



An Alternative Diagnostic Procedure for Meta-Regression

Ali Abuzaid^{1,*}, Enass Abed², Abdu Atta³, Esam Mahdi³

¹*Department of Mathematics, Faculty of Science, Al Azhar University - Gaza, Palestine*

²*Department of Applied Statistics, Faculty of Economics and Administrative Sciences, Al Azhar University - Gaza, Palestine*

³*Department of Mathematics, Statistics and Physics, School of Arts and Sciences, Qatar University, Doha, Qatar*

Abstract This paper proposes an alternative procedure for detecting outliers in meta-regression using the penalized maximum likelihood with smoothly clipped absolute deviation penalty function. The coordinate descent algorithm is implemented to estimate the parameters where the cross-validation criterion is used to determine the tuning parameter. Extensive simulation experiments demonstrate the usefulness of our proposed procedure as well as its improved power performance compared to previous procedures. Simulation results demonstrate that the performance has a direct relationship with the number of studies and an inverse relationship with the heterogeneity between studies. An illustrative application with real data, implementing the proposed procedure and others, is given.

Keywords Coordinate Descent Algorithm, Effect-size, Meta-analysis, Outliers, Penalized Maximum Likelihood

AMS subject classification: 62J20, 62J99

DOI: 10.19139/soic-2310-5070-864

1. Introduction

The effect-size is considered to be the unit of analysis and the fundamental quantity of interest in meta-analysis. It measures the strength, magnitude, and the direction of a relationship or difference between groups [9]. Although, the effect-size does not depend on the aspects of the study design that may vary from study to study, such as the sample size, the effect of the intervention depends on the nature of the collected data whether it is continuous, dichotomous, ordinal, counts, or time-to-event [14]. Meta-analysis is subject to bias due to the existence of extreme effect size values, which often called “outliers”. The presence of outliers may complicate the analysis and influence its validity, accuracy, and robustness. To account for the presence of possible outliers, various methods have been proposed to literature [4, 5, 12, 7, 2]. In many cases, a deletion diagnostic method has been used to identify the outliers. This method is simply based on deleting one or multiple observations at the same time and then compute the deletion statistics that measure the effect of those deleted observations on parameter estimates or residuals. The popular diagnostic, graphical and numerical, methods that are used for detecting the outliers in the meta-analysis are given in the book of [13]. The problem of outliers and influential points in mixed-effect model have been studied by [21]. They developed a method based on the impact of excluding studies on various statistics such as the summary estimate, heterogeneity and residuals.

Penalized Maximum Likelihood Estimation (PMLE) is widely used for the stabilization of the estimators as well as detecting outliers. [18] discussed the possibility of detecting outliers in regression models based on the mean shift model which predicts the dependent variable by the usual linear model plus an outlier term γ_i , where if $\gamma_i = 0$ then the i th observation is a consistent observation and not an outlier, otherwise it is an outlier. PMLE is

*Correspondence to: ALI ABUZAIID (Email: a.abuzaid@alazhar.edu.ps). Department of Mathematics, Al-Azhar University-Gaza. Jamal Abdl Naser St., Gaza 1277, Gaza Strip, Palestine.

used for outlier detection in different statistical models, including general profiles monitoring [23], and exponential regression for censored data [22].

In this paper, we develop an alternative procedure to detect the outliers in meta-regression models utilizing the penalized maximum likelihood. We utilize the coordinate descent algorithm to estimate the parameters, and the cross-validation criterion is used to determine the tuning parameter. We demonstrate via simulation experiments the usefulness of our proposed procedure as well as its improved power performance compared to previous procedures.

The rest of this paper is organized as follows. Section 2 reviews the main effect models. Section 3 discusses outliers and influential diagnostics in meta-regression. Section 4 proposes an alternative procedure for outlier detection in meta-regression based on the PMLLE technique. The power of performance of the proposed procedure is investigated and compared with other available procedures via simulation as described in section 5. Finally, in section 6, we discussed an illustrative application on educational meta-analysis.

2. Meta-regression models

Suppose y_1, y_2, \dots, y_k are the observed effect size estimates in a set of k independent studies, where

$$y_i = \theta_i + e_i, \quad i = 1, 2, \dots, k, \quad (1)$$

where θ_i is the true effect in the i th study and e_i is the error within the studies.

If $\theta_i = \theta$, then the true effect is homogeneous. This leads to the *fixed-effect-model*, where $e_i \sim \mathcal{N}(0, \sigma_i^2)$, $y_i \sim \mathcal{N}(\theta, \sigma_i^2)$, and σ_i^2 is the sampling variance of the i th study and it is assumed to be known [9]. On the other hand, if $\theta_i \neq \theta$, then the true effect is heterogeneous, which leads to the *random-effects-model*.

Consider $\theta_i = \mu + u_i$, where $u_i \sim \mathcal{N}(0, \tau^2)$ is the error between studies, so that $y_i \sim \mathcal{N}(\mu, \tau^2 + \sigma_i^2)$, μ is the mean of the true effect, and τ^2 is the amount of heterogeneity between the studies. If $\tau^2 = 0$, then it implies the homogeneity among the true effects, thus $\mu = \theta$.

In some situations, the effect size is moderated by known variables and can be formulated by

$$\theta_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + u_i, \quad i = 1, 2, \dots, k, \quad (2)$$

where X_{ij} is the value of the j th moderator variable for the i th study, β_0 is the expected effect size when $X_{ij} = 0$ for, $j = 1, 2, \dots, p$, and β_j is a coefficient explain how the j th moderator influences the size of the true effect for a one-unit increase in X_{ij} . Since the true effect sizes are now considered to be a function of both fixed and random effects, model (2) is typically called the mixed-effects model.

