Annals of the Institute of Statistical Mathematics (2022) 74:93–112
https://doi.org/10.1007/s10463-021-00793-4

Empirical likelihood meta-analysis with publication bias correction under Copas-like selection model

Mengke Li[1] · Yukun Liu[1] · Pengfei Li[2] · Jing Qin[3]

Received: 28 May 2020 / Revised: 14 November 2020 / Accepted: 18 January 2021 /
Published online: 27 March 2021

Abstract

Meta-analysis is commonly used to synthesize multiple results from individual studies. However, its validation is usually threatened by publication bias and between-study heterogeneity, which can be captured by the Copas selection model. Existing inference methods under this model are all based on conditional likelihood and may not be fully efficient. In this paper, we propose a full likelihood approach to meta-analysis by integrating the conditional likelihood and a marginal semi-parametric empirical likelihood under a Copas-like selection model. We show that the maximum likelihood estimators (MLE) of all the underlying parameters have a jointly normal limiting distribution, and the full likelihood ratio follows an asymptotic central chi-square distribution. Our simulation results indicate that compared with the conditional likelihood method, the proposed MLEs have smaller mean squared errors and the full likelihood ratio confidence intervals have more accurate coverage probabilities. A real data example is analyzed to show the advantages of the full likelihood method over the conditional likelihood method.

Keywords Copas selection model · Empirical likelihood · Meta-analysis · Publication bias · Trim-and-fill method

The online version of this article contains supplementary material.

Supplementary Information The online version supplementary material available at https://doi.org/10.1007/s10463-021-00793-4.

✉ Yukun Liu
ykliu@sfs.ecnu.edu.cn

[1] KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200241, China

[2] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo N2L 3G1, Ontario, Canada

[3] National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

1 Introduction

Meta-analysis or systematic review is the statistical technique of collecting and synthesizing multiple published scientific results from individual studies. The most important advantage of meta-analysis over a single study is that it usually has higher statistical power and can answer research questions that cannot be answered by a single study (Jackson et al. 2011). Since formally introduced by (Glass 1976) to evaluate the effectiveness of psychological therapies, it has become increasingly important and popular in many fields of research, including medical, social, and biological sciences (Egger et al. 2001; Cooper et al. 2009; Koricheva et al. 2012).

However, as the basis of meta-analysis, published scientific results may not be representative of those from all relevant studies, both published and unpublished, due to the so-called publication bias, which is a well-recognized threat to the validation of the results of a meta-analysis (Rothstein 2008; Jin et al. 2015). Typically, the studies used in the meta-analysis are biased toward those which report statistically significant positive findings. A standard meta-analysis may arrive at a misleading conclusion that is biased toward significance or positivity (Rosenthal 1979) .

To retrieve valid inferences, it is necessary and important to detect and correct for publication bias in meta-analysis. Many approaches have been developed in the literature for this purpose. They can generally be divided into two categories. The first category includes the funnel plot (Light and Pillemer 1984) and related graphical methods (Galbraith 1988; Egger et al. 1997; Sterne et al. 2000, 2001). The funnel plot is a plot of effect estimates from individual studies versus their precisions, an asymmetric plot indicating potential publication bias. The most well-known statistical test based on funnel plot asymmetry is (Duval and Tweedie 2000a, b)'s trim-and-fill test, which is a rank-based data augmentation technique. By formalizing the use of funnel plots, it estimates and adjusts for the numbers and outcomes of missing studies. However, simulation studies have found that the trim-and-fill method may detect "missing" studies in a substantial proportion of meta-analyses, even in the absence of bias. In other words, it would add and adjust for nonexistent studies in response to funnel plot asymmetry arising from nothing more than random variation (Sterne et al. 2001).

The second category includes methods based on parametric models of the selection mechanism. In these methods, parametric distributions are imposed to characterize the underlying publication mechanism by which effect estimates are selected to be observed (Rothstein et al. 2006) . The most popular method is the Copas selection model (Copas and Li 1997; Copa 1999; Copas and Shi 2000, 2001), which is derived from the Heckman two-stage regression model (Copas and Li 1997). Assuming the most extreme studies are missing, the trim-and-fill method often produces excessively conservative inference (Scharzer et al. 2010). By contrast, the Copas model is more flexible, because it not only characterizes the heterogeneity and within-variation of individual effect estimates, but also allows the probability of selection to depend on both the effect estimate and its standard error.

Much attention has been paid to the Copas model in recent years. After analyzing 157 meta-analyses with binary outcomes, (Carpenter et al. 2009) concluded

that the Copas selection model provided a useful summary in 80% of meta-analyses. (Scharzer et al. 2010) demonstrated by empirical evaluation indicated that the Copas selection model is preferable to the trim-and-fill method for selection bias in meta-analysis. (Mavridis et al. 2013) implemented the Bayesian method for model fitting under the Copas model. This method offers great flexibility to incorporate in the model prior information on the extent and strength of selection. An unavoidable difficulty that blocks the wide application of the Copas selection model is the frequent non-convergence in maximizing the likelihood of the observed data. A possible reason for this dilemma is that the data often contains very little information about the underlying parameters (Copas and Shi 2001). To overcome this problem, (Ning et al. 2017) re-casted the biased-sampling problem as a missing data problem and proposed an EM algorithm to calculate the maximum likelihood estimators (MLE).

The main goal of meta-analysis is to estimate the overall effect size θ after adjusting publication bias. As an index of publication bias, the number of unreported studies, N_u, can also be important to researchers (Rosenthal 1979; Fragkos et al. 2017). If one can estimate or determine N_u in some way from the available data, one can then compare this with one's knowledge about the field (Gleser and Olkin 1996). To the best of our knowledge, the existing developments on the Copas selection model are all based on the conditional likelihood which is the conditional joint distribution of data given that they are observed or published. Statistical inferences based on conditional likelihood are generally less efficient than those based on full likelihood when we are estimating θ. Under the Copas selection model, the usual point estimator for $N = N_u + n$, the total number of studies of interest, is the inverse-probability-weighting estimator (Mavridis et al. 2013), whose asymptotical normality is used to construct Wald confidence intervals for N. However such intervals may have poor coverage accuracy, and its lower bound can be even less than the sample size n, which is clearly absurd (Liu et al. 2017, 2018).

In this paper, we formally show that all the underlying parameters are identifiable if publication bias exists or equivalently the parameter ρ is not equal to zero. Assuming publication bias exists, we focus on the estimation of effect size θ and the total number of studies, N. Motivated by the weaknesses of the conditional likelihood methods under the Copas selection model, we propose a full likelihood method for meta-analysis by integrating the conditional likelihood and a marginal semi-parametric likelihood. We make the same model assumptions as (Ning et al. 2017) , use (Owen 1990)'s empirical likelihood (EL) to handle the nonparametric distribution of the standard errors of the individual effect sizes, and finally derive the marginal semi-parametric likelihood. We show that the proposed MLEs of all the underlying parameters have a jointly normal limiting distribution, and the full likelihood ratio follows an asymptotic central chi-square distribution. In particular, we propose to construct confidence intervals for the effect size and the total number of studies and test the existence of publication bias by the corresponding full likelihood ratio tests. Our simulation results indicate that the full maximum likelihood estimators have smaller mean squared errors than the conditional-likelihood-based estimators. Also the full likelihood ratio confidence intervals for the

effect size and the total number of studies have more accurate coverage probabilities than the Wald intervals under the conditional likelihood.

The paper proceeds as follows. In Sect. 2, we introduce the Copas-like model of (Ning et al. 2017), present our semi-parametric full likelihood method, and investigate the large-sample properties of the MLE and the likelihood ratio test. An algorithm to calculate the proposed MLEs is also provided. Section 3 contains simulation results. Section 4 is devoted to two real-life data analyses. We end in Sect. 5 with some discussions. For clarity, all proofs are postponed to the supplementary material.

