



Annamaria Guolo\* and Tania Erika Pesantez Cabrera

# A SIMEX approach for meta-analysis of diagnostic accuracy studies with attention to ROC curves

<https://doi.org/10.1515/ijb-2022-0012>

Received January 24, 2022; accepted October 10, 2022; published online October 27, 2022

**Abstract:** Bivariate random-effects models represent an established approach for meta-analysis of accuracy measures of a diagnostic test, which are typically given by sensitivity and specificity. A recent formulation of the classical model describes the test accuracy in terms of study-specific Receiver Operating Characteristics curves. In this way, the resulting summary curve can be thought of as an average of the study-specific Receiver Operating Characteristics curves. Within this framework, the paper shows that the standard likelihood approach for inference is prone to several issues. Small sample size can give rise to unreliable conclusions and convergence problems deeply affect the analysis. The proposed alternative is a simulation-extrapolation method, called SIMEX, developed within the measurement error literature. It suits the meta-analysis framework, as the accuracy measures provided by the studies are estimates rather than true values, and thus are prone to error. The methods are compared in a series of simulation studies, covering different scenarios of interest, including deviations from normality assumptions. SIMEX reveals a satisfactory strategy, providing more accurate inferential results if compared to the likelihood approach, while avoiding convergence failure. The approaches are applied to a meta-analysis of the accuracy of the ultrasound exam for diagnosing abdominal tuberculosis in HIV-positive subjects.

**Keywords:** diagnostic test; likelihood inference; multivariate meta-analysis; ROC curve; SIMEX.

Bivariate random-effects models represent a well-established approach for meta-analysis of diagnostic accuracy studies [1–4]. Accuracy of a diagnostic test is typically measured in terms of sensitivity, that is, the conditional probability of testing positive in diseased subjects, and specificity, that is, the conditional probability of testing negative in nondiseased subjects. The bivariate random-effects approach allows a joint modelization of sensitivity and specificity while accounting for measurement errors affecting the study-specific measures of accuracy of the test. In this way, the approach outperforms the criticized traditional method developed in Littenberg and Moses [5] that constructs a summary Receiver Operating Characteristic (ROC) curve based on the regression of the difference between sensitivity and specificity on their sum [2]. Inference under the bivariate random-effects model specification has the aim of providing an overall measure of sensitivity and/or specificity, diagnostic odds ratio and summary ROC curves. Typically, the aim is attained via a standard likelihood-based approach, although the literature warns against several issues that can affect inferential conclusions. Issues include small sample size, model misspecification and convergence problems. See, e.g., Guolo [6] and Takwoingi et al. [7] for illustrations. More sophisticated solutions based, for example, on pseudo-likelihood [8], copula models [9] or non-parametric approaches [10] have been recently proposed. Guolo [11] suggests the use of SIMEX, a simulation-extrapolation approach developed within the

---

\*Corresponding author: **Annamaria Guolo**, Department of Statistical Sciences, University of Padova, Padova, Italy,  
E-mail: [annamaria.guolo@unipd.it](mailto:annamaria.guolo@unipd.it)

**Tania Erika Pesantez Cabrera**, Department of Statistical Sciences, University of Padova, Padova, Italy,  
E-mail: [taniaerika.pesantezcabrera@studenti.unipd.it](mailto:taniaerika.pesantezcabrera@studenti.unipd.it)

measurement error literature, that is flexible enough to cover different measurement error structures, and that is characterized by an effortless implementation with standard software, see, e.g., Carroll et al. [12].

Despite the chosen approach for estimation, the formulation of the original bivariate random-effects approach [1, 2] has been criticized with respect to the interpretation of the resulting summary ROC curve. Hamza et al. [3] notice that the model does not necessarily allow to interpret the resulting summary ROC curve as a kind of average or overall ROC representative of the study-specific ROC curves. In order to obtain such an average ROC curve, Hamza et al. [3] propose a modification of the traditional bivariate random-effects model starting from the ROC study-specific curves and under extra assumptions on the test accuracy measures. The model is such that the estimated summary ROC curve can be considered a real overall summary ROC curve. Hamza et al. [3] suggest a likelihood-based approach for inference, but the performance of the proposed methodology is not deeply investigated. Within this framework, the aim of the paper is twofold. First of all, the paper investigates the performance of the likelihood-based strategy under a Normal approximation for the random-effects distribution in the model proposed in Hamza et al. [3]. Drawbacks of the likelihood-based approach in case of small sample size and computational problems ask for an alternative solution for inference on the accuracy measures of a test. To this aim, the paper develops a SIMEX-type methodology. The likelihood-based solution and the SIMEX approach are investigated and compared in terms of accuracy of the inferential procedures and from the computational point of view. Simulation studies for comparison are designed in order to reflect the real data-generating mechanism of the meta-analysis study level. Different scenarios are taken into account involving small to moderate sample size, increasing accuracy of the diagnostic test, deviations from the normality assumptions on the distribution of the unknown test accuracy measures. The last case deserves special attention, as model misspecifications are known to affect the likelihood-based procedures. SIMEX, conversely, is a functional method making no assumption on the distribution of the unknown mismeasured quantities. Thus, it is expected not to suffer from deviations from normality. The performance of the competing methods is further evaluated on a real meta-analysis about the accuracy of ultrasound exam for the diagnosis of abdominal tuberculosis in HIV-positive subjects.

## 1 Bivariate random-effects model

Consider a meta-analysis of  $K$  independent studies about the accuracy of a diagnostic test. Let  $n_{i11}$ ,  $n_{i10}$ ,  $n_{i01}$ ,  $n_{i00}$  be the number of true positives, false positives, false negatives and true negatives, respectively, from study  $i$ ,  $i = 1, \dots, K$ , obtained comparing the results from the diagnostic test to a reference test, assumed to be a gold standard [4]. Let  $n_{i1} = n_{i11} + n_{i01}$  be the number of diseased subjects and let  $n_{i0} = n_{i10} + n_{i00}$  be the number of free-diseased subjects, see Table 1.

Consider the sensitivity  $Se_i$  and the specificity  $Sp_i$  as diagnostic accuracy measures from study  $i$ ,  $i = 1, \dots, K$ . The estimates of sensitivity and specificity obtained from Table 1 are  $\widehat{Se}_i = n_{i11}/(n_{i11} + n_{i01})$  and  $\widehat{Sp}_i = n_{i00}/(n_{i00} + n_{i10})$ , respectively. Keeping with much of the literature, meta-analysis of diagnostic accuracy studies can be carried out on logit transformations of  $Se_i$  and  $1 - Sp_i$ , namely,

$$\eta_i = \log\left(\frac{Se_i}{1 - Se_i}\right) \quad \text{and} \quad \xi_i = \log\left(\frac{1 - Sp_i}{Sp_i}\right), \quad i = 1, \dots, K,$$

**Table 1:** Reference table for the  $i$ th study.

		Reference test		
		Positives	Negatives	
Diagnostic test	Positives	$n_{i11}$	$n_{i10}$	
	Negatives	$n_{i01}$	$n_{i00}$	
Total		$n_{i1}$	$n_{i0}$	$n_i$

whose estimates are given by

$$\hat{\eta}_i = \log\left(\frac{n_{i11}}{n_{i01}}\right) \quad \text{and} \quad \hat{\xi}_i = \log\left(\frac{n_{i10}}{n_{i00}}\right), \quad i = 1, \dots, K.$$

## 1.1 Classical formulation

The classical bivariate random-effects model developed in Reitsma et al. [1] and in Arends et al. [2] has a hierarchical structure, distinguishing a between-study level and a within-study level. The between-study model considers a Normal joint distribution for  $(\eta_i, \xi_i)^\top$ ,

$$\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \eta \\ \xi \end{pmatrix}, \Sigma_1\right), \quad \text{with} \quad \Sigma_1 = \begin{pmatrix} \sigma_\eta^2 & \sigma_{\xi\eta} \\ \sigma_{\xi\eta} & \sigma_\xi^2 \end{pmatrix}, \quad (1)$$

where  $\eta$  and  $\xi$  are the means over the studies,  $\sigma_\eta^2$  and  $\sigma_\xi^2$  are the between-study covariances and  $\sigma_{\xi\eta}$  is the covariance between  $\eta$  and  $\xi$ . The covariance is typically positive as sensitivity and 1-specificity tend to be positively correlated. The within-study model relates the observed pair  $(\hat{\eta}_i, \hat{\xi}_i)^\top$  to  $(\eta_i, \xi_i)^\top$ . For computational convenience, a Normal distribution is adopted, namely,

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \Big| \begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix}, S_i\right), \quad \text{where} \quad S_i = \begin{pmatrix} n_{i1}^{-1} \hat{S}e_i^{-1} (1 - \hat{S}e_i)^{-1} & 0 \\ 0 & n_{i0}^{-1} \hat{S}p_i^{-1} (1 - \hat{S}p_i)^{-1} \end{pmatrix}. \quad (2)$$

The variance/covariance matrix  $S_i$  is diagonal, as sensitivity and specificity are evaluated on different subjects, with non-zero entries are estimated from each study. Under the previous Normal distributions, it follows that, marginally,

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \eta \\ \xi \end{pmatrix}, \Sigma_1 + S_i\right). \quad (3)$$

Likelihood-based inference on the parameter vector  $(\eta, \xi, \sigma_\eta^2, \sigma_\xi^2, \sigma_{\xi\eta})^\top$  takes advantage of the closed-form of the likelihood function. The model resulting from (1) to (2) is strictly connected to the model suggested in Rutter and Gatsonis [13] within a Bayesian framework, although with a different parameterization [14]. Nevertheless, the implementation of (3) is more convenient [2].

