Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression

I. Kosmidis
Department of Statistical Science
University College London
London, WC1E 6BT, U.K
i.kosmidis@ucl.ac.uk

A. Guolo
Department of Statistical Sciences
University of Padova
Via Cesare Battisti, 241, 35121 Padova, Italy
annamaria.guolo@unipd.it

C. Varin
Department of Environmental Sciences, Informatics and Statistics,
Ca' Foscari University of Venice
Via Torino 150, 30170 Venezia Mestre, Italy
cristiano.varin@unive.it

September 2, 2015

Abstract

Random-effects models are frequently used to synthetize information from different studies in meta-analysis. While likelihood-based inference is attractive both in terms of limiting properties and in terms of implementation, its application in random-effects meta-analysis may result in misleading conclusions, especially when the number of studies is small to moderate. The current paper shows that methodology designed to reduce the asymptotic bias of the maximum likelihood estimator of the variance component can also yield substantial improvement on inference about the mean effect size. The results are derived for the more general framework of random-effects meta-regression, which allows the mean effect size to vary with study-specific covariates.

Keywords: *Bias reduction; Heterogeneity; Meta-analysis; Penalized likelihood; Random effects; Restricted maximum likelihood*

1 Introduction

Meta-analysis is a widely applicable approach to combine information from different comparable studies about a common effect of interest. One of the major topics of debate in meta-analysis is how to best deal with the heterogeneity across studies. The popular formulation for meta-analysis described in DerSimonian & Laird (1986) accounts for the between-study heterogeneity via a random-effects specification. The meta-analytic estimator of the mean effect size is a weighted average of the study-specific estimators, with weights depending on the unknown random-effect variance. There is ample evidence that frequentist inference, typically carried out relying on first-order asymptotic arguments, can result in misleading conclusions, especially in the not unusual setting of a limited number of studies (e.g., van Houwelingen et al., 2002). The same considerations apply to the random-effects meta-regression model that is a direct extension of random-effects meta-analysis allowing for study-specific covariates to further describe the heterogeneity across studies.

Several proposals have been discussed to account for the finite number of studies, including modification of the limiting distribution of test statistics (Knapp & Hartung, 2003), restricted maximum likelihood (Viechtbauer, 2005) and second-order asymptotics (Guolo, 2012). Recently, Zeng & Lin (2015), on the pages of this journal, proposed a double bootstrap approach that outperfoms several alternatives in terms of empirical coverage probability of confidence intervals for the mean effect size.

The current paper studies the extent of the bias of the maximum likelihood estimator of the random-effect variance and introduces a bias-reducing penalized likelihood that yields a substantial improvement on the estimation of the random-effect variance. The bias-reducing penalized likelihood in the present setting is closely related to the approximate conditional likelihood of Cox & Reid (1987) and restricted maximum likelihood for inference on the random-effects variance. The associated penalized deviance can be used for inference about the fixed-effect parameters. Real-data examples and a simulation study illustrate the improvement in finite-sample performance against various alternative methods from the recent literature.

2 Random-effects meta-regression and meta-analysis

Suppose that there are K studies about a common effect of interest, each of them providing pairs of summary measures $(y_i, \hat{\sigma}_i^2)$, where y_i is the study-specific estimate of the effect, and $\hat{\sigma}_i^2$ is the associated estimation variance $(i = 1, \ldots, K)$. As with much of the literature in meta-analysis, assume that the within-study variances $\hat{\sigma}_i^2$ are estimated accurately enough to be considered as known and equal to the values reported in each study.

In some situations, the pairs $(y_i, \hat{\sigma}_i^2)$ may be accompanied by study-specific covariates, aimed at describing the heterogeneity across studies. The random-effects meta-regression model postulates that $y = (y_1, \ldots, y_K)^\top$ are realizations of random variables $Y = (Y_1, \ldots, Y_K)^\top$ that are independent conditionally on independent random effects $U = (U_1, \ldots, U_K)^\top$. The conditional distribution of Y_i given $U_i = u_i$ is assumed to be $N(u_i + x_i^\top \beta, \hat{\sigma}_i^2)$, where $x_i = (x_{i1}, \ldots, x_{ip})^\top$ is the p-vector that collects study-specific covariates, and β indicates the corresponding p-dimensional vector of effects. Typically $x_{i1} = 1$ and the random effect U_i is assumed to be distributed according to $N(0, \psi)$ $(i = 1, \ldots, K)$, where ψ describes the between-study heterogeneity. In matrix notation, and conditionally on $U = u$, the random-effects meta-regression model is

$$Y = X\beta + u + \epsilon, \tag{1}$$

where X is the model matrix of dimension $K \times p$ with $x_i^\top$ in its ith row, and $\epsilon = (\epsilon_1, \ldots, \epsilon_K)^\top$ is a vector of independent errors each with a $N(0, \hat{\sigma}_i^2)$ distribution. Under this specification, the marginal distribution of Y is multivariate normal with mean $X\beta$ and variance $\hat{\Sigma} + \psi I_K$,, where I_K is the $K \times K$ identity matrix and $\hat{\Sigma} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2)$. The random-effects meta-analysis model is a meta-regression model where X is a column of ones.