2 Full likelihood approach and its properties

2.1 Model setup

Let N be the total number of studies of interest, including published and unpublished. For study i, let θ_i^* denote the estimated effect size and s_i^* the estimated standard variance of θ_i^*. We make the same assumptions on data as (Ning et al. 2017). Specifically, suppose that (θ^*, s_i^*) $(1 \le i \le N)$ are independent and identically distributed (IID) and that θ^* and s_i^* are also independent of each other. Given s_i^*, (DerSimonian and Laird 1986) modeled θ_i^* by a random effect model

$$\theta_i^* = \theta + \tau u_i + s_i^* \epsilon_i, \tag{1}$$

where the random effect u_i and the error ϵ_i are independent, and both follow the standard normal distribution. Here θ is the underlying effect size and τ^2 describes the between-study heterogeneity. To characterize the publishing mechanism, (Copas and Li 1997) proposed a separate selection model that uses a latent variable

$$Z_i = \gamma_1 + \gamma_2 / s_i^* + \delta_i, \tag{2}$$

where (ϵ_i, δ_i) are independent and identically distributed from a bivariate standard normal distribution with correlation coefficient ρ. Study i is assumed to be published if $Z_i > 0$. Suppose there are n studies published with estimated effect sizes and estimated standard variances $\{(\theta_1, s_1), \ldots, (\theta_n, s_n)\}$. We wish to estimate the effect size θ, heterogeneity τ, the total number N of studies, the marginal distribution function $F(x)$ of s_i^*, and the marginal distribution $G(x)$ of θ_i^*, after adjusting for publication bias.

Lemma 1 *If $\rho \neq 0$, the parameters $\gamma_1, \gamma_2, \rho, \tau$ and θ are all identifiable.*

Unlike (Ning et al. 2017), we assume that $\rho \neq 0$ in this paper; therefore, the observations (θ_i, s_i)'s constitute a biased sample of the study of interest, or equivalently publication bias exists. Lemma 1 implies that all the parameters are identifiable in this case.

2.2 Full semi-parametric likelihood

Let pr denote the probability density/mass function of a continuous/discrete random variable. We assume that the observations $(\theta_1, s_1), \ldots, (\theta_n, s_n)$ are independent given n, the number of observations. It follows that the full likelihood is

$$\tilde{L} = \text{pr}\{n, (\theta_i, s_i), i = 1, \ldots, n\} = \text{pr}(n) \times \prod_{i=1}^{n} \text{pr}(\theta_i, s_i).$$

Without otherwise statement, we use the same notation to denote both a random element and its realization, whose meanings can be clear from the context. For example, $\text{pr}(z)$ denotes the density function of a random element z at z.

According to its generating process, (θ_i, s_i) has the same distribution as $\{(\theta_i^*, s_i^*)|Z_i > 0\}$. Thus formally the full likelihood can be written as

$$\begin{aligned}\tilde{L} &= \text{pr}(n) \times \prod_{i=1}^{n} \text{pr}(\theta_i^* = \theta_i, s_i^* = s_i | Z_i > 0) \\ &= \text{pr}(n) \times \prod_{i=1}^{n} \frac{\text{pr}(Z_i > 0|\theta_i^* = \theta_i, s_i^* = s_i)\text{pr}(\theta_i^* = \theta_i|s_i^* = s_i)\text{pr}(s_i^* = s_i)}{\text{pr}(Z_i > 0)}.\end{aligned} \tag{3}$$

In contrast, the commonly used conditional likelihood (Copas and Shi 2001) is the conditional density function of θ_i's given n, s_i's, and that they are observed, that is,

$$L_c = \prod_{i=1}^{n} \frac{\text{pr}(Z_i > 0|\theta_i^* = \theta_i, s_i^* = s_i)\text{pr}(\theta_i^* = \theta_i|s_i^* = s_i)}{\text{pr}(Z_i > 0|s_i^* = s_i)}. \tag{4}$$

These two likelihoods have a close relationship:

$$\tilde{L} = \text{pr}(n) \times \prod_{i=1}^{n} \text{pr}(s_i^* = s_i|Z_i > 0) \times L_c.$$

Obviously, $\text{pr}(n) = \binom{N}{n}\alpha^n(1-\alpha)^{N-n}$ with $\alpha = \text{pr}(Z_i > 0)$. Lemma 2 presents the other conditional probabilities and densities in (3).

Lemma 2 *Let $\phi(x)$ and $\Phi(x)$ denote the standard normal density and distribution functions, and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \rho, \tau, \theta)^\top$. We have*

$$\text{pr}(Z_i > 0|\theta_i^* = \theta_i, s_i^* = s_i) = \Phi\{v_i(\boldsymbol{\gamma})\}, \tag{5}$$

$$\text{pr}(\theta_i^* = \theta_i|s_i^* = s_i) = \frac{1}{\sqrt{2\pi(\tau^2 + s_i^2)}} \exp\left\{-\frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)}\right\}, \tag{6}$$

$$\text{pr}(Z_i > 0|s_i^* = s_i) = \Phi(\gamma_1 + \gamma_2/s_i), \tag{7}$$

where

$$v_i(\boldsymbol{\gamma}) = \frac{\gamma_1 + (\gamma_2/s_i) + \rho s_i(\theta_i - \theta)/(\tau^2 + s_i^2)}{\sqrt{1 - \rho^2 s_i^2/(\tau^2 + s_i^2)}}. \tag{8}$$

With the formulae in (5–7), the conditional log-likelihood, i.e., $\log(L_c)$, becomes

$$\ell_c(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \left\{ \log \Phi\{v_i(\boldsymbol{\gamma})\} - \frac{1}{2}\log(\tau^2 + s_i^2) - \frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)} - \log \Phi(\gamma_1 + \gamma_2/s_i) \right\}.$$

These formulae also imply that the full log-likelihood is

$$\tilde{\ell} = \log \binom{N}{n} + (N-n)\log(1-\alpha) + \sum_{i=1}^{n} \left[\log\{\Phi(v_i(\boldsymbol{\gamma}))\} - \frac{1}{2}\log(\tau^2 + s_i^2) - \frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)} \right] + \sum_{i=1}^{n} \log\{\mathrm{pr}(s_i^* = s_i)\}.$$

We use (Owen 1988, 1990)'s EL method to handle the distribution function of s_i^*. Let $p_i = \mathrm{pr}(s_i^* = s_i)$. Since $\mathrm{pr}(Z_i > 0 | s_i^* = s_i) = \Phi(\gamma_1 + \gamma_2/s_i)$, we have

$$\alpha = \mathrm{pr}(Z_i > 0) = \mathbb{E}\{\mathrm{pr}(Z_i > 0 | s_i^* = s_i)\} = \int \Phi(\gamma_1 + \gamma_2/s)\mathrm{pr}(s_i^* = s)ds.$$

Hence the feasible p_i's satisfy

$$p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i \Phi(\gamma_1 + \gamma_2/s_i) = \alpha. \tag{9}$$

With p_i in place of $\mathrm{pr}(s_i^* = s_i)$, the maximizer of $\tilde{\ell}$ with respect to p_i's under the constraints in (9) is $p_i = n^{-1}[1 + \lambda\{\Phi(\gamma_1 + \gamma_2/s_i) - \alpha\}]^{-1}$, where λ is the solution to

$$\sum_{i=1}^{n} \frac{\Phi(\gamma_1 + \gamma_2/s_i) - \alpha}{1 + \lambda\{\Phi(\gamma_1 + \gamma_2/s_i) - \alpha\}} = 0. \tag{10}$$

Accordingly, we have the profile log EL (up to a constant)

$$\ell(N, \alpha, \boldsymbol{\gamma}) = \log \binom{N}{n} + (N-n)\log(1-\alpha) + \sum_{i=1}^{n} \left[\log\{\Phi(v_i(\boldsymbol{\gamma}))\} - \frac{1}{2}\log(\tau^2 + s_i^2) - \frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)} \right] - \sum_{i=1}^{n} \log[1 + \lambda\{\Phi(\gamma_1 + \gamma_2/s_i) - \alpha\}].$$

2.3 Estimation and asymptotics

We propose to estimate $(N, \alpha, \boldsymbol{\gamma})$ by the MLEs $(\widehat{N}, \widehat{\alpha}, \widehat{\boldsymbol{\gamma}}) = \arg\max \ell(N, \alpha, \boldsymbol{\gamma})$. Define the likelihood ratio function to be $R(N, \alpha, \boldsymbol{\gamma}) = 2\{\ell(\widehat{N}, \widehat{\alpha}, \widehat{\boldsymbol{\gamma}}) - \ell(N, \alpha, \boldsymbol{\gamma})\}$. This section investigates the asymptotical properties of the MLEs and the likelihood ratio test statistic.

