources such as validation data sets, replications or auxiliary variables are available to estimate the covariance matrix of the measurement errors. However, these sources may not always be accessible in practice. This limits the usability of such methods.

In this paper, we proposed a method to take the correlated measurement error structure into account. We worked on a bivariate classical measurement error model with Gaussian errors to estimate the error variance matrix. Our method is flexible, since it does not rely on external information sources and does not make distributional assumptions on the unobserved covariates, apart from some minor assumptions on the support of these variables. Instead, these covariate distributions are approximated by using bivariate Bernstein polynomials. Thus, our method can be applied to a wide range of contexts. Once the covariance matrix is estimated, any correction method can be used to estimate the regression model correctly.

Simulation studies indicate the good finite sample performance of the proposed method in various settings. Moreover, our method performs well in a logistic regression model in which the SIMEX method is used to correct for the measurement errors. Analysis of the data from the Framingham Heart Study allowed us to assess the effects of both the contaminated and the error-free covariates on the risk of coronary heart disease.

In our simulations, polynomials up to degree 6 are sufficient, and degree 6 is actually rarely selected. However, if for a given data set higher degrees would be required, the computation time can become problematic. In order to overcome this issue, a two-step estimation scheme can be used in that case. In the first step, one could estimate the parameters (a_j, b_j, σ_j), $j = 1, 2$, related to the marginal error distributions by using the method of [5]. Note that the error U_j $(j = 1, 2)$ is independent of X_j and has a univariate normal distribution with mean zero and variance σ_j^2. In addition, X_j has compact support $[b_j, a_j + b_j]$. Hence, we are exactly in the setting needed for applying the method of [5]. Then, in the second step, the estimates $(\hat{a}_j, \hat{b}_j, \hat{\sigma}_j)$, $j = 1, 2$ are plugged into the log-likelihood function (4.1), and the error covariance σ_{12} and the parameters θ_m related to the bivariate distribution f_{X_1,X_2} can be estimated. Although esti-

mation of all parameters in one step is accurate, the two-step estimation scheme could be a practical solution when the computation time for higher degrees becomes excessive. In order to reduce the computation time in optimization, such methods are not uncommon. See for example, [20] and [30].

The method in this paper can be considered as an important step in the literature on correlated measurement errors. A number of extensions can be studied in the future. First of all, we assumed that the measurement errors of two covariates are correlated. A natural extension would be to consider the case where instead of having two covariates X_1 and X_2 subject to measurement error, we have $p \geq 2$ covariates of which the errors are correlated. When the corresponding errors follow a multivariate normal distribution, the $p \times p$ variance-covariance matrix of these errors needs to be identified and estimated. In the bivariate case the critical parameter in the identification of the model, is the correlation ρ between U_1 and U_2. In the multivariate case, there will be several (partial) correlations. Although we expect that the identification and estimation of this matrix will be feasible, it will be technically more challenging.

A second possible extension is regarding the identifiability of model (1.2)-(1.3)-(2.1) when the assumed bivariate normal distribution of the errors is replaced by another parametric family of distributions. [39] mention in their Remark 2.3 a few examples of univariate parametric families under which the univariate measurement error model is identified. However, they do not give necessary and sufficient conditions, i.e. they do not give a full characterization of the class of parametric families that make the model identifiable. It would be useful to develop such an identification result, first in the univariate case, and if successful, also in the more challenging bivariate or multivariate case.

Software

Software in the form of R code and complete documentation are available on request from the corresponding author.

Acknowledgments

The authors like to thank Aurélie Bertrand (ISBA, UCLouvain, Belgium) and François Portier (S^2A,Télécom Paris Tech) for very helpful discussions.

Supplementary Material

Supplementary material for estimation of the variance matrix in bivariate classical measurement error models
(doi: 10.1214/22-EJS1996SUPP; .pdf). The supplement contains the additional simulations regarding $n = \{300, 4000\}$.

References

[1] Aitkin, M., and Rocci, R. (2002). A general maximum likelihood analysis of measurement error in generalized linear models. *Statistics and Computing*, **12**(2), 163–174. MR1897515

[2] Bărbosu, D. (2000). Some generalized bivariate Bernstein operators. *Mathematical Notes, Miskolc.*, **1**, 1, 3–10. MR1793257

[3] Berkson, J. (1950). Are there two regressions? *Journal of the American Statistical Association*, **45**, 164–180.

