

Parameter estimation in the random effects meta regression model

Dissertation

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

der Fakultät Statistik

der Technischen Universität Dortmund

vorgelegt von

Thomas W. D. Möbius

Dortmund, 2014

Thomas W. D. Möbius
Fakultät Statistik
Technische Universität Dortmund
44221 Dortmund

Betreuer und erster Gutachter:

PD Dr. Guido Knapp
Fakultät Statistik
Technische Universität Dortmund
44221 Dortmund

Zweiter Gutachter:

Prof. Dr. Katja Ickstadt
Fakultät Statistik
Technische Universität Dortmund
44221 Dortmund

Vorsitzender:

Prof. Dr. Jörg Rahmenführer
Fakultät Statistik
Technische Universität Dortmund
44221 Dortmund

Datum der Disputation: 25. August, 2014

© 2014 Thomas W. D. Möbius

Can you not see how you spend your limited time on earth?

– Professor Proton

Contents

Notation	iii
1 Preliminaries	1
1.1 Introduction	1
1.2 <i>Example:</i> Efficacy of the Bacillus Calmette–Guérin (BCG) vaccine	3
1.3 The <code>metagen</code> software package	6
2 The random effects meta regression model	9
2.1 Introduction	9
2.2 Model description – distribution free	9
2.3 Linear projections	11
2.4 Quadratic forms, inner products, and the squared length of the residual vector	12
2.5 Expectations of residual squared lengths	16
2.6 Residual squared lengths as functions of the heterogeneity	17
2.7 <i>Example:</i> Studying the BCG vaccine efficacy	21
2.8 Model description – with Gaussian responses	24
2.9 Likelihood functions and their derivatives	25
2.10 Likelihood function of a contrast and its derivatives	30
2.11 Independence of projected random vectors	32
2.12 Distributions of residual squared lengths	32
2.13 <i>Extension:</i> Mean responses as summary statistics	34
2.14 <i>Extension:</i> Binomial responses, risk ratios, and their limits	35
3 Theoretical background about generalised inference	39
3.1 Introduction	39
3.2 Basic concepts of generalised inference for confidence intervals	39
3.3 Higher dimensions: generalising generalised inference	41
3.4 Discussion	44
4 Inference on the heterogeneity parameter	45
4.1 Introduction	45
4.2 Methods of moments	45
4.3 Interval estimation for Gaussian responses	51
4.4 Methods based on maximum likelihood	52
4.5 Methods based on generalised inference	53
4.6 Performance study of inference methods about the heterogeneity	58
4.7 <i>Example:</i> Heterogeneity of clinical trials studying BCG vaccine efficacy	74

5	Inference on the regression coefficients	77
5.1	Introduction	77
5.2	Theoretical considerations	77
5.3	Point estimation of the regression coefficients	80
5.4	Interval estimation for Gaussian responses	80
5.5	Methods based on generalised inference	85
5.6	Performance study of inference methods about the regression coefficients	97
5.7	<i>Example:</i> Regressing BCG vaccine efficacy on geographic location	112
6	Discussion	117
	Appendix: Software documentation	119
	List of figures	173
	List of tables	177
	References	179

Notation

\mathbb{R}	the real line.
$\mathbb{R}_{\geq 0}$	the closed non-negative real half-line.
$\mathbb{R}_{\leq 0}$	the closed non-positive real half-line.
$\mathbb{R}_{> 0}$	the open strictly positive real half-line.
$\mathbb{R}_{< 0}$	the open strictly negative real half-line.
$(c)_{\mathbb{N}}$	for $c \in \mathbb{R}$, the scalar c rounded to the nearest integer.
$(c)_{\geq}$	for $c \in \mathbb{R}$, the scalar c truncated at 0 from below, $(c)_{\geq} = \max(0, c)$.
\mathbb{R}^k	the k -vector space of reals.
$\mathbb{R}^{p \times k}$	the vector space of $k \times p$ matrices with elements in \mathbb{R} .
$\mathbb{R}^{k \times k}$	the vector space of squared $k \times k$ matrices with elements in \mathbb{R} .
1_k	the k -vector in which each component is equal to the multiplicative unity in \mathbb{R} , hence $1_k \in \mathbb{R}^k$ and $1_k = (1, \dots, 1)$.
0_k	the k -vector in which each component is equal to the additive unity in \mathbb{R} , hence $0_k \in \mathbb{R}^k$ and $0_k = (0, \dots, 0)$.
δ_k	for a scalar $\delta \in \mathbb{R}$, the k -vector in which each component is equal to δ , hence $\delta_k \in \mathbb{R}^k$ and $\delta_k = (\delta, \dots, \delta)$.
\mathbb{G}_k	the subspace of units in $\mathbb{R}^{k \times k}$, i.e., each $A \in \mathbb{G}_k$ is of full rank and invertible. Under matrix multiplication $(\mathbb{G}_k, \circ, I_k)$ is a group.
I_k	the identity matrix and unity of \mathbb{G}_k .
$[\tau]_k$	for a scalar $\tau \in \mathbb{R}$, the $k \times k$ -matrix which has $\tau \cdot 1_k$ on its diagonal and zeros otherwise.
$[\delta]$	for a vector $\delta \in \mathbb{R}^k$, the $k \times k$ -matrix which has δ on its diagonal and zeros otherwise.
$\Omega_{\tau\delta}^{-1}$	$= [\tau]_k + [\delta]$.
A^{\top}	for a matrix $A \in \mathbb{R}^{k \times p}$, the transpose of A . Hence, $A^{\top} \in \mathbb{R}^{p \times k}$.
$\text{vec } A$	$\in \mathbb{R}^{kp}$, columns of a matrix $A \in \mathbb{R}^{k \times p}$ stacked as a vector.
$\text{rk } A$	$\in \mathbb{R}$, the rank of a matrix A .
$\text{tr } A$	$\in \mathbb{R}$, the trace of a matrix A .
$\text{diag } A$	$\in \mathbb{R}^k$, the diagonal of square matrix $A \in \mathbb{R}^{k \times k}$.
A^{-1}	for a matrix $A \in \mathbb{G}_k$, the inverse of A in \mathbb{G}_k .
A^+	for a matrix $A \in \mathbb{R}^{k \times k}$, the Moore-Penrose inverse of A .
$A \subset B$	for sets A and B , A is a subset of B .
$A \leq B$	for linear spaces A and B , A is a linear subspace of B .

$\langle A \rangle$	for a set of vectors $A = \{x_1, \dots, x_p\}$, the linear hull of the these vectors.
$\langle A \rangle$	for a matrix $A \in \mathbb{R}^{k \times p}$, the linear hull of the p columns of A .
$S \simeq T$	for linear spaces S and T , S and T are isomorph.
$S \hookrightarrow T$	for linear spaces S and T , S is embeddable in T .
x_j	for a vector $x \in \mathbb{R}^p$, the j th element of the vector x .
A_i	for a matrix $A \in \mathbb{R}^{k \times p}$, the i th row of the matrix A .
A_{ij}	for a matrix A , the element in the i th row and j th column of A , i.e., the j th element of A_i .
$f(\alpha)$	the image of α after applying the function f to α .
$f[\alpha]$	the image space of α after applying the function f to each element of α , i.e., $f[\alpha] = \{f(\beta) : \beta \in \alpha\}$.
b_A	$(x, y) \mapsto x^\top A y$ for $x, y \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times k}$. Usually, $A \in \mathbb{G}_k$ and b_A denotes a bilinear form on \mathbb{R}^k .
q_A	$x \mapsto x^\top A x$ for $x \in \mathbb{R}^k$ and $A \in \mathbb{R}^{k \times k}$. In particular, $q_A(x) = b_A(x, x)$, and q_A is called a quadratic form.
$\langle x, y \rangle_V$	$= b_{V^{-1}}(x, y)$.
$\ x\ _V$	$= \sqrt{b_{V^{-1}}(x, x)} = \sqrt{\langle x, x \rangle_V}$.
$d_V(x, y)$	$= \ x - y\ _V$.
$\langle x, y \rangle_{\tau\delta}$	$= \langle x, y \rangle_{\Omega_{\tau\delta}^{-1}}$.
$\ x\ _{\tau\delta}$	$= \ x\ _{\Omega_{\tau\delta}^{-1}}$.
$d_{\tau\delta}$	$= d_{\Omega_{\tau\delta}^{-1}}$.
(S, b)	for a linear space S and an inner product b on S , the inner product space defined by S and b .
$D_\beta f(\beta)$	the derivative/gradient of $f: \mathbb{R}^p \rightarrow \mathbb{R}^k$, $\beta \mapsto f(\beta)$ with respect to β .
$\text{Hess}_\beta f(\beta)$	the Hessian matrix of $f(\beta)$ with respect to β .
$\mathfrak{N}(\theta, \sigma^2)$	the Gaussian distribution function with mean $\theta \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{>0}$.
$\mathfrak{N}_k(\theta, V)$	the k -variate Gaussian distribution function with mean $\theta \in \mathbb{R}^k$ and variance-covariance matrix $V \in \mathbb{G}_k$.
\mathfrak{X}_p^2	the chi-squared distribution function with p degrees of freedom.
\mathfrak{T}_p	the Student's t-distribution function with p degrees of freedom.
$\text{Binom}(n, p)$	the binomial distribution function with n number of trials and success probability p .

$\mathfrak{F}_{\beta\tau\delta}$	the distribution function of a random effects meta regression model.
\mathfrak{z}_γ	the γ -quantile of a standard Gaussian distribution function.
$\mathfrak{r}_{p,\gamma}^2$	the γ -quantile of a chi-squared distribution function with p degrees of freedom.
$\mathfrak{t}_{p,\gamma}$	the γ -quantile of a Student's t-distribution function with p degrees of freedom.
X	depending on the context, usually a random element following some generic distribution, say \mathfrak{F} .
Y	usually a random element with sample space \mathbb{R}^k and following a random effects meta regression model.
x	usually an element in the sample space of a random element X .
y	usually an element in the sample space \mathbb{R}^k of a random element Y .
$\mathbb{E}(X)$	for a random element X , the expectation of X .
$\mathbb{V}(X)$	for a random element X , the variance-covariance of X .
$\mathbb{B}_\theta(X)$	for a random element X , the bias of X with respect to θ .
mean(x)	for $x = (x_1, \dots, x_k)$, the (empirical) mean of x .
var(x)	for $x = (x_1, \dots, x_k)$, the (empirical) variance of x .
$X \sim Y$	for random elements X and Y , X and Y are equally distributed.
$X \sim \mathfrak{F}$	for a random elements X and distribution function \mathfrak{F} , \mathfrak{F} is a distribution function of X .
$X \sim (\theta, V)$	for a random k -vector X , for $\theta \in \mathbb{R}^k$, and $V \in \mathbb{G}_k$, X has mean θ and variance-covariance V .
$X \sim (\theta, \sigma^2)$	for a random variable X , for $\theta \in \mathbb{R}$, and $\sigma \in \mathbb{R}_{>0}$, X has mean θ and standard deviation σ .
$X_n \xrightarrow{P} \theta$	for a random element X , X is consistent for θ .
$\mathfrak{F}_n \rightsquigarrow \mathfrak{F}$	for a sequence of distribution functions \mathfrak{F}_n , $n \in \mathbb{N}$, \mathfrak{F}_n converges in distribution to \mathfrak{F} .
$X_n \rightsquigarrow \mathfrak{F}$	for a sequence of random elements X_n with distribution functions \mathfrak{F}_n , $\mathfrak{F}_n \rightsquigarrow \mathfrak{F}$.
X	$\in \mathbb{R}^{k \times p}$, frequently denotes the design matrix of a regression model.
$V_{\tau\delta}$	$= (X^\top \Omega_{\tau\delta} X)$.
$B_{\tau\delta}$	$= V_{\tau\delta} X^\top \Omega_{\tau\delta} = (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta}$.
$H_{\tau\delta}$	$= X B_{\tau\delta} = X (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta}$.
$E_{\tau\delta}$	$= I_k - H_{\tau\delta} = I_k - X (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta}$.

$$q_{y\delta}(\tau) = \|E_{\tau\delta}y\|_{\tau\delta}^2.$$
$$p_{y\delta}(\eta) \quad \text{the inverse function of } q_{y\delta} = q_{y\delta}(\tau).$$

ANOVA analysis of variance.