The selection of effect model depends on the variation in the true effect sizes (heterogeneity). There are several tests and indexes of heterogeneity [14]. The most common test is the Q test, and given by

$$Q = \sum_{i=1}^k w_i (y_i - \hat{\theta}),$$

where $w_i = 1/\sigma_i^2$. Under the assumption of homogeneity Q test has chi-square distribution with $k - 1$ degrees of freedom. Alternatively, one can use the I^2 index

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%,$$

where df is the degree of freedom of Q test, and I^2 index is the percentage ranging from 0% to 100%.

3. Outliers and diagnostics in meta-regression

In the context of meta-analysis, outlier refers to a study which does not fit the model of other studies, where we cannot explain them by an obvious characteristic of the study [13]. Some of the outlier values influence the results of the analysis and these values also called "influential points".

Despite of limited interest on the problem of outliers in meta-analysis, most of the published work considered fixed and random-effect models [13, 4, 16, 12, 7]. Viechtbauer and Cheung [21] discussed the extension of some of the outliers and influential points detection procedures from the multiple regression to the meta-regression in the following subsection.

3.1. Outliers in meta-regression

The adaption of several types of residuals from multiple linear regression to meta-regression is a helpful tool to detect outliers as follows:

1. **Studentized residual**, which is the ratio of the raw residual e_i and its standard error, and given by

$$t_i = (y_i - \hat{\mu}_i) / \sqrt{(1 + h_i)(\sigma_i^2 + \hat{\tau}^2)},$$

where $\hat{\mu}_i$ is the predicted average true effect size for the i th study, and h_i is the leverage of the i th study. If an outlier exist, then the estimate of τ^2 will be inflated and $\hat{\mu}_i$ will be pulled toward the y_i value.

2. **Studentized deleted residuals**, which is given by

$$t_{i(-i)} = (y_i - \hat{\mu}_{i(-i)}) / \sqrt{\sigma_i^2 + \hat{\tau}_{i(-i)}^2 + \text{Var}(\hat{\mu}_{i(-i)})},$$

where $\hat{\mu}_{i(-i)}$ is the predicted average true effect size for the i th study based on the model that actually excludes the i th study during the model fitting.

The studentized method may be a simple method, however when there are multiple outliers, it may fail.

3.2. Identifying influential studies in meta-regression

The authors of [8] proposed a deletion one row approach to detect possible influential points based on some of the statistical measures of the fitted linear regression model, by comparing the value of statistical measure for full data with the measure after the deletion of one point at time. Viechtbauer and Cheung [21] adopted the known deletion diagnostics from linear regression to the context of meta-analysis to identify possible influential studies. The following are the popular diagnostic tests implemented in the **metafor** R package [20]:

1. **DFFITs**: it measures the difference between the predicted mean effect for the i th study with and without the i th study included in the model, and it is given by

$$DFFITs_i = (\hat{\mu}_i - \hat{\mu}_{i(-i)}) / \sqrt{h_i(\sigma_i^2 + \hat{\tau}_{i(-i)}^2)}.$$

2. **Cook's distance**: It considers the change in the fitted values when the i th study is excluded, and it is given by

$$D_i = \left(\hat{\beta} - \hat{\beta}_{(-i)} \right)^T \left(\mathbf{X}^T \tilde{\mathbf{W}} \mathbf{X} \right) \left(\hat{\beta} - \hat{\beta}_{(-i)} \right),$$

where $\hat{\beta}$ and $\hat{\beta}_{(-i)}$ are the vectors of parameter estimates for full data and after deleting the i th study, respectively, \mathbf{X} is the $k \times (p+1)$ design matrix that includes one's in the first column (corresponding to β_0) and the values of the moderators in the other p columns, and $\tilde{\mathbf{W}}$ is a diagonal matrix $\tilde{\mathbf{W}} = \text{diag}[1/(\sigma_1^2, \hat{\tau}^2), 1/(\sigma_2^2, \hat{\tau}^2), \dots, 1/(\sigma_k^2, \hat{\tau}^2)]$. The statistic D_i follows chi-squared distribution with degree of freedom $(p+1)$, where p is the number of coefficients included in a mixed-effects model. If D_i exceeds the value χ_{p+1}^2 at a desirable level of significance, then the i th study is a suspected outlier and should be further examined.

3. **DFBETAS**: The change in the parameter estimates can also be measured as each study is deleted in turn. Difference in coefficients betas (DFBETAS) is calculated using

$$DFBETAS_i = (\hat{\beta}_j - \hat{\beta}_{j(-i)}) / \sqrt{(\mathbf{X}^T \widetilde{\mathbf{W}}_{(-i)} \mathbf{X})_{jj}^{-1}},$$

where $(\mathbf{X}^T \widetilde{\mathbf{W}}_{(-i)} \mathbf{X})_{jj}^{-1}$ is the value of the $(j + 1)$ th diagonal element of the matrix $(\mathbf{X}^T \widetilde{\mathbf{W}} \mathbf{X})^{-1}$. For small to medium data sets, a value of $DFBETAS_i$ larger than 1 could indicate an influential point.

4. **COVRATIO**: A change in the variance-covariance matrix of the overall effect size estimates can be quantified by

$$COVRATIO_i = \det [\text{Var}(\hat{\beta}_{(-i)})] / \det [\text{Var}(\hat{\beta})],$$

where $\text{Var}(\hat{\beta}) = (\mathbf{X}^T \widetilde{\mathbf{W}} \mathbf{X})^{-1}$. Removing the i th study will yield a more precise effect size estimate when $COVRATIO_i$ is less than 1.

5. **R_i**: The change in the estimate of the heterogeneity, it can be measured with the exclusion of each study, R_i is calculated as a percent of change as follows

$$R_i = \frac{\hat{\tau}^2 - \hat{\tau}^2_{(-i)}}{\hat{\tau}^2} \times 100\%.$$

If the i th study is an influential point, then its removal will cause a large decrease for heterogeneity (residual).