Parameter β is naturally esimated by weighted least squares as

$$\hat{\beta}(\psi) = \{X^\top W(\psi) X\}^{-1} X^\top W(\psi) Y\,, \tag{2}$$

with $W(\psi) = (\hat{\Sigma} + \psi I_K)^{-1}$. Inference on β critically depends on the availability of an accurate estimate of the between-study variance ψ. A large body of applications has resorted in using the DerSimonian & Laird (1986) estimator $\hat{\psi}_{\mathrm{DL}} = \max\{0, (Q - n + p)/A\}$, where $Q = (y - X\hat{\beta}_{\mathrm{F}})^\top \hat{\Sigma}^{-1}(y - X\hat{\beta}_{\mathrm{F}})$ is the Cochran statistic, with $\hat{\beta}_{\mathrm{F}} = \hat{\beta}(0)$ the estimate of β in the fixed-effects model obtained by dropping the random effect u in (1), and $A = \mathrm{tr}(\hat{\Sigma}^{-1}) - \mathrm{tr}\{(X^\top \hat{\Sigma}^{-1} X)^{-1} X^\top \hat{\Sigma}^{-2} X\}$. Inference on β is based on the fact that under model (1), $\hat{\beta}(\psi)$ has an asymptotic normal distribution with mean β and variance $X^\top W(\psi) X$. The considerable loss of efficiency of $\hat{\psi}_{\mathrm{DL}}$ (Hardy & Thompson, 1996; Guolo, 2012) has motivated the use of several likelihood-based approaches under both frequentist and Bayesian perspectives (see, e.g., van Houwelingen et al., 2002; Guolo & Varin, 2015, for recent reviews).

The log-likelihood function for $\theta = (\beta^\top, \psi)^\top$ in model (1) is

$$\ell(\theta) = -\frac{1}{2}\log|W(\psi)| - \frac{1}{2}R(\beta)^\top W(\psi) R(\beta), \tag{3}$$

where $|W(\psi)|$ denotes the determinant of $W(\psi)$ and $R(\beta) = y - X\beta$. A calculation of the gradient $s(\theta)$ of $\ell(\theta)$ shows that the maximum likelihood estimator $\hat{\theta}_{\mathrm{ML}} = (\hat{\beta}_{\mathrm{ML}}^\top, \hat{\psi}_{\mathrm{ML}})^\top$ for θ results from the solution of the system of equations

$$\begin{cases} s_{(\beta)}(\theta) = X^\top W(\psi) R(\beta) = 0_p \\ s_{(\psi)}(\theta) = R^\top(\beta) W(\psi)^2 R(\beta) - \mathrm{tr}\,[W(\psi)\}] = 0 \end{cases}, \tag{4}$$

where $s_{(\beta)}(\theta) = \nabla_\beta \ell(\theta)$ and $s_{(\psi)}(\theta) = \partial\ell(\theta)/\partial\psi$, so that $\hat{\beta}_{ML} = \hat{\beta}(\hat{\psi}_{ML})$. As is observed in Zeng & Lin (2015), a major drawback of maximum likelihood inference is the poor performance of associated procedures based on first-order asymptotics, when the number of studies K is small to moderate.

3 Bias reduction

3.1 Bias-reducing penalized likelihood

Using the results in Kosmidis & Firth (2009, 2010) and some algebraic effort, the first term in the expansion of the bias function (first-order bias) of the maximum likelihood estimator is found to be $b(\theta) = \{0_p^\top, b_{(\psi)}(\psi)\}^\top$, where 0_p denotes a p-dimensional vector of zeros and

$$b_{(\psi)}(\psi) = -\frac{\mathrm{tr}\{W(\psi)H(\psi)\}}{\mathrm{tr}\{W(\psi)^2\}}\,. \tag{5}$$

In the above expression, $H(\psi) = X\{X^\top W(\psi) X\}^{-1} X^\top W(\psi)$ is the 'hat' matrix. The derivation of the first-order bias is sketched in the Appendix.

The non-zero entries of $W(\psi)$ and the diagonal entries of $H(\psi)$ are all necessarily positive. Hence, equation (5) demonstrates that the maximum likelihood estimator of ψ is subject to downwards bias, which, as also noted elsewhere (Viechtbauer, 2005), has direct consequences on the performance of first-order procedures for hypothesis tests and confidence intervals for β. Specifically, a downwards bias of $\hat{\psi}$ has the effect of over-estimating the non-zero entries of $W(\psi)$, and hence, over-estimating the information matrix

$$F(\theta) = E_\theta\{I(\theta)\} = \begin{bmatrix} X^\top W(\psi) X & 0_p \\ 0_p^\top & \frac{1}{2}\mathrm{tr}\left\{W(\psi)^2\right\} \end{bmatrix},$$

where $I(\theta) = -\partial^2\ell(\theta)/\partial\theta\partial\theta^\top$ is the observed information matrix on θ. This, in turn, can result in hypothesis tests with large Type I error and confidence intervals or regions with actual coverage sensibly lower than the nominal level.