Parameter estimation is typically performed via maximum likelihood or restricted maximum likelihood [2]. Despite the feasibility of the approach, the literature warns against unreliable inferential conclusions due to small sample size, convergence problems, deviations from normality assumptions [6, 7].

When the maximum likelihood estimate or restricted maximum likelihood estimate of the parameter vector  $(\eta, \xi, \sigma_\eta^2, \sigma_\xi^2, \sigma_{\xi\eta})^\top$  is available, the overall estimate of sensitivity  $Se$  and specificity  $Sp$  are obtained by back-transforming the estimates of  $\eta$  and  $\xi$ . The associated standard errors can be obtained using the delta method. Other measures of diagnostic test accuracy include the positive likelihood ratio  $LR+ = Se/(1 - Sp)$ , the negative likelihood ratio  $LR- = (1 - Se)/Sp$  and the diagnostic odds ratio  $dOR = Se/(1 - Se) \times Sp/(1 - Sp)$ . A useful description of the diagnostic test is provided through the summary ROC curve. As Arends et al. [2] illustrate, the summary ROC curve can be obtained after characterizing the bivariate Normal model via an appropriate line or relationship between sensitivity and specificity and then by transforming the line to the ROC space. Common choices are the regression line of  $\eta_i$  over  $\xi_i$  or the alternative regression line of  $\xi_i$  over  $\eta_i$ .

## 1.2 Alternative formulation

Criticism towards the use of the classical specification of the summary ROC curve as in Arends et al. [2] is expressed in Hamza et al. [3], who note that the bivariate model does not make assumptions on the study

specific curves. Consequently, the chosen summary ROC curve does not correspond to an average summary ROC curve or an overall representative summary ROC curve for the ROCs of the different studies. For this reason, Hamza et al. [3] suggest a new formulation of the bivariate model, providing a summary ROC curve that is an average of the study specific ROC curves. For each study  $i$ , the model assumes a linear relationship between  $\eta_i$  and  $\xi_i$ , given by  $\eta_i = \alpha_i + \beta\xi_i$ , with  $\alpha_i \sim \mathcal{N}(\alpha, \sigma_\alpha^2)$  and  $\xi_i \sim \mathcal{N}(\xi, \sigma_\xi^2)$ ,  $i = 1, \dots, K$ . The relationship implies that the ROC curves have a different intercept  $\alpha_i$  but the same slope  $\beta$ , in the  $(\xi, \eta)$  space. The between-study model is a modification of (1)

$$\begin{pmatrix} \eta_i \\ \xi_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \alpha + \beta\xi \\ \xi \end{pmatrix}, \Sigma_2\right), \quad \text{where} \quad \Sigma_2 = \begin{pmatrix} \sigma_\alpha^2 + \beta^2\sigma_\xi^2 + 2\beta\sigma_{\alpha\xi} & \sigma_{\alpha\xi} + \beta\sigma_\xi^2 \\ \sigma_{\alpha\xi} + \beta\sigma_\xi^2 & \sigma_\xi^2 \end{pmatrix}. \quad (4)$$

Given the within-study model (2), the marginal model is

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \alpha + \beta\xi \\ \xi \end{pmatrix}, \Sigma_2 + S_i\right), \quad (5)$$

with variance/covariance matrix

$$\Sigma_2 + S_i = \begin{pmatrix} s_{\eta_i}^2 + \sigma_\alpha^2 + \beta^2\sigma_\xi^2 + 2\beta\sigma_{\alpha\xi} & \sigma_{\alpha\xi} + \beta\sigma_\xi^2 \\ \sigma_{\alpha\xi} + \beta\sigma_\xi^2 & s_{\xi_i}^2 + \sigma_\xi^2 \end{pmatrix}.$$

Model (5) is similar to (3), with a different parameterization characterized by the fixed-effect component  $\beta$  entering the variance/covariance matrix  $\Sigma_2 + S_i$ , and one parameter more. Thus, model (5) is not identifiable. In order to guarantee identifiability, further assumptions are needed. A first solution is setting the correlation between  $\alpha_i$  and  $\xi_i$  to zero, namely,  $\sigma_{\alpha\xi} = 0$ . Hamza et al. [3] underline that such an assumption would imply that investigators in selecting the level of  $\xi_i$  are not lead by the intercept of their specific ROC curve. Under the assumption  $\sigma_{\alpha\xi} = 0$ ,  $\beta$  is the slope of the regression line of  $\eta_i$  on  $\xi_i$ . Then, the regression of  $\eta_i$  on  $\xi_i$  produces the average line over the studies and, in the ROC space, the associated summary ROC curve can be interpreted as an average ROC curve. A second assumption considers the correlation between  $\eta_i$  and  $\alpha$  is zero, namely,  $\sigma_{\alpha\eta} = 0$ . In this case,  $\sigma_{\eta\xi} = -\sigma_\alpha^2/\beta$  and  $\beta$  is the slope of the regression of  $\xi_i$  on  $\eta_i$ . Then, the regression of  $\xi_i$  on  $\eta_i$  produces the average line over the studies and, in the ROC space, the associated SROC curve can be interpreted as an average ROC curve. Alternative formulations are possible, including linear combinations of  $\eta_i$  and  $\xi_i$ , see Arends et al. [2] and Hamza et al. [3]

### 1.3 Measurement errors

The information available from the studies included in the meta-analysis is affected by error, as the observed  $\hat{\eta}_i$  and  $\hat{\xi}_i$  are estimated versions of the true  $\eta_i$  and  $\xi_i$ . This is quite a common problem in meta-analysis, given the nature of the available data that are summary information obtained from each study, see Arends et al. [15], Ghidry et al. [16], Guolo [17]. Not accounting for measurement errors can result in misleading inference, as a huge literature testifies, see, e.g., Keogh et al. [18] and Shaw et al. [19]. Consequences of ignoring the presence of measurement errors in variables are various, with negligible to relevant effects on inferential conclusions. The most famous effect in simple linear regression models with mismeasured covariate is the slope biased towards zero, a phenomenon known as attenuation. Attenuation occurs in case of classical and additive measurement error [20], that is, when the observed measure  $X^*$  is the sum of the true unobserved covariate  $X$  plus an error component  $U$ , where  $U$  is independent of  $X$ , with zero mean and constant variance. In case of non-linear models or complex error structures, effects are unpredictable, see, e.g., Chapter 3 in Carroll et al. [12]. In meta-analysis of diagnostic accuracy studies, an attenuated slope of the regression line of the summary ROC curve is the result of the measurement errors affecting  $\hat{\eta}_i$  and  $\hat{\xi}_i$ , see Arends et al. [2]. In the context of meta-analysis of diagnostic accuracy studies, the measurement error problem is somehow peculiar. Commonly, measurement error impacts inference when a covariate is affected, while errors in the measure of

the response variable of a regression model give rise to less concern. In case of diagnostic accuracy studies, there is no clear definition of the response and the covariate, that is, the role of  $\eta_i$  and  $\xi_i$  is not undoubtedly established, as  $\eta_i$  and  $\xi_i$  can act as response or covariate according to the specific regression model chosen for the summary ROC curve, see, e.g., Arends et al. [2]. Only when a specific summary ROC curve is chosen, then the role of  $\eta_i$  and  $\xi_i$  in terms of response variable or covariate is defined.

The likelihood approach based on the random-effects bivariate model in (3) implicitly accounts for the measurement error in  $(\hat{\eta}_i, \hat{\xi}_i)^\top$  through the specification of the within-study model that defines a relationship between the error prone  $(\hat{\eta}_i, \hat{\xi}_i)^\top$  and the true unobserved  $(\eta_i, \xi_i)^\top$  [2, 21, 22]. A similar conclusion holds with respect to the model formulation proposed in Hamza et al. [3] giving rise to (5). Nevertheless, the performance of the likelihood approach with reference to model (5) has not been evaluated in detail, neither alternative solutions inspired by the measurement error literature have been explored. In the following section, we adapt the SIMEX methodology, a simulation-based technique used to estimate and correct for measurement error in regression models, to the model in Hamza et al. [3] and investigate its performance compared to that of the likelihood approach in a series of simulation studies.