For ease of presentation, we use $\boldsymbol{\gamma}_{12}$ and $\boldsymbol{\gamma}_{45}$ to denote $(\gamma_1, \gamma_2)^\top$ and $(\tau, \theta)^\top$, respectively, and define

$$
\begin{aligned}
f_1(\theta_i, s_i; \boldsymbol{\gamma}) =& \mathrm{pr}(Z_i > 0 | \theta_i^* = \theta_i, s_i^* = s_i) = \boldsymbol{\Phi}\{v_i(\boldsymbol{\gamma})\}, \\
f_2(\theta_i, s_i; \boldsymbol{\gamma}_{45}) =& \mathrm{pr}(\theta_i^* = \theta_i | s_i^* = s_i) = \{2\pi(\tau^2 + s_i^2)\}^{-\frac{1}{2}} \exp\left\{ -\frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)} \right\}, \\
f_3(s_i; \boldsymbol{\gamma}_{12}) =& \mathrm{pr}(Z_i > 0 | s_i^* = s_i) = \boldsymbol{\Phi}(\gamma_1 + \gamma_2 / s_i).
\end{aligned}
$$

Let $(N_0, \alpha_0, \gamma_0)$ be the truth of $(N, \alpha, \boldsymbol{\gamma})$ with $\boldsymbol{\gamma}_0 = (\gamma_{10}, \gamma_{20}, \rho_0, \tau_0, \theta_0)^\top$. Throughout the paper, we use $\boldsymbol{\gamma}_{12,0}$ and $\boldsymbol{\gamma}_{45,0}$ to denote the truths of $\boldsymbol{\gamma}_{12,0}$ and $\boldsymbol{\gamma}_{45,0}$, respectively. Define $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\top$ for a matrix or vector $\mathbf{A}$, and $\mathbf{A}^{\oplus 2} = \mathbf{A} + \mathbf{A}^\top$ for a square matrix $\mathbf{A}$, $\mathbf{E}_{12} = (\mathbf{I}_2, \mathbf{0}_{2\times 3})^\top$, and $\mathbf{E}_{45} = (\mathbf{0}_{2\times 3}, \mathbf{I}_2)^\top$ with $\mathbf{I}_k$ the $k \times k$ identity matrix. We use $\nabla_{\boldsymbol{\gamma}}$ to denote the differentiation operator with respect to $\boldsymbol{\gamma}$. Let $\varphi_1 = \mathbb{E}[\{f_3(s_i^*; \boldsymbol{\gamma}_{12,0})\}^{-1}]$, $\boldsymbol{\varphi}_2 = \mathbb{E}\{\nabla_{\boldsymbol{\gamma}_{12}} \log f_3(s_i^*; \boldsymbol{\gamma}_{12,0})\}$, $\mathbf{F}_1 = (\mathbf{I}_4, \mathbf{0}_{4\times 3})$ and $\mathbf{F}_2 = (\mathbf{0}_{5\times 2}, \mathbf{I}_5)$. Define

$$
\boldsymbol{\Omega} = \mathbf{F}_2^\top \mathbf{V}_c \mathbf{F}_2 + \mathbf{F}_1^\top \tilde{\mathbf{V}}_m \mathbf{F}_1, \tag{11}
$$

where

$$
\begin{aligned}
\mathbf{V}_c =& \mathbb{E}\frac{\{\nabla_{\boldsymbol{\gamma}} f_1(\theta_i^*, s_i^*; \boldsymbol{\gamma}_0)\}^{\otimes 2}}{f_1(\theta_i^*, s_i^*; \boldsymbol{\gamma}_0)} + \mathbb{E}\left[\mathbf{E}_{45} \int \frac{\{\nabla_{\boldsymbol{\gamma}_{45}} f_2(t, s_i^*; \boldsymbol{\gamma}_{45,0})\}^{\otimes 2}}{f_2(t, s_i^*; \boldsymbol{\gamma}_{45,0})} f_1(t, s_i^*; \boldsymbol{\gamma}_0) dt \mathbf{E}_{45}^\top\right] \\
&+ \mathbb{E}\left[\int \nabla_{\boldsymbol{\gamma}} f_1(t, s_i^*; \boldsymbol{\gamma}_0) \nabla_{\boldsymbol{\gamma}_{45}^\top} f_2(t, s_i^*; \boldsymbol{\gamma}_{45,0}) dt \mathbf{E}_{45}^\top\right]^{\oplus 2} - \mathbb{E}\left[\mathbf{E}_{12} \frac{\{\nabla_{\boldsymbol{\gamma}_{12}} f_3(s_i^*; \boldsymbol{\gamma}_{12,0})\}^{\otimes 2}}{f_3(s_i^*; \boldsymbol{\gamma}_{12,0})} \mathbf{E}_{12}^\top\right]
\end{aligned}
$$

and

$$
\tilde{\mathbf{V}}_m = \begin{pmatrix}
\frac{\alpha_0}{1-\alpha_0} & \frac{1}{1-\alpha_0} & \mathbf{0} \\
\frac{1}{1-\alpha_0} & \frac{1-\varphi_1}{(1-\alpha_0)(1-\alpha_0\varphi_1)} & \frac{\boldsymbol{\varphi}_2^\top}{1-\alpha_0\varphi_1} \\
\mathbf{0} & \frac{\boldsymbol{\varphi}_2}{1-\alpha_0\varphi_1} & -\frac{\alpha_0 \boldsymbol{\varphi}_2^{\otimes 2}}{1-\alpha_0\varphi_1}
\end{pmatrix}.
$$

Theorem 1 *Assume Conditions C*1 *and C*2 *in the supplementary materials*, $\rho_0 \neq 0$, *and that the matrix* $\boldsymbol{\Omega}$ *defined in* (11) *is positive definite. As* $N_0 \to \infty$, *the following results hold.*

(1) $N_0^{1/2}(\widehat{N}/N_0 - 1,\ \widehat{\alpha} - \alpha_0, (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)^\top) \xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Omega}^{-1})$, *where* $\xrightarrow{d}$ *stands for convergence in distribution.*

(2) $N_0^{1/2}(\widehat{N}/N_0 - 1) \xrightarrow{d} N(0,\ \sigma^2)$, *and* $N_0^{1/2}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(\mathbf{0},\ \mathbf{V}^{-1})$, *where* σ^2 *is the* $(1, 1)$ *element of* $\boldsymbol{\Omega}^{-1}$ *and* $\mathbf{V}^{-1}$ *is the down-right* 5×5 *submatrix of* $\boldsymbol{\Omega}^{-1}$.

(3) *The likelihood ratio* $R(N_0, \alpha_0, \boldsymbol{\gamma}_0) = 2\{\ell(\widehat{N}, \widehat{\alpha}, \widehat{\boldsymbol{\gamma}}) - \ell(N_0, \alpha_0, \boldsymbol{\gamma}_0)\} \xrightarrow{d} \chi_7^2$.

The proof of result (3) of Theorem 1 (See the supplementary material) implies that the likelihood ratio statistic of testing any subvector of $(N, \alpha, \boldsymbol{\gamma})$ also follows an asymptotic central chi-square distribution. This result can be used to construct likelihood ratio confidence intervals for any of the parameters θ, τ, N, ρ and α with asymptotically correct coverage probabilities.