An estimator that corrects for the first-order bias of $\hat{\theta}_{\mathrm{ML}}$ results by solving the adjusted score equations $s^*(\theta) = s(\theta) - F(\theta)b(\theta) = 0_{p+1}$ (Firth, 1993; Kosmidis & Firth, 2009). In the specific case of the random-effects meta-regression model, $s^*_{(\beta)}(\theta) = s_{(\beta)}(\theta)$ and, hence, the adjusted score equation for β is exactly the same as in (4). On the other hand, the score equation for ψ is adjusted to

$$s^*_{(\psi)}(\theta) = R^\top(\beta)W(\psi)^2R(\beta) - \mathrm{tr}\left[W(\psi)\{I_K - H(\psi)\}\right] = 0\,. \tag{6}$$

The adjusted score functions $s^*_{(\beta)}(\theta)$ and $s^*_{(\psi)}(\theta)$ can also be obtained as the derivatives of the penalized log-likelihood function

$$\ell^*(\theta) = \ell(\theta) - \frac{1}{2}\log\left|F_{(\beta\beta)}(\psi)\right|\,, \tag{7}$$

where $\ell(\theta)$ is as in (3), $F_{(\beta\beta)}(\psi) = X^\top W(\psi)X$ is the β-block of the information matrix $F(\theta)$, and $\left|F_{(\beta\beta)}(\psi)\right|$ denotes the determinant of $F_{(\beta\beta)}(\psi)$. So, the maximum penalized likelihood estimator $\hat{\theta}_{\mathrm{MPL}}$ of θ is also a reduced-bias one.

Notice here that for $\beta = \hat{\beta}(\psi)$, (7) reduces both to the logarithm of the approximate conditional likelihood of Cox & Reid (1987) for inference on ψ, when β is treated as a nuisance component, and to the restricted log-likelihood function (Harville, 1977). In particular, the restricted log-likelihood function is constructed to reduce underestimation of variance components in finite samples as a consequence of failing to account for the simultaneous estimation of the fixed effects β. Smyth & Verbyla (1996) and Stern & Welsh (2000) have shown the equivalence of the restricted log-likelihood with approximate conditional likelihood in the more general context of inference on variance components in normal linear mixed models.

3.2 Estimation

Given a starting value $\psi^{(0)}$ for ψ, the following iterative process has a stationary point that maximizes (7). At the jth iteration ($j = 1, 2, \ldots$), a new candidate value $\beta^{(j+1)}$ for β is obtained by calculation of the weighted least squares estimator (2) at $\psi = \psi^{(j)}$, and, then, a candidate value for $\psi^{(j+1)}$ is computed through a line search for solving the adjusted score equation (6) evaluated at $\beta = \beta^{(j+1)}$. The iteration is repeated until either the candidate values do not change across iterations or the adjusted score functions are sufficiently close to zero, when the value of $\hat{\theta}_{\mathrm{MPL}}$ is returned.

Example 3.1: *Meat consumption data.* Larsson & Orsini (2014) investigate the association between meat consumption and relative risk of all-cause mortality. The data set comprises 16 prospective studies, eight of them about unprocessed read meat consumption and the remaining eight about processed meat consumption. The current example considers a meta-regression model with covariate the binary indicator of the type of meat consumption. The DerSimonian & Laird estimate of ψ is $\hat{\psi}_{\mathrm{DL}} = 0.57 \times 10^{-2}$, the maximum likelihood estimate is $\hat{\psi}_{\mathrm{ML}} = 0.85 \times 10^{-2}$and the maximum penalized likelihood estimate is $\hat{\psi}_{\mathrm{MPL}} = 1.18 \times 10^{-2}$. The corresponding estimates of β are almost identical with $\hat{\beta}_{\mathrm{ML}} = (0.10, 0.11)^\top$, $\hat{\beta}_{\mathrm{MPL}} = (0.09, 0.11)^\top$ and $\hat{\beta}_{\mathrm{DL}} = (0.11, 0.10)^\top$, where the first element in each vector of estimates corresponds to the intercept parameter and the second to the parameter for meat consumption. Figure 1 displays the boxplots for the distribution of the various estimators of ψ, as calculated from 10 000 samples simulated under the maximum likelihood fit. The dashed line is the value of the parameters used in the simulation and the point inside each box is the average of the estimates for the corresponding method. As expected, the maximum likelihood estimator of ψ is subject to downwards bias, while the other estimators manage to almost fully compensate for that bias. The distribution of the DerSimonian & Laird estimator of ψ appears to have a heavier right tail than the maximum