## 2 SIMEX

SIMEX is a general-purpose simulation-extrapolation technique developed to estimate and reduce bias due to measurement error [23–26]. SIMEX is well suited to deal with errors having a classical and additive structure, that is, when the error prone variable  $W$  is an unbiased measure of the unobserved variable  $X$ , such that  $W = X + U$ , with  $E(U|X) = 0$ . Actually, it can be easily extended to scenarios where measurement error can be simulated, as, for example, when the measurement error variance is known or approximately known. SIMEX consists of a simulation step followed by an extrapolation step. In the simulation step, a resampling-like strategy is used to generate  $M$  datasets with increasing measurement error. Each dataset is used to estimate the parameters of interest. In the extrapolation step, the relationship between the estimates and the amount of additional error is obtained and used to extrapolate the corrected SIMEX estimate of the parameter to the case of no measurement error.

In the following we illustrate the application of SIMEX to meta-analysis of diagnostic accuracy studies, when the reference model is the bivariate random-effects model according to the formulation in Hamza et al. [3]. The application of SIMEX within this framework is inspired by Guolo [11], who applied SIMEX in the traditional bivariate random-effects model for meta-analysis of diagnostic accuracy studies, and by Guolo [17], who considered SIMEX in meta-analysis including information about the underlying risk of disease in healthy subjects.

Let  $W_i(\lambda) = (\hat{\xi}_i, \hat{\eta}_i)^\top$  and let  $X_i = (\xi_i, \eta_i)^\top$  be the vector of error-prone covariates and the vector of true variables, respectively. We consider the measurement error model relating  $W_i$  to  $X_i$  being the within-study model (2). The measurement error structure perfectly suits the assumption in the original SIMEX methodology development, as we deal with classical and additive errors, with approximately known variance/covariance matrix  $S_i$ . The components of  $W_i$  can act as response variable or as covariate when the relationship useful to obtain the summary ROC curve is specified. Accordingly, the SIMEX methodology is called double SIMEX, following Holcomb [27] who firstly investigated the use of SIMEX in case of measurement error affecting both the response variable and the covariate in regression models.

In the simulation step,  $B$  datasets with additional error are generated. For fixed  $\lambda \geq 0$ , let  $W_{b,i} = W_i + \sqrt{\lambda} S_i^{-1/2} U_{b,i}$  be the  $i$ th observation of the  $b$ th dataset,  $b = 1, \dots, B$ , where  $S_i$  is the variance/covariance matrix of  $W_i$  and the pseudo errors  $U_i^b, i = 1, \dots, K$  are simulated from a Gram–Schmidt process [27] in order to guarantee zero mean and unit variance. Such a choice reduces the Monte Carlo variance of the SIMEX estimates, see Chapter 5 in Carroll et al. [12]. Variable  $W_{b,i}$  can be seen as a remeasurement of  $W_i$ , with variance  $\text{Var}(W_i) = \sigma_u^2$ . Consider that  $\text{Var}(W_i|X_i) = \sigma_u^2$  and that

$$\text{Var}\{W_{b,i}(\lambda)|X_i\} = (1 + \lambda) \text{Var}\{W_i(\lambda)|X_i\} = (1 + \lambda)\sigma_u^2.$$

Since  $E\{W_{b,i}(\lambda)|X_i\} = X_i$ , the mean squared error of  $W_{b,i}$

$$\text{MSE}\{W_{b,i}(\lambda)\} = (1 + \lambda)\sigma_u^2$$

converges to zero as  $\lambda \rightarrow -1$ . Let  $\hat{\theta}_b(\lambda)$  be the estimator of  $\theta$  using data  $W_{b,i}(\lambda)$  for fixed  $\lambda$  obtained by a chosen approach, e.g., the method of moments, and let  $\hat{\theta}(\lambda)$  be the sample mean of  $\hat{\theta}_b(\lambda)$ . The peculiar feature of model (5), with the fixed-effects components  $\beta$  entering the variance/covariance matrix, implies that the method-of-moments estimate of  $\theta_b(\lambda) = (\alpha, \beta, \xi, \sigma_\alpha^2, \sigma_\xi^2)^\top$  based on  $(\hat{\eta}_{i,b}, \hat{\xi}_{i,b})^\top$  from the  $b$ -th dataset is not a useful choice, differently from the classical framework in model (3) investigated in Guolo [11]. We suggest to estimate  $\theta_b(\lambda)$  by maximizing the log-likelihood function

$$\begin{aligned} \ell_b(\theta) = & -\frac{1}{2}(\log\sigma_\xi^2 + \log\sigma_\alpha^2) - \frac{1}{2\sigma_\alpha^2\sigma_\xi^2} \left\{ \sigma_\xi^2(\hat{\eta}_{i,b} - \alpha - \beta\xi)^2 \right. \\ & \left. + (\sigma_\alpha^2 + \beta^2\sigma_\xi^2)(\hat{\xi}_{i,b} - \xi)^2 - 2\beta\sigma_\xi^2(\hat{\eta}_{i,b} - \alpha - \beta\xi)(\hat{\xi}_{i,b} - \xi) \right\}. \end{aligned} \quad (6)$$

Differently from the classical bivariate random-effects model in Reitsma et al. [1] and Arends et al. [2], the maximum likelihood estimate of  $\theta$  in (6) is not available in closed form. An iterative algorithm for maximization can be used with starting values available, for example, from the method-of-moments type estimators.

In the extrapolation step, a relationship between  $\hat{\theta}(\lambda)$  and  $\lambda$  is defined and used to obtain the SIMEX estimate of  $\theta$ ,  $\theta_{\text{SIMEX}}$ , as the extrapolation to the case  $\lambda = -1$ . In practice, two popular solutions are the linear extrapolation function and the quadratic extrapolation function. The quadratic extrapolation function is usually preferable given its numerical stability, see Chapter 3 in Carroll et al. [12]. The SIMEX estimate of  $\theta$  is the SIMEX estimate of each component, after applying the extrapolation function to each set of  $B$  estimates of the parameters resulting from the simulation step. The resulting SIMEX estimator  $\hat{\theta}_{\text{SIMEX}}$  is a consistent estimator of  $\theta$  with asymptotically Normal distribution [25]. The computation of the standard error of  $\hat{\theta}_{\text{SIMEX}}$  can be obtained via a similar approach using to derive the SIMEX estimate. Let  $s_b^2(\lambda)$  be the variance/covariance matrix estimate of  $\hat{\theta}_b$ , obtained by means of the sandwich estimator or the inverse of the observed Fisher Information matrix. Let  $s^2(\lambda)$  be the average of the  $B$  variance estimates,  $s^2(\lambda) = B^{-1} \sum_{b=1}^B s_b^2(\lambda)$ , and let  $s_\Delta^2(\lambda)$  be the sample variance/covariance matrix of the terms  $\hat{\theta}_b$ ,  $b = 1, \dots, B$ . The variance/covariance matrix of the SIMEX estimator is obtained by extrapolating  $s^2(\lambda) - s_\Delta^2(\lambda)$  to the case  $\lambda = -1$ , see Stefanski and Carroll [24] and Appendix B.4 in Carroll et al. [12].

## 3 Simulation study

### 3.1 Set-up

The performance of SIMEX is evaluated in a series of simulation studies and compared to that of the likelihood approach based on a Normal approximation for the within-study model, as described in Section 1.

Data are generated according to a two-step approach. In the first step, values of  $\eta_i$  and  $\xi_i$  are generated for each study included in the meta-analysis, starting from the between-study model (4) under the identifiability assumption  $\sigma_{\alpha\xi} = 0$ . Values of parameters  $\alpha$ ,  $\beta$ ,  $\xi$  are chosen in order to reflect a high accuracy of the test ( $\alpha = 3.5, \beta = 0.25, \xi = -2.2$ ), a medium accuracy of the test ( $\alpha = 1.76, \beta = 0.25, \xi = -1.5$ ) and a low accuracy of the test ( $\alpha = 0.83, \beta = 0.25, \xi = -0.85$ ). The choice ensures different values of sensitivity/specificity, namely,  $Se \in \{0.95, 0.8, 0.65\}$  and  $Sp \in \{0.9, 0.82, 0.7\}$ , respectively. Values of the variance components  $\sigma_\xi^2$  and  $\sigma_\alpha^2$  are chosen in way to consider increasing

levels of correlation  $\rho = \beta\sigma_\xi^2 \left\{ (\sigma_\alpha^2 + \beta^2\sigma_\xi^2) \sigma_\xi^2 \right\}^{-1/2}$  between  $\eta_i$  and  $\xi_i$ , namely,  $\rho = 0.222$  ( $\sigma_\xi^2 = 0.5, \sigma_\alpha^2 = 0.6$ ),  $\rho = 0.438$  ( $\sigma_\xi^2 = 1.9, \sigma_\alpha^2 = 0.5$ ) and  $\rho = 0.800$  ( $\sigma_\xi^2 = 4, \sigma_\alpha^2 = 0.141$ ). In the second step, given the values of  $(\eta_i, \xi_i)^\top$ , values of true positives  $n_{11i}$  and false positives  $n_{10i}$  for study  $i$  are simulated from a Binomial distribution, namely,

$$n_{11i} \sim \text{Bin}\{n_{1i}, (1 + e^{-\eta_i})^{-1}\}, \quad n_{10i} \sim \text{Bin}\{n_{0i}, (1 + e^{-\xi_i})^{-1}\}. \quad (7)$$

The number of positives  $n_{1i}$  and the number of negatives  $n_{0i}$  are drawn from a Uniform distribution on  $[40, 200]$ , respectively. The simulation study considers an increasing number  $K$  of studies included in the meta-analysis,  $K \in \{10, 20, 50\}$ . The simulation experiment has been repeated 1000 times for each combination of the parameters.