The proposed full likelihood method can conveniently provide consistent estimators for the marginal distributions of s_i^* and θ_i^*. Given $\widehat{\alpha}$ and $\widehat{\gamma}$, the MLE of the distribution function $F(x)$ of s_i^* is

$$\widehat{F}(s) = \sum_{i=1}^{n} \widehat{p}_i I(s_i \le s) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{1 + \widehat{\lambda}\{\boldsymbol{\Phi}(\widehat{\gamma}_1 + \widehat{\gamma}_2/s_i) - \widehat{\alpha}\}} I(s_i \le s),$$

where $\widehat{\lambda}$ is the solution to Eq. (10) with $(\gamma_1, \gamma_2, \alpha)$ replaced by $(\widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\alpha})$. To estimate the marginal distribution $G(t)$ of θ_i^*, we rewrite $G(t)$ as

$$\begin{aligned} G(t) &= \int_{-\infty}^{t} \int \mathrm{pr}(\theta_i^* = r | s_i^* = s) dF(s) dr \\ &= \int_{-\infty}^{t} \int \frac{1}{\sqrt{\tau^2 + s_i^2}} \phi\left(\frac{r - \theta}{\sqrt{\tau^2 + s^2}} \right) dF(s) dr \\ &= \int \boldsymbol{\Phi}\left(\frac{t - \theta}{\sqrt{\tau^2 + s^2}} \right) dF(s), \end{aligned}$$

where the second equality follows from Eq. (6). Based on the MLE $\widehat{F}(x)$ of $F(x)$, we immediately obtain the MLE of $G(t)$,

$$\widehat{G}(t) = \int \boldsymbol{\Phi}\left(\frac{t - \widehat{\theta}}{\sqrt{\widehat{\tau}^2 + s^2}} \right) d\widehat{F}(s) = \sum_{i=1}^{n} \widehat{p}_i \boldsymbol{\Phi}\left(\frac{t - \widehat{\theta}}{\sqrt{\widehat{\tau}^2 + s_i^2}} \right).$$

By Theorem 1, $\widehat{\boldsymbol{\gamma}}$ and $\widehat{\alpha}$ are consistent; therefore, $\widehat{F}$ and $\widehat{G}$ are also consistent estimators of F and G, respectively.

2.4 Comparison with conditional likelihood

In the literature, the conditional likelihood $\ell_c(\boldsymbol{\gamma})$ is usually used to estimate the underlying parameters for the Copas-like model. Let $\tilde{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \ell_c(\boldsymbol{\gamma})$ be the conditional

MLE of $\boldsymbol{\gamma}$. Then N can be estimate by the inverse probability weighting estimator or the MLE

$$\tilde{N} = \sum_{i=1}^{n} \frac{1}{f_3(s_i;\tilde{\boldsymbol{\gamma}}_{12})} = \sum_{i=1}^{n} \frac{1}{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_i)}.$$

Theorem 2 *Assume Condition C1 in the supplementary materials,* $\rho_0 \neq 0$, *and that* $\mathbf{V}_c$ *is positive definite. Then as* $N_0 \to \infty$, *(i)* $N_0^{1/2}(\tilde{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}_c^{-1})$, *and (ii)* $N_0^{1/2}(\tilde{N}/N_0 - 1) \xrightarrow{d} N(0, \sigma_c^2)$, *where* $\sigma_c^2 = \varphi_1 - 1 + \boldsymbol{\varphi}_2^{\top}\mathbf{E}_{12}\mathbf{V}_c^{-1}\mathbf{E}_{12}^{\top}\boldsymbol{\varphi}_2$.

We may wonder whether the proposed MLEs have efficiency gain over the conditional MLEs in terms of asymptotical variance. Unfortunately the answer is negative.

Proposition 1 *With the symbols used in Theorems 1 and 2,* $\sigma^2 = \sigma_c^2$ *and* $\mathbf{V} = \mathbf{V}_c$.

This proposition indicates that to estimate N and $\boldsymbol{\gamma}$, the conditional MLEs and proposed MLEs have the same asymptotic normal distribution. The proposed full likelihood estimation procedure has no efficiency improvement over the conditional likelihood estimation procedure. Even so, the proposed full likelihood method still has several advantages over the conditional likelihood method. First, although the resulting point estimators are asymptotical equivalent, the interval estimators based on these two methods generally have quite different finite-sample performances. Our simulation results indicate that the proposed likelihood ratio interval usually has better coverage accuracy than the conditional-likelihood-based Wald interval. Second, the proposed likelihood ratio interval is free from variance estimation, which however is inevitable for the conditional-likelihood-based Wald interval. Third, the Wald interval estimators may have so small lower bounds that are even less than the number of studies observed in the meta-analysis, which is clearly unreasonable. By contrast, the proposed likelihood ratio interval estimators never suffer from such an embarrassment. Finally, under the conditional likelihood method, similar to $\tilde{N}$, $F(x)$ and $G(t)$ are also estimated by their inverse probability weighting estimators

$$\tilde{F}(s) = \frac{1}{\tilde{N}} \sum_{i=1}^{n} \{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_i)\}^{-1} I(s_i \leq s) = \frac{\sum_{i=1}^{n} \{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_i)\}^{-1} I(s_i \leq s)}{\sum_{j=1}^{n} \{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_j)\}^{-1}},$$

$$\tilde{G}(t) = \frac{\sum_{i=1}^{n} \{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_i)\}^{-1} \boldsymbol{\Phi}\{(t - \tilde{\theta})/\sqrt{\tilde{\tau}^2 + s_i^2}\}}{\sum_{j=1}^{n} \{\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_j)\}^{-1}}.$$

Because of inverse probability weighting, it is well known that the numerical performance of inverse probability weighting estimators can be quite unstable when some of $\boldsymbol{\Phi}(\tilde{\gamma}_1 + \tilde{\gamma}_2/s_i)$ are close to 0. The use of EL in our full likelihood estimator considerably mitigates this embarrassment. The maximization of the empirical likelihood $\prod_{i=1}^{n} p_i$ greatly prevents the occurrence of extreme small weights and thus leads to more stable numerical performances of $\widehat{F}(s)$. This numerical advantage of

the proposed MLE over the inverse probability weighting estimator was also noticed by Han (2014) .

2.5 Calculation of MLEs

Maximizing the full likelihood $\ell(N, \alpha, \boldsymbol{\gamma})$ is computationally challenging because the function $\ell(N, \alpha, \boldsymbol{\gamma})$ takes maximum over a very flat plateau and its maximization often produces non-convergence results. If $\boldsymbol{\gamma}_{12}$ is fixed, then the maximization can be stably obtained. This phenomenon has been acknowledged by (Copas and Shi 2001) and also observed by (Ning et al. 2017) in maximizing the conditional likelihood ℓ_c. To overcome this problem, (Ning et al. 2017) proposed an expectation-maximization algorithm after recasting the bias sampling issue as a missing data problem. Their simulation studies indicate that the expectation-maximization algorithm usually produces stable estimates for $\boldsymbol{\gamma}$. Let $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_{12}, \tilde{\rho}, \tilde{\tau}, \tilde{\theta})$ be the conditional estimate of $\boldsymbol{\gamma}$ calculated by (Ning et al. 2017) 's expectation-maximization algorithm. Since direct maximization with respect to $\boldsymbol{\gamma}_{12}$ is very unstable, we propose to maximize our full log-likelihood $\ell(N, \alpha, \boldsymbol{\gamma})$ by fixing $\boldsymbol{\gamma}_{12} = \tilde{\boldsymbol{\gamma}}_{12}$.