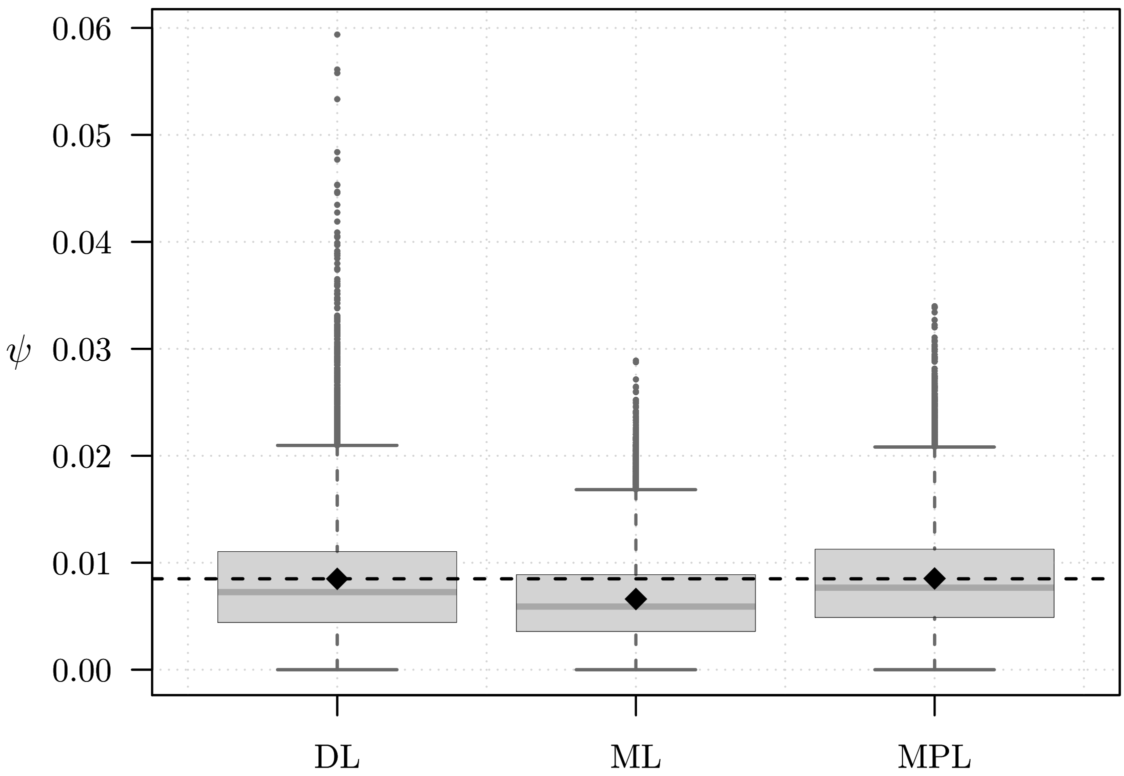

Figure 1: Boxplot for the DerSimonian and Laird (DL), maximum likelihood (ML) and maximum penalized likelihood (MPL) estimators of ψ as calculated from 10 000 simulated samples under the maximum likelihood fit in Example 3.1. The horizontal dashed line is the parameter value used for the simulation and the point inside each box is the average of the estimates for the corresponding method.

penalized likelihood estimator, which links to the findings of past studies on its loss of efficiency (e.g., Viechtbauer, 2005).

3.3 Penalized likelihood inference

Apart from a way to improve the bias in the estimation of ψ, the profiles of the penalized likelihood function can be used to construct confidence intervals and regions, and carry out hypothesis tests for β. If $\beta = (\gamma^\top, \delta^\top)^\top$, and the $\hat{\gamma}_{\mathrm{MPL},\delta}$ and $\hat{\psi}_{\mathrm{MPL},\delta}$ are the estimators from maximising (7) for fixed δ, then the penalized deviance $2\{\ell^*(\hat{\gamma}_{\mathrm{MPL}}, \hat{\delta}_{\mathrm{MPL}}, \hat{\psi}_{\mathrm{MPL}}) - \ell^*(\hat{\gamma}_{\mathrm{MPL},\delta}, \delta, \hat{\psi}_{\mathrm{MPL},\delta})\}$ has the usual limiting χ^2_q distribution, where $q = \dim(\delta)$. To show that, note that the adjustment to the scores is additive and $O(1)$ (see § 3.1), and, hence, the extra terms depending on it and its derivatives in the asymptotic expansion of the profile penalized likelihood disappear as information increases.

4 Simulation study

Following Zeng & Lin (2015), the proposed bias-reducing penalized likelihood is compared with alternative methods using the simulation set-up of Brockwell & Gordon (2001). The study-specific effects y_i are simulated from the random-effect meta-analysis with true effect $\beta = 0.5$ and variance $\hat{\sigma}_i^2 + \psi$ ($i = 1, \ldots, K$), where the $\hat{\sigma}_i^2$ are independently generated from a χ^2_1 distribution