REML restricted maximum likelihood.

BCG Bacillus Calmette–Guérin.

GPLv3 GNU General Public License Version 3, 29 June 2007.

1 Preliminaries

1.1 Introduction

Meta analysis aims to combine the effect estimates of various related studies, trials, or experiments. It is a highly important statistical tool with various areas of application and has been a field of active research for many years, (Hedges and Olkin, 1985), (Whitehead, 2002), (Hartung et al., 2008).

When combining different outcomes into an overall analysis, one cannot expect that each study specific outcome will be centred around the very same value. Such differences in location occur due to individual study specific design features. Usually, such differences get accounted for by including a heterogeneity parameter into the model. Sometimes, however, these differences can be explained by known study specific covariates. Additionally to adding another variance component, one can account for these location differences by including these study specific covariates into the model. This approach also allows for understanding the rationale behind the observed differences. Berkey et al. (1995) called the resulting model the *random effects meta regression model*.

Depending on the underlying scientific question, different model parameters of the random effects meta regression model are of interest. Either the interest lies in accessing the amount of heterogeneity itself, as Hartung and Knapp (2005) and Knapp et al. (2006), or one is interested in the regression coefficients and, thus, aims to study the causes that may explain the occurring heterogeneity. The latter has been studied by Thompson and Sharp (1999) who approached the problem from a Bayesian and a classical weighted least squares point of view. Knapp and Hartung (2003) proposed an adjustment to this weighted least squares approach which greatly improved its performance. One objective in this dissertation is to test whether a construction based on generalised inference principles may again improve the performance for statistical inference especially when dealing with small sample sizes. Generalised inference principles have already been successfully applied to inference on the overall treatment effect and between study variance in a meta analysis setting with Gaussian distributed responses in (Tian, 2008a) and (Tian, 2008b).

The outline of this work is as follows: Section 1.2 presents a data set that will be used as an example throughout the text to emphasize certain aspects of the presented methods and their theory. Additionally, the text is accompanied by a general purpose R-package called `metagen`, which provides a range of methods for random effects models and also facilitates the set-up and realisation of large scale simulation studies evaluating the performance of these methods. Some general information about `metagen` are discussed in Section 1.3.

Chapter 2 introduces the random effects meta regression model. Some frequently used notation and important linear projections on the sample space of the random effects model are introduced in Section 2.3. Related quadratic forms and their induced inner products on the sample space are discussed in Section 2.4. In Sections 2.5 to 2.11, fundamental results are laid out allowing to unify many of the subsequently presented inferential methods into a common framework. The developed functionals are applied to real world data in Section 2.7. Sections 2.2 through 2.7 present results about the random effects meta regression model, which do not rely on any particular parametric distributional assumptions. Section 2.8 describes the model in its most frequent form with Gaussian distributed error terms. Likelihood functions of the model are carried out in Sections 2.9 and 2.10. The distributions of previously introduced quadratic forms are discussed in Section 2.12 consequently laying the ground for the constructions of confidence intervals based on asymptotic statistical theory. Some extensions to the random effects meta regression model itself are discussed in Sections 2.13 and 2.14.

The *generalised inference* principle was coined under this term by Tsui and Weerahandi (1989). The paper was followed a few years later by (Weerahandi, 1993) and (Weerahandi, 1995). An introduction to the basic ideas of this principle are summarised Chapter 3, in particular in Sections 3.1 and 3.2. Section 3.3 is a contribution to extending the theory of the generalised principle from one to multidimensional parameters of interest, which is original and has not been published yet. Some additional discussion is found in Section 3.4.

All inferential methods, point and interval estimates, concerning the heterogeneity parameter of the random effects meta regression model are gathered in Chapter 4. After an introduction in Section 4.1, different method of moments estimators for the heterogeneity parameter are discussed in Section 4.2. Each of these estimators were originally proposed for the random effects meta analysis model that does not include any covariance terms. Extensive work went into establishing a general framework in which each of these estimators could be integrated. The iterative generalisations suggested in this chapter are new and have not been published or discussed previously. Their almost apparent appearance in the theory are ramifications of the earlier developed framework of Chapter 2. Confidence intervals for the heterogeneity parameter are discussed in Section 4.3. Maximum likelihood estimators for the heterogeneity parameter are introduced in Section 4.4. A new genuine method for point and interval estimation of the heterogeneity parameter based on generalised inference principles is introduced in Section 4.5.

The first of the two performance studies of this dissertation is presented in Section 4.6. This section starts with an outline of the set-up of the corresponding simulation study which will also be used in the performance study of Section 5.6. It is the first time this kind of experimental design is applied to the performance study of inferential methods of a parametric model such as this dissertation's random effects

model. Results of the performance study are discussed for the presented point and interval estimates for the heterogeneity parameter. An exemplary application of the methods to real data can be found in Section 4.7.

Inference on the regression coefficients of the random effects meta regression model are put together in Chapter 5. After an introduction and some theoretical considerations in Sections 5.1 and 5.2, point estimators for the regression coefficients are defined in Section 5.3. Three important interval estimates for the regression coefficients based on asymptotic statistical theory are introduced in Section 5.4 including the current state-of-the-art estimator by Knapp and Hartung (2003) and including a discussion about inference on the full coefficient vector.

Section 5.5 develops new point and new interval estimates for the regression coefficients based on generalised inference principles. Different approaches and strategies are discussed including computational issues and generalisations to higher dimensions. Most of the content of Section 5.5 is genuine, though, some of the results got published in (Friedrich and Knapp, 2013) prior to finalising this dissertation.

The second of the two simulation studies is summarised in Section 5.6. The general set-up follows the same experimental design as in Section 4.6. Here, each possible combination of inferential methods for the regression coefficients, point and interval estimators, are discussed and evaluated in this section. Knapp and Hartung (2003) have defined, in fact, two different interval estimates, though, only one of which is commonly applied and implemented in statistical software. The other has not yet been discussed in the literature. The simulation study of this chapter also presents results about this interval estimator including some surprising results. The chapter ends with an application of the discussed methods to real data in Section 5.7.

The dissertation concludes with some final remarks and a discussion of open problems in Chapter 6.

1.2 *Example*: Efficacy of the Bacillus Calmette–Guérin (BCG) vaccine

Throughout the text, a data set is used for illustrative purposes that is well discussed in the literature, e.g., (Berkey et al., 1995), (Knapp and Hartung, 2003), and (Friedrich and Knapp, 2013). The data consist of a combination of 13 clinical trials which evaluated the efficacy of the Bacillus Calmette–Guérin (BCG) vaccine for the prevention of tuberculosis.

Even though BCG is in use as a vaccine in humans as early as 1921, the reasons for its highly variable efficacy are still a topic of current research. One conjecture is that the presence of certain environmental mycobacteria provides a certain level of natural immunity against tuberculosis within an exposed population, and that

Trial	Author	Year	Vaccinated		Not vaccinated		Ab- solute
			Dis- eased	Not diseased	Dis- eased	Not diseased	Lati- tude
A	Aronson	1948	4	119	11	128	44
B	Ferguson & Simes	1949	6	300	29	274	55
C	Rosenthal et al.	1960	3	228	11	209	42
D	Hart & Sutherland	1977	62	13536	248	12619	52
E	Frimodt-Moller et al.	1973	33	5036	47	5761	13
F	Stein & Aronson	1953	180	1361	372	1079	44
G	Vandiviere et al.	1973	8	2537	10	619	19
H	TPT Madras	1980	505	87886	499	87892	13
I	Coetzee & Berjak	1968	29	7470	45	7232	27
J	Rosenthal et al.	1961	17	1699	65	1600	42
K	Comstock et al.	1974	186	50448	141	27197	18
L	Comstock & Webster	1969	5	2493	3	2338	33
M	Comstock et al.	1976	27	16886	29	17825	33

Table 1.1 The results of 13 clinical trials evaluating the efficacy of the BCG vaccine.

these mycobacteria are more likely to be found in tropical environments, (Ginsberg, 1998). For this reason, the distance of a clinic to the equator may serve as a potential influential covariate, and can, therefore, be treated as a surrogate for the environmental effect on efficacy in the analysis. The surrogate is measured by the absolute value of the latitude of the geographic location of a clinic.

The plan is to regress on this covariate, and thus, asking the question whether the presence of such mycobacteria do indeed interfere with the efficacy of the BCG vaccine.

The data are put together in Table 1.1, which is accessible via the `metafor` package of Viechtbauer (2010). The data are unbalanced in many aspects. Figure 1.1, for example, shows how unbalanced the data are with respect to differences in total study size. Study sizes range from a minimum of 262 in trial A to 176782 in trial H. But the trials not only differ in total sample size but also in the proportion of vaccinated and non-vaccinated subjects. Whereas most trials are relatively balanced between the two groups, Figure 1.2 shows that in some trials the vaccinated group is considerably larger than the non-vaccinated. See, for example, trials K and G. Other studies, on the other hand, are very balanced, such as trial H. All this makes the data difficult to access for statistical inference methods. Figure 1.3 shows a

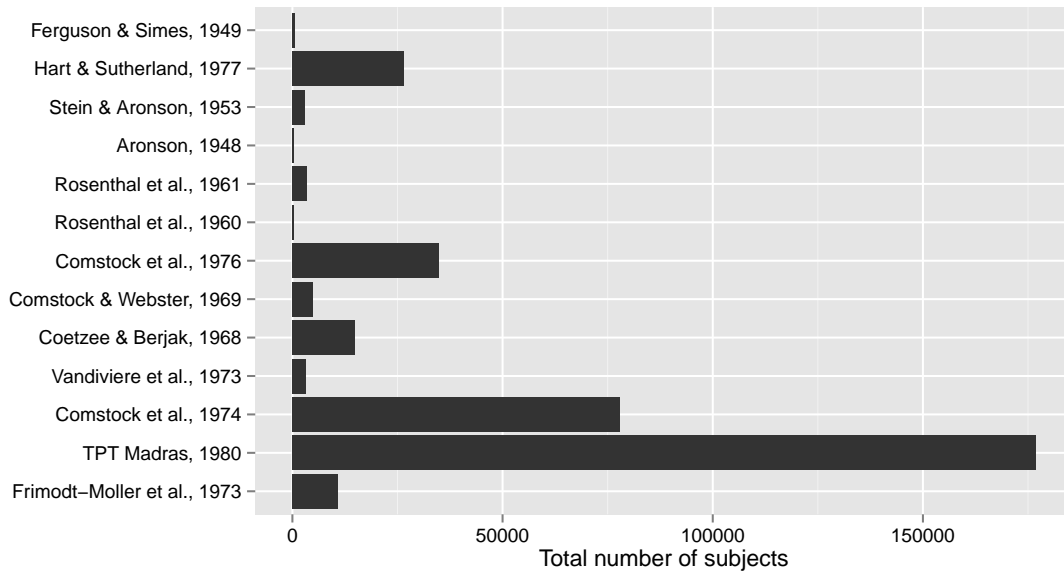


Figure 1.1 Differences in total study sizes of 13 clinical trials evaluating the efficacy of the BCG vaccine.

