COMSTA: 6671 | Model 3G | pp. 1–26 (col. fig: NIL)

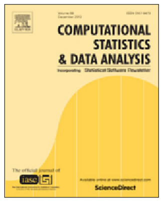
ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda



Robust estimation and confidence interval in meta-regression models

Dalei Yu [a], Chang Ding [a], Na He [b], Ruiwu Wang [c], Xiaohua Zhou [d], Lei Shi [a,*]

[a] School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, China
[b] Industry and Commerce Administration of Yunnan Province, Kunming 650228, China
[c] Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an 710072, China
[d] Department of Biostatistics, The University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 14 July 2017
Received in revised form 22 May 2018
Accepted 11 August 2018
Available online xxxx

Keywords:
Confidence interval
Meta-regression model
Outlier
Random effect
Robust estimation
Second-order stochastic expansion

ABSTRACT

Meta-analysis provides a quantitative method for combining results from independent studies with the same treatment. However, existing estimation methods are sensitive to the presence of outliers in the datasets. In this paper we study the robust estimation for the parameters in meta-regression, including the between-study variance and regression parameters. Huber's rho function and Tukey's biweight function are adopted to derive the formulae of robust maximum likelihood (ML) estimators. The corresponding algorithms are developed. The asymptotic confidence interval and second-order-corrected confidence interval are investigated. Extensive simulation studies are conducted to assess the performance of the proposed methodology, and our results show that the robust estimators are promising and outperform the conventional ML and restricted maximum likelihood estimators when outliers exist in the dataset. The proposed methods are applied in three case studies and the results further support the eligibility of our methods in practical situations.

1. Introduction

Meta-analysis provides a full and comprehensive summary of related studies which have addressed a similar question (Fiocco et al., 2012). In medical research, results from a single research project rarely provide exhaustive scientific evidence (Kicinski et al., 2015), therefore meta-analysis has been extensively used in the research areas such as clinical trials, biomedical studies, evidence based medicine, social science, education and so on. Among various models, the random effects meta-regression model has become a routinely used procedure in combining the results of a set of independent studies and assessing heterogeneity in effect sizes by introducing the between-study variance or heterogeneity (Sidik and Jonkman, 2005; Sutton and Higgins, 2008). When the distribution of the effect size is specified, the maximum likelihood (ML) estimator (Hardy and Thompson, 1996; Thompson and Sharp, 1999) and the restricted maximum likelihood (REML) estimator (Morris, 1983; Berkey et al., 1995; Thompson and Sharp, 1999) are usually adopted to conduct the parameter estimation in random effects meta-regression. Other parameter estimation methods are also studied extensively in literatures, see for example Sutton and Higgins (2008), Friedrich and Knapp (2013), Veroniki et al. (2016) and the references therein for details.

The current study is motivated by the fact that many meta-analyses will include at least a few studies yielding observed effects that appear to be outlying or extreme (Viechtbauer, 2010). In Section 2, we provide a motivating example where the data at hand are deemed to be contaminated by outlying trials. The presence of outlying studies could substantially

* Corresponding author.
E-mail address: shi_lei65@hotmail.com (L. Shi).

https://doi.org/10.1016/j.csda.2018.08.010

Please cite this article in press as: Yu D., et al., Robust estimation and confidence interval in meta-regression models. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.010.

alter the conclusions in a meta-analysis (Gumedze and Jackson, 2011) and the issue of outlying studies in meta-analysis has been paid increasing attention in recent years. As pointed out by an anonymous reviewer, one of the most commonly used methods for handling outliers is to windorize them, using Tukey's criteria of outliers as beyond 1.5 interquartile range from the lower or upper quartile of the distribution (Lipsey and Wilson, 2001). Focusing on the detection and identification of outliers, Viechtbauer and Cheung (2010) extended standard diagnostic procedures developed for linear regression analysis to the meta-analytic fixed- and random/mixed effects models. Moreover, Zhang et al. (2015) proposed several Bayesian outlier detection measures for detecting and handling trial-level outliers in network meta-analysis. Recently, Mavridis et al. (2016) developed a forward search algorithm for identifying outlying and influential studies in meta-analysis models. In addition to the detection of outlying studies, it is also important to accommodate the effects of these studies appropriately. In general, it is untenable to exclude studies from a meta-analysis on the basis of their results (Deeks et al., 2011). Besides, the identification and understanding of the reason for unusual study results can in itself lead to further understanding of the subject area (Gumedze and Jackson, 2011).