For ease of exposition, we first re-express the profile empirical log-likelihood as

$$\ell(N, \alpha, \boldsymbol{\gamma}) = h_1(N, \alpha) + h_2(\boldsymbol{\gamma}) + \min_{\lambda} h_3(\boldsymbol{\gamma}_{12}, \alpha, \lambda),$$

where

$$\begin{aligned} h_1(N, \alpha) &= \log \binom{N}{n} + (N - n) \log(1 - \alpha), \\ h_2(\boldsymbol{\gamma}) &= \sum_{i=1}^{n} \left[\log\{\Phi(v_i)\} - \frac{1}{2} \log(\tau^2 + s_i^2) - \frac{(\theta_i - \theta)^2}{2(\tau^2 + s_i^2)} \right], \\ h_3(\alpha, \boldsymbol{\gamma}_{12}, \lambda) &= -\sum_{i=1}^{n} \log[1 + \lambda\{\Phi(\gamma_1 + \gamma_2 / s_i) - \alpha\}]. \end{aligned}$$

To avoid the non-definition problem in EL, we adopt (Owen 1990)'s calculation strategy and replace the $\log(\cdot)$ function in h_3 by

$$\log_*(z) = \begin{cases} \log(z) & z > 1/c_n \\ -\log(c_n) - 1.5 + 2zc_n - 0.5z^2c_n^2 & z \le 1/c_n \end{cases}, \tag{12}$$

where c_n is a pre-specified large number and is usually chosen to be n.

We propose to maximize $\ell(N, \alpha, \boldsymbol{\gamma})$ via the following algorithm:

Step 1 Calculate $\tilde{h}_1(\alpha) = \max_N h_1(N, \alpha)$, and $\tilde{h}_3(\alpha, \boldsymbol{\gamma}_{12}) = \min_\lambda h_3(\alpha, \boldsymbol{\gamma}_{12}, \lambda)$.
Step 2 Let $h_{23}(\alpha, \boldsymbol{\gamma}) = h_2(\boldsymbol{\gamma}) + \tilde{h}_3(\alpha, \boldsymbol{\gamma}_{12})$. Calculate $\tilde{h}_{23}(\alpha) = \max_{\boldsymbol{\gamma}} h_{23}(\alpha, \boldsymbol{\gamma})$.
Step 3 Let $h_{123}(\alpha) = \tilde{h}_1(\alpha) + \tilde{h}_{23}(\alpha)$. Calculate $\max_\alpha h_{123}(\alpha)$ and the maximizer $\hat{\alpha}$.
Step 4 Calculate $\widehat{N} = \arg\max_N \ h_1(N, \hat{\alpha})$ and $\widehat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} h_{23}(\boldsymbol{\gamma}, \hat{\alpha})$.

As $\boldsymbol{\gamma}$ is a 5-variate vector, we implement the optimizations with respect to $\boldsymbol{\gamma}$ by the R command `nlminb`. All the other optimizations in the above algorithm are with respect to a scalar variable and can be quickly solved by the R command `optimize`.

3 Simulations

3.1 Simulation settings

We carry out simulations to investigate the finite-sample performance of the proposed full likelihood method (Full Likelihood or FL) and compare it with the conditional likelihood method (Conditional Likelihood or CL) implemented by (Ning et al. 2017) 's expectation-maximization algorithm. Consistent estimators are needed for σ_c^2 and $\mathbf{V}_c$ when we apply the conditional likelihood method to construct Wald-type intervals for N and $\boldsymbol{\gamma}$. Following (Ning et al. 2017) , we adopt the estimation procedure by (Loui 1982) together with (Ning et al. 2017) 's expectation-maximization algorithm to obtain consistent variance estimators for $\tilde{N}$ and $\tilde{\boldsymbol{\gamma}}$.

We generate study-specific variance s_i^{*2} from the square of a normal random variable $N(0.25, 0.5)$ and generate (ϵ_i, δ_i)'s from a bivariate standard normal distribution with correlation ρ_0. Given N_0 and $\boldsymbol{\gamma}_0$, we calculate θ_i^* and Z_i from models (1) and (2) with $\boldsymbol{\gamma}_0$ in place of $\boldsymbol{\gamma}$. The (θ_i^*, s_i^*)'s with $Z_i > 0$ constitute a simulated sample. We consider two choices of N_0, 50 and 100, and two choices of $\boldsymbol{\gamma}_{12,0}$, $(-0.6, 0.8)$ and $(-1, 0.6)$. As $\boldsymbol{\gamma}_{12,0}$ changes from $(-0.6, 0.8)$ to $(-1, 0.6)$, the publishing rate decreases from 80% to 64%, publication bias getting more and more severe. To examine the effects of effect size, publication bias and heterogeneity on the performances of the full likelihood and conditional likelihood methods, we consider two scenarios for $(\theta_0, \tau_0, \rho_0)$: (1) $\theta_0 = 0.4$, $\tau_0 = 0.5$, $\rho_0 = 0.2, 0.8$, and (2) $\theta_0 = 0.2$, $\rho_0 = 0.2$, $\tau_0 = 0.5, 1$.

For each parameter combination, we generate 1000 simulation samples and calculate the full likelihood and conditional likelihood point estimates for N, θ, and τ based on each sample. The simulated bias (BIAS), standard deviation (SD) and root mean square error (RMSE) of these two type estimators are then obtained. We also calculate the simulated coverage probabilities of the proposed likelihood ratio confidence intervals and the conditional-likelihood-based Wald confidence intervals for the three parameters at the 95% confidence level. These simulation results are given in Tables 1 and 2, corresponding to scenarios (1) and (2), respectively.

3.2 Simulation results

We first examine the results on point estimation. The full likelihood estimators of all the four parameters θ, τ, ρ and N have obviously smaller RMSEs than the conditional likelihood estimators in almost all scenarios (58 out of 64). In the rest 8 scenarios, although the full likelihood estimators do not win, their performances are nearly the same as the conditional likelihood estimators in terms of RMSE. These

Table 1 Simulation results for scenario (1) with $\tau_0 = 0.5$ and different choices of ρ. All numbers have been multiplied by 100

$\gamma_{12,0}$	ρ_0	N_0	PAR	Conditional likelihood				Full likelihood			
				BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP
(−0.6, 0.8)	0.2	50	θ	− 3.44	11.93	12.42	95.3	0.22	11.58	11.59	94.3
			τ	− 1.81	14.23	14.34	89.9	− 0.96	8.88	8.93	94.0
			ρ	21.09	47.95	52.38	100	− 2.02	52.49	52.53	92.0
			N	23.11	428.47	429.09	89.8	− 44.31	428.66	430.95	93.4
		100	θ	− 3.39	7.85	8.55	95.3	− 0.07	8.12	8.12	95.1
			τ	− 2.36	16.02	16.19	90.3	− 0.91	6.14	6.20	94.5
			ρ	20.51	37.87	43.06	100	− 0.85	36.25	36.26	93.6
			N	28.56	609.81	610.48	90.0	− 38.71	609.98	611.21	94.3
	0.8	50	θ	3.08	13.64	13.98	93.5	− 0.35	11.69	11.70	92.4
			τ	− 3.17	13.85	14.21	96.3	− 2.32	8.34	8.65	94.2
			ρ	− 15.37	57.47	59.49	100	− 8.49	49.56	50.37	96.3
			N	207.55	438.85	485.45	87.1	138.78	439.21	460.62	95.6
		100	θ	2.01	8.93	9.15	95.1	− 0.48	8.99	9.01	93.4
			τ	− 1.35	9.39	9.49	95.3	− 1.50	6.31	6.48	95.9
			ρ	− 12.00	40.66	42.40	100	− 4.34	35.83	36.09	94.8
			N	464.97	650.34	799.46	91.3	396.54	650.59	761.92	94.0
(−1, 0.6)	0.2	50	θ	− 4.69	14.64	15.36	93.8	− 0.004	13.38	13.37	92.6
			τ	− 1.02	13.79	13.82	92.3	− 2.40	9.33	9.64	94.1
			ρ	13.91	49.36	51.28	100	2.11	42.98	43.03	93.5
			N	99.71	769.12	775.55	89.9	18.24	770.62	770.84	92.8
		100	θ	− 4.17	10.56	11.35	92.5	− 0.18	9.12	9.12	94.3
			τ	0.15	11.02	11.02	91.7	0.89	6.59	6.65	93.5
			ρ	14.13	38.18	40.71	99.9	1.01	30.35	30.37	93.3
			N	255.43	1003.34	1035.35	90.1	174.44	1004.56	1019.59	94.8
	0.8	50	θ	− 16.75	15.38	22.74	92.0	− 0.98	13.77	13.81	91.5
			τ	− 2.53	10.76	11.05	94.3	− 1.99	9.29	9.50	94.0
			ρ	38.75	82.42	91.08	100	23.84	34.84	42.22	96.7
			N	308.42	830.88	886.28	88.1	229.04	831.31	862.29	91.1
		100	θ	− 8.75	11.80	14.69	92.3	− 1.25	8.85	8.94	92.7
			τ	− 14.32	8.07	16.45	94.7	− 0.82	6.19	6.24	94.4
			ρ	− 3.94	68.64	68.76	100	− 2.72	28.50	28.64	95.2
			N	549.06	1248.90	1364.27	92.7	469.78	1249.01	1334.43	93.3