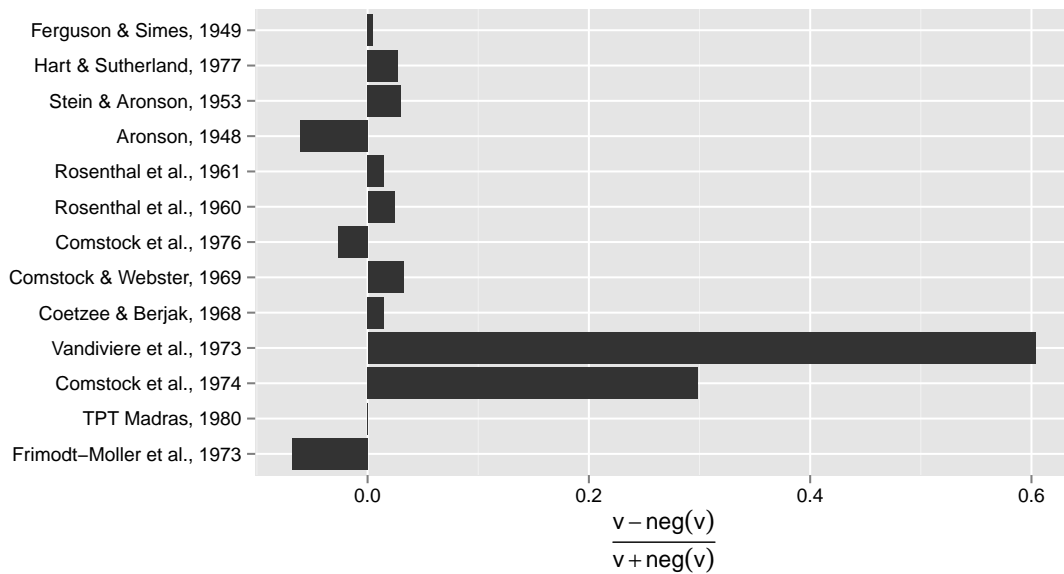


Figure 1.2 Differences in subject assignments of 13 clinical trials evaluating the efficacy of the BCG vaccine. The measure of balance is $\frac{v - \text{neg}(v)}{v + \text{neg}(v)}$ where v is the number of vaccinated and $\text{neg}(v)$ the number of non-vaccinated subjects in a study.

forest plot of the data together with 95% confidence intervals for the logarithm of the relative risks of the studies. The studies in all plots are sorted bottom to top by

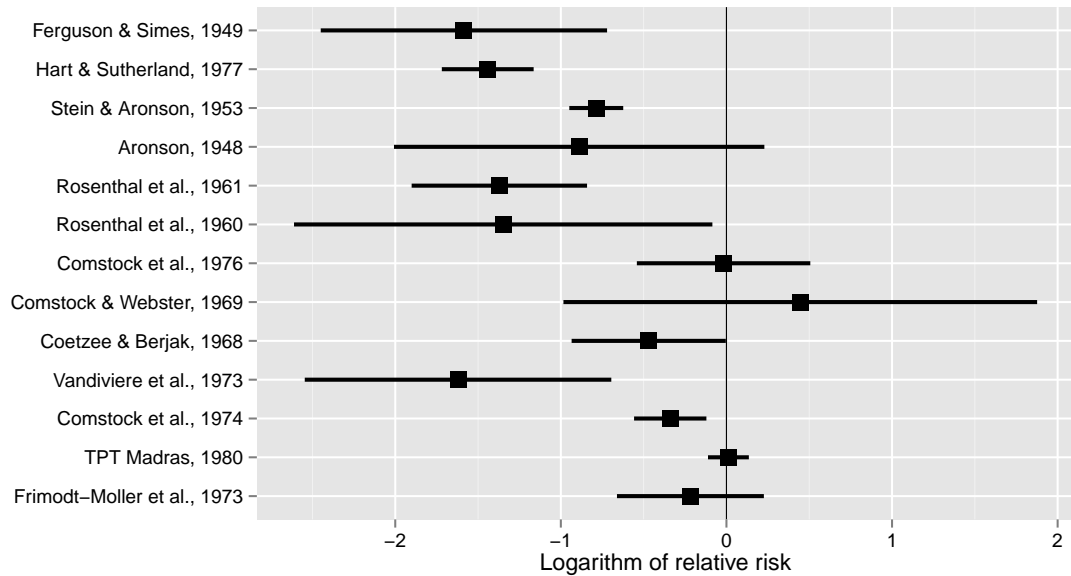


Figure 1.3 Forest plot of 13 clinical trials evaluating the efficacy of the BCG vaccine. Trials are sorted top to bottom by their absolute latitude. Intervals show 95% confidence intervals for the logarithm of relative risk.

their absolute latitude with the highest absolute latitude at the top. Thus, clinical trials at the lower end of the plots are closer to the equator than studies at the upper end. As Figure 1.3 shows, due to the discussed sample size differences of the clinical trials, the differences in accuracy of the estimates are large. The plot also shows, however, that there is nonetheless good reason to suspect a trend in efficacy with respect to latitude.

A random effects meta regression model is used to give suitable answers for these kind of questions. This model is often used to analyse data such as the example given here, in which the model is used to compare the relative risks of two different treatments. Such data consists of counts of different binary random variables such as in Table 1.1. Even though each respected count follows a binomial distribution, the relative risks ratios can be analysed in a meta regression model since the logarithm of the relative risks are approximately Gaussian distributed.

1.3 The metagen software package

Scientific practice advocates to accompany any scientific work with the means and necessary tools that allow for its reproducibility. Although, the emphasis of this work is primarily of theoretic character – to situate common inference techniques and state-of-the-art inferential methods in a general theoretical framework and to

develop new inferential methods for the random effects meta regression model – a considerable and comparable amount of work went into the development of a software package that implements important aspects of the presented theory.

Thus, the text is accompanied by a software package with the name `metagen`, (Möbius, 2014). The package is implemented for the widespread statistical software environment R, (R Development Core Team, 2010). It is published freely available under the GPLv3 license and including all source code at CRAN:

<http://cran.r-project.org/web/packages/metagen/index.html>.

The code-base itself is hosted at GitHub:

<http://00tau.github.io/metagen/>.

This website also includes links on how to contribute to the package, or how to simply submit new feature requests.

The package `metagen` contains all functions and tools needed to reproduce all results, figures, plots, and tables of this document. The latest version hosted on CRAN can easily be installed by opening an R-session and executing:

```
install.packages("metagen", dependencies=TRUE)
```

The text frequently contains small software excursions which contain fully functional code samples that explain how to use the `metagen` package. The following excursion, e.g., shows how to reproduce the sample plots of Section 1.2.

Excursion 1.1 The following lines reproduce Figure 1.1, Figure 1.2, and Figure 1.3, which show differences between study sizes, unbalances in group assignments, and a forest plot.

```
library(metagen)
bcg <- bcgVaccineData()
plotStudySizes(bcg)
plotStudyUnbalance(bcg)
plotStudyForest(bcg)
```

If desirable, this makes it possible to use the package in an interactive fashion, side-by-side, next to reading this text.

The package can, however, not only be used as a supporting tool for this particular text but `metagen` is, in fact, a fully fledged statistical software package that enables to perform any kind of desirable *random effects meta regression* or *meta*

analysis. The software sections by themselves can be used as short tutorials and contain all necessary information on how set up such an analysis. The software package is, of course, fully documented and all functions contain informative manual pages that integrate fully into R's native help system. The reference manual, which is also provided with the distribution of **metagen**, can be found in the Appendix.

Apart from to the inferential methods, the package also contains functions and algorithms that allow to reproduce the simulation studies included in this document. The package allows to set up custom simulation studies with free parameter choice. The package is written in a modular way allowing the user to extend the built-in functionality by own inferential methods. The purpose was to establish a tool box that allows researchers working with meta regression and meta analysis problems to test their own inferential methods in a quick, non-distracting, and standardised way to common state-of-the-art inferential methods.

This part of the software is, of course, also fully documented in the same fashion and accessible via R's native help system. Thus, the purpose of the accompanying package **metagen** are threefold: (i) acknowledging good scientific practise, thus, allowing the reproducibility of all mentioned results, (ii) implementing an interface for analysing data within a meta regression and meta analysis framework, and, finally, (iii) allowing to compare new inference methods in a modular and standardised way to known methods of the field.

2 The random effects meta regression model

2.1 Introduction

Before developing the random effects regression model with Gaussian responses, a mathematical description of the model is developed in term of its moments first. For the k -variate random vector Y with mean $\theta \in \mathbb{R}^k$ and variance-covariance matrix $V \in \mathbb{R}^{k \times k}$, write

$$Y \sim (\theta, V). \quad (2.1)$$

This is standard notation which simply defines a relation on the set of random elements with sample space \mathbb{R}^k and the space $\mathbb{R}^k \times \mathbb{V}$ where \mathbb{V} shall denotes the open, convex cone of positive definite matrices in $\mathbb{R}^{k \times k}$. A distribution free model can be understood as the \sim -pre-image of a pair (θ, V) . Thus, the model is identified with its expectation and variance-covariance structure.

Let $1_k \in \mathbb{R}^k$ denote the k -vector that only consists of 1's. If $c \in \mathbb{R}^k$, the notation $[c]$ shall be used for the diagonal matrix $C \in \mathbb{R}^{k \times k}$ which has c on its diagonal and zeros otherwise. For a scalar $c \in \mathbb{R}$, the notation $[c]_k$ is used for the diagonal matrix $C \in \mathbb{R}^{k \times k}$ which has $c1_k$ on its diagonal and zeros otherwise. In particular, $[c]_k = [c1_k] = cI_k$, and

$$[\tau]_k = \begin{pmatrix} \tau & & \\ & \ddots & \\ & & \tau \end{pmatrix}, \quad [\delta] = \begin{pmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_k \end{pmatrix},$$

and

$$([\tau]_k + [\delta])^{-1} = \begin{pmatrix} \frac{1}{\tau + \delta_1} & & \\ & \ddots & \\ & & \frac{1}{\tau + \delta_k} \end{pmatrix}$$

for $\tau \in \mathbb{R}$ and $\delta \in \mathbb{R}^k$.