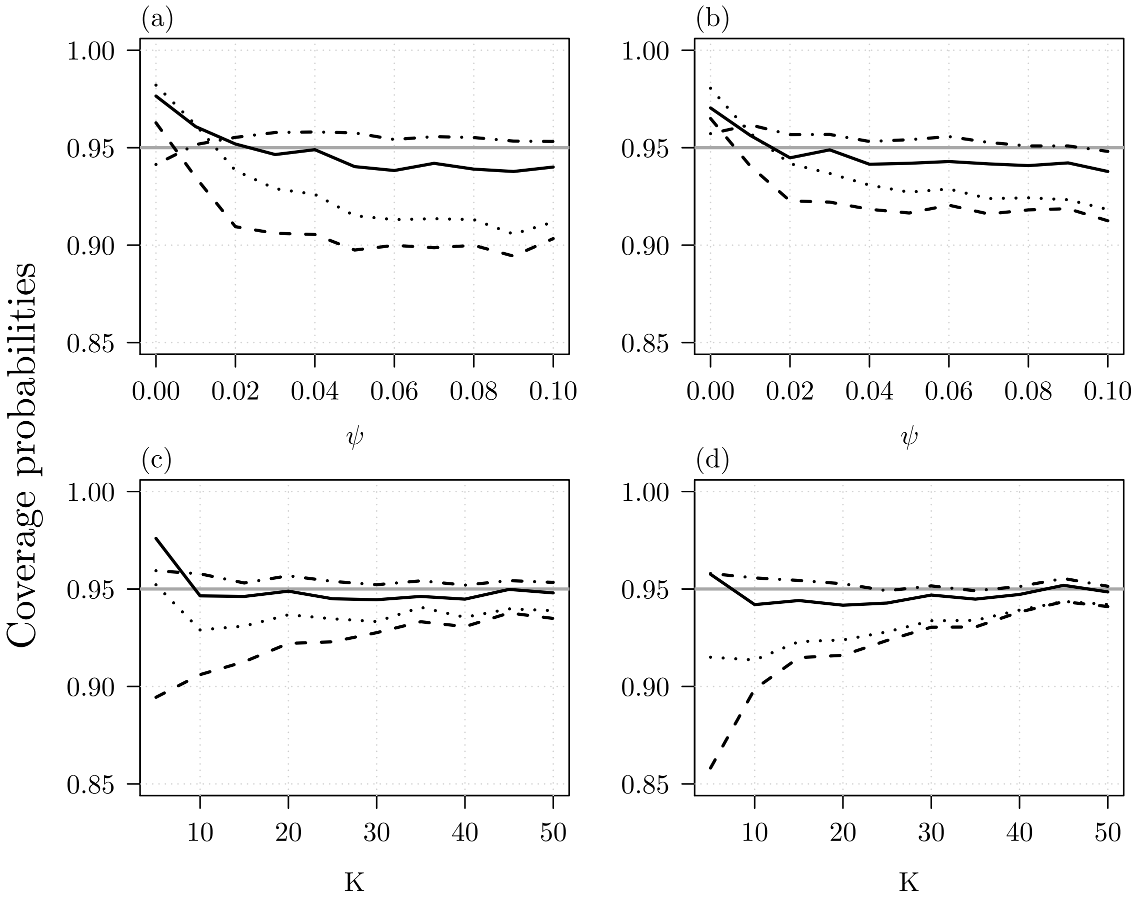

Figure 2: Empirical coverage probabilities of nominal 95% confidence intervals for increasing values of ψ, when (a) $K = 10$ and (b) $K = 20$ and for increasing values of K when (c) $\psi = 0.03$ and (d) $\psi = 0.07$. The plotted curves correspond to the proposed penalized likelihood method (solid), the DerSimonian & Laird method (dashed), the Zeng & Lin double resampling method (dotted), and the Skovgaard's statistic (dotted-dashed). The solid, grey horizontal line is the nominal level.

multiplied by 0.25 and then restricted to the interval $(0.09, 0.6)$. The between-study variance ψ varies from 0 to 0.1 and the number of studies K from 5 to 50. For each considered combination of ψ and K, 10 000 data sets are simulated.

The results in Zeng & Lin (2015, Section 5) show that the double resampling approach therein outperforms several existing methods in terms of the empirical coverage probabilities of confidence intervals at nominal level 95%. The methods considered in Zeng & Lin (2015) include profile likelihood (Hardy & Thompson, 1996), modified DerSimonian & Laird (see Sidik & Jonkman, 2002, and Knapp & Hartung, 2003, for the method, and Copas, 2003, for a critique), quantile approximation (Jackson & Bowden, 2009) and the approach described in Henmi & Copas (2010). The present simulation study takes advantage of the availability of these previous simulation results, and Figure 2 compares the performance of double resampling confidence interval with that from the profile penalized likelihood. Remarkably, the profile penalized likelihood confidence interval has empirical coverage that is sensibly closer to the nominal level than double resampling. It should be emphasised that the equivalence shown in § 3.1 is between the penalized likelihood for $\beta = \hat{\beta}(\psi)$ and the approximate conditional likelihood for inference on ψ only. Hence, this equivalence does not relate and cannot be used to explain the good finite-sample performance of profile penalized likelihood confidence intervals for β.