The simulation study examines the robustness of the methods against model misspecifications, by investigating deviations from the normality assumption of the random-effect. In a first case, values of  $\xi_i$  are generated from a Skew-Normal distribution [28] in order to account for skewness of the distribution, with skewness parameter assuming values in  $\{0.4, 0.55, 0.65\}$ . Two additional cases have been considered, namely, a Student  $t$  distribution in order to account for heavy tails of the distribution and the logarithm of a Pareto distribution. Increasing value of the degrees of freedom for the Student  $t$  distribution has been considered,  $\{3, 6, 10\}$ . With respect to the Pareto distribution, the shape parameter is fixed to 3, while the scale parameter assumes values in  $\{0.008, 0.16, 0.31\}$ . For each scenario, the choice of the parameters values guarantees a mean of  $\xi$  in  $\{-2.2, -1.5, -0.85\}$ , in order to generate a high accuracy scenario, a medium accuracy scenario and a low accuracy scenario.

When applying the likelihood approach, the sandwich estimate of the variance is adopted in order to account for potential model misspecifications. The application of SIMEX considers  $\lambda$  assuming values on a grid,  $\lambda \in \{0.0, 0.5, 1, 1.5, 2\}$ , a number  $B = 100$  of remeasured data generated using the Gram–Schmid process, and the quadratic extrapolating function, given its numerical stability, see, e.g., Chapter 5 in Carroll et al. [12]. In the simulation step, the estimation of the parameters is performed through the optimization of the likelihood function (6), with initial values obtained from a method-of-moments strategy. The  $R$  programming language [29] has been used for analysis. The software for implementing SIMEX is available as Supplementary Material.

The competing methods are examined in terms of bias, estimated standard error and standard deviation of the estimators of the parameters  $\alpha, \beta, \xi, \sigma_\xi^2, \sigma_\alpha^2$  and in terms of the 95% Wald-type confidence interval for the estimators of  $\alpha, \beta, \xi$ . The measures of test accuracy given by the diagnostic odds ratio dOR, the positive likelihood ratio LR+ and the negative likelihood ratio LR– are provided as well. The methods are evaluated also in terms of convergence problems. Successful convergence is intended as meeting the criterion convergence (e.g., difference between current and updated estimates less than 0.0001) and positive definite variance/covariance matrix. The results under non-convergence are excluded when summarising the simulation results.

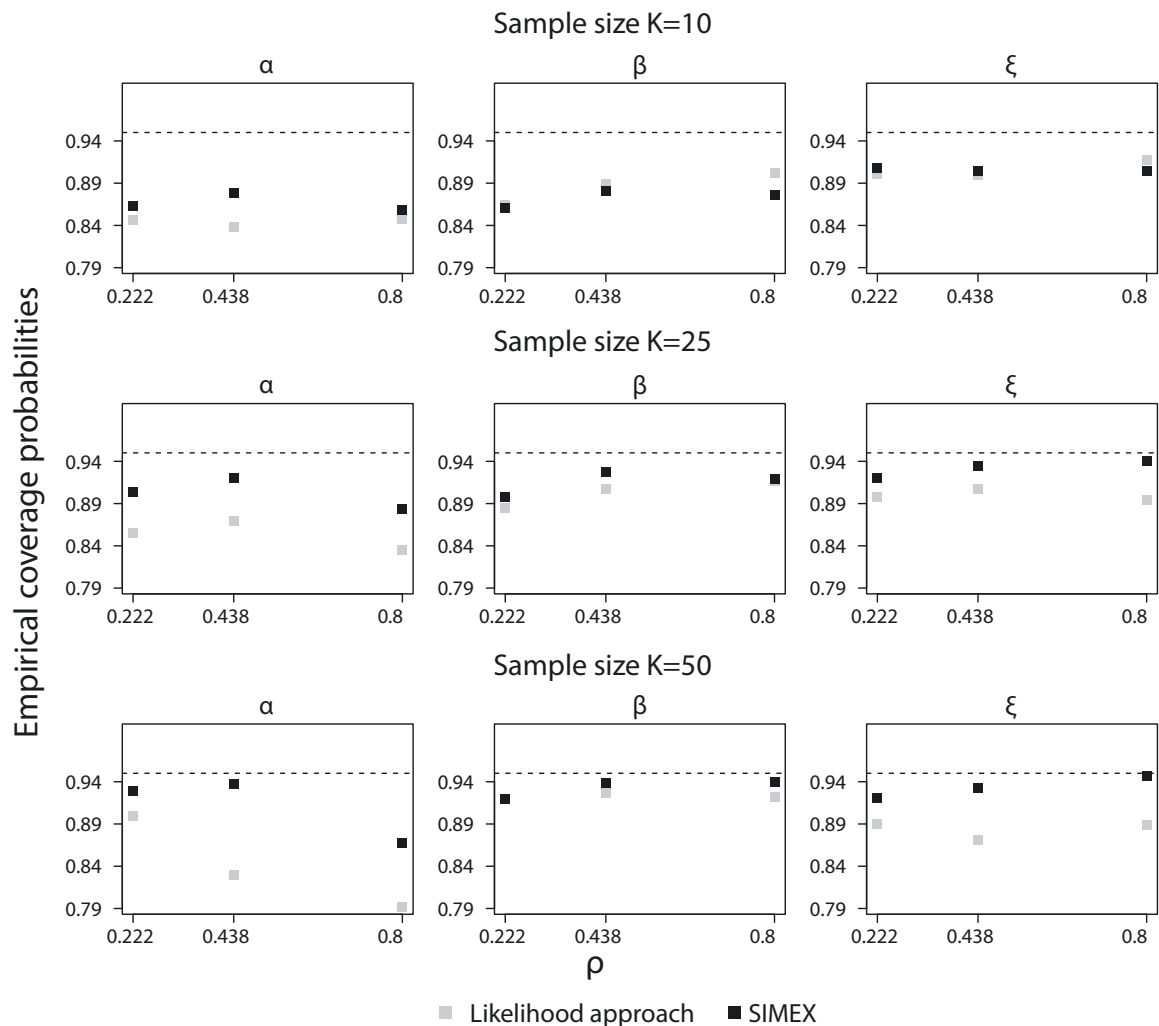
## 3.2 Results

Under a Normal specification for the distribution of  $\xi_i$ , the likelihood-based approach tends to provide estimators of  $\alpha, \beta, \xi$  more biased than SIMEX, as illustrated in Table A1 for the high accuracy scenario and the low accuracy scenario, and in Table S1 in the Supplementary Material for the medium accuracy scenario, when  $K = 10$ . The bias is larger under the high accuracy scenario, and it tends to decrease moving to the medium accuracy case (Table S1 in the Supplementary Material) and to the low accuracy case. As expected from a theoretical point of view, the bias decreases as the sample size increases, see Tables S2 and S3 in the Supplementary Material referred to  $K = 25$  and  $K = 50$ . For both the methods the bias is larger when the correlation  $\rho$  between  $\eta_i$  and  $\xi_i$  close to the upper boundary level, with the effect being more pronounced for the likelihood approach.

Similar results are experienced for the estimators of the variance components, see Table A2 for  $K = 10$  and Tables S4 and S5 in the Supplementary Material for larger  $K$ . The bias is even larger when relying on the likelihood approach than it was with respect to the regression coefficients  $\alpha, \beta, \xi$ , especially for small sample size and large correlation  $\rho$ . Results are more biased in case of high accuracy of the test.