4. Penalized maximum likelihood estimation for outliers detection

In regression analysis, when the number of covariates is larger than the sample size, the maximum likelihood estimation is not working well. To overcome such a problem, one way is to reduce the number of explanatory variables looking for all possible subsets of these explanatory variables [17, 3]. Variable selection can be carried out by imposing a penalty on the likelihood function, where, instead of maximizing $\ell(\theta | \mathbf{x}) = \ln[\mathcal{L}(\theta | \mathbf{x})]$, we maximize the function $M(\theta) = \ell(\theta | \mathbf{x}) - \lambda P(\theta)$, where M is the objective function, λ is a scalar controls the tradeoff between the two parts, and P is a function that penalizes what one would consider less realistic values of the unknown parameters. Penalized Maximum Likelihood Estimation (PMLE) is widely used for stabilization of the estimators as well as detecting outliers. The performance of (PMLE) depends on the value of the tuning parameter. One way of selecting the tuning parameter, λ is the cross validation (CV), which uses the leave-one-out method [19]. Here, we propose an alternative efficient method to identify the outliers based on the penalized maximum likelihood estimation.

4.1. Model formulation

Suppose $\mathbf{y} = (y_1, y_2, \dots, y_k)^T$ is the vector of observed effect sizes and $\mathbf{x} = (x_{1i}, x_{2i}, \dots, x_{ki})^T$ is the vector of moderates for $i = 1, 2, \dots, p$. Suppose β is the unknown meta-regression parameters, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)^T$ is the vector of errors (homogeneous and heterogeneous errors), and τ^2 is the variance between studies so $\gamma_{(k \times 1)}$ incorporate into meta-regression to indicate whether the i th observation is an outlier (the non-zero components of γ will correspond to the outliers).

The interest here is to estimate β with $(p + 1)$ parameters, γ with k parameters, and τ^2 , so the number of the unknown parameters is $k + p + 2$ which is greater than the sample size k . Therefore, a possible approach of this problem is by introducing a penalty, instead of maximizing $\ell(\beta, \gamma, \tau^2)$, we maximize the following function

$$M(\beta, \gamma, \tau^2) = \ell(\beta, \gamma, \tau^2) - \lambda(\beta, \gamma, \tau^2).$$

For simplicity, we will work with the case of one moderator (i.e., $p = 1$), so we have to estimate $k + 3$ parameters. Consider the simple mixed-effect model

$$Y = \beta_0 + \beta_1 X + \gamma + e,$$

where γ is unknown shifted parameter corporate to indicate whether the i th observation is an outlier or not. Since $Y \sim \mathcal{N}(\beta_0 + \beta_1 X + \gamma, \tau^2 + \sigma_i^2)$, the log likelihood function will be given by

$$\ell(f(\beta_0, \beta_1, \gamma)) = -\frac{k}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^k \ln(\tau^2 + \sigma_i^2) - \frac{1}{2} \sum_{i=1}^k \frac{y_i - \beta_0 - \beta_1 x_i - \gamma_i}{\tau^2 + \sigma_i^2}. \quad (3)$$

Hence, the penalized log likelihood will be given by

$$M(\beta_0, \beta_1, \gamma) = \ell(f(\beta_0, \beta_1, \gamma)) - \lambda P(\beta_0, \beta_1, \gamma). \quad (4)$$

4.2. Parameters estimation

The smoothly clipped absolute deviation (SCAD) penalty function and its first derivative, respectively, are given by

$$SCAD_{\lambda}(\theta) = \begin{cases} \lambda |\theta|, & \text{for } |\theta| \leq \lambda, \\ \frac{-(\theta^2 - 2a\lambda|\theta| + \lambda^2)}{2(a-1)}, & \text{for } \lambda < |\theta| \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2}, & \text{for } |\theta| > a\lambda, \end{cases}$$

and

$$SCAD'_{\lambda}(\theta) = \begin{cases} \text{sgn}(\theta)\lambda, & \text{for } |\theta| \leq \lambda, \\ \text{sgn}(\theta)\frac{(a\lambda - |\theta|)}{a-1}, & \text{for } \lambda < |\theta| \leq a\lambda, \\ 0, & \text{for } |\theta| > a\lambda. \end{cases}$$

where $\text{sgn}(x) = 1$ if $x > 0$, and -1 if $x < 0$ [15]. The SCAD function is used where it is a nonconvex continuous and differentiable on $(-\infty, 0) \cup (0, \infty)$, but not differentiable at 0. Its derivative vanishes outside the interval $[-a\lambda, a\lambda]$ and it produces estimators with continuity, sparsity, and unbiasedness for the large coefficients [22]. Thus, using the SCAD penalty, the penalized log likelihood in (4), we get

$$M_{\lambda}(\beta, \gamma, \tau^2) = \ell(\beta, \gamma, \tau^2) - k \sum_{i=1}^k SCAD_{\lambda_k} \left(\left| \hat{\gamma}_i^{(0)} \right| \right) |\gamma_i|,$$

where λ_k is used to emphasize the dependence of λ on k . Note that maximizing M_{λ} is equivalent to the minimizing

$$M_{\lambda}(\beta, \gamma, \tau^2) = -\ell(\beta, \gamma, \tau^2) + k \sum_{i=1}^k SCAD_{\lambda_k} \left(\left| \hat{\gamma}_i^{(0)} \right| \right) |\gamma_i|.$$

However, the minimizing procedure is nontrivial, we use an iterative algorithm based on β and γ . The algorithm we utilize, to obtain the estimator of γ for each iteration, is the coordinate descent algorithm. This algorithm is one of the most efficient algorithms in processing large-scale data due to its simple operation and fast convergence [22]. The procedure of coordinate descent algorithm proceeds as follows:

Step 1: Set an initial estimate $\hat{\beta}^{(0)}$, which is obtained by the mixed-effect-model ignoring possible outliers, and set $m = 1$.