2.2 Model description – distribution free

The random effects meta regression model is a hierarchical model that models the univariate effect measures of a number of different studies which may differ

in their location and their variability. Let $Y = (Y_1, \dots, Y_k)^\top$ be an independent random k -vector where each Y_j corresponds to the outcome of a single study. It is assumed that Y is independent. For variance components $\delta_j, \tau \in \mathbb{R}_{>0}$, location component $\beta \in \mathbb{R}^p$, and a vector of covariates $x_j \in \mathbb{R}^k$, $j = 1, \dots, k$, assume that

$$Y_j | \theta_j \sim (\theta_j, \delta_j), \quad (2.2)$$

$$\theta_j \sim (x_j^\top \beta, \tau). \quad (2.3)$$

In (2.2), $Y_j | \theta_j$ denotes the response of the j th study with expected value θ_j and variance δ_j . The parameter δ_j is called the *within study* variances and the vector $\delta = (\delta_1, \dots, \delta_k)^\top$ the *heteroscedasticity* parameter vector of the model. A random effect is introduced when assuming that the studies are themselves a random selection of otherwise also possible studies which locations, i.e., their expected responses, may vary. This random effect is modelled by the *heterogeneity* parameter τ in (2.3), also called the *between study* variance. We assume that there exists a linear relation between $\mathbb{E}(\theta_j)$ and the covariates x_j , in signs: $\mathbb{E}(\theta_j) = x_j^\top \beta$, for some unknown *regression coefficients* $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$. Each *covariate* x_j^\top equals the j th row of a *design* matrix $X \in \mathbb{R}^{k \times p}$ with rank p strictly less than $k - 1$. It is common to define the design matrix X in the form

$$X = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{k2} & \cdots & x_{kp} \end{pmatrix}. \quad (2.4)$$

Defined in this way, the first component β_1 of the regression coefficient vector β equals the intercept of the model. In the special case of $p = 1$, $\beta \in \mathbb{R}$, and $X = \mathbf{1}_k \in \mathbb{R}^{k \times 1}$, the model in (2.2) and (2.3) reduces to the common *random effects meta analysis model*. Thus, all results of this text concerning the random effects meta regression model inherit results which also apply to conventional meta analysis.

From the rank restriction, $\text{rk}(X) = p < k - 1$, follows that X has full column rank, and, thus, its induced linear map is injective. Hence, if either the δ_j 's or τ are known parameters, the above model would be identifiable for β . The model is, however, not identifiable for the full set $(\beta^\top, \tau, \delta_1, \dots, \delta_k)$ of parameters. Usually, the within study variances are reported together with the estimated effects as part of the summary statistics of the respected studies, thus, enabling to draw inference on τ and β .

Definition 2.1 Let $X \in \mathbb{R}^{k \times p}$ with $\text{rk}(X) = p < k - 1$. The *random effects meta regression model* is defined as

$$(X\beta, [\tau]_k + [\delta]) \quad (2.5)$$

for $\beta \in \mathbb{R}^p$, $\tau \in \mathbb{R}_{>0}$ and $\delta \in \mathbb{R}_{>0}^k$.

This is the reference model for most parts of this text. As said before, for theoretical considerations, the variance components δ need to be known. In practice, this is hardly the case, and each δ_j will usually be set equal to some summary statistic that is reported together with the respected study. An extension to the random effects meta regression model will be discussed in Section 2.13 that allows to model this uncertainty. The model described in Section 2.13 will also be used to simulate data for the performance studies in order to provide a realistic impression of the performance of the upcoming methods. Another modelling strategy is discussed in Section 2.14 which models the underlying data generating process of data such as the exemplary data presented in Section 1.2.

2.3 Linear projections

For notational brevity, define

$$\Omega_{\tau\delta} := \begin{pmatrix} \frac{1}{\tau + \delta_1} & & \\ & \ddots & \\ & & \frac{1}{\tau + \delta_k} \end{pmatrix}. \quad (2.6)$$

In other words, $\Omega_{\tau\delta}^{-1} = [\tau]_k + [\delta]$ and $Y \sim (X\beta, \Omega_{\tau\delta}^{-1})$. Define

$$V_{\tau\delta} := (X^\top \Omega_{\tau\delta} X)^{-1} \in \mathbb{R}^{p \times p}, \quad (2.7)$$

$$B_{\tau\delta} := V_{\tau\delta} X^\top \Omega_{\tau\delta} = (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta} \in \mathbb{R}^{p \times k}, \quad (2.8)$$

$$H_{\tau\delta} := X B_{\tau\delta} = X (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta} \in \mathbb{R}^{k \times k}, \quad (2.9)$$

$$E_{\tau\delta} := I_k - H_{\tau\delta} = I_k - X (X^\top \Omega_{\tau\delta} X)^{-1} X^\top \Omega_{\tau\delta} \in \mathbb{R}^{k \times k}. \quad (2.10)$$

The matrices $B_{\tau\delta}$, $H_{\tau\delta}$ and $E_{\tau\delta}$ are linear functions acting on \mathbb{R}^k , the image space of Y . In particular, each of these functions is a statistic of the random Y . For brevity, and only if the values of τ and δ are clear from the context, the subscripts may be skipped, yielding $V = V_{\tau\delta}$, $B = B_{\tau\delta}$, $H = H_{\tau\delta}$, $E = E_{\tau\delta}$ and $\Omega = \Omega_{\tau\delta}$. Note

that $\mathbb{E}(B_{\tau\delta}Y) = \beta$ and $\mathbb{E}(E_{\tau\delta}Y) = 0$. Hence, $B_{\tau\delta}y$ is a unbiased linear estimator of β .

The matrices $H_{\tau\delta}$ and, consequently, $E_{\tau\delta}$ are of particular importance. For any τ and δ , the matrix $H_{\tau\delta}$ is idempotent, in signs: $HH = H$. In other words, H is a projection matrix. Note that if $\tau = 0$ and $\delta = 1_k$, then H_{01} and E_{01} would also be symmetric, and, therefore, denote orthogonal projections.

By construction, the columns of $E_{\tau\delta}$ span the kernel of $H_{\tau\delta}$. In other words, E projects onto the kernel space of H , and H onto the kernel of E . Let $\langle E \rangle$ denote the column space or, equivalently, the image space of E . Then the sample space \mathbb{R}^k of Y can uniquely be decomposed, as for any $b \in \langle H \rangle \cap \langle E \rangle$, it is $b = HEb$, since H and E are idempotent. Hence, $b = H(I - H)b = Hb - Hb = 0$. Thus, each $y \in \mathbb{R}^k$ can uniquely be written as $y = u + w$ such that $u \in \langle H \rangle$ and $v \in \langle E \rangle$, and the decomposition is given by $u = Hy$ and $w = y - Hy = Ey$. In particular, the sample space can be written as the direct sum $\mathbb{R}^k = \langle H_{\tau\delta} \rangle \oplus \langle E_{\tau\delta} \rangle$,

Thus, both $H_{\tau\delta}$ and $E_{\tau\delta}$ are projections in which H projects along the subspace $\langle E \rangle$ onto $\langle H \rangle$ and E projects along the subspace $\langle H \rangle$ onto $\langle E \rangle$. They are, in this sense, complementary. Interestingly, each $H_{\tau\delta}$, independent of the choice of τ and δ , projects into the same subspace of \mathbb{R}^k , namely: the space spanned by the columns of the design matrix X . In fact, each $H_{\tau\delta}$ is a function of $B_{\tau\delta}$, namely $H_{\tau\delta} = X \circ B_{\tau\delta}$, and, since X is injective, their kernels coincide, in signs: $\ker B_{\tau\delta} = \ker H_{\tau\delta} = \langle E \rangle_{\tau\delta}$.

The model assumption state that $\mathbb{E}(Y)$ is a linear function of β , in particular: $\mathbb{E}(Y) = X\beta$ for the injective linear map X . Thus, the parameter space \mathbb{R}^p can be thought of as being embedded in \mathbb{R}^k via the map $X: \mathbb{R}^p \hookrightarrow \mathbb{R}^k$. This suggests to identify the parameter space \mathbb{R}^p with the linear subspace $\langle X \rangle$ of \mathbb{R}^k . On the other hand, the map $B_{\tau\delta}$ maps the sample space \mathbb{R}^k to \mathbb{R}^p , which we have just identified with $\langle X \rangle$. Thus, under the embedding X , the projection $H_{\tau\delta} = X \circ B_{\tau\delta}$ and the map $B_{\tau\delta}$ are in principle the same map.

If $y \in \mathbb{R}^k$ and β is estimated by $B_{\tau\delta}y$, we are, in fact, projecting along $\ker B$ onto the image $X[\mathbb{R}^p]$ of X , namely $\langle X \rangle$. The image $H_{\tau\delta}y$ is called the *estimated responses* or the *fit* of the model. The goodness of fit of $H_{\tau\delta}y$ is judged by evaluating the projection onto its complement, namely the *residuals* $E_{\tau\delta}y$.

2.4 Quadratic forms, inner products, and the squared length of the residual vector

The goodness of a fit for an estimate of β is evaluated by a function of the residual vector $E_{\tau\delta}y$. This is often a function of the squared length of this vector in \mathbb{R}^k with respect to a suitable metric. The function that maps a vector to its squared length is called a quadratic form.

Let $A \in \mathbb{R}^{k \times k}$ be a matrix. Any A defines a bilinear form

$$b_A: (x, y) \mapsto x^\top Ay \quad (2.11)$$

and a corresponding quadratic form

$$q_A: y \mapsto y^\top Ay \quad (= \langle y, Ay \rangle) \quad (2.12)$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical inner product on \mathbb{R}^k . If A is symmetric, then so is b_A . The bilinear form b_A is not degenerated if and only if $\det A \neq 0$. We say A , b_A and q_A are *positive definite (semidefinite)*, if $q_A(y) > 0$ ($q_A(y) \geq 0$) for all $y \in \mathbb{R}^k$, $y \neq 0$. If A is symmetric and positive definite, its bilinear form b_A defines an inner product on \mathbb{R}^k , making (\mathbb{R}^k, b_A) into an inner product space with induced norm and metric. Of particular interest are the positive definite variance-covariance matrices of (non-degenerated) random vectors. Say, V is the symmetric and positive definite variance-covariance matrix of a non-degenerated random vector Y . For notational clarity the induced inner product, norm, and induced distance of $b_{V^{-1}}$ are written as

$$\begin{aligned} \langle x, y \rangle_V &:= x^\top V^{-1}y &&= b_{V^{-1}}(x, y), \\ \|y\|_V &:= \sqrt{\langle x, y \rangle_V} &&= \sqrt{q_{V^{-1}}(y)}, \\ d_V(x, y) &:= \sqrt{(x - y)^\top V^{-1}(x - y)} &&= \sqrt{q_{V^{-1}}(x - y)}. \end{aligned}$$

The latter distance d_V is often called a *Mahalanobis distance* in acknowledgement of Mahalanobis (1936).