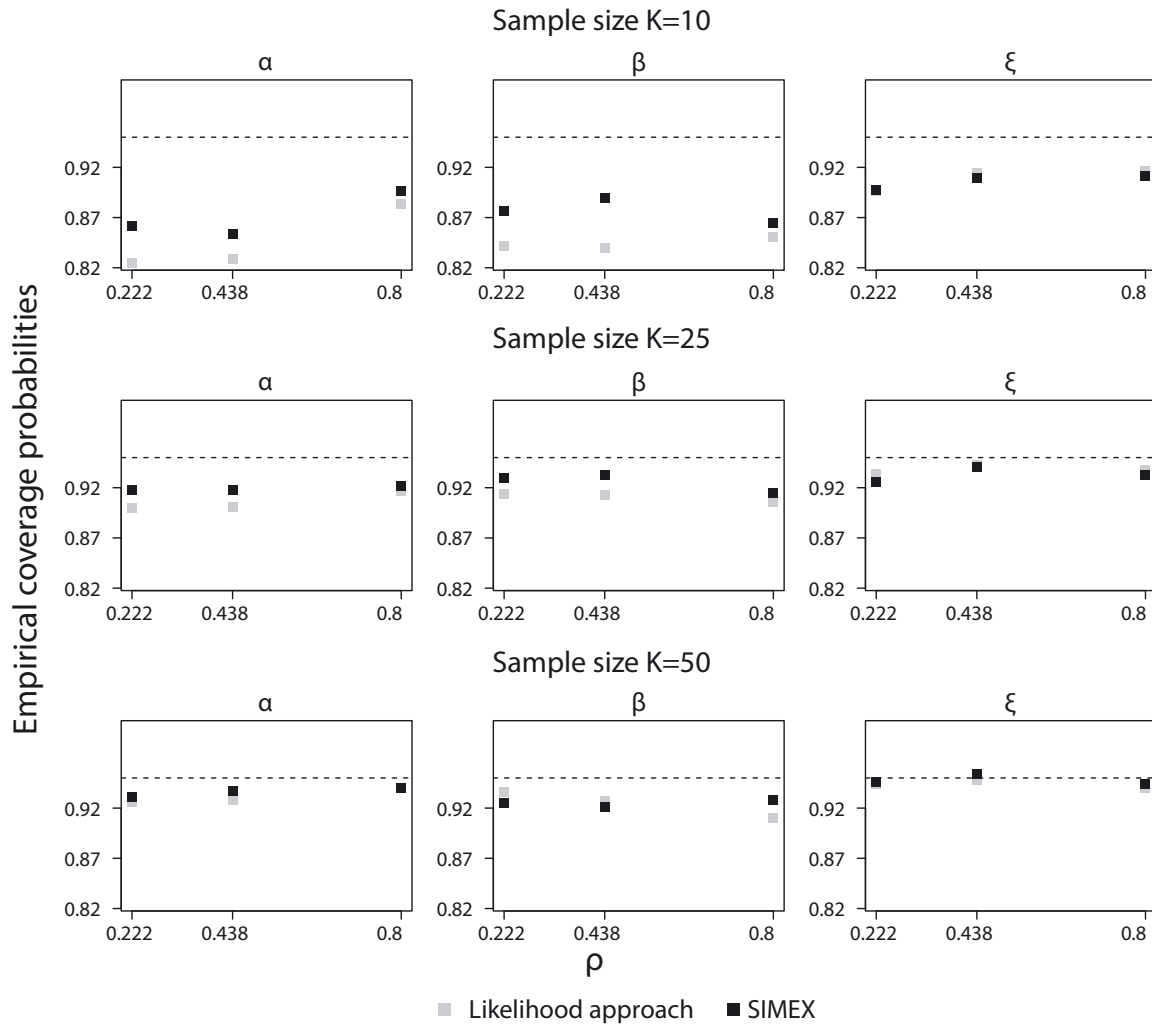
The likelihood approach provides larger standard errors of the estimators than SIMEX, mainly with reference to the estimators of  $\alpha$  and  $\beta$ , while smaller results are provided with reference to the estimator of the variance component  $\sigma_{\xi}^2$ . For both the cases, bias tends to reduce moving to scenarios with lower accuracy.

Figures 1 and 2 report the empirical coverage probability at nominal level 0.95 for the estimators of  $\alpha, \beta, \xi$ , under high accuracy or low accuracy of the diagnostic test, for increasing  $\rho$  and sample size  $K$ . The results for the medium accuracy case are reported in Figure S1 in the Supplementary Material.



**Figure 1:** Empirical coverage probabilities of confidence intervals for the estimators of  $\alpha, \beta, \xi$ , based on the likelihood approach under a normal specification of the random-effects (grey points) and SIMEX (black points). Values of  $(\eta_i, \xi_i)^T$  are generated from a bivariate normal distribution, under high accuracy scenario. Results are reported for increasing correlation and sample size  $K$ , on the basis of 1000 replicates. The grey horizontal line is the target 0.95 nominal level.



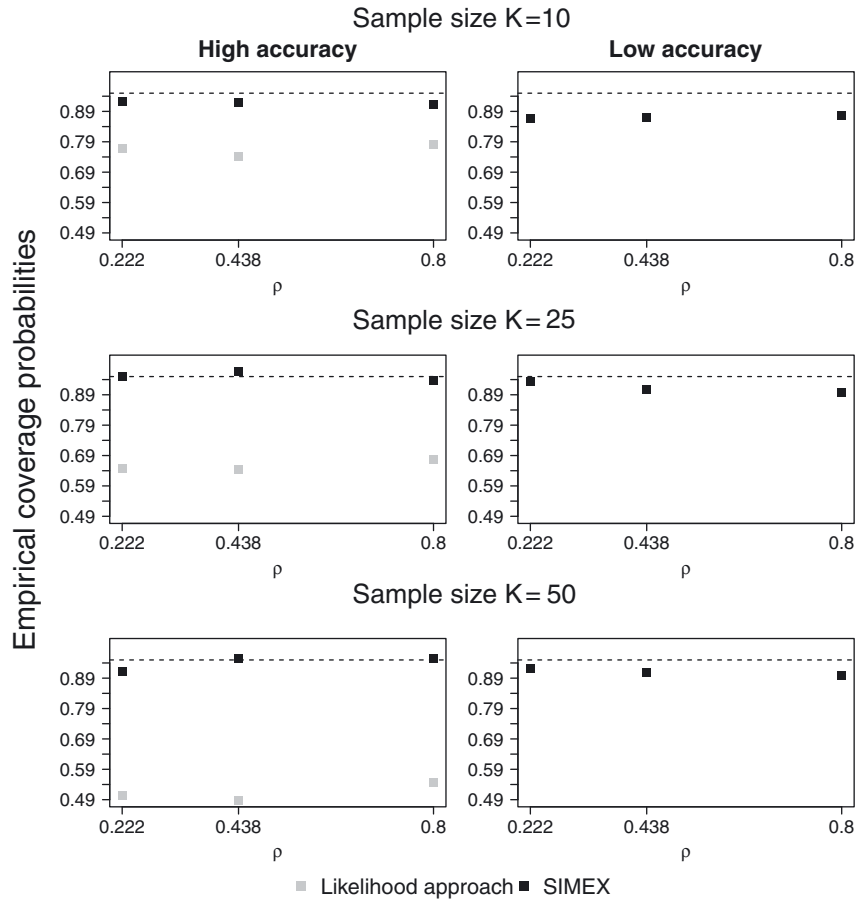


**Figure 2:** Empirical coverage probabilities of confidence intervals for the estimators of  $\alpha, \beta, \xi$ , based on the likelihood approach under a normal specification of the random-effects (grey points) and SIMEX (black points). Values of  $(\eta_i, \xi_i)^T$  are generated from a bivariate normal distribution, under low accuracy scenario. Results are reported for increasing correlation and sample size  $K$ , on the basis of 1000 replicates. The grey horizontal line is the target 0.95 nominal level.

Results show that the likelihood approach tends to underestimate the target level, especially in case of small sample size and under the high accuracy scenario. See, for example, the low empirical coverage probability for the estimator of  $\alpha$  in case of high accuracy scenario in Figure 1. Increasing the sample size does not help, as the standard error of the estimators decreases faster than the associated bias, implying that the confidence intervals are centered on values far from the true ones. Results from SIMEX are more satisfactory. Empirical coverage probability for all the examined parameters tend to be closer to the target level, under different values of  $\rho$ , sample size  $K$  and under either the high accuracy scenario or the low accuracy scenario. Such a superior performance with respect to the likelihood approach is more evident for the high accuracy case, with increasing  $K$ , see Figure 1, bottom panels.

Figure 3 and Figure S2 in the Supplementary Material report the empirical coverage probabilities at nominal level 0.95 for the positive and negative likelihood ratios and for the diagnostic odds ratio, respectively.

With reference to the diagnostic odds ratio, results from SIMEX notably outperform the likelihood approach, which suffer from empirical coverage probabilities substantially lower than the target level, mainly



**Figure 3:** Empirical coverage probabilities of confidence intervals for the estimators of diagnostic odds ratio (dOR), based on the likelihood approach under a normal specification of the random-effects (grey points) and SIMEX (black points). Values of  $(\eta_i, \xi_i)^T$  are generated from a bivariate normal distribution, under high and low accuracy scenario. Results are reported for increasing correlation and sample size  $K$ , on the basis of 1000 replicates. The grey horizontal line is the target 0.95 nominal level.

for the high accuracy scenario, see Figure 3. In case of low accuracy of the test, the performance of the competing methods is similar. A more satisfactory performance of SIMEX is experienced also in terms of positive and negative likelihood ratio, with the discrepancy in favour of SIMEX being more evident for large  $K$  and under the high accuracy scenario. Analogous results for the medium accuracy scenario are reported in Figure S3 in the Supplementary Material.