PAR parameter, *SD* standard deviation, *BIAS* bias, *RMSE* root mean square error, *CP* coverage probability at 95% confidence level

observations indicate that the proposed full likelihood method has clear advantages over the traditional conditional likelihood method. Meanwhile the efficiency gain of the full likelihood estimators increases as the effect size θ increases (from 0.2 to 0.4) or the publication bias becomes more severe (as ρ increases from 0.2 to 0.8). When estimating θ, N and ρ, the full likelihood estimators usually have smaller absolute

Table 2 Simulation results for scenario (2) with $\rho_0 = 0.2$ and different choices of τ. All numbers have been multiplied by 100

$\gamma_{12,0}$	τ_0	N_0		Conditional likelihood				Full likelihood			
			PAR	BIAS	SD	RMSE	CP	BIAS	SD	RMSE	CP
$(-0.6, 0.8)$	0.5	50	θ	– 3.24	11.31	11.76	95.3	0.006	11.77	11.77	93.7
			τ	1.96	11.88	12.04	93.4	– 2.02	8.61	8.85	95.4
			ρ	18.34	52.73	55.85	100	– 2.22	50.08	50.13	92.8
			N	161.76	442.90	471.52	87.2	93.72	443.32	453.12	94.3
		100	θ	– 3.87	7.81	8.71	94.7	0.29	8.06	8.06	94.3
			τ	0.27	8.82	8.82	91.8	– 0.39	6.23	6.25	94.8
			ρ	18.54	30.77	35.93	100	– 2.51	35.50	35.59	92.6
			N	284.71	596.57	661.03	87.3	217.02	596.78	635.01	95.0
	1	50	θ	– 10.33	21.92	24.23	93.4	0.59	20.69	20.70	94.5
			τ	0.26	17.55	17.55	92.2	2.39	14.00	14.20	93.9
			ρ	37.32	91.12	98.46	100	– 2.39	61.37	61.42	93.8
			N	153.37	449.37	474.83	87.4	84.35	449.61	457.45	94.9
		100	θ	– 2.05	12.59	12.76	97.7	0.006	10.44	10.44	94.6
			τ	– 0.43	10.11	10.10	94.7	– 0.39	9.88	9.85	94.8
			ρ	9.05	52.60	53.37	100	– 0.91	48.30	48.30	93.6
			N	134.07	590.23	605.26	90.3	66.17	590.48	594.18	94.8
$(-1, 0.6)$	0.5	50	θ	– 4.52	14.65	15.33	91.7	– 0.42	12.87	12.88	93.7
			τ	– 1.79	18.14	18.23	90.5	– 2.19	9.40	9.65	93.2
			ρ	16.47	61.13	63.31	99.9	2.80	43.07	43.16	92.4
			N	33.16	703.52	704.30	86.7	– 47.01	704.76	706.31	93.3
		100	θ	– 5.66	12.09	13.35	92.6	1.07	8.65	8.72	95.2
			τ	1.54	11.32	11.42	91.9	– 0.96	6.11	6.18	95.5
			ρ	18.41	48.78	52.14	99.7	2.42	35.34	35.43	91.9
			N	292.59	1027.42	1068.27	89.7	210.83	1028.20	1049.59	95.0
	1	50	θ	– 5.72	24.20	24.87	95.0	1.91	22.33	22.41	94.0
			τ	– 0.58	18.19	18.20	93.1	– 2.37	14.80	14.91	94.6
			ρ	15.83	74.37	76.04	100	– 3.42	55.56	55.66	93.9
			N	46.74	685.31	686.90	89.6	– 33.69	686.24	687.01	94.7
		100	θ	– 7.08	15.94	17.45	94.9	– 0.28	16.28	16.28	95.0
			τ	0.08	12.89	12.89	92.7	– 1.41	10.36	10.46	95.3
			ρ	15.54	48.86	51.27	100	– 1.49	40.88	40.91	93.3
			N	113.87	1041.61	1047.81	91.8	33.62	1042.29	1042.83	93.7

PAR parameter, *SD* standard deviation, *BIAS* bias, *RMSE* root mean square error, *CP* coverage probability at 95% confidence level

bias although their standard deviations are very close to those of the conditional likelihood estimators. In particular for ρ, the full likelihood estimator corrects most part of the conditional likelihood estimator's bias. While when estimating τ, the proposed full likelihood estimator usually has smaller standard deviation, although the full likelihood and conditional likelihood estimators have very close biases. In all

cases, as N_0 increases from 50 to 100, the values of BIAS, SD and RMSE decrease as expected when the parameters are θ, τ and ρ. When N is of interest, in almost all cases, the relative BIAS, relative SD and relative RMSEs of both estimators decrease as N_0 increases.

For interval estimation, the proposed likelihood ratio interval or the full likelihood interval for the four parameters almost always has more accurate and more reliable coverage probabilities than the conditional-likelihood-based Wald interval or the conditional likelihood interval. For example, in the cases of $\boldsymbol{\gamma}_{12,0} = (-0.6, 0.8)^{\top}$, $\theta = 0.4$, and $\rho_0 = 0.8$, the coverage probabilities of the conditional likelihood interval for N are only 87.1% and 91.3% when the true value of N is 50 and 100, respectively. By contrast, the corresponding numbers of the full likelihood interval are 95.6% and 94.0%, respectively, which are much more desirable. The advantage of the full likelihood interval is more evident in the estimation of ρ. The coverage probabilities of the full likelihood intervals varies from 92.1% to 95.3%, which are reasonable. However, the coverage probabilities of the conditional likelihood interval for ρ are almost always 100%, indicating that this interval is too wide to be practically useful. In addition, the full likelihood interval exhibits more robust performance than the conditional likelihood interval in terms of coverage probability as the simulation setting varies.

3.3 Comparison in QQ-plots

To get more insights about the better performance of the full likelihood intervals over the conditional likelihood intervals, we consider the following four hypothesis testing problems: $H_{01} : \theta = \theta_0$, $H_{02} : \tau = \tau_0$, $H_{30} : \rho = \rho_0$, and $H_{04} : N = N_0$. We generate 1000 samples from the Copas-like model with $\theta_0 = 0.2, \tau_0 = 1, \rho_0 = 0.2$, $\boldsymbol{\gamma}_{12,0} = (-0.6, 0.8)$ and $N_0 = 50$. For each of the four hypothesis testing problems, 1000 likelihood ratio statistics and Wald test statistics were calculated. The qq-plots of the sign-roots of the 1000 likelihood ratio statistics and Wald test statistics versus the standard normal quantiles are displayed in Fig. 1.