For reference, two theorems from (Graybill, 1961) are stated. The first is a necessary and sufficient condition for a symmetric matrix to be positive definite.

Theorem 2.2 Let $A \in \mathbb{R}^{k \times k}$ be symmetric. Then A is positive definite if and only if there exists some $B \in \mathbb{G}_k$ such that $A = B^\top B$.

Here, \mathbb{G}_k denotes the *unit group* of $k \times k$ squared matrices of full rank. The set \mathbb{G}_k becomes a group under matrix multiplication. In the statistical context, the next theorem is interesting as many statistics contain constructions such as AA^\top and $A^\top A$ for all kinds of matrices A .

Theorem 2.3 Let $p < k$ and $X \in \mathbb{R}^{k \times p}$. Then $X^\top X \in \mathbb{R}^{k \times k}$ and $XX^\top \in \mathbb{R}^{p \times p}$ are squared matrices.

- a) If $\text{rk } X = p$, then $X^\top X$ is positive definite and XX^\top is positive semidefinite.
- b) If $\text{rk } X < p$, then $X^\top X$ and XX^\top are positive semidefinite.

The random effects meta regression model has the variance-covariance matrix $\Omega_{\tau\delta}^{-1}$. Again, to avoid notational abuses, the respected induced inner product and norm of the inner product $b_{\Omega_{\tau\delta}}$ shall be written as $\langle x, y \rangle_{\tau\delta}$ and $\|y\|_{\tau\delta}$ instead of the awkward looking $\langle x, y \rangle_{\Omega_{\tau\delta}^{-1}}$ and $\|y\|_{\Omega_{\tau\delta}^{-1}}$. Hence, $\langle x, y \rangle_{\tau\delta} = x^\top \Omega_{\tau\delta} y$ and $\|y\|_{\tau\delta}^2 = y^\top \Omega_{\tau\delta} y$.

If τ and δ are known or estimated with sufficient accuracy, we may want to analyse any observed data not in the canonical inner product space \mathbb{R}^k but in \mathbb{R}^k equipped by an adjusted inner product, namely one that reflects the variance-covariance structure of Y . Depending on the underlying hypothesis, it is suitable to either work in canonical

$$(\mathbb{R}^k, \langle \cdot, \cdot \rangle), \quad (2.13)$$

or in

$$(\mathbb{R}^k, \langle \cdot, \cdot \rangle_{[\hat{\delta}]^{-1}}) \quad (2.14)$$

and an estimate $\hat{\delta}$ of δ , or even in the Hilbert space

$$(\mathbb{R}^k, \langle \cdot, \cdot \rangle_{([\hat{\tau}]_k + [\hat{\delta}])^{-1}}) \quad (2.15)$$

for an additional suitable estimate $\hat{\tau}$ of τ .

When studying quadratic forms as in (2.12), the interest obviously lies in the case in which A is a non-trivial projection, i.e., in which $AA = A$ but $\text{rk } A < k$. In case A is additionally symmetric, A is positive semidefinite and $q_A(y) = \|Ay\|^2$, $y \in \mathbb{R}^k$. In other words, q_A maps each vector to the squared length of its projected vector Ay in \mathbb{R}^k . Let $B \in \mathbb{G}_k$ be the matrix of a basis transformation. Here, B should be thought of as the inverse of the positive definite covariance-variance matrix of a non-degenerated random vector. Then $q_A \circ B = q_{B^\top AB}$. Again, a theorem from (Graybill, 1961) is stated.

Theorem 2.4 Let $A, B \in \mathbb{R}^{k \times k}$ such that $B \in \mathbb{G}_k$ and A is positive definite (semidefinite). Then also $B^\top AB$ is positive definite (semidefinite).

The next theorem also deals with idempotent matrices and their definiteness.

Theorem 2.5

- a) Any symmetric and idempotent matrix not of full rank is positive semidefinite.
- b) Any symmetric and idempotent matrix of full rank is equal to the identity matrix and, therefore, positive definite.

Proof. If A is symmetric and idempotent, then $x^\top Ax = x^\top A^\top Ax = \|Ax\|^2 \geq 0$. Hence, A is positive semidefinite. If A is of full rank, then $A \in \mathbb{G}_k$. Multiplying both sides of $AA = A$ by A^{-1} yields the claim. \square

The following is a very interesting result about the idempotent matrices $H_{\tau\delta}$ and $E_{\tau\delta}$ for non-trivial choices of τ and δ . In trivial cases, such as $\tau = 1$ and $\delta = 0_k$, these matrices are symmetric and, thus, constitute very simple forms by Theorem 2.5. In general, however, $E_{\tau\delta}$ and $H_{\tau\delta}$ are not symmetric. The following result is the main reason that it is nevertheless possible to successfully work with these matrices and to use them to build inferential methods.

The proof is not too difficult which might be the reason that it cannot be found in the literature.

Theorem 2.6 Let $\tau > 0$ and $\delta \geq 0$. Then $\Omega_{\tau\delta}H_{\tau\delta}$ and $\Omega_{\tau\delta}E_{\tau\delta}$ are symmetric, in signs

$$(\Omega_{\tau\delta}H_{\tau\delta})^\top = \Omega_{\tau\delta}H_{\tau\delta}, \quad (\Omega_{\tau\delta}E_{\tau\delta})^\top = \Omega_{\tau\delta}E_{\tau\delta}.$$

In particular,

$$E_{\tau\delta}^\top \Omega_{\tau\delta} E_{\tau\delta} = E_{\tau\delta}^\top \Omega_{\tau\delta}. \quad (2.16)$$

Proof. $(\Omega H)^\top = (\Omega X (X^\top \Omega X)^{-1} X^\top \Omega)^\top = \Omega X (X^\top \Omega X)^{-1} X^\top \Omega = \Omega H$. Hence, $\Omega H = (\Omega H)^\top = H^\top \Omega^\top = H^\top \Omega$. Also $(\Omega E)^\top = (\Omega(I - H))^\top = (\Omega - \Omega H)^\top = \Omega - (\Omega H)^\top = \Omega - \Omega H = \Omega(I - H) = \Omega E$. In particular, $E^\top \Omega E = E^\top (\Omega E)^\top = E^\top E^\top \Omega = (EE)^\top \Omega = E^\top \Omega$, since E is idempotent. And, $E^\top \Omega = E^\top \Omega^\top = (\Omega E)^\top = \Omega E$. \square

Suppose there is no heteroscedasticity between the single study responses, i.e., suppose $\tau > 0$ and $\delta = 0_k$. Then the random effects meta regression model reduces to a simple linear regression model with variance component τ . Applying the linear operator $E = E_{10}$ to the observed data y yields the *residuals* Ey . In standard linear regression analysis the (squared) norm of the residual vector equals $\|Ey\|^2 = (Ey)^\top (Ey)$. In contrast to mathematics, it is common in statistical literature to give a mathematical entity a name that reflects its canonical construction instead of its properties. In this case, the squared residual norm $\|Ey\|^2$ is called the *error sum of squares* and commonly abbreviated by *SSE*: let e_i^\top denote the i th row of $E = E_{10}$. Then

$$\|Ey\|^2 = (Ey)^\top (Ey) = \sum_i (e_i^\top y)^2 \quad (2.17)$$

which is indeed a sum of squares. When dealing with heteroscedasticity, one may work with a weighted quadratic instead, namely $\|EY\|_{\delta\tau}^2$ for $E = E_{10}$:

$$\|EY\|_{\delta\tau}^2 = (EY)^\top \Omega_{\tau\delta}(EY) = \sum_i \frac{1}{\tau + \delta_i} (e_i^\top y)^2. \quad (2.18)$$

This is consistent with the special case of the simple linear regression model, since

$$\|EY\|_{\tau 0}^2 = \frac{1}{\tau} \|EY\|^2 \quad (2.19)$$

is just a scaled version of the same measure.

2.5 Expectations of residual squared lengths

Of interest is the squared norm or length of the residual vector $E_{\tau\delta}Y$ for different values of τ or δ and for different norms on \mathbb{R}^k . In this section, the particular interest lies in the expected value of this squared length as it is frequently used for building statistics for hypothesis testing and inferential methods.

Theorem 2.7 Let $Y \sim (\theta, V)$ for some $\theta \in \mathbb{R}^k$ and positive definite $V \in \mathbb{G}_n$. Let $A \in \mathbb{R}^{k \times k}$. Then $\mathbb{E}_{\theta V}(q_A(Y)) = q_A(\theta) + \text{tr}(AV)$. In particular, if θ lies in the kernel of A , then

- a) $\mathbb{E}(\|AY\|^2) = \text{tr}(AV)$.
- b) For any positive definite $U \in \mathbb{G}_k$, $\mathbb{E}(\|AY\|_U^2) = \text{tr}(A^\top U^{-1}AV)$.

Proof. The claim follows from $\mathbb{E}(Y^\top AY) = (\mathbb{E}Y)^\top A(\mathbb{E}Y) + \text{tr}(AV) = \theta^\top A\theta + \text{tr}(AV)$ which holds true for any arbitrary squared matrix $A \in \mathbb{R}^{k \times k}$.

- a) When $A\theta = 0$, i.e., when θ lies in the kernel of A , then $q_A(\theta) = \theta^\top A\theta = 0$.
- b) By definition $\|AY\|_U^2 = (AY)^\top U^{-1}(AY) = Y^\top (A^\top U^{-1}A)Y = q_{A^\top U^{-1}A}(Y)$. Also, $q_{A^\top U^{-1}A}(\theta) = \theta^\top (A^\top U^{-1}A)\theta = (A\theta)^\top U^{-1}(A\theta) = 0$, since $A\theta = 0$. By a) follows the claim. □

In the context of the random effects meta regression model, this yields the following corollary.

Corollary 2.8 Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$ and fixed $\tau \geq 0$ and $\delta > 0$. Let $E' = E_{\tau'\delta'}$ for some choice of $\tau' \geq 0$ and $\delta' > 0$. Then

- a) $\mathbb{E}_{\tau\delta} (\|E'Y\|^2) = \tau(k - p) + \text{tr}(E'[\delta])$.
- b) $\mathbb{E}_{\tau\delta} (\|EY\|_{\tau\delta}^2) = k - p$.
- c) $\mathbb{E}_{\tau\delta} (\|E'Y\|_{\tau'\delta'}^2) = \tau \text{tr}(E'^\top \Omega') + \text{tr}(E'^\top \Omega' [\delta])$ with $\Omega' = ([\tau']_k + [\delta'])^{-1}$.
- d) If $\tau' = 0$, then $\mathbb{E}_{\tau\delta} (\|E'Y\|_{0\delta'}^2) = \tau \text{tr}(E'^\top [\delta']^{-1}) + \text{tr}(E'^\top [\delta']^{-1} [\delta])$.