Step 2: Let $\hat{\gamma}^{(m)} = \text{argmin}_{\gamma} M_k(\hat{\beta}^{(m-1)}, \gamma)$, where the coordinate descent algorithm is adopted to do the maximization. In particular, for $i = 1, 2, \dots, k$,

$$\hat{\gamma}_i^{(m)} = \text{argmin}_{\gamma} M_k(\hat{\beta}^{(m-1)}, (\hat{\gamma}_1^{(m)}, \dots, \hat{\gamma}_{i-1}^{(m)}, \gamma_i, \hat{\gamma}_{i+1}^{(m-1)}, \dots, \hat{\gamma}_k^{(m-1)})^T),$$

and

$$\hat{\gamma}^{(m)} = (\hat{\gamma}_1^{(m)}, \dots, \hat{\gamma}_k^{(m)})^T.$$

Step 3: Let $\hat{\beta}^{(m)} = \operatorname{argmin}_{\gamma} M_k(\beta, \hat{\gamma}^{(m)})$,

Step 4: Update m by $m+1$ and repeat steps 2 and 3 until the prespecified convergence criterion is met.

Once the convergence is satisfied, we get the estimates of γ , and then we can judge the values of the coefficients for every study, whether they are outliers or not, (i.e., if the i th coefficient in γ is closed to 0 then the observation is not outlier, otherwise, it is a suspected outlier).

The estimation of model parameters including the estimation of tuning parameters as well as the implication of the coordinate descent algorithm has conducted by using R software, with partial use of **ncvreg** R package [10]. The proposed procedure can be extended straightforward to other meta-regression models with another settings including model type and the number of predictors.

5. Simulation study

The purpose of our simulations is to demonstrate the usefulness of our proposed procedure as well as its improved power performance compared to previous procedures. We conducted simulation studies to investigate the power of performance of the five outliers and influential diagnostic procedures, namely studentized deleted residuals, DFFITS values, Cook's distances, COVRATIO, and penalized maximum likelihood method [11].

The settings of the simulation study are summarized as follows: the study of specific effects y_i is simulated from the random-effect meta-analysis with $\mu = 0.5$, true effect $\beta_1 = 0.5$, $\beta_2 = 1$, and variance $(\hat{\sigma}_i^2 + \tau^2)$, $i = 1, 2, \dots, k$, where the $\hat{\sigma}_i^2$ are independently generated from a χ_1^2 distribution multiplied by 0.25 and then restricted to the interval (0.009, 0.6). Four values of the number of studies $k = 10, 20, 30$, and 50, and six values of $\tau^2 = 0, 0.2, 0.4, 0.6, 0.8$ and 1. The contamination of the generated meta-data with an outlier at position d is presented as follow $y_d^* = y_d + \omega$, where y_d^* is the contaminated value of the effect y_d at the d position and ω is the contamination level where $0 \leq \omega \leq 1$. The simulated data are obtained using the function **rma.glm()** from the **metafor** R package [20]. The mixed-effect-model is fitted and then the five outlier detection procedures are obtained. The power of performance is investigated by calculating the percentage of correct detection of the contaminated study at position d .

The results of simulation study reveal that the performance of all considered statistics has a direct relationship with the level of contamination ω . Furthermore, the proposed PMLM method outperforms the other statistics, followed by COVRATIO as shown in Figure 1.

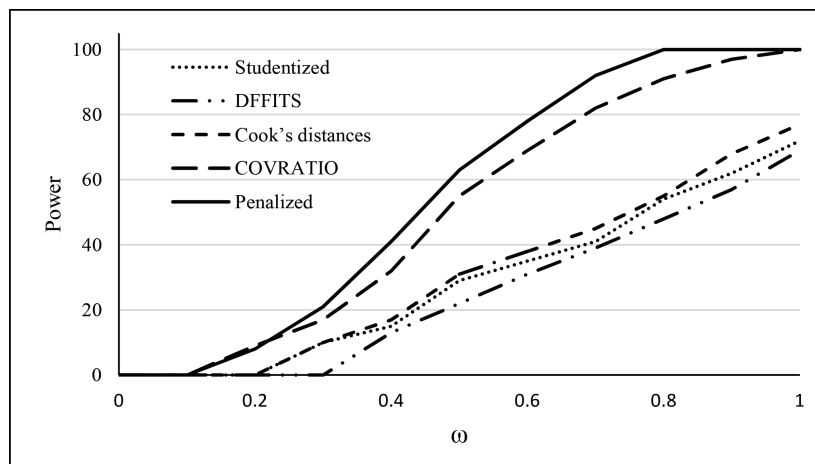


Figure 1. The power performance of outlier detection procedures for $k = 30$ and $\tau^2 = 0.4$.

The power of performance of the PMLE procedure has a direct relationship with the sample size, while it has an inverse relationship with the value of τ^2 as shown in Figure 2 and Figure 3. Same behavior has been observed for other values of k and τ^2 .

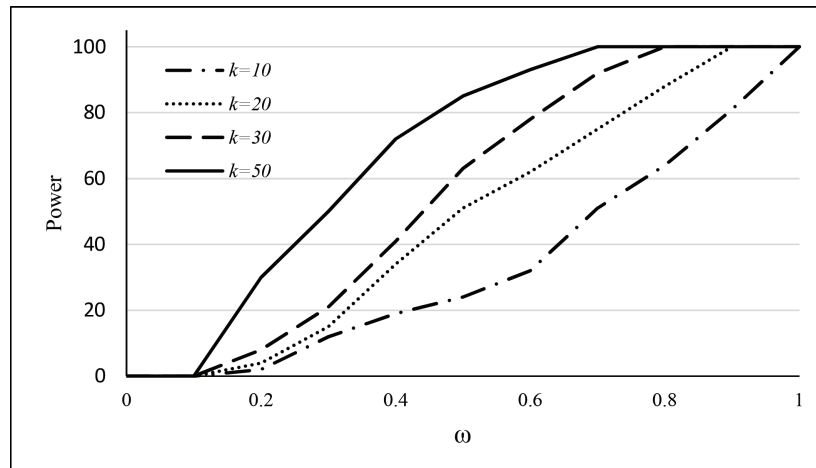


Figure 2. The power performance of PMLE procedure for $\tau^2 = 0.4$ and different values of k .