Figure 2 also includes a symmetric normal-theory confidence interval based on the classical DerSimonian & Laird estimator $\hat{\beta}(\psi_{\rm DL})$ and its estimated variance $\sum_{i=1}^{K} 1/(\hat{\sigma}_i + \hat{\psi}_{\rm DL})$. Not surprisingly, the empirical coverage of this confidence interval is grossly smaller than the nominal confidence level. Lastly, Figure 2 also includes results for the confidence intervals based on the Skovgaard's statistic (see Guolo, 2012, for details) that is designed to produce second-order accurate p-values for tests on the mean effect size, and is implemented in the R (R Core Team, 2015) package metaLik (Guolo & Varin, 2012). The Skovgaard's statistic yields empirical coverages slightly closer to the nominal level for a wider range of values for ψ than the penalized likelihood. Nevertheless, the penalized likelihood does not only offer a device for obtaining reasonable p-values and confidence intervals for β as the Skovgaard's statistic does, but also allows for bias-reduced estimation of the degree of heterogeneity ψ, which is a quantity of interest in the medical literature. Furthermore, the construction of confidence intervals for β based on the Skovgaard's approach requires numerical inversion of the statistic with potential instabilities due to its discontinuity around zero.

5 Examples

Example 5.1: *Local anesthesia data.* Ambulatory hysteroscopy is a useful instrument to identify intrauterine pathologies. Cooper et al. (2010) perform a meta-analysis about the efficacy of different types of local anesthesia used to control pain during hysteroscopy. The data considered here refer to the use of paracervical anesthesia and consist of information from five randomized controlled trials, expressed in terms of standardized mean difference of pain scores measured at the time of hysteroscopy and as vasovagal episodes. The DerSimonian & Laird estimate of ψ is $\hat{\psi}_{\rm DL} = 1.08$, while the maximum likelihood estimate is the sensibly larger $\hat{\psi}_{\rm ML} = 2.31$. The maximum penalized likelihood estimate is even larger, $\hat{\psi}_{\rm MPL} = 2.93$.

The DerSimonian & Laird method strongly supports the significance of the local anesthesia efficacy with a p-value of 0.007, as the double resampling approach does with a p-value of 0.018. Converse conclusion is obtained using the penalized deviance, which returns a p-value equal to 0.137. The result is also confirmed by the Skovgaard's statistic, with a p-value of 0.158.

Example 5.2: *Meat consumption data (cont.).* Continuing from Example 3.1, the DerSimonian & Laird method indicates weak evidence for a higher risk associated to the consumption of red processed meat than unprocessed meat, with a p-value of 0.054. In contrast, the penalized deviance and the Skovgaard's statistic both agree that there is no evidence of difference between the two types of meat, with p-values of 0.132 and 0.145, respectively.

6 Final remarks

An alternative estimator of ψ that corrects for the first-order bias is $\hat{\psi}_{BC} = \hat{\psi} - b_{(\psi)}(\hat{\psi})$ (see Efron, 1975, for proof that the new estimator is free from first-order bias), and as such has no associated inference function like (7) for the maximum penalized likelihood estimator.

The impact of using the penalized likelihood for estimation and inference in random-effects meta-analysis and meta-regression is more profound for small to moderate number of studies. As the number of studies increases the log-likelihood derivatives dominate the bias-reducing adjustment in (6) in terms of asymptotic order, and, so, inference based on the penalized likelihood becomes indistinguishable from likelihood-based inference.

Appendix

Sketch derivation of the expression for the first-order bias

The first-order bias of $\hat{\theta}_{(ML)}$ has the form $b(\theta) = -\{F(\theta)\}^{-1} A(\theta)$ (Kosmidis & Firth, 2010), where $A(\theta)$ has components

$$A_t(\theta) = -\frac{1}{2}\mathrm{tr}\left[\{F(\theta)\}^{-1}\{P_t(\theta) + Q_t(\theta)\}\right] \quad (t = 1, \ldots, p+1)\,.$$

In the above expression, $P_t(\theta) = E_\theta\{s(\theta)s(\theta)^\top s_t(\theta)\}$ and $Q_t(\theta) = E_\theta\{-I(\theta)s_t(\theta)\}$. The model assumptions imply that $E_\theta\{R_i(\beta)^m\}$ is 0 if m is odd and $(m-1)!!/w_i(\psi)^{m/2}$ if m is even, where $w_i(\psi) = 1/(\hat{\sigma}_i^2 + \psi)$, and $(m-1)!!$ denotes the double factorial of $m-1$ ($m = 1, 2, \ldots; i = 1, \ldots, K$). Using this fact, direct matrix calculations give that

$$P_t(\theta) + Q_t(\theta) = 0_{(p+1)\times(p+1)} \quad (t = 1, \ldots, p)\,; \quad P_{p+1}(\theta) + Q_{p+1}(\theta) = \begin{bmatrix} X^\top W(\psi)^2 X & 0_p \\ 0_p^\top & 0 \end{bmatrix},$$

where $0_{p\times p}$ is the $p \times p$ zero matrix. Hence, $A_t(\theta) = 0$ for $t \in \{1, \ldots, p\}$ and $A_{p+1}(\theta) = \mathrm{tr}\left[\left\{X^\top W(\psi) X\right\}^{-1} X^\top W(\psi)^2 X\right] = \mathrm{tr}\{W(\psi)H(\psi)\}$. Plugging the expressions for the components of $A(\theta)$ in the expression for $b(\theta)$ gives $b(\theta) = \{0_p^\top, b_{(\psi)}(\psi)\}^\top$, where $b_{(\psi)}(\psi)$ is as in (5).