Proof. Note that $X\beta$ lies in the kernel of E' for any τ' and δ' , since $E'(X\beta) = (E'X)\beta = 0$. For a) note that $E'([\tau]_k + [\delta]) = E' \cdot \tau + E'[\delta]$. Hence,

$$\text{tr}((E'([\tau]_k + [\delta]))) = \tau \text{tr} E' + \text{tr}(E'[\delta]) = \tau(k - p) + \text{tr}(E'[\delta]). \quad (2.20)$$

From Theorem 2.7 follows that

$$\mathbb{E}_{\tau\delta} (\|E'Y\|_{\tau'\delta'}^2) = \text{tr}(E'^\top ([\tau']_k + [\delta'])^{-1} E'([\tau]_k + [\delta])). \quad (2.21)$$

Let $\Omega' = ([\tau']_k + [\delta'])^{-1}$ and $\Omega = ([\tau]_k + [\delta])^{-1}$. Then from the right hand side follows

$$\begin{aligned} E'^\top \Omega' E' \Omega^{-1} &= E'^\top \Omega' E'([\tau]_k + [\delta]) \\ &= E'^\top \Omega'([\tau]_k + [\delta]) \\ &= \tau \cdot E'^\top \Omega' + E'^\top \Omega'[\delta] \end{aligned}$$

Taking the traces of these matrices yields c). Setting $\tau' = 0$, yields the claim in d). For the particular case in which $\tau' = \tau$ and $\delta' = \delta$, it also follows from Theorem 2.6 that

$$\text{tr}(E^\top \Omega E \Omega^{-1}) = \text{tr}(E^\top \Omega \Omega^{-1}) = \text{tr}(E^\top) = k - p. \quad (2.22)$$

This shows the claim in b). □

2.6 Residual squared lengths as functions of the heterogeneity

The previous sections have shown that

$$\|E_{\tau\delta} Y\|_{\tau\delta}^2 \quad (2.23)$$

is a frequently appearing term in the study of the random effects meta regression model. As mentioned, the within-study variances δ are usually considered known

and being part of the summary statistics of the studies regarded for the meta regression or analysis. Henceforth, (2.23) will be treated as a function in τ , namely

$$q_y: \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}, \quad \tau \mapsto q_y(\tau) := \|E_{\tau\delta}y\|_{\tau\delta}^2. \quad (2.24)$$

This idea goes back to Iyer et al. (2004) who suggested to view q_y as a function in τ first. The notation not only suggests that δ is known but also that δ is fixed for different values of y . This is reasonable as the within study variances are assumed as being inherited from the respected study designs and not something that would depend on the random outcome Y . See Section 2.14 for a discussion. The main result of this section is a matrix result by Khatri (1966) that allows q_y to be rewritten in a mathematically more accessible way. This step is not only of theoretical interest but, when used in actual inference methods, crucial to obtain numerically stable results from this function.

The result is comparably old and also not particularly well known. For this reason, a proof is provided. The result is then applied to $q_y(\tau)$ and its first derivative is calculated. This allows to deduce some fundamental properties of q_y .

For $A \in \mathbb{R}^{k \times p}$, the *Moore-Penrose inverse* is the unique matrix $A^+ \in \mathbb{R}^{p \times k}$ such that

$$\begin{aligned} A &= AA^+A, \\ A^+ &= A^+AA^+, \\ (AA^+)^{\top} &= AA^+, \\ (A^+A)^{\top} &= A^+A. \end{aligned}$$

In particular, the Moore-Penrose inverse is a *pseudo inverse*, and $\text{tr } AA^+ = \text{rk } A$.

Harville (1977) defined an *error contrast* to be a linear combination u of the random Y such that $u^{\top}Y$ has expectation 0, in signs: $\mathbb{E}(u^{\top}Y) = 0$. In the case of the random effects meta regression model, it is sufficient for an error contrast that $u^{\top}X = 0$. If $K \in \mathbb{R}^{k \times q}$ is a full rank matrix such that the column space $\langle K \rangle$ is spanning the kernel of X^{\top} , then each column of K is an error contrast of Y and each error contrast lies in $\langle K \rangle$. The *error contrast matrix* K can be found, e.g., by a singular values decomposition of X^{\top} .

If X is of full rank, then the kernel of X^{\top} is of rank $k - p$. Thus, the maximum number of linear independent error contrasts is $k - p$. Inference using the restricted maximum likelihood approach is based on the distribution of $K^{\top}Y$ instead of Y .

Lemma 2.9 For $q \leq p \leq k$, let $X \in \mathbb{R}^{k \times p}$ and $K \in \mathbb{R}^{k \times q}$ such that the columns of K are spanning the kernel of $\ker X^\top$, in signs: $\ker X^\top = \langle K \rangle$. If K is of full rank, then

$$I_k - XX^+ = KK^+. \quad (2.25)$$

Proof. It is not difficult to see that for any $A \in \mathbb{R}^{m \times n}$, it is $A = 0$ if and only if $\text{tr} AA^\top = 0$. So, let $A := I_k - XX^+ - KK^+$, and let us show that $\text{tr} AA^\top = 0$. Now, since I_k , XX^+ and KK^+ are symmetric and idempotent, also A is symmetric and idempotent. Hence, $\text{tr} AA^\top = \text{tr} A = k - \text{tr} XX^+ - \text{tr} KK^+ = k - \text{rk} X - \text{rk} K$. Since the columns of K form a basis of $\ker X^\top$, it is K of full rank and $\text{rk} K = k - \text{rk} X^\top = k - \text{rk} X$. This completes the proof. \square

If A is of full rank, then $AA^+ = A(A^\top A)^{-1}A^\top$. Thus, for a full rank matrix X in Lemma 2.9, it is possible to replace XX^+ by $X(X^\top X)^{-1}X^\top$, and KK^+ can be replaced by $K(K^\top K)^{-1}K^\top$. The following result is due to (Khatrı, 1966).

Lemma 2.10 Let $X \in \mathbb{R}^{k \times p}$ and let $K \in \mathbb{R}^{k \times (k-p)}$ be such that the columns of K form a basis of the kernel of X^\top . If X is of full rank, then for any positive definite $V \in \mathbb{G}_k$,

$$V^{-1} - V^{-1}X(X^\top V^{-1}X)^{-1}X^\top V^{-1} = K(K^\top VK)^{-1}K^\top. \quad (2.26)$$

Proof. Since V is positive definite, then so is V^{-1} positive definite. Thus, V^{-1} has a unique principle square root $V^{-\frac{1}{2}}$. Since $V^{-\frac{1}{2}} \in \mathbb{G}_k$, it is also a matrix of basis transformation. If X and K fulfil the requirements of Lemma 2.9, then so do $V^{-\frac{1}{2}}X$ and $V^{\frac{1}{2}}K$. Substituting $V^{-\frac{1}{2}}X$ for X and $V^{\frac{1}{2}}K$ for K in Lemma 2.9 results in

$$I_k - V^{-\frac{1}{2}}X(X^\top V^{-1}X)^{-1}XV^{-\frac{1}{2}} = V^{\frac{1}{2}}K(K^\top VK)K^\top V^{\frac{1}{2}}. \quad (2.27)$$

Multiplying the above equation by $V^{-\frac{1}{2}}$ from both sides yields the claim. \square

The important thing to notice is that the matrix V loses its inverse operation on the right hand side of (2.26). This makes this result computationally so appealing. Applying the result to $q_y(\tau)$ makes it possible to prove the following equation.

Theorem 2.11 Let $K \in \mathbb{R}^{k \times k-p}$ which columns (k_1, \dots, k_{k-p}) form a basis of $\ker X^\top$. Then

$$q_y(\tau) = y^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y. \quad (2.28)$$

Proof. Note that $q_y(\tau) = y^\top E_{\tau\delta}^\top \Omega_{\tau\delta} E_{\tau\delta} y$. In order to proof the claim, Lemma 2.10 is applied to the term $E_{\tau\delta}^\top \Omega_{\tau\delta} E_{\tau\delta}$. By construction, K is of full rank and its columns span the kernel of X^\top . Also, Ω^{-1} is obviously of full rank and positive definite. From Theorem 2.6 it follows that $E^\top \Omega E = \Omega E$. Hence, $E^\top \Omega E = \Omega E = \Omega(I - H) = \Omega(I - X^\top (X^\top \Omega X)^{-1} X^\top \Omega) = \Omega - \Omega X (X^\top \Omega X)^{-1} X^\top \Omega$. Applying Lemma 2.10 yields the claim. \square

The above theorem shows that $q_y(\tau)$ only depends on τ and δ via the term $([\tau]_k + [\delta]) = \Omega_{\tau\delta}^{-1}$ and this at *one* single place only. Since $\tau \rightarrow \Omega_{\tau\delta}^{-1}$ is smooth and each τ is mapped to a positive definite matrix, the function $\tau \mapsto q_y(\tau)$ is also smooth on $\mathbb{R}_{>0}$. The following lemma calculates its first derivative.

Lemma 2.12 Let $X \in \mathbb{R}^{k \times p}$ and let $K \in \mathbb{R}^{k \times (k-p)}$ be such that the columns of K form a basis of the kernel of X^\top . If X is of full rank, then q_y is smooth on $\mathbb{R}_{>0}$ and

$$D_\tau q_y(\tau) = -y^\top G y \quad (2.29)$$

for the positive definite $G := A^\top A \in \mathbb{G}_k$ with $A := K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top$.

Proof. First note that $D_\tau(\Omega_{\tau\delta}^{-1}) = I_k$. Therefore, $D_\tau(K^\top \Omega_{\tau\delta}^{-1} K) = K^\top (D_\tau \Omega_{\tau\delta}^{-1}) K = K^\top K$. Hence,

$$\begin{aligned} D_\tau(K^\top \Omega_{\tau\delta}^{-1} K)^{-1} &= -(K^\top \Omega_{\tau\delta}^{-1} K)^{-1} D_\tau(K^\top \Omega_{\tau\delta}^{-1} K) (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} \\ &= -(K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1}. \end{aligned}$$

And this yields,

$$\begin{aligned} D_\tau q_y(\tau) &= D_\tau (y^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y) \\ &= y^\top K D_\tau (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y \\ &= -y^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y. \end{aligned}$$

\square

Lemma 2.12 shows that, in particular, $D_\tau q_y(\tau)$ is again a quadratic form. Corollary 2.13 studies the convergence of q_y .

Corollary 2.13 Let $X \in \mathbb{R}^{k \times p}$ and let $K \in \mathbb{R}^{k \times (k-p)}$ be such that the columns of K form a basis of the kernel of X^\top . If X is of full rank, then q_y is smooth on $\mathbb{R}_{>0}$, strictly monotone decreasing, and converging to 0 as τ goes to ∞ .

Proof. Since $G \in \mathbb{G}_k$ in (2.29) is positive definite, $D_\tau q_y(\tau)$ is strictly negative and, therefore, q_y is strictly decreasing. Now, let $c_\delta(\tau) := \frac{1}{\tau + \max_j \delta_j}$. Then $c_\delta(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$. Also, $q_y(\tau) = y^\top K(K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y \leq y^\top K(K^\top \cdot c_\delta(\tau) \cdot K)^{-1} K^\top y = c_\delta(\tau) \cdot y^\top K(K^\top K)^{-1} K^\top y \rightarrow 0$ as $\tau \rightarrow \infty$. \square

The expected value of the random $q_Y(\tau)$ is

$$\mathbb{E}_\tau(q_Y(\tau)) = k - p. \quad (2.30)$$

This already indicates that the point τ_0 for which $q_y(\tau_0) = k - p$ is of particular importance. In fact, (2.30) shows that τ_0 is a method of moments estimator for τ , which is commonly known as the *Mandel–Paule* estimator in meta analysis first introduced in (Mandel and Paule, 1970) and extended in (Paule and Mandel, 1982).