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

[5] Bertrand, A., Van Keilegom, I., and Legrand, C. (2019). Flexible parametric approach to classical measurement error variance estimation without auxiliary data. *Biometrics*, **75**, 297–307. MR3953730

[6] Blackwell, M., Honaker, J., and King, G. (2017). A unified approach to measurement error and missing data: Overview and aplications. *Sociological Methods and Research*, **46**(3), 303–341. MR3671517

[7] Blank, M.D., Breland, A., Enlow, P.T., Duncan, C., Metzger, A., and Cobb, C.O. (2016). Measurement of smoking behavior: Comparison of self-reports to returned cigarette butts and toxicant levels. *Experimental and Clinical Psychopharmacology*, **24**(5), 348–355.

[8] Brunner, J., and Austin, P.C. (2009). Inflation of Type I error rate in multiple regression when independent variables are measured with error. *The Canadian Journal of Statistics*, **37**(1), 33–46. MR2509460

[9] Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods, and Applications*. Boca Raton: CRC Press. MR2682774

[10] Carroll, R.J. (2014). Estimating the distribution of dietary consumption patterns. *Statistical Science*, **29**(1), 2–8. MR3201840

[11] Carroll, R.J., Gallo, P., and Gleser, L. (1985). Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. *Journal of the American Statistical Association*, **80**, 929–932. MR0819596

[12] 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**(433), 242–250. MR1394078

[13] Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition*. Chapman & Hall, Boca Raton. MR2243417

[14] Carroll, R.J., and Stefanski, L. (1990). Approximate quasilikelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, **85**, 652–663. MR1138349

[15] Cole, S.R., Chu, H., and Greenland, D. (2006). Multiple-imputation for measurement-error correction. *International Journal of Epidemiology*, **35**, 1074–1081.

[16] Cook, J. R., and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, **89**, 1314–1328. MR1379467

[17] Datta, G.S., Torabi, M., Rao, J.N.K., and Liu, B. (2018). Small area estimation with multiple covariates measured with errors: A nested error linear regression approach of combining multiple surveys. *Journal of Multivariate Analysis*, **167**, 49–59. MR3830633

[18] Day, N.E., Wong, M.Y., Bingham, S., Khaw, K.T., Luben, R., Michels, K.B., Welch, A., and Wareham, N.J. (2004). Correlated measurement error: implications for nutritional epidemiology. *International Journal of Epidemiology*, **33**(6), 1373–1381.

[19] Fibrinogen Studies Collaboration (2009). Correcting for multivariate measurement error by regression calibration in meta-analyses of epidemiological studies. *Statistics in Medicine*, **28**(7), 1067–1092. MR2662198

[20] Ford, J.A., and Moghrabi, I.A. (1994). ulti-step quasi-Newton methods for optimization. *Journal of Computational and Applied Mathematics*, **50**, 305–323. MR1284271

[21] Fraser, G.E., and Stram, D.O. (2001). Regression calibration in studies with correlated variables measured with error. *American Journal of Epidemiology*, **154**(9), 836–844.

[22] Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley. MR0898653

[23] Guolo, A. (2008). Robust techniques for measurement error correction: a review. *Statistical Methods in Medical Research*, **17**(6), 555–580. MR2654662

[24] Gustafson, P. (2004). *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton: Chapman and Hall, CRC. MR2005104

[25] Hammersley, J. M., and Handscomb, D. C. (1964). *Monte Carlo Methods*. London: Methuen & Co. MR0223065

[26] Kekeç, E., and Van Keilegom, I. (2022) *Supplement to "Estimation of the variance matrix in bivariate classical measurement error models"* DOI: 10.1214/22-EJS1996SUPP.

[27] Kipnis, V., Freedman, L.S., Carroll, R.J., and Douglas, M. (2016). A bivariate measurement error model for semicontinuous and continuous variables: application to nutritional epidemiology. *Biometrics*, **72**(1), 106–115. MR3500579

[28] Lesaffre, E., and Lawson, A.B. (2012). *Bayesian Biostatistics*. John Wiley & Sons. MR3236827

[29] Michels, K.B., Bingham, S.A., Luben, R., Welch, A.A., Day, N.E. (2004). The effect of correlated measurement error in multivariate models of diet. *American Journal of Epidemiology*, **160**(1), 59–67.