References

Brockwell, S. E. & Gordon, I. N. (2001). A comparison of statistical methods for meta-analysis. *Stat. Med.* **20**, 825–40.

Cooper, N. A. M., Khan, K. S. & Clark T. J. (2010). Local anaesthesia for pain control during outpatient hysteroscopy: systematic review and meta-analysis. *Brit. Med. J.* **340**, c1130.

Copas, J. (2003) Letters to the editor: a simple confidence interval for meta-analysis. Sidik K., Jonkman J. N. (2002) *Statistics in Medicine* **21**, 3153–3159. *Stat. Med,* **22**, 2667–68.

Cox, D. R. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *J. R. Statist. Soc. B* **49**, 1–18.

DerSimonian, R. & Laird, N. (1986). Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–88.

Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency) (with discussion). *Ann. Stat.* **3**, 1189–1217.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27–38.

Guolo, A. (2012). Higher-order likelihood inference in meta-analysis and meta-regression. *Stat. Med.* **31**, 313–27.

Guolo, A. & Varin, C. (2012). The R package metaLik for likelihood inference in meta-analysis. *J. Stat. Softw.* **50**, 1–14.

Guolo, A. & Varin, C. (2015). Random-effects meta-analysis: The number of studies matters. *Stat. Methods Med. Res.*, to appear.

Hardy, R. J. & Thompson, S. G. (1996). A likelihood approach to meta-analysis with random effects. *Stat. Med.* **15**, 619–29.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Stat. Ass.* **72**, 320–38.

Henmi, M. & Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. *Stat. Med.* **29**, 2969–83.

van Houwelingen, H.C., Arends, L.R. & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat. Med.* **21**, 589–624.

Jackson, D. & Bowden, J. (2009). A re-evaluation of the 'quantile approximation method' for random effects meta-analysis. *Stat. Med.* **28**, 338–48.

Knapp, G. & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Stat. Med.* **22**, 2693–710.

Kosmidis, I. & Firth, D. (2009). Bias reduction in exponential family nonlinear models. *Biometrika* **96**, 793–804.

Kosmidis, I. & Firth, D. (2010). A generic algorithm for reducing bias in parametric estimation. *Electron. J. Stat.* **4**, 1097–112.

Larsson, S. C. & Orsini, N. (2014) Red meat and processed meat consumption and all-cause mortality: A meta-analysis. *Am. J. Epidemiol.* **179**, 282–89.

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

Searle, S. R., Casella, G. & McCulloch, C. E. (1982). *Variance Components.* New York: Wiley.

Smyth, G. K. & Verbyla, A. P. (1996). A conditional likelihood apporach to residual maximum likelihood estimation in generalized linear models. *J. R. Statist. Soc. B* **58**, 565–72.

Stern, S. E. & Welsh, A. H. (2000). Likelihood inference for small variance components. *Can. J. Stat.* **83**, 319–30.

Sidik, K. & Jonkman, J. N. (2002). A simple confidence interval for meta-analysis. *Stat. Med.* **21**, 3153–9.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *J. Educ. Behav. Stat.* **30**, 261–93.

Zeng, D. & Lin, D. Y. (2015). On random-effects meta-analysis. *Biometrika*, to appear.

Supplementary material for

Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression

I. Kosmidis
Department of Statistical Science
University College London
London, WC1E 6BT, U.K
i.kosmidis@ucl.ac.uk

A. Guolo
Department of Statistical Sciences
University of Padova
Via Cesare Battisti, 241, 35121 Padova, Italy
annamaria.guolo@unipd.it

C. Varin
Department of Environmental Sciences, Informatics and Statistics,
Ca' Foscari University of Venice
Via Torino 150, 30170 Venezia Mestre, Italy
cristiano.varin@unive.it

September 2, 2015

R version, package details and other functions

The current report reproduces the real-data analyses that appear in the examples of main text of the paper "Improving the accuracy of likelihood-based inference in meta-analysis and meta-regression" by I. Kosmidis, A. Guolo and C. Varin.

The resultant output in the current report has been produced using R version 3.2.1 (R Core Team, 2015), and the R package `metaLik` version 0.42.0 (Guolo & Varin, 2012).