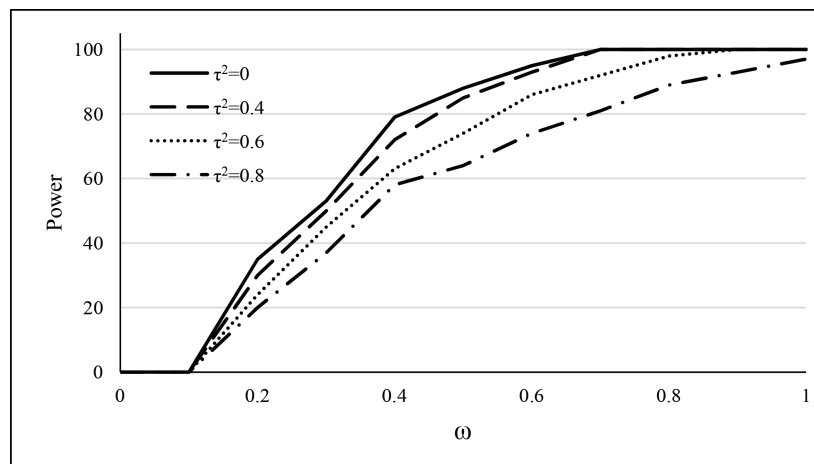


Figure 3. The power performance of PMLE procedure for $k = 50$ and different values of τ^2 .

6. Illustrative application

The authors of [6] conducted a meta-analysis of 48 school-based writing-to-learn programs, aimed to investigate the effectiveness of writing to learn interventions on academic achievement. Their results revealed that writing can has a small and positive impact on conventional measures of academic achievement. Recently,[1, 21] considered the analysis of 26 studies as a subset of the 48 studies to illustrate their procedures of detecting outliers and influential points. Here, we consider the same dataset consisting of 26 studies, and 6 variables summarized as follows:

Year: The years between 1980 and 1998 of conducting the study.

College: A dummy variable indicates whether the sample consisted of high-school or college students. There are 17 (65.4%) of the studies were conducted in colleges and 9 (34.6%) of the studies were conducted in schools.

Length: The intervention in weeks between 1 and 24 weeks, with mean 9.62 weeks and standard deviation 6.88 weeks.

Meta: A dummy variable indicates whether the intervention incorporated prompts for meta-cognitive reflection. There are 19 (73.1%) of the studies were the intervention incorporated prompts for meta-cognitive reflection and 7 (26.9%) of the studies were the intervention did not incorporated prompts for meta-cognitive reflection.

The effectiveness, y_i : It reflects the effect size of each study, where it was quantified in terms a continuous effect size of the standardized mean difference d_i . With mean 0.30 and standard deviation 0.39 and ranging between -0.32 and 1.46. Figure 4 shows the forest plot of the effect size y_i presented by black boxes at the center of 95% confidence intervals with whiskers. The weight of the black box indicates the precision, where the larger symbols indicate greater precision of the effect-size determined from its corresponding study.

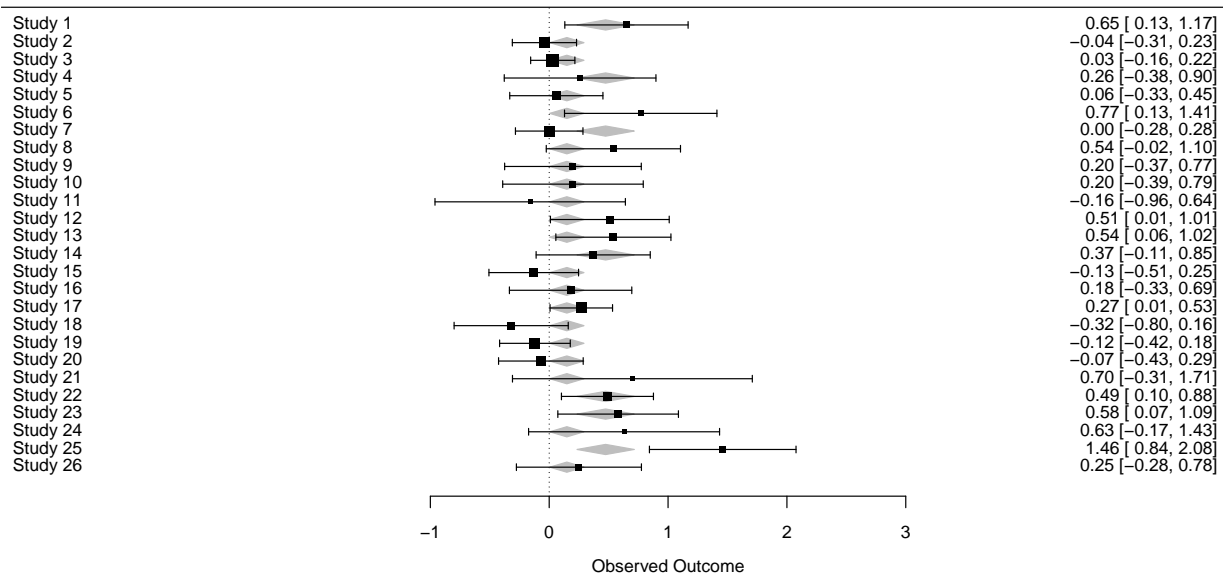


Figure 4. A forest plot of the effect size y_i with 95% confidence intervals.

The sampling variance, σ_i^2 : The variance of the standardized mean difference.