Substantial differences between the methods occur in terms of failure rate of the estimation process. The likelihood approach suffers from computational problems for large values of  $\rho$  and under the high accuracy case, see the larger values of the failure rate in Table A1. Issues are mostly related to the estimate of the parameters on the boundary of the parameter space. Such a result is in line with previous studies in the literature, see, for example, Diaz et al. [30], Chen et al. [8], Guolo [11], Takwoingi et al. [7]. As expected, increasing the sample size is helpful in reducing the computational issues. See the very low failure rate for  $K = 50$  in Table S3 in the Supplementary Material. Conversely, no convergence problems have been encountered when applying SIMEX, independently of the sample size  $K$ , the correlation  $\rho$  or the level of accuracy of the test.

Results in case of misspecification of the distribution of  $\xi_i$ , considering either a Skew-Normal distribution, a Student  $t$  distribution or a log Pareto distribution, are reported in Tables S6–S11 and Figures S4–S14 in the Supplementary Material. The behavior of the likelihood approach and SIMEX is similar

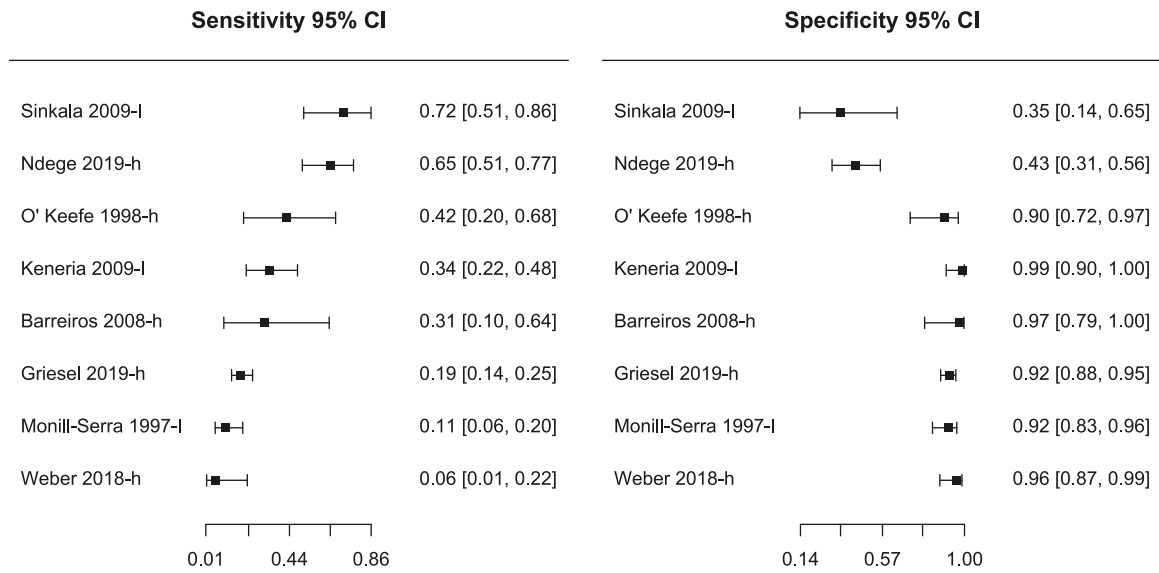
to that under the Normal distribution. A larger bias of the likelihood-based estimation of the parameters of interest with respect to SIMEX is still present, especially in case of small sample size. See, for example, the large bias of the likelihood-based estimator of  $\xi$  and of the variance components under a Student  $t$  distribution for small sample size  $K = 10$  and the large bias of the likelihood-based estimator of  $\alpha$  under a log Pareto distribution for small sample size  $K = 10$  (Tables S6 and S7). A discrepancy between the estimated standard error and the standard deviation of the parameter estimators is expected as a consequence of the model misspecification. The effects of misspecification of the distribution of  $\xi_i$  are evident in terms of empirical coverage probability, with results from the likelihood approach sometimes very far from the target level. See the empirical coverage probability close to 50% for the estimator of  $\xi$  under the Student  $t$  distribution for the high accuracy case in Figure S5. The advantages of SIMEX include empirical coverage probabilities that tend to have a quite stable behaviour close to the target level, under the misspecification scenarios, see Figures S4–S7. As for the Normal case, such a behavior is not affected by variations of the sample size  $K$  as well as variations of the accuracy of the diagnostic test.

Figures S8–S14 in the Supplementary Material report the empirical coverage probabilities of the positive likelihood ratio, the negative likelihood ratio, and the diagnostic odds ratio in case of misspecification of the distribution of  $\xi_i$ . Differences between the competing methods are more evident under the Skew-Normal distribution and the Student  $t$  distribution for the high accuracy case, see Figures S8 and S11. In these cases, the likelihood-based approach provides empirical coverage probabilities substantially far from the target level, while the performance of SIMEX is less affected. Notable improvements of SIMEX over the likelihood solution are experienced under a Student  $t$  distribution for  $\xi_i$  with a small number of degrees of freedom, especially when estimating the diagnostic odds ratio, see Figure S11. Differences between the methods reduce moving from the high accuracy scenario to the low accuracy scenario. Increasing the sample size does not help the likelihood approach to improve on empirical coverage probabilities. This result is an expected consequence of the poor behaviour of the likelihood approach when estimating the regression components  $\alpha, \beta, \xi$ . Less marked differences in terms of empirical coverage probabilities between the likelihood approach and SIMEX appear under a log Pareto distribution for  $\xi$ , for increasing values of the scale parameter, see Figure S14.

The comparison of the methods with respect to the convergence problems still highlights a large failure rate for the likelihood approach, with evidence in case of small sample size, while no computational issues affect SIMEX, see Table S6 in the Supplementary Material for  $K = 10$ . The failure rate for the likelihood approach reduces when increasing the sample size, see Tables S8 and S10 in the Supplementary Material.

## 4 Data example

Tuberculosis is an infection involving mostly the lungs. Nowadays, it remains a major cause of ill health, being one of the top ten causes of death worldwide, mainly affecting adult men [e.g., 31]. In HIV-positive people with advanced immunosuppression, tuberculosis is challenging to detect, since symptoms are similar to those presented by other pulmonary infections. As a consequence, a substantial portion of tuberculosis cases remained undiagnosed at death. Nevertheless, autopsy studies indicate a very high proportion of tuberculosis in HIV patients [32]. Within the population having tuberculosis associated with HIV, up to 50% of the cases is estimated to correspond to extrapulmonary tuberculosis including abdominal or disseminated tuberculosis [e.g., 33]. van Hoving et al. [34] evaluate the accuracy of the abdominal ultrasound exam to diagnose the infection from abdominal or disseminated tuberculosis in HIV-positive subjects. The reference standard is represented by a bacteriological confirmation, usually associated to clinical diagnosis based on X-ray abnormalities or suggestive histology. We consider the data referred to ascites on abdominal ultrasound for tuberculosis detection, including eight studies, for a total of 891 subjects. The forest plot for the estimates of sensitivity and specificity provided by each study included in the meta-analysis is reported in Figure 4. Univariate meta-analysis provides an estimate of the overall sensitivity and an estimate of the overall specificity equal to 0.342 (standard error 0.09) and 0.817 (standard error 0.091), respectively.

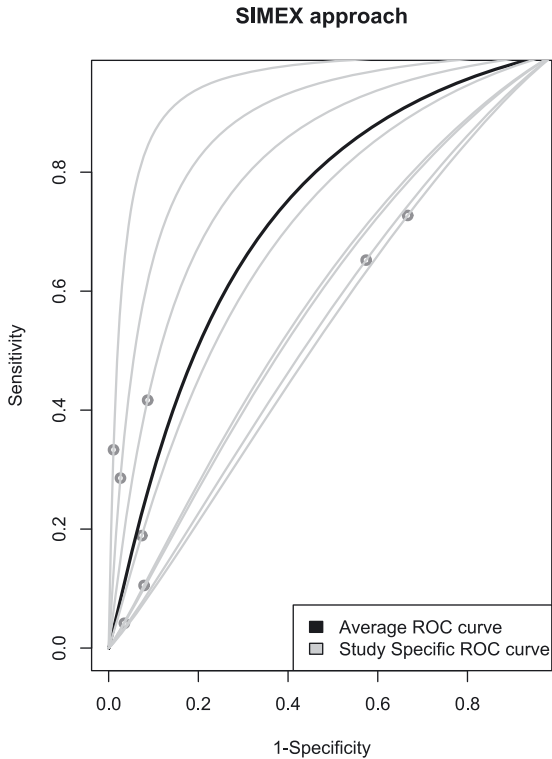


**Figure 4:** Forest plot of sensitivity and specificity for the meta-analysis of abdominal ultrasound examination for diagnosing tuberculosis in HIV-positive subjects [34].