Since q_y is strictly monotone decreasing, it has an inverse. Let p_y denote this inverse of q_y but defined on the whole real line \mathbb{R} , namely

$$p_y(\eta) = \begin{cases} q_y^{-1}(\eta) & : 0 < \eta < q_y(0) \\ 0 & : \text{otherwise} \end{cases}$$

for $\eta \in \mathbb{R}$. Then, e.g., $\tau_0 = p_y(k - p)$ is a Mandel–Paule type estimator for τ .

2.7 Example: Studying the BCG vaccine efficacy

In Section 1.2, data evaluating the efficacy of the BCG vaccine was introduced as an exemplary data set that is commonly analysed using a random effects meta regression model. Section 2.14 will discuss how to specifically analyse this kind of data in the context of the random effects meta regression model. The residual squared length as a function of τ , namely $q_y(\tau)$, is shown in Figure 2.1 for this data. Its inverse function $p_y(\eta)$ is shown to the right of this plot.

The function q_y hits the y -axis at the intercept of $q_y(0) \approx 30.73$. Equivalently, p_y hits the x -axis at $p_y(30.73) \approx 0$. Both functions are strictly decreasing. The function q_y is converging to zero as τ increases. For p_y it holds that $p_y(\eta) = 0$ for $\eta \geq q_y(0) \approx 30.73$.

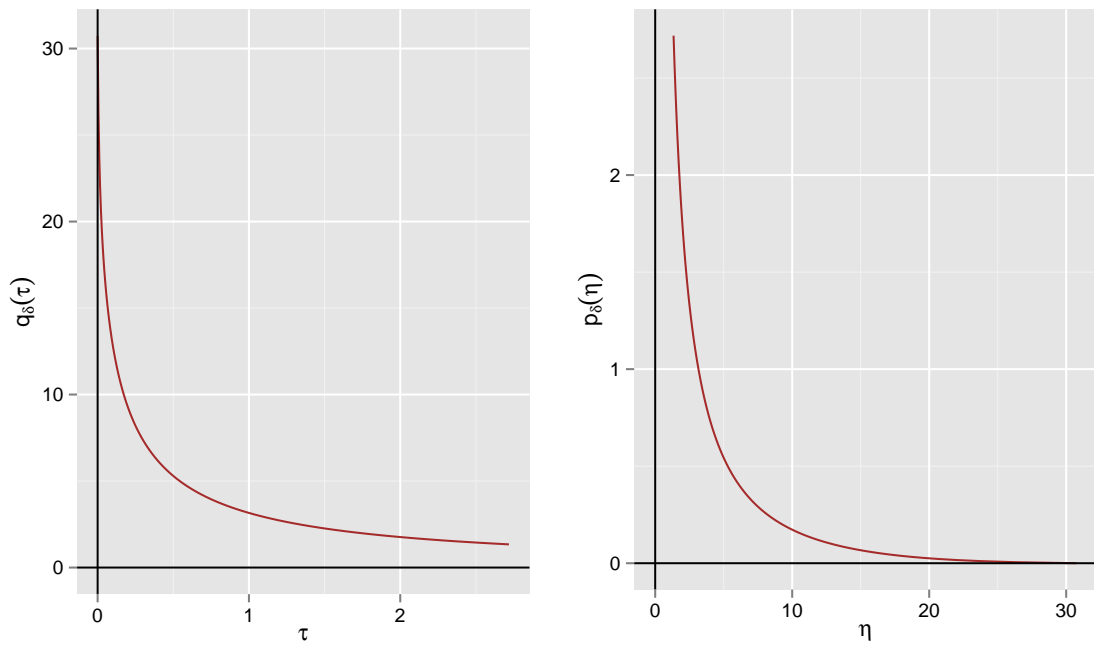


Figure 2.1 For the BCG vaccine efficacy data, the squared length of the residual vector $q_y = q_y(\tau)$ as a function of the unknown heterogeneity τ in the data is plotted for different values of τ . On the right, the inverse function $p_y = p_y(\eta)$ of q_y is plotted for different values of η .

Excursion 2.14 In the `metagen` package, the functions `qfunc` and `pfunc` can be used to generate the functions q_y and p_y respectively.

```
library(metagen)
bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
x <- cbind(1, bcg$x)

qfunction <- qfunc(y, d, x)
pfunction <- pfunc(y, d, x)

# The intercept of the qfunction
qfunction(0)
```

In order to get an easy and quick view of their dynamical range and properties, there also exists a ready to use function which will produce plots of q_y and p_y . These plots can help evaluating the output of inferential methods based on these functions.

```
plots <- plotStudyQfuncPfunc(y=y, d=d, x=x, n=500)
```

where $n = 500$ tells the algorithm at how many different points the functions shall be evaluated. The larger n , the higher will be the resolution of the plots. The above call will in fact generate a list of plots which can be accessed by calling its respected elements.

```
plots$plotQ
plots$plotP
```

For example, the Mandel–Paule type estimator for τ , which is defined in Definition 4.4, can be calculated in the following way.

```
k <- dim(x)[1]; p <- dim(x)[2]
pfunction(k-p)
```

This yields $p_y(k - p) = p_y(11) \approx 0.1421$.

2.8 Model description – with Gaussian responses

When assuming Gaussian distributed study effects, the random effects meta regression model is a Gaussian-Gaussian hierarchical model that models the univariate effect measures of a number of different studies. Again, let $Y = (Y_1, \dots, Y_k)^\top$ be a random k -vector where each Y_j corresponds to the outcome of a single study. It is now assumed that each Y and θ follow Gaussian distributions.

$$Y_j | \theta_j \sim \mathfrak{N}(\theta_j, \delta_j), \quad (2.31)$$

$$\theta_j \sim \mathfrak{N}(x_j^\top \beta, \tau), \quad (2.32)$$

where \mathfrak{N} denotes the distribution of a univariate Gaussian distributed random element. If $(Y^\top, \theta_1, \dots, \theta_k)^\top$ is independent and follows a multivariate Gaussian distribution, then the marginal model for Y yields the *random effects meta regression model*.

Let \mathfrak{N}_k denote the distribution function of a k -variate Gaussian distributed random k -vector. Reformulating Definition 2.1 including the above distribution assumptions, the random effects meta regression model can be defined as follows.

Definition 2.15 Let $X \in \mathbb{R}^{k \times p}$ with $\text{rk}(X) = p < k - 1$. For any $\beta \in \mathbb{R}^p$, $\tau \in \mathbb{R}_{\geq 0}$ and $\delta \in \mathbb{R}_{> 0}^k$, let

$$\mathfrak{F}_{\beta\tau\delta} := \mathfrak{N}_k(X\beta, [\tau]_k + [\delta]). \quad (2.33)$$

Then the family of distribution functions

$$\mathfrak{F} = (\mathfrak{F}_{\beta\tau\delta} : (\beta, \tau, \delta) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{> 0}^k) \quad (2.34)$$

is called the *random effects meta regression model*.

In other words, it is assumed that there exists some $(\beta, \tau, \delta) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{> 0}^k$ such that

$$Y \sim \mathfrak{F}_{\beta\tau\delta}. \quad (2.35)$$

In case $p = 1$, $\beta \in \mathbb{R}$ and $X = \mathbf{1}_k$, the above model reduces to the random effects meta analysis model.

2.9 Likelihood functions and their derivatives

Following Palais (2001) and Hartl (2010), define $\mathbf{t} = 2\pi$. For a single study response Y_j , the Lebesgue density $\mathbf{f}_{\beta\tau\delta,j}$ of the random effects meta regression model is

$$\mathbf{f}_{\beta\tau\delta,j}(y) = \mathbf{t}^{-\frac{1}{2}} (\tau + \delta_j)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (\tau + \delta_j)^{-1} (x_j^\top \beta - y)^2\right). \quad (2.36)$$

Since Y is independent, the joint density $\mathbf{f}_{\beta\tau\delta} = D \mathfrak{F}_{\beta\tau\delta}$ of (2.33) is

$$\begin{aligned} \mathbf{f}_{\beta\tau\delta}(y) &= \mathbf{t}^{-\frac{k}{2}} \left(\prod_{j=1}^k (\tau + \delta_j) \right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{j=1}^k (\tau + \delta_j)^{-1} (x_j^\top \beta - y)^2\right) \\ &= \mathbf{t}^{-\frac{k}{2}} (\det \Omega_{\tau\delta}^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (X\beta - y)^\top \Omega_{\tau\delta} (X\beta - y)\right) \\ &= \mathbf{t}^{-\frac{k}{2}} (\det \Omega_{\tau\delta}^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \|X\beta - y\|_{\tau\delta}^2\right) \end{aligned} \quad (2.37)$$

as $\det \Omega_{\tau\delta}^{-1} = \prod_j (\tau + \delta_j)$. Let us have a closer look at this function, in particular, at the exponent of the exponential in (2.37). This exponent can be rewritten in terms of the operators defined on page 11. Note that

$$\begin{aligned} &(\beta - By)^\top V^{-1} (\beta - By) \\ &= \beta^\top V^{-1} \beta - 2y^\top B^\top V^{-1} \beta + y^\top B^\top V^{-1} B y \\ &= \beta^\top V^{-1} \beta - 2y^\top \Omega X \beta + y^\top \Omega H y, \end{aligned}$$

since

$$B^\top V^{-1} = \Omega X (X^\top \Omega X)^{-1} (X^\top \Omega X) = \Omega X,$$

and

$$B^\top V^{-1} B = \Omega X (X^\top \Omega X)^{-1} X^\top \Omega = \Omega H.$$

Hence,

$$(X\beta - y)^\top \Omega (X\beta - y) \quad (2.38)$$

$$\begin{aligned} &= \beta^\top X^\top \Omega X \beta - 2y^\top \Omega X \beta + y^\top \Omega y \\ &= \beta^\top V^{-1} \beta - 2y^\top \Omega X \beta + y^\top \Omega H y - y^\top \Omega H y + y^\top \Omega y \\ &= (\beta - By)^\top V^{-1} (\beta - By) + y^\top (E^\top \Omega E) y \\ &= \|\beta - By\|_{V_{\tau\delta}} + \|Ey\|_{\tau\delta}. \end{aligned} \quad (2.39)$$

Here, (2.39) follows from Theorem 2.6 and

$$y^\top \Omega y - y^\top \Omega H y = y^\top (\Omega - \Omega H) y = y^\top \Omega E y = (E y)^\top \Omega (E y). \quad (2.40)$$

The function $q_{y\delta}(\tau) = \|E_{\tau\delta} y\|_{\tau\delta}$ has turned up a few times in previous sections. Thus, the joint density of the random effects meta regression model is equal to

$$f_{\beta\tau\delta}(y) = \mathbf{t}^{-\frac{k}{2}} (\det \Omega_{\tau\delta}^{-1})^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \|\beta - B_{\tau\delta} y\|_{V_{\tau\delta}}\right) \exp\left(-\frac{1}{2} \|E_{\tau\delta} y\|_{\tau\delta}\right) \quad (2.41)$$

In particular, this shows that $B_{\tau\delta} y$ is a sufficient statistic for β in case δ and τ are known parameters by the factorisation theorem. As said before, for any theoretical considerations, it is reasonable to assume the heteroscedasticity δ to be known. Thus, the conditional (with respect to δ) log-likelihood function can be written in the following two forms, namely

$$\begin{aligned} l(\beta, \tau|y, \delta) &= \ln f_{\beta\tau\delta}(y) \\ &= -\frac{1}{2} \left(k \ln \mathbf{t} + \sum_j \ln(\tau + \delta_j) + \|X\beta - y\|_{\tau\delta}^2 \right) \\ &= -\frac{1}{2} \left(k \ln \mathbf{t} + \sum_j \ln(\tau + \delta_j) + \|\beta - B_{\tau\delta} y\|_{V_{\tau\delta}} + \|E_{\tau\delta} y\|_{\tau\delta} \right). \end{aligned} \quad (2.42)$$

Likelihood methods aim to estimate the unknown parameters by maximising the likelihood function conditional on the observed data y and any known parameters, here δ . With regards to β , two cases can be distinguished: τ known and τ to be estimated from the data as well.