We fit the mixed-effects model for this data considering the effect-size with three explanatory variables named as “length of intervention” and the two dummy variables “college and meta-cognition”. In order to determine the appropriate meta-regression model, the test of heterogeneity Q is conducted. The observed variation $Q = 44.14$ which exceeds $\chi_{25,0.05}^2 = 37.652$. Furthermore, the expected variation (df) equals number of studies minus 1, ($df = 25$). Thus, $Q - df = 21.14$ which is greater than 0. Thus, this is a significant indication of the existence of variations between the studies. Furthermore, the estimated variance of the observed effect-size is $\hat{\tau}^2 = 0.0472$. Thus, the mixed-effects model will be used to fit the data.

Table 1 shows the coefficients estimates of the fitted model associated with its 95% confidence intervals. The estimates are not significantly differ from zero where the 95% confidence interval includes zeros. Therefore, we conclude that none of the estimated coefficients are significant. Figure 5 shows the residuals plots, where there are two points are inconsistent with other residuals, which are candidate to be outliers.

Table 1. Estimation of mixed-effect model coefficients.

Variable	Estimate	95% confidence interval
Intercept	0.12	(-0.13 to 0.37)
Length	0.01	(-0.01 to 0.03)
Meta-cognition	0.24	(-0.06 to 0.54)
College	-0.1	(-0.37 to 0.16)

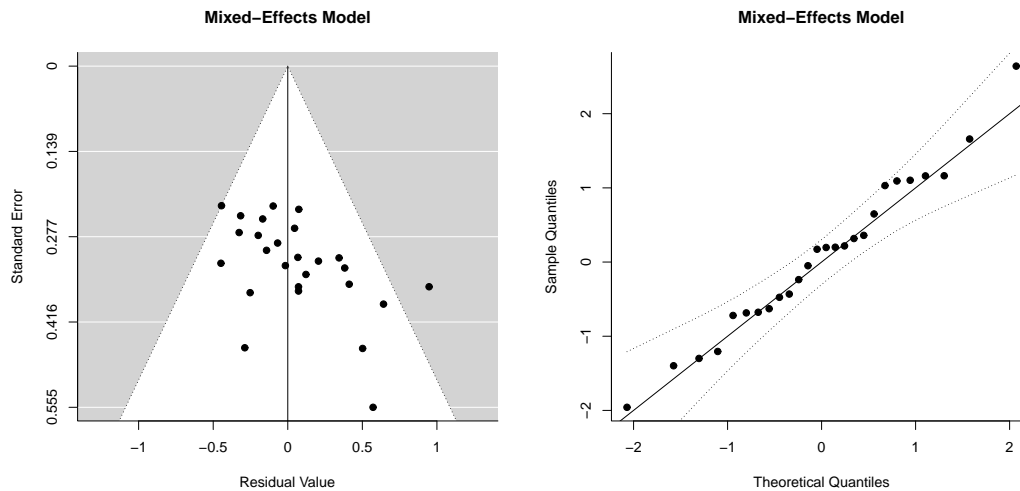


Figure 5. Plots of the residuals of mixed-effect model.

We apply the outlier detection procedures using the six statistics (a) studentized deleted residuals; (b) DFFITS values; (c) Cook's distances; (d) COVRATIO (e) $\hat{\tau}$ deleted and (f) QE deleted. All statistics reveal that two studies with numbers 7 and 25 are inconsistent with other studies (Figure 6), hence, these two studies can be considered as outliers. Figure 7 shows the plots of DFBETAS for four coefficients of the mixed-effect model. Intercept coefficient reveals the influence of studies with numbers 7, 15, 17, 19, and 25. The coefficient of length of interventions reveals the influence of studies with numbers 6, 8, 12, 13, 15, 19, 20, 22, 2, 24, and 25. The coefficient of meta-cognition reveals the influence of studies with numbers 7, 22 and 25. The coefficient of college reveals the influence of studies with numbers 7, 8, 12, 13, 15, 19, 22, and 25.

In order to apply the PMLE, data set need to be reconstructed via adding 26 dummy variables in matrix form such as S_{ij} , where $i, j = 1, 2, \dots, 26$ and $S_{ij} = 1$ for $i = j$ and $S_{ij} = 0$ otherwise. The SCAD penalty function is used by fixing the scalar $a = 3.7$ according to [22].

Figure 8 shows the cross-validation for different values of tuning parameter λ , it is obvious that the value of λ that minimizes the cross-validation error can be obtained by $\exp(-2.5) = 0.082$ with four selected variables. For further illustration, Figure 9 shows the way how variables are entered the model one at the time and at any given value of λ , several coefficients are zero.

Using the optimum tuning parameter λ and fitting the meta-regression model we have the coefficient estimates as listed in Table 2. There are four non zero coefficients. These coefficients are corresponding to, length of intervention, S_7 , S_{18} , and S_{25} . Therefore, three studies with numbers 7, 18 and 25 are defined as outliers, where the highest magnitude belongs to study number 25.

The results of different outliers' detection procedures are consistent, and there are three studies with numbers 7, 18 and 25 are outliers. In order to determine the appropriate meta-regression model, the test of heterogeneity Q is conducted based on the reduced data set. The observed variation $Q = 24.12$ which is less than the value of $\chi_{22, 0.05}^2 = 33.924$. Furthermore, the estimated variance of the observed effect size is $\hat{\tau}^2 = 0.0111$. Thus, there is