Data have been examined through the likelihood approach and SIMEX, using a method-of-moments type estimate as starting value in the optimization process. The application of the likelihood approach did not reach convergence as the estimate of the variance components reaches the boundary of the parameter space, and the resulting variance/covariance matrix is not positive-definite. Changes in the optimization algorithm and changes of the starting values do not solve the non-convergence problem. Results are thus reported only for SIMEX, whose application has no convergence issues. The estimates of the parameters  $\alpha$ ,  $\beta$ ,  $\xi$ ,  $\sigma_{\xi}^2$ ,  $\sigma_{\alpha}^2$  in model (5) and the associated standard errors are reported in Table 2, under the identifiability assumption  $\sigma_{\alpha\xi} = 0$  (left column) and under the identifiability assumption  $\sigma_{\alpha\eta} = 0$  (right column). Results for sensitivity and specificity, likelihood ratios and diagnostic odds ratio are reported as well. The corresponding standard errors are evaluated using the delta method. Differences between the scenarios include a larger estimate of the fixed-effects components  $\alpha$  and  $\beta$  and a larger estimate of  $\sigma_{\alpha}^2$  under the identifiability assumption  $\sigma_{\alpha\eta} = 0$ .

**Table 2:** Estimates and associated standard error of parameters  $\alpha$ ,  $\beta$ ,  $\xi$ ,  $\sigma_{\xi}^2$ ,  $\sigma_{\alpha}^2$ , sensitivity, specificity, LR+, LR-, dOR obtained from SIMEX for the analysis of abdominal ultrasound for diagnosing tuberculosis in HIV-positive subjects [34] under identifiability assumptions  $\sigma_{\alpha\xi} = 0$  and  $\sigma_{\alpha\eta} = 0$ .

Estimate	$\sigma_{\alpha\xi} = 0$	$\sigma_{\alpha\eta} = 0$
$\alpha$	0.329 (0.547)	1.561 (1.047)
$\beta$	0.535 (0.204)	1.103 (0.424)
$\xi$	-2.226 (0.595)	-2.214 (0.599)
$\sigma_{\xi}^2$	2.727 (1.405)	2.783 (1.441)
$\sigma_{\alpha}^2$	0.810 (0.438)	1.763 (1.636)
Sensitivity	0.297 (0.094)	0.293 (0.091)
Specificity	0.902 (0.353)	0.901 (0.353)
LR+	3.048 (1.531)	2.976 (2.369)
LR-	0.779 (0.080)	0.784 (0.079)
dOR	3.913 (1.707)	3.796 (1.699)



**Figure 5:** Study-specific ROC curves and estimated average ROC curve from SIMEX, under the assumption  $\sigma_{\alpha\eta} = 0$ , for the abdominal ultrasound examination for diagnosing tuberculosis in HIV-positive subjects [34].

Within the same scenario, larger standard errors affect the estimators of the parameters if compared to the case  $\sigma_{\alpha\eta} = 0$ . There are no substantial variations in terms of measures of the diagnostic test accuracy with respect to the identifiability assumptions. Under the assumption  $\sigma_{\alpha\xi} = 0$ , the estimated sensitivity is 0.297 and estimated specificity is 0.902, while under the assumption  $\sigma_{\alpha\eta} = 0$ , the estimated sensitivity is 0.293 and estimated specificity is 0.901. The study-specific ROC curves and the estimated average ROC curve obtained from SIMEX applied to model (5) under the identifiability assumption  $\sigma_{\alpha\xi} = 0$  are reported in Figure 5.

## 5 Conclusions

This paper considered a random-effects structure for the meta-analysis model following the specification in Hamza et al. [3] for the evaluation of accuracy of a diagnostic test. The model specification allows the resulting summary ROC curve as an average of the ROC curves from the studies included in the meta-analysis. As an alternative to the classical likelihood approach based on a Normal approximation for the logit transformation of sensitivity and specificity, the paper investigated the applicability of SIMEX, a simulation-based approach derived from the measurement error literature. Both the solutions take into account the presence of errors due to the fact that the available information is a summary measure of the true unknown study-specific accuracy of the diagnostic test. The performance of the methods has been compared in a series of simulation studies exploring different scenarios, without and against violations of model assumptions, in particular concerning the normality distribution for the random effects components. SIMEX tends to outperform the likelihood approach in terms of bias of the estimators and in terms of empirical coverage probabilities, especially in case of scenarios characterized by high accuracy of the diagnostic test, small sample size and correlation between (logit) sensitivity and specificity close to boundary of the parameter space. Such a performance is even more evident in case of departures from the normality assumption for the random-effects components, where SIMEX is not seriously affected by skewness or kurtosis of the distribution of the (logit) specificity.

In the mentioned scenarios, the likelihood-based analysis tends to provide unreliable conclusions, a result in line with previous investigations in the literature [7, 8, 11].

The straightforward implementation of the likelihood approach as well as the negligible computational effort required for likelihood evaluation and maximization is paid in terms of convergence issues. The large failure rate experienced for small sample size, large correlation between (logit) sensitivity and specificity and under model misspecification substantially reduces the appealing of the method. Conversely, the failure rate of SIMEX is zero in most of the examined scenarios and close to zero under violations of the identifiability assumptions of the model. The computational effort of SIMEX is slightly superior to that required by the likelihood approach as a consequence of the first simulation-based step of SIMEX.

The likelihood approach considered in the paper is based on the commonly adopted approximate Normal distribution for the logit transformation of sensitivity and specificity at the within-study level. A natural alternative substitutes the approximate model with the exact Binomial distribution for the number of true positives and false negatives, see Hamza et al. [37, 38]. The resulting likelihood function is not in closed-form, with the need of numerical evaluation of integrals. Simulation studies in the literature, e.g. [7, 11, 39], within the classical bivariate random-effects model [2] show that such a specification can be preferable in terms of accuracy of the results, but at the price of additional substantial convergence problems.

The approaches examined in this paper can be extended to include study-level specific covariates. In case of no mismeasured covariates, straightforward extensions involve modifications of the mean in the between-study model (4), with no substantial complications. In case of mismeasured covariates, instead, such an extension needs to be accompanied by a proper model for the measurement error structure, similarly to that relating  $(\hat{\eta}_i, \hat{\xi}_i)^\top$  to  $(\eta_i, \xi_i)^\top$ . With reference to SIMEX, modifications useful to include mismeasured covariates is straightforward from a theoretical point of view, and it would affect only the simulation step. However, the number of remeasured datasets has to substantially increase to guarantee the results having an acceptable precision. Thus, the total computational effort of SIMEX might be not negligible. We refer the interested reader to Carroll et al. [12, Section 5.3] for details.

As a Referee pointed out, a Bayesian approach for inference might be considered. In this case, the Bayesian solution would enter the simulation step of the SIMEX strategy, useful to estimate the vector of the parameters by maximizing the log-likelihood function (6) which is not in closed-form. Such a choice, however, could increase the computational burden, when MCMC instrument are involved.

This paper focused on meta-analysis for evaluating a diagnostic test in terms of its capability to distinguish between diseased and non-diseased subjects, that is, to classify the subjects with respect to one threshold. However, test performance can be evaluated at multiple thresholds, or at ordered categories. The multiple thresholds extension of the examined model [3, 35] would lead to an overall summary ROC curve and to summary sensitivity and specificity for each threshold. Likelihood inference under the Normal specification for the random-effects component in the multiple thresholds extension of the model in Arends et al. [2] is shown to perform poorly and suffering from non-convergence problems [36]. Investigating the applicability of SIMEX in case of multiple thresholds, its capability to provide satisfactory results as in the one threshold case, while guaranteeing high convergence rates, is an interesting topic of feature research.

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Research funding:** None declared.

**Conflict of interest statement:** The authors declare no conflicts of interest regarding this article.

## Appendix A

**Table A1:** Bias of the estimators of  $\alpha, \beta, \xi$ , associated standard error (SE), standard deviation (SD) and failure rate obtained from the likelihood approach under a normal specification of  $(\eta_i, \xi_i)^T$  and SIMEX, by distinguishing high accuracy scenario and low accuracy scenario. Values of  $(\eta_i, \xi_i)^T$  are generated from a bivariate normal distribution. Results are reported for increasing correlation and sample size  $K = 10$ , on the basis of 1000 replicates.