Clearly, the qq-plots of the sign-roots of the likelihood ratio statistics are all quite close to the identity line for all the four hypothesis testing problems. This implies that N(0, 1) and χ^2_1 are desirable approximates to the finite-sample distributions of the sign-root and itself of the likelihood ratio statistic. It also explains why the full likelihood intervals for the four parameters always have very nice coverage accuracy. By contrast, the qq-plots of the Wald test statistic are not that close to the identity line. The departure becomes larger and larger from H_{01} to H_{02} to H_{04}, which explains the poorer and poorer coverage probabilities (from 93.4% to 92.2% to 87.4%) of the conditional likelihood interval. For H_{03}, the Wald statistic has much larger lower quantiles and much smaller upper quantiles compared with the standard normal. This makes the coverage probability of the resulting conditional likelihood interval unacceptably large since its construction is based on the Wald statistic calibrated by the standard normal.

In summary, the limiting χ^2 distribution always approximates much better to the finite-sample distribution of the likelihood ratio statistic than the standard normal to

Table 3 Meta-analysis results of the lung cancer data and the premature birth data

	Conditional likelihood		Full likelihood	
	Est	CI	Est	CI
Premature birth data				
θ	– 0.476	[– 0.662, – 0.289]	– 0.476	[– 0.760, – 0.244]
exp(θ)	0.621	[0.516,0.748]	0.621	[0.468,0.784]
τ	0	[– 0.240, – 0.240]	0	[– 0.201, 0.484]
ρ	– 0.977	[– 2.543, – 0.586]	– 0.837	[– 1.000, – 0.025]
N	23	[7, 39]	16	[13, 20]

Est estimate value, *CI*, 95% confidence intervals, θlog-odds ratio, τ between-study heterogeneity

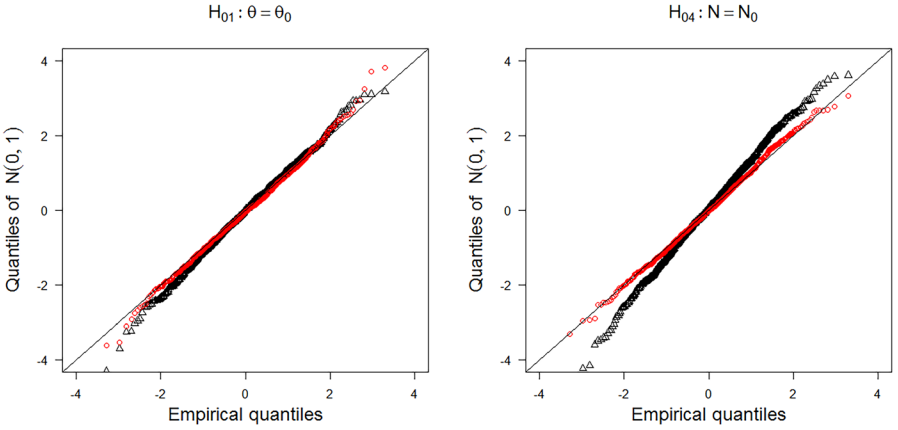

Fig. 1 QQ-plots of the Wald statistic (triangle) and the sign-root of the proposed likelihood ratio statistic (circle) versus N(0, 1). The test statistics are calculated based on 1000 samples from the Copas-like selection model with $N_0 = 50$, $\theta_0 = 0.2$, $\tau_0 = 1$, $\boldsymbol{\gamma}_{12,0} = (-0.6, 0.8)$, and $\rho_0 = 0.2$

that of the Wald statistic. This makes the resulting full likelihood intervals always have more, sometimes far more, accurate coverage probabilities than the conditional likelihood intervals.

4 Premature birth data

For further comparison of the full likelihood and conditional likelihood methods, we apply them to a meta-analysis, in both of which θ stands for log-odds ratio. The data for meta-analysis comes from (Copas and Jackson 2004). It consists of the results of 14 randomized clinical trials concerning the use of prophylactic corticosteroids in cases of premature birth. The treatment is administered to the mother in order to improve the chance of the infant's survival if a birth is anticipated to be premature.

Table 3 reports the analysis results of the full likelihood and conditional likelihood methods.

The funnel plot of this data, shown in the right panel of Fig. 2, looks quite asymmetric, indicating that publication bias does exist. This also coincides with the observation from our likelihood ratio confidence interval for ρ at the 95% level, $[-1, -0.025]$, which excludes 0. However the Wald confidence based on the conditional likelihood is $[-2.543, 0.586]$, seemingly supporting $\rho = 0$, which is clearly unreliable. In addition, both confidence intervals for τ include zero, which provides certain evidence for the nonexistence of between-study heterogeneity.

The parameter θ denotes the underlying log-odds ratio comparing the probability of death in the treated group with that for a parallel sample of controls. Both the full likelihood and conditional likelihood methods give the same point estimate, – 0.476, for θ with similar intervals [– 0.760, – 0.244] and [– 0.662, – 0.289]. The log-odds ratio estimate – 0.476 implies that the use of prophylactic corticosteroids in cases of premature birth can reduce morality by as large as 28%. Hence our meta-analysis provides strong support for the use of prophylactic corticosteroids in cases of premature birth.

For point estimation of N, the full likelihood and conditional likelihood estimates are 23 and 16, respectively. However for interval estimation, again the lower bound (7) of the Wald interval is less than the number of observed studies (13), while that of the likelihood ratio interval is 13, which makes more sense.

In the presence of publication bias, the published studies, $\{(\theta_i, s_i) : i = 1, 2, \ldots, n\}$, constitute a biased sample of all studies; the empirical distributions of s_i^*'s and θ_i^*'s,

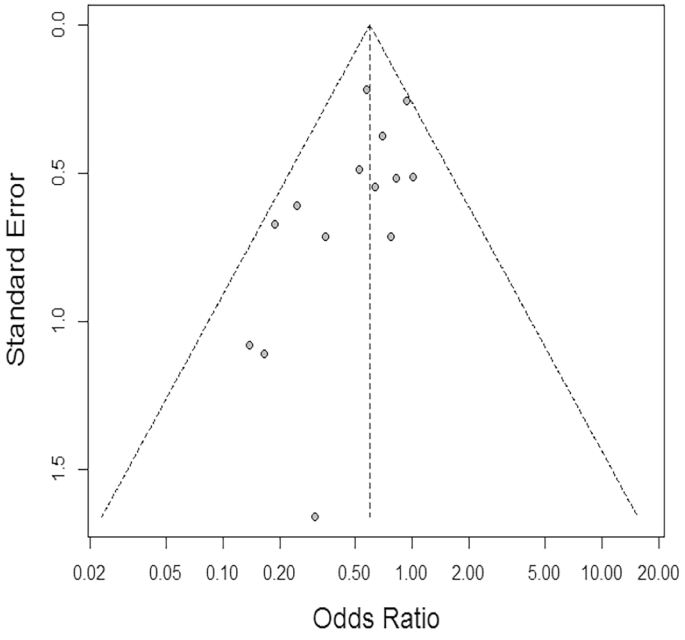

Fig. 2 Funnel plot of the premature birth data

respectively, are inconsistent estimators of the underlying population distributions, $F(s)$ and $G(t)$. Theoretically both the proposed full likelihood and the conditional likelihood methods can correct publication bias and consistently estimate $F(s)$ and $G(t)$. In Sect. 2, we have presented the MLEs $\widehat{F}(s)$ and $\widehat{G}(t)$ based on the full likelihood method, and the inverse probability weighting estimators $\tilde{F}(s)$ and $\tilde{G}(t)$ based on the conditional likelihood method. Figure 3 displays the empirical distributions, the full likelihood and conditional likelihood estimators for both $F(s)$ and $G(t)$ based on the premature birth data. As publication bias very likely exists in this data, we observe that the full likelihood and conditional likelihood estimates are away from the empirical distribution.

5 Discussion

To correct publication bias in meta-analysis, we propose a full likelihood semi-parametric approach under the Copas-like selection model of (Ning et al. 2017) . We have demonstrated the advantages of the proposed full likelihood method over the commonly used conditional likelihood and Wald-type method by theoretical and numerical studies. We show that the full MLEs have smaller mean squared errors than the conditional-likelihood-based estimators. The full likelihood ratio confidence intervals for the effect size and the total number of studies have more accurate coverage probabilities than the Wald intervals under the conditional likelihood.