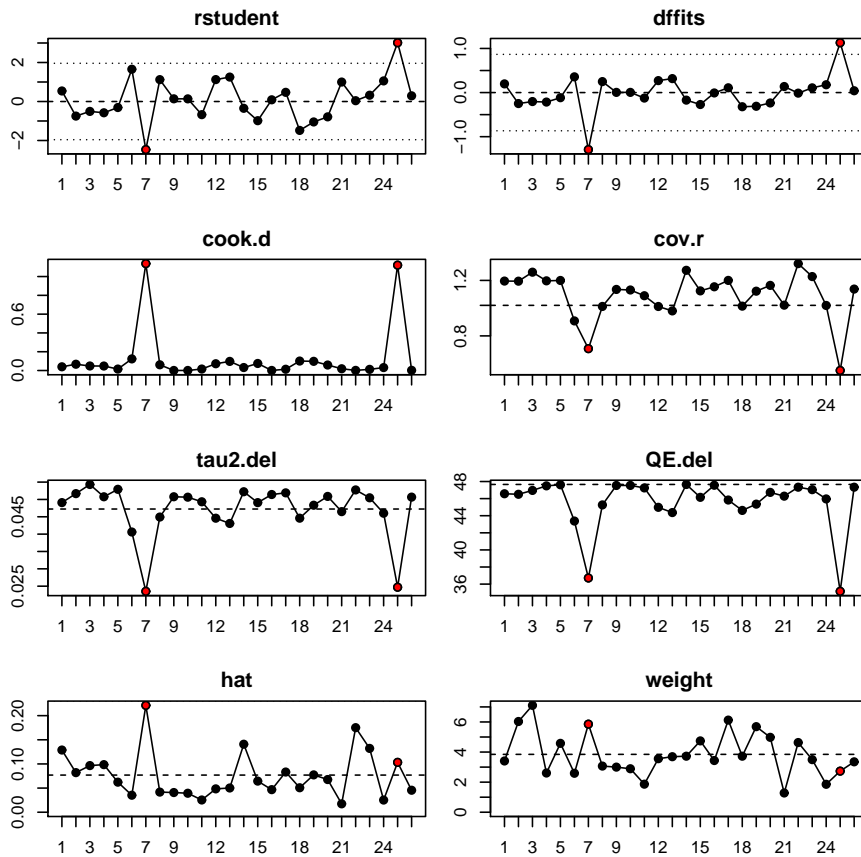


Figure 6. Plot of the (a) studentized deleted residuals; (b) DFFITS values; (c) Cook's distances; and (d) COVRATIO (e) $\hat{\tau}$ deleted and (f) QE deleted.

Table 2. MLE of mixed-effect meta-regression coefficients.

(Intercept)	Length	Meta	college	S_1	S_2
0.185	0.008	0.00	0.00	0.00	0.00
S_3	S_4	S_5	S_6	S_7	S_8
0.00	0.00	0.00	0.00	0.047	0.00
S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
0.00	0.00	0.00	0.00	0.00	0.00
S_{15}	S_{16}	S_{17}	S_{18}	S_{19}	S_{20}
0.00	0.00	0.00	-0.249	0.00	0.00
S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	S_{26}
0.00	0.00	0.00	0.00	1.062	0.00

Table 3. Estimation of mixed-effect model coefficients for reduced data.

Variable	Estimate	95% confidence interval
Intercept	0.21	(-0.10 to 0.44)
Length	0.03	(-0.02 to 0.04)
Meta-cognition	0.37	(0.12 to 0.62)
College	-0.13	(-0.33 to 0.07)

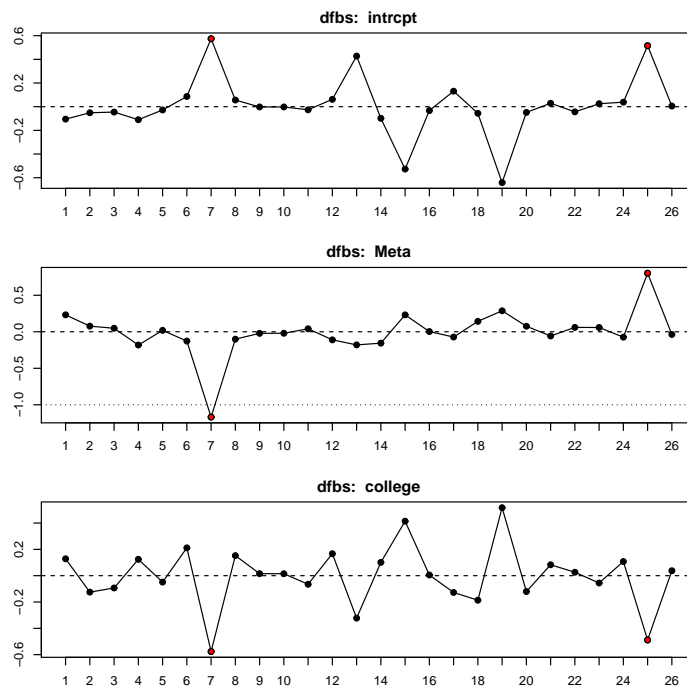


Figure 7. Plot of the DFBETAs of the mixed-effect model.

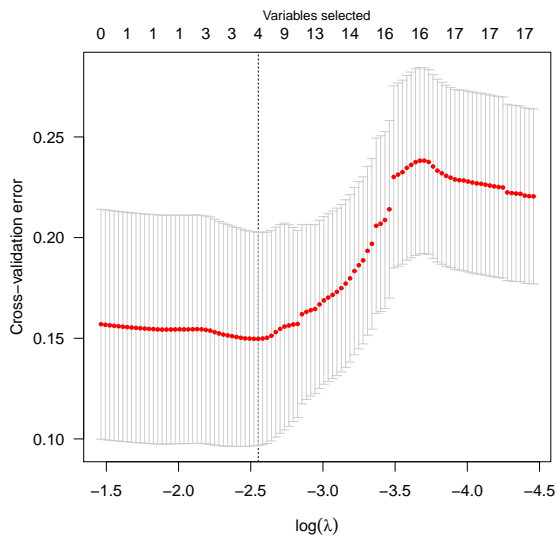


Figure 8. Cross-validation for different values of tuning parameter λ .