[30] Moreno Alamo, A. C., and Costa Alberto, L. F. (2015). A multi-step optimization approach for power flow with transient stability constraints. 2015 IEEE Eindhoven PowerTech. 1-6. doi: 10.1109/PTC.2015.7232666.

[31] Muff, S., Ott, M., Braun, J., and Held, L. (2017). Bayesian

two-component measurement error modelling for survival analysis using INLA—A case study on cardiovascular disease mortality in Switzerland. *Computational Statistics & Data Analysis*, **113**, 177–193. MR3662399

[32] Nakamura, T. (1990). Corrected score functions for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, **77**, 127–137. MR1049414

[33] Patriota, A.G., and Bolfarine, H. (2010). Measurement error models with a general class of error distribution. *Statistics*, **44**(2), 119–127. MR2674412

[34] R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[35] Rabe-Hesketh, S., Pickles, A., and Skrondal, A. (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*, **3**, 215–232. MR2005474

[36] Rosner, B., Spiegelman, D., and Willett, W.C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology*, **132**(4), 734–745.

[37] Rosner, B., Willett, W.C., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, **8**(9), 1051–1069. MR0882774

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

[39] Schwarz, M., and Van Bellegem, S. (2010). Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, **80** (3-4), 236–241. MR2575451

[40] Song, X., Davidian, M., and Tsiatis, A.A. (2001). An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics*, **1**(1), 1–32. MR1844844

[41] Spiegelman, D. (2004). Commentary: Correlated errors and energy adjustment–where are the data? *International Journal of Epidemiology*, **33**(6), 1387–1388.

[42] Spiegelman, D., Carroll, R.J., and Kipnis, V. (2001). Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Statistics in Medicine*, **20**(1), 139–160.

[43] Spiegelman, D., McDermott, A., and Rosner, B. (1997a). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *American Journal of Clinical Nutrition*, **65** (suppl. 4), 1179S–1186S.

[44] Spiegelman, D., Schneeweiss, S., and McDermott, A. (1997b). Measurement error correction for logistic regression models with an 'alloyed gold standard'. *American Journal of Epidemiology*, **145**, 184–196.

[45] Spiegelman, D., Zhao, B., and Kim, J. (2005). Correlated errors in

biased surrogates: Study designs and methods for measurement error correction. *Statistics in Medicine*, **24** (11), 1657–1682. MR2137643

[46] Stancu, D.D. (1963). Evaluation of the remainder term in approximation formulas by Bernstein polynomials. *Mathematics of Computation*, **17** (83), 270–278. MR0179524

[47] Stefanski, L.A., and Carroll, R.J. (1985). Covariate measurement error in logistic regression. *Annals of Statistics*, **13**, 1335–1351. MR0811496

[48] Stefanski, L.A., and Carroll, R.J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, **74**, 703-7-16. MR0919838

[49] Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society, Series B*, **41** (2), 276–278.

[50] The Framingham Heart Study Dataset by Boston University and the National Heart, Lung, and Blood Institute. Retrieved on May 14, 2018 from www.kaggle.com.

[51] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25. MR0640163

[52] Wong, M.Y., Day, N.E., and Wareham, N.J. (1999). Measurement error in epidemiology: The design of validation studies II: Bivariate situation. *Statistics in Medicine*, **18**, 2831–2845.

[53] Yi, Y.G. (2017). *Statistical Analysis with Measurement Error or Misclassification*. Springer Nature, LLC. MR3676914

[54] Zhang, S., Midthune, D., Guenther, P.M., Krebs-Smith, S.M., Kipnis, V., Dodd, K.W., Buckman, D.W., Tooze, J.A., Freedman, L.S., and Carroll, R.J. (2011a). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Annals of Applied Statistics*, **5**(2B), 1456–1487. MR2849782

[55] Zhang, S., Krebs-Smith, S.M., Midthune, D., Perez, A., Buckman, D.W., Kipnis, V., Freedman, L.S., Dodd, K.W., and Carroll, R.J. (2011b). Fitting a bivariate measurement error model for episodically consumed dietary components. *The International Journal of Biostatistics*, **7**(1), 1–32. MR2753569