The file `functionsMPL.R` that accompanies the current report (at the time of writing the file can be also be obtained from `http://www.ucl.ac.uk/~ucakiko/files/functionsMPL.R`) contains a function to maximize the penalized likelihood in meta-regression settings (see expression (7) in the main text and the function `BiasFit` in `functionsMPL.R`), a function to perform hypothesis tests for the parameters of a meta-regression model using the profiles of the penalized likelihood (see §3.3 of the main text and the function `lrtest` in `functionsMPL.R`), an implementation of the double resampling approach in Zeng & Lin (2015) for hypothesis testing in meta-analysis, and other helper functions for the above.

```
require(metaLik)
source("functionsMPL.R")
```

Meat consumption data

```
########################
## Meat consumption data
larsson <- data.frame(y = c(-0.3425,  0.2546,  0.1740,  0.1655,
                            -0.0834,  0.0953,  0.2151,  0.3988,
                             0.0488,  0.1484,  0.2231,  0.2390,
                             0.1823,  0.3577,  0.0583,  0.1484),
                      sigma2 = c(0.017224, 0.001271, 0.000663, 0.005027,
                                 0.003383, 0.003603, 0.062186, 0.118504,
                                 0.071613, 0.000310, 0.000501, 0.001160,
                                 0.000759, 0.005087, 0.031266, 0.023078),
                      type = c(rep("a", 8), rep("b", 8)))

## Fit the model using metaLik
m1 <- metaLik(y ~ type, data = larsson, sigma2 = sigma2)

## Example 1
## The various estimates for psi
estimates1 <- BiasFit(m1)
psi1_ml <- estimates1$ML[3]
psi1_dl <- estimates1$DL[3]
psi1_mpl <- estimates1$MPL[3]
psi1_estimates <- c(psi1_ml, psi1_dl, psi1_mpl)
names(psi1_estimates) <- c("ML", "DL", "MPL")
round(psi1_estimates, 4)

##     ML     DL    MPL
## 0.0085 0.0057 0.0118

## The various estimates for beta
beta1_ml <- estimates1$ML[1:2]
beta1_dl <- estimates1$DL[1:2]
beta1_mpl <- estimates1$MPL[1:2]
beta1_estimates <- data.frame(ML = beta1_ml, DL = beta1_dl, MPL = beta1_mpl)
round(beta1_estimates, 2)

##               ML   DL  MPL
## (Intercept) 0.10 0.11 0.09
## typeb       0.11 0.10 0.11

## Example 3
## p-values for testing the effect of meat consumption
pvalue1_dl <- 2 * (1 - pnorm(m1$DL[2] / sqrt(m1$vcov.DL[2, 2])))
pvalue1_pd <- lrtest(m1, what = 2, type = "penloglik", null = 0.0, optMethod = "BFGS")$pvalue
pvalue1_Skovgaard <- test.metaLik(m1, param = 2, value = 0, print = FALSE)$pvalue.rskov
pvalues1 <- c(pvalue1_dl, pvalue1_pd, pvalue1_Skovgaard)
names(pvalues1) <- c("DerSimonian Laird", "Penalized deviance", "Skovgaard")
round(pvalues1, 3)

##  DerSimonian Laird Penalized deviance          Skovgaard
##              0.054              0.132              0.145
```

Local anesthesia data

```
########################
## Local anesthesia data
cooper <- data.frame(y = c(0.00, -1.71, -0.19, -0.58, -4.27),
                     sigma2 = c(0.03959288, 0.07731804, 0.02265332, 0.01759683, 0.16040842))

## Fit the model using metaLik
m2 <- metaLik(y ~ 1, data = cooper, sigma2 = sigma2)

## Example 3
## The various estimates for psi
estimates2 <- BiasFit(m2)
psi2_ml <- estimates2$ML[2]
psi2_dl <- estimates2$DL[2]
psi2_mpl <- estimates2$MPL[2]
psi2_estimates <- c(psi2_ml, psi2_dl, psi2_mpl)
names(psi2_estimates) <- c("ML", "DL", "MPL")
round(psi2_estimates, 2)

##   ML   DL  MPL
## 2.31 1.08 2.93

## p-values for testing local anesthesia efficacy
pvalue2_dl <- 2*pnorm(m2$DL / sqrt(m2$vcov.DL))
pvalue2_dr <- double.resampling(0.0,  m2, B = 1000, myseed = 123)
pvalue2_pd <- lrtest(m2, what = 1, type = "penloglik", null = 0.0, optMethod = "BFGS")$pvalue
pvalue2_Skovgaard <- test.metaLik(m2, param = 1, value = 0, print = FALSE)$pvalue.rskov
pvalues2 <- c(pvalue2_dl, pvalue2_dr, pvalue2_pd, pvalue2_Skovgaard)
names(pvalues2) <- c("DerSimonian Laird", "Double resampling", "Penalized deviance", "Skovgaard")
round(pvalues2, 3)

##  DerSimonian Laird  Double resampling Penalized deviance          Skovgaard
##              0.007              0.018              0.137              0.158
```

References

Guolo, A. & Varin, C. (2012). The R package metaLik for likelihood inference in meta-analysis. *J. Stat. Softw.* **50**, 1–14.

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

Zeng, D. & Lin, D. Y. (2015). On random-effects meta-analysis. *Biometrika*, to appear.