Our full likelihood method is built on Copas's selection model (2). It is therefore interesting to check whether this model is correct or to conduct a sensitivity analysis on our method. (Almalik et al. 2020) have shown through simulations that the conditional likelihood method is not robust against misspecification of the selection mechanism. In particular, they found that the conditional likelihood method performs well when model (2) is correct, and it performs poorly otherwise. Because

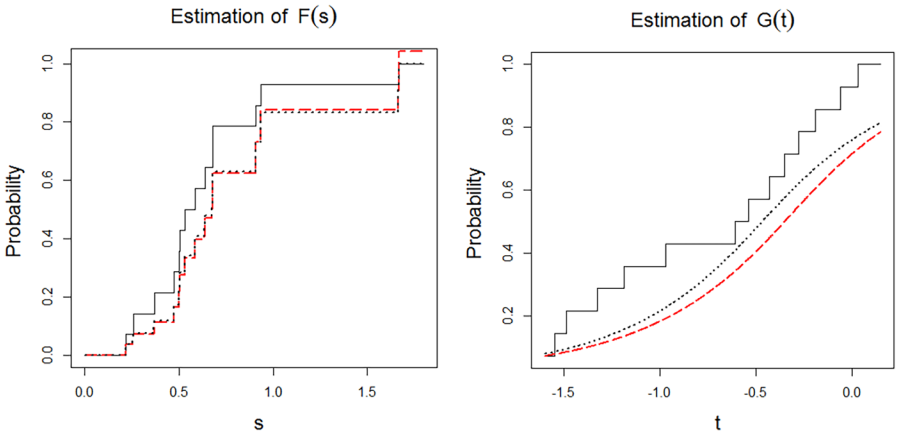

Fig. 3 Display of empirical distribution, the FL distribution estimate (dashed line) and the CL distribution estimate (dotted line) for the premature birth data

our full likelihood method is built on model (2) and integrates the conditional likelihood, we believe that our method may also be somewhat sensitive to the misspecification of the selection mechanism. We may leave this issue as our future research topic.

A key issue in the implementation of our method is that the maximization of the full likelihood is numerically very difficult, as the data contain little information about γ_1 and γ_2. This problem exists also in the maximization of the condition likelihood as pointed out by (Ning et al. 2017). To then end, we fix γ_1 and γ_2 to be their maximum conditional likelihood estimates that are calculated with (Ning et al. 2017)'s expectation-maximization algorithm. In doing so, the resulting parameter estimates are in essence different from the true maximum likelihood estimates. Similar to the conditional likelihood (see (Ning et al. 2017)), the full likelihood function seems to be a very flat plateau around its maximum. This also implies that the replacement of the true MLEs γ_1 and γ_2 with their conditional MLEs does not lead much change in the full likelihood ratio test statistics with respect to parameters other than γ_1 and γ_2.

6 Supplementary material

The Supplementary Material contains detailed proofs for Lemma 1, Theorems 1 and 2, and Proposition 1.

Acknowledgements The authors thank the Editor, the Associate Editor and an anonymous referee for helpful comments and suggestions that have led to significant improvements in the paper. Dr. Liu's research was supported by the National Natural Science Foundation of China (11771144, 11971300, 11871287), the State Key Program of the National Natural Science Foundation of China (71931004), the Natural Science Foundation of Shanghai (19ZR1420900, 17ZR1409000), the development fund for Shanghai talents , the 111 project (B14019), and the Fundamental Research Funds for the Central Universities. Dr. Li was supported in part by the Natural Sciences and Engineering Research Council of Canada grant number RGPIN-2015-06592.

References

Almalik, O., Zhan, Z., Van den Heuvel, E. (2020). Copas' method is sensitive to different mechaisms of publication bias. arXiv:2007.15955.

Carpenter, J. R., Schwarzer, G., Rucker, G., Kunstler, R. (2009). Empirical evaluation showed that the Copas selection model provided a useful summary in 80% of meta-analysis. *Journal of Clinical Epidemiology, 62,* 624–631.

Cooper, H., Hedges, L. V., Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). New York: Russell Sage Foundation.

Copas, J. B. (1999). What works?: Selectivity models and meta-analysis. *Journal of the Royal Statistical Society*: *Series A, 162,* 95–109.

Copas, J. B., Jackson, D. (2004). A bound for publication bias based on the fraction of unpublished studies. *Biometrics, 60*(1), 146–153.

Copas, J. B., Li, H. G. (1997). Inference for non-random samples. *Journal of the Royal Statistical Society*: *Series B, 59,* 55–95.

Copas, J. B., Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics, 1,* 247–262.

Copas, J. B., Shi, J. Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research, 10,* 251–265.

DerSimonian, R., Laird, N. (1986). Meta-analysis in clinical trials. *Contrilled Clinical Trials, 7,* 177–188.

Duval, S., Tweedie, R. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95,* 89–98.

Duval, S., Tweedie, R. (2000b). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463.

Egger, M., Smith, G. D., Altman, D. G. (2001). *Systematic reviews in health care*: *Meta-analysis in context* (2nd ed.). London: BMJ.

Egger, M., Smith, G. D., Schineider, M., Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634.

Fragkos, K. C., Tsagris, M., Frangos, C. C. (2017). Exploring the distribution for the estimator of Rosenthal's 'fail-safe' number of unpublished studies in meta-analysis. *Communications in Statistics-Theory and Methods, 46*(11), 5672–5684.

Galbraith, R. (1988). A note on the graphical presentation of estimated odds ratios from several clinical trials. *Statistics in Medicine, 7,* 889–894.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.

Gleser, L. J., Olkin, I. (1996). Models for estimating the number of unpublished studies. *Statistics in Medicine, 15,* 2493–2507.

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association, 109,* 1159–1173.

Jackson, D., Riley, R., White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine, 30,* 2481–2498.

Jin, Z. C., Zhou, X. H., He, J. (2015). Statistical method for dealing with publication bias in meta-analysis. *Statistics in Medicine, 34,* 343–360.

Koricheva, J., Gurevitch, J., Mengersen, K. (2012). *The handbook of meta-analysis in ecology and evolution*. Princeton: Princeton University Press.

Light, R., Pillemer, D. (1984). *Summing up*: *The science of reviewing research*. Cambridge: Harvard University Press.

Liu, Y., Li, P., Qin, J. (2017). Maximum empirical likelihood estimation for abundance in a closed population from capture recapture data. *Biometrika, 104,* 527–543.

Liu, Y., Liu, Y., Li, P., Qin, J. (2018). Full likelihood inference for abundance from continuous-time capture-recapture data. *Journal of the Royal Statistical Society*: *Series B*, *80*(5), 995–1014.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*: *Series B*, *44,* 226–233.

Mavridis, D., Sutton, A., Cipriani, A., Salanti, G. (2013). A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Statistics in Medicine, 32*(1), 51–66.

Ning, J., Chen, Y., Piao, J. (2017). Maximum likelihood estimation and EM algorithm of Copas-like selection model for publication bias correction. *Biostatistics, 18*(3), 495–504.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika, 75,* 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics, 18,* 90–120.

Rosenthal, R. (1979). The "file drawer problem'' and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rothstein, H. R. (2008). Publication bias as a threat to the validity of meta-analytic results. *Journal of Experimental Criminology, 4,* 61–81.

Rothstein, H. R., Sutto, A. J., Borenstein, M. (2006). *Publication bias in meta-analysis*: *Prevention, assessment and adjustments*. Sussex: Wiley.

Scharzer, G., Carpenter, J., Rucker, G. (2010). Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology, 63,* 282–288.

Sterne, J. A., Gavaghan, D., Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal Clinical Epidemiology, 53,* 1119–1129.

Sterne, J. A. C., Egger, M., Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal, 323,* 101–105.