An alternative approach is to depart from simply deleting outlying studies but accommodate the outlying studies via robust estimation. Baker and Jackson (2008) proposed a model with long-tailed distribution for random effects, where the problematical outlying result is adjusted with a reduced weight. Instead of using the normally distributed random effects, Lee and Thompson (2008) proposed an innovative parametric model for random-effects distributions based on t-distribution and skew extensions to the normal and t-distributions, implemented using Markov Chain Monte Carlo methods. The new model accounts for the potential skewing and heavy tails in random-effects distributions. Also focusing on the flexible random effects distribution, Beath (2014) proposed a new meta-regression model based on finite mixture of a normal distributions, where a finite mixture of outliers and non-outliers is modeled. Beaths (2014) model provides a number of advantages over traditional methods of outlier detection and robustness, and is shown to be useful in practical meta-analysis. All the aforementioned studies fall into the class of "random effects contamination models", where the heavy-tailed random effects distribution is involved to achieve robustness (R package for implementing these methods are available at https://cran.r-project.org/web/packages/metaplus/). However, the random effects contamination models are deemed to be more complicated than the usually used meta-regression model. Therefore more computationally intensive methods are employed to conduct parameter estimation. For example, Lee and Thompson (2008) adopted Markov Chain Monte Carlo methods to implement parameter estimation, and R package "metaplus" employs expectation–maximization algorithm and adaptive Gaussian quadrature to overcome the numerical difficulties induced by heavy-tailed random effects distribution. More importantly, these more sophisticated models pose theoretical difficulties to the study of finite sample properties and thus it is in general very difficult (if not impossible) to study the improved confidence interval in finite sample settings (like Section 3.2 in current paper). Gumedze and Jackson (2011) considered a random effects variance shift model for detecting and accommodating outliers in meta-analysis, where a bootstrap-based likelihood ratio test statistic is used to check whether a study has potential inflated variance and can be considered as an outlier. Based on the confidence distribution, Xie et al. (2011) established a unifying framework for meta-analysis and applied it to study robust estimation. The above two studies are mainly focusing on the estimation of overall effect, where no moderators (study-level variables) are involved. Sidik and Jonkman (2006) developed an adjusted sandwich formula for the asymptotic variance of the overall effect estimator. Moreover, Hedges et al. (2010) provided an estimator of the covariance matrix of meta-regression coefficients that are robust against the potential correlation between effect sizes. It should be noted that Sidik and Jonkman (2006) and Hedges et al. (2010) are all focusing on the robust estimations of standard error which are robust to misspecification of the correlation structure of the effect size or poor estimation of τ^2 and are not designed for handling outliers.

In a closely related but different field, the robust estimation in linear mixed-effects model (LMM) has been studied extensively and a rich literature on approaches of robust LMMs has emerged in the past two decades. For example, Richardson and Welsh (1995) defined the robust maximum likelihood and robust restricted maximum likelihood functions based on Huber's rho function and the corresponding robust estimators were studied accordingly. Richardson (1997) extended the bounded influence estimation into the domain of LMMs, where the quadratic function in the normal likelihood is replaced by the Huber's rho function with a bounded derivative. Richardson (1997) also established the corresponding asymptotic properties of the proposed estimators. Staudenmayer et al. (2009) developed and implemented a LMM using the t-distribution and incorporated the outlier robustness into general design LMMs. Book-length treatments of LMMs with the flexible generalized skew-elliptical random effects can be found in Ma et al. (2004). Recently, Koller (2016) introduced an R package, robustlmm, to robustly fit LMMs. The package unifies a number of existing robust estimation methods in LMMs (e.g. the mixed-effects model with t-distribution, the Design Adaptive Scale estimation by Koller and Stahel (2011), and so on) and provides a very useful robust alternative for modeling the data with complicated correlation structure. However, the robust estimation methods in LMMs cannot be used directly in meta-regression. The reason is that *meta-regression model and* LMM *are essentially different*. Unlike in LMMs where repeated measurements are available, in meta-regression, we are focusing on effect size estimates and the original raw data are not typically available for analysis. Moreover, in meta-regression, the variances of random error terms are assumed to be known, whereas the variances of random error in LMMs are unknown and need to be estimated from the observable data. Additionally, existing studies about robust estimation in LMMs are mainly focusing on the asymptotic properties, but in many practical applications, the meta-analysis sample size is relatively small (Sidik and Jonkman, 2006) and therefore the corresponding finite sample properties are also of research interest. The above arguments indicate that we cannot apply the existing robust procedure in LMMs directly to handle the outlying studies in meta-analysis/regression. It is necessary to study the robust estimation in meta-regression, *independently*.

In current study, we aim to develop a new class of robust methods in the context of meta-regression models. There are two major contributions.

Please cite this article in press as: Yu D., et al., Robust estimation and confidence interval in meta-regression models. Computational Statistics and Data Analysis (2018), https://doi.org/10.1016/j.csda.2018.08.010.