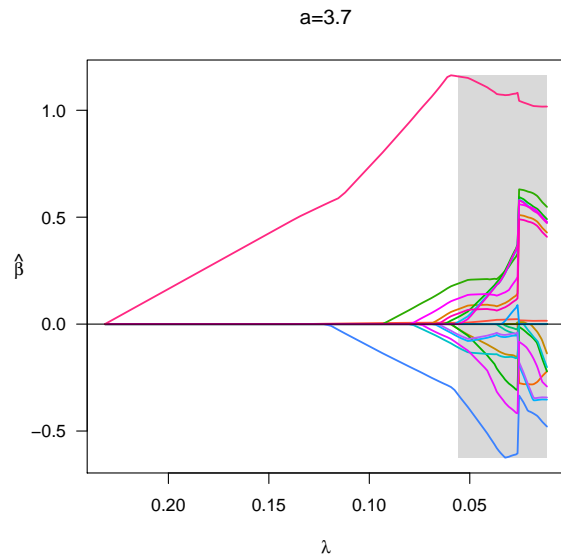


Figure 9. Coefficient estimates at different values of tuning parameter.

no significant indication of residual heterogeneity. The estimates of model coefficients for the reduced data are given in Table 3. The estimates show that meta-cognitive and college prompts variables become significant.

7. Conclusion

The problem of diagnostics in meta-analysis still need more investigation. The proposed procedure was introduced as an alternative to other available procedures considering the whole data set and it could identify multiple outliers in one run. Extensive simulation experiments demonstrate the usefulness of the proposed procedure as well as its improved power performance compared to previous procedures.

REFERENCES

1. E. Abed, *On the Detection of Outliers in Meta-regression*, Master Thesis, Al Azhar University - Gaza, Palestine, 2018.
2. H. Alavi Majid, Kh. Najafi Ghobadi, A. Akbarzadeh Baghban, N. Ahmadi, and E. Sajjadi, *Detecting and accommodating outliers in meta-analysis for evaluating effect of Albendazole on Ascaris lumbricoides infection*, Iran Red. Crescent Medical Journal, vol. 16, no. 5, e17648, 2014.
3. Z. Y. Algamal, *Variable Selection in Count Data Regression Model based on Firefly Algorithm*, Statistics, Optimization, & Information Computing, vol. 7, pp. 520-529, 2019.
4. RD. Baker, and D. Jackson, *A new approach to outliers in meta-analysis*. Healthcare Management Science, vol. 11, pp. 121–131, 2008.
5. R. Baker, and D. Jackson, *New models for describing outliers in meta-analysis*, Journal of Research Synthesis Method, vol. 7, no. 3, pp. 314–328, 2016.
6. R. Bangert-Drowns, M. Hurley, and B. Wilkinson, *The Effects of School-Based Writing-to-Learn Interventions on Academic Achievement*. A Meta-Analysis, vol. 74, pp. 29-58, 2004.
7. K. Beath, *A finite mixture method for outlier detection and robustness in meta-analysis*. Research Synthesis Methods, vol. 5, no. 4, pp. 285–293, 2014.
8. D. A. Belsley, E. Kuh, and R. E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. : New York: John Wiley & Sons, 1980 .
9. M. Borenstein, L. V. Hedges, J.P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*. Chichester: John Wiley Sons, 2009.
10. P. Breheny, *ncvreg: Regularization Paths for SCAD and MCP Penalized Regression Models*. R package version 3.11-1, 2019, URL <https://cran.r-project.org/web/packages/ncvreg/>.
11. S. E. Brockwell, and I. R. Gordon, *A comparison of statistical methods for meta-analysis*. Statistics in Medicine, vol. 20, pp. 825–840, 2001.
12. Gumedze F. N., and D. Jackson *A random effects variance shift model for detecting and accommodating outliers in meta-analysis*. BMC Medical Research Methodology, vol. 11, no. 19, pp. 1-9, 2011.
13. L. V. Hedges, and I. Olkin, *Statistical Method for Meta-Analysis*. Academic Press, INC. Harcourt Brace Jovanovich, Publishers, 1985.
14. J. P. T. Higgins, and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions*, Cochrane Book Series. John Wiley Sons, 2008.
15. J. Huang, and H. Xie, *Asymptotic oracle properties of SCAD-penalized least squares estimators*, Particles, Processes and Inverse Problems, vol. 55, pp. 149-166, 2007.
16. K. L. Lee, and S. G. Thompson, *. Flexible parametric models for random-effect distributions*. Statistics in Medicine, vol. 27, pp. 418–434, 2008.
17. S. Saleh, and A. H. Abuzaid, *Alternative Robust Variable Selection Procedures in Multiple Regression*, Statistics, Optimization & Information Computing, vol. 7, no. 4, pp. 816-825, 2019.
18. Y. She, and A. Owen, *Outlier Detection Using Nonconvex Penalized Regression*. Journal of the American Statistical Association, vol. 106, no. 494, pp. 626-639, 2011.
19. M. Ueki, and K. Fueda, *Optimal tuning parameter estimation in maximum penalized method*. Annals of the Institute of statistical Mathematics, vol. 63, pp. 413-438, 2010.
20. W. Viechtbauer, *metafor-Meta-Analysis*, Package for R, version 2.1-0, 2019. URL <http://cran.r-project.org/package=metafor>.
21. W. Viechtbauer, and M. C. Cheung, *Outlier and influence diagnostics for meta-analysis*. Journal of Research Synthesis Method, vol. 1, pp. 112-125, 2010.
22. J. Zhang, Y. Liu, and Y. Wu, *Exponential regression for censored data with outliers*. Journal of Statistical Computation and Simulation, vol. 86, pp. 431-442, 2015.
23. C. Zou, S. Tseng, and Z. Wang, *Outlier detection in general profiles using penalized regression method*, IE Transactions, vol. 45, no. 2, pp. 106-117, 2014.