	High accuracy scenario						Low accuracy scenario					
	Likelihood approach			SIMEX approach			Likelihood approach			SIMEX approach		
	$\alpha$	$\beta$	$\xi$	$\alpha$	$\beta$	$\xi$	$\alpha$	$\beta$	$\xi$	$\alpha$	$\beta$	$\xi$
$\rho = 0.222$												
True	3.500	0.250	-2.200	3.500	0.250	-2.200	0.830	0.250	-0.850	0.830	0.250	-0.850
Bias	-0.192	0.003	0.072	-0.031	-0.055	-0.058	0.004	0.021	0.007	-0.002	-0.019	-0.021
SE	1.245	0.681	0.229	0.899	0.382	0.242	0.419	0.408	0.214	0.386	0.353	0.219
SD	1.399	0.766	0.233	1.122	0.474	0.268	0.607	0.624	0.235	0.484	0.461	0.245
Failure	0.200	0.200	0.200	0.000	0.000	0.000	0.008	0.008	0.008	0.000	0.000	0.000
$\rho = 0.438$												
True	3.500	0.250	-2.200	3.500	0.250	-2.200	0.830	0.250	-0.850	0.830	0.250	-0.850
Bias	-0.164	0.016	0.150	0.077	-0.012	-0.032	0.014	0.018	0.025	0.027	-0.002	-0.024
SE	0.774	0.327	0.425	0.517	0.200	0.423	0.256	0.171	0.407	0.265	0.169	0.422
SD	0.641	0.276	0.417	0.639	0.248	0.466	0.345	0.235	0.419	0.332	0.205	0.448
Failure	0.210	0.210	0.210	0.000	0.000	0.000	0.003	0.003	0.003	0.000	0.000	0.000
$\rho = 0.800$												
True	3.500	0.250	-2.200	3.500	0.250	-2.200	0.830	0.250	-0.850	0.830	0.250	-0.850
Bias	-0.123	0.021	0.183	0.112	0.010	0.014	-0.005	0.015	0.062	0.011	-0.004	0.005
SE	0.845	0.269	0.568	0.318	0.111	0.570	0.142	0.076	0.568	0.146	0.071	0.596
SD	0.371	0.136	0.580	0.395	0.133	0.619	0.163	0.094	0.603	0.164	0.087	0.654
Failure	0.520	0.520	0.520	0.000	0.000	0.000	0.045	0.045	0.045	0.000	0.000	0.000

**Table A2:** Bias of the estimators of  $\sigma_{\xi}^2, \sigma_{\alpha}^2$ , associated standard error (SE) and standard deviation (SD), obtained from the likelihood approach under a Normal specification of  $(\eta_i, \xi_i)^T$  and SIMEX, by distinguishing high accuracy scenario and low accuracy scenario. Values of  $(\eta_i, \xi_i)^T$  are generated from a bivariate Normal distribution. Results are reported for increasing correlation and sample size  $K = 10$ , on the basis of 1000 replicates.

	High accuracy scenario				Low accuracy scenario			
	Likelihood approach		SIMEX approach		Likelihood approach		SIMEX approach	
	$\sigma_{\xi}^2$	$\sigma_{\alpha}^2$	$\sigma_{\xi}^2$	$\sigma_{\alpha}^2$	$\sigma_{\xi}^2$	$\sigma_{\alpha}^2$	$\sigma_{\xi}^2$	$\sigma_{\alpha}^2$
$\rho = 0.222$								
True	0.500	0.600	0.500	0.600	0.500	0.600	0.500	0.600
Bias	-0.131	-0.308	0.113	0.100	-0.079	-0.167	0.008	-0.074
SE	0.195	0.263	0.285	0.334	0.187	0.198	0.230	0.238
SD	0.232	0.237	0.334	0.351	0.236	0.238	0.273	0.260
$\rho = 0.438$								
True	1.900	0.500	1.900	0.500	1.900	0.500	1.900	0.500
Bias	-0.519	-0.238	-0.042	0.118	-0.289	-0.122	-0.014	-0.038
SE	0.745	0.374	0.849	0.297	0.683	0.173	0.851	0.209
SD	0.716	0.231	0.815	0.314	0.801	0.218	0.956	0.242
$\rho = 0.800$								
True	4.000	0.141	4.000	0.141	4.000	0.141	4.000	0.141
Bias	-1.171	-0.062	-0.637	0.193	-0.829	-0.038	-0.265	0.019
SE	1.512	0.420	1.527	0.170	1.316	0.061	1.687	0.074
SD	1.459	0.090	1.477	0.191	1.552	0.073	1.753	0.087

## References

1. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
2. Arends LR, Hamza TH, van Houwelingen JC, Heijnenbroek-Kal MH, Hunink MGM, Stijnen T. Bivariate random effects meta-analysis of ROC curves. *Med Decis Making* 2008;28:621–38.
3. Hamza TH, Arends LR, van Houwelingen HC, Stijnen T. Multivariate random effects meta-analysis of diagnostic tests with multiple thresholds. *BMC Med Res Methodol* 2009;9:73.
4. Ma X, Nie L, Cole SR, Chu H. Statistical methods for multivariate meta-analysis of diagnostic tests: an overview and tutorial. *Stat Methods Med Res* 2016;25:1596–619.
5. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making* 1993;13:313–21.
6. Guolo A, Varin C. Random-effects meta-analysis: the number of studies matters. *Stat Methods Med Res* 2017;26:1500–18.
7. Takwoingi Y, Guo B, Riley RD, Deeks JJ. Performance of methods for meta-analysis of diagnostic test accuracy with few studies or sparse data. *Stat Methods Med Res* 2017;26:1896–911.
8. Chen Y, Liu Y, Ning J, Nie L, Zhu H, Chu H. A composite likelihood method for bivariate meta-analysis in diagnostic systematic reviews. *Stat Methods Med Res* 2017;26:914–30.
9. Nikoloulopoulos AK. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Stat Med* 2015;34:3842–65.
10. Zapf A, Hoyer A, Kramer K, Kuss O. Nonparametric meta-analysis for diagnostic accuracy studies. *Stat Med* 2015;34:3831–41.
11. Guolo A. A double SIMEX approach for bivariate random-effects meta-analysis of diagnostic accuracy studies. *BMC Med Res Methodol* 2017;17:6.
12. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton: Chapman and Hall/CRC Press; 2006.
13. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865–84.



14. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JAC. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2006;8:239–51.
15. Arends LR, Hoes AW, Lubsen J, Grobbee DE, Stijnen T. Baseline risk as predictor of treatment benefit: three clinical meta-re-analyses. *Stat Med* 2000;19:3497–518.
16. Ghidry W, Stijnen T, van Houwelingen HC. Modelling the effect of baseline risk in meta-analysis: a review from the perspective of errors-in-variables regression. *Stat Methods Med Res* 2013;22:307–23.
17. Guolo A. The SIMEX approach to measurement error correction in meta-analysis with baseline risk as covariate. *Stat Med* 2014;33:2062–76.
18. Keogh RH, Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: part 1 - basic theory and simple methods of adjustment. *Stat Med* 2020;39:2197–231.
19. Shaw PA, Gustafson P, Carroll RJ, Deffner V, Dodd KW, Keogh RH, et al. STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: part 2 - more complex methods of adjustment and advanced topics. *Stat Med* 2020;39:2232–63.
20. Buonaccorsi JP. *Measurement error: models, methods and applications*. Boca Raton: Chapman and Hall/CRC Press; 2010.
21. McInosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med* 1996;15:1713–28.
22. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589–624.
23. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. *J Am Stat Assoc* 1994;89:1314–28.
24. Stefanski LA, Cook JR. Simulation-extrapolation: the measurement error jackknife. *J Am Stat Assoc* 1995;90:1247–56.
25. Carroll RJ, Küchenhoff H, Lombard F, Stefanski LA. Asymptotics for the SIMEX estimator in nonlinear measurement error models. *J Am Stat Assoc* 1996;91:242–50.
26. Apanasovich TV, Carroll RJ, Maity A. SIMEX and standard error estimation in semiparametric measurement error models. *Electron J Stat* 2009;3:318–48.
27. Holcomb JP. Regression with covariates and outcome calculated from a common set of variables measured with error: estimation using the SIMEX method. *Stat Med* 1999;18:2847–62.
28. Azzalini A. A class of distributions which includes the normal ones. *Scand J Stat* 1985;12:171–8.
29. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: <https://www.R-project.org/>.
30. Diaz M. Performance measures of the bivariate random effects model for meta-analyses of diagnostic accuracy. *Comput Stat Data Anal* 2015;83:82–90.
31. World Health Organization. *Global tuberculosis report 2021*. Geneva, Switzerland: Licence: CC BY-NC-SA 3.0 IGO; 2021.
32. Gupta RK, Lucas SB, Fielding KL, Lawn SD. Prevalence of tuberculosis in post-mortem studies of HIV-infected adults and children in resource-limited settings: a systematic review and meta-analysis. *AIDS* 2015;29:1987–2002.
33. Sharma SK, Mohan A, Kadiravan T. HIV-TB co-infection: epidemiology, diagnosis & management. *Indian J Med Res* 2005;121:550–67.
34. van Hoving DJ, Griesel R, Meintjes G, Takwoingi Y, Maartens G, Ochodo EA, et al. Abdominal ultrasound for diagnosing abdominal tuberculosis or disseminated tuberculosis with abdominal involvement in HIV-positive individuals. *Cochrane Database Syst Rev* 2019;9:CD012777.
35. Riley RD, Takwoingi Y, Trikalinos T, Guha H, Biswas A, Ensor J, et al. Meta-analysis of test accuracy studies with multiple and missing thresholds: a multivariate-normal model. *J Biometrics Biostat* 2014;5:196.
36. Guolo A, To DK. A pseudo-likelihood approach for multivariate meta-analysis of test accuracy studies with multiple thresholds. *Stat Methods Med Res* 2021;30:204–20.
37. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *J Clin Epidemiol* 2008;61:41–51.
38. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Med Decis Making* 2008;28:639–49.
39. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol* 2006;59:1331–2.