Since the conditional likelihood l is smooth, a necessary condition for l to attain its maximum, is that its gradient vanishes. In signs, $\nabla l = 0$. This gradient is now being computed.

Lemma 2.16 The partial derivatives of the log-likelihood function $l(\beta, \tau|y, \delta)$ with respect to β and τ are

a) $D_\beta l(\beta, \tau) = -(X\beta - y)^\top \Omega_{\tau\delta} X,$

b) $D_\tau l(\beta, \tau) = \frac{1}{2} ((X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y) - \text{tr} \Omega_{\tau\delta}).$

In particular, $\nabla l(\beta, \tau) = (D_\beta l(\beta, \tau), D_\tau l(\beta, \tau)).$

Proof. Equation (2.42) will be used for calculations. Then,

$$\begin{aligned}
D_\beta l(\beta, \tau) &= -\frac{1}{2} D_\beta ((X\beta - y)^\top \Omega_{\tau\delta} (X\beta - y)) \\
&= -(X\beta - y)^\top \Omega_{\tau\delta} D_\beta (X\beta - y) \\
&= -(X\beta - y)^\top \Omega_{\tau\delta} X \quad \in \mathbb{R}^{1 \times p},
\end{aligned}$$

since $D_\beta f(\beta)^\top A f(\beta) = 2f(\beta)^\top A D_\beta f(\beta)$, and $D_\beta \beta = I_p$, and $D_\beta (X\beta - y) = X$. Moreover,

$$\begin{aligned}
D_\tau ((X\beta - y)^\top \Omega_{\tau\delta} (X\beta - y)) &= (X\beta - y)^\top (D_\tau \Omega_{\tau\delta}) (X\beta - y) \\
&= -(X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y), \quad (2.43)
\end{aligned}$$

since $D_\tau \left(\frac{1}{\tau + \delta_j} \right) = -\frac{1}{(\tau + \delta_j)^2}$ for any $j = 1, \dots, k$ and, therefore, $D_\tau \Omega_{\tau\delta} = -\Omega_{\tau\delta}^2$. Then, note that

$$D_\tau \left(\sum_{j=1}^n \ln(\tau + \delta_j) \right) = \sum_{j=1}^n \frac{1}{\tau + \delta_j} = \text{tr } \Omega_{\tau\delta} \quad (2.44)$$

Hence,

$$\begin{aligned}
D_\tau l(\beta, \tau) &= -\frac{1}{2} \left(D_\tau \left(\sum_j \ln(\tau + \delta_j) \right) + D_\tau ((X\beta - y)^\top \Omega_{\tau\delta} (X\beta - y)) \right) \\
&= -\frac{1}{2} (\text{tr } \Omega_{\tau\delta} - (X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y)). \quad (2.45)
\end{aligned}$$

□

Let $(\hat{\beta}, \hat{\tau})$ be such that the gradient of $l = l(\beta, \tau)$ vanishes. In order to show that l obtains an optimum, it is sufficient to prove that l is locally convex in a neighbourhood of $(\hat{\beta}, \hat{\tau})$. This can be checked by evaluating the Hessian matrix of l at $(\hat{\beta}, \hat{\tau})$ and testing whether this matrix is negative definite.

Lemma 2.17 The Hessian matrix of l is equal to

$$\text{Hess } l = \begin{pmatrix} -X^\top \Omega_{\tau\delta} X & X^\top \Omega_{\tau\delta}^2 (X\beta - y) \\ (X\beta - y)^\top \Omega_{\tau\delta}^2 X & \frac{1}{2} \text{tr } \Omega_{\tau\delta}^2 - (X\beta - y)^\top \Omega_{\tau\delta}^3 (X\beta - y) \end{pmatrix}.$$

In particular, this shows that $\text{Hess } l$ is negative definite if and only if

$$\det \text{Hess } l < 0. \quad (2.46)$$

Proof. The Hessian matrix of l is equal to

$$\text{Hess } l(\beta, \tau) = \begin{pmatrix} D_{\beta_1\beta_1} & \cdots & D_{\beta_1\beta_p} & D_{\beta_1\tau} \\ \vdots & & \vdots & \vdots \\ D_{\beta_p\beta_1} & \cdots & D_{\beta_p\beta_p} & D_{\beta_p\tau} \\ D_{\tau\beta_1} & \cdots & D_{\tau\beta_p} & D_{\tau\tau} \end{pmatrix} l(\beta, \tau). \quad (2.47)$$

Now, by Lemma 2.16,

$$\begin{aligned} D_{\beta\beta}l &= D_{\beta} \left(D_{\beta}^{\top} l \right) = -D_{\beta} X^{\top} \Omega (X\beta - y) \\ &= -X^{\top} \Omega X. \end{aligned}$$

Also, $(D_{\beta_1\tau}, \dots, D_{\beta_p\tau})^{\top} = (D_{\tau\beta_1}, \dots, D_{\tau\beta_p}) = D_{\tau} D_{\beta}$, and since $D_{\tau} \Omega_{\tau\delta} = -\Omega_{\tau\delta}^2$, it follows that

$$\begin{aligned} D_{\tau} D_{\beta} l &= -D_{\tau} \left((X\beta - y)^{\top} \Omega_{\tau\delta} X \right) \\ &= (X\beta - y)^{\top} \Omega_{\tau\delta}^2 X. \end{aligned} \quad (2.48)$$

Before evaluating $D_{\tau\tau}$, note again that $D_{\tau} \text{tr } \Omega = D_{\tau} \sum_j \frac{1}{\tau + \delta_j} = -\sum_j \frac{1}{(\tau + \delta_j)^2} = -\text{tr } \Omega^2$, and $D_{\tau} \Omega^2 = -2\Omega^3$, since $D_{\tau} \frac{1}{(\tau + \delta_j)^2} = -2 \frac{1}{(\tau + \delta_j)^3}$. Again by Lemma 2.16,

$$\begin{aligned} D_{\tau\tau} l &= \frac{1}{2} \left(D_{\tau} \left((X\beta - y)^{\top} \Omega_{\tau\delta}^2 (X\beta - y) \right) - D_{\tau} (\text{tr } \Omega_{\tau\delta}) \right) \\ &= \frac{1}{2} \text{tr } \Omega_{\tau\delta}^2 - (X\beta - y)^{\top} \Omega_{\tau\delta}^3 (X\beta - y). \end{aligned} \quad (2.49)$$

Since $X^{\top} \Omega_{\tau\delta} X$ is positive definite, all minors of $\text{Hess } l$ up to p are negative. Hence, $\text{Hess } l$ is negative definite if and only if its determinant is negative. \square

For any potential candidate $(\hat{\beta}, \hat{\tau})$, the condition $\det \text{Hess } l < 0$ needs to be checked in order for $(\hat{\beta}, \hat{\tau})$ to be a valid maximum likelihood estimate, (Viechtbauer, 2005).

Corollary 2.18

- a) Conditional on observed data y , known τ , and known δ , the likelihood function of the random effects meta regression model is maximised for the unique β that is a solution to the system of linear equations

$$(X^\top \Omega_{\tau\delta} X) \cdot \beta = X^\top \Omega_{\tau\delta} y. \quad (2.50)$$

- b) Conditional on observed data y and known δ , a necessary condition for the likelihood function of the random effects meta regression model to obtain its maximum is that τ and β are solutions to the system of equations

$$\beta = B_{\tau\delta} y, \quad (2.51)$$

$$\tau = \frac{\sum_{j=1}^k \frac{1}{(\tau + \delta_j)^2} ((X\beta - y)^2 - \delta_j)}{\text{tr } \Omega_{\tau\delta}^2}. \quad (2.52)$$

The first equation (2.51) is a system of linear equations, the second (2.52) is non-linear and can only be written implicitly.

Proof.

- a) From Lemma 2.16a), $D_\beta l(\beta) = 0$ if and only if $0 = (X\beta - y)^\top \Omega_{\tau\delta} X = \beta^\top X^\top \Omega_{\tau\delta} X - y^\top \Omega_{\tau\delta} X$ if and only if $(X^\top \Omega_{\tau\delta} X)\beta = X^\top \Omega_{\tau\delta} y$. As usual, let $V_{\tau\delta}^{-1} = (X^\top \Omega_{\tau\delta} X)$. Since X is of full rank, $V^{-1} \in \mathbb{G}_k$ and $\beta = V_{\tau\delta} X^\top \Omega_{\tau\delta} y = B_{\tau\delta} y$ is a unique solution to this system of linear equations.
- b) It is $\nabla l = 0$ if and only if $D_\beta l = 0$ and $D_\tau l = 0$. We have just seen above that $D_\beta l = 0$ if and only if $\beta = B y$. For the implicit formula for the heterogeneity, the following trick will be applied:

$$\begin{aligned} \text{tr } \Omega &= \sum_j \frac{1}{\tau + \delta_j} \\ &= \sum_j \frac{\tau + \delta_j}{(\tau + \delta_j)^2} \\ &= \sum_j \frac{\tau}{(\tau + \delta_j)^2} + \sum_j \frac{\delta_j}{(\tau + \delta_j)^2} \\ &= \tau \text{tr } \Omega^2 + \sum_j \frac{\delta_j}{(\tau + \delta_j)^2}. \end{aligned}$$

Applying the above to Lemma 2.16b),

$$\begin{aligned}
D_\tau l(\beta, \tau) = 0 &\Leftrightarrow 0 = (X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y) - \text{tr} \Omega_{\tau\delta} \\
&\Leftrightarrow 0 = (X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y) - \tau \text{tr} \Omega_{\tau\delta}^2 - \sum_{j=1}^k \frac{\delta_j}{(\tau + \delta_j)^2} \\
&\Leftrightarrow \tau = \frac{(X\beta - y)^\top \Omega_{\tau\delta}^2 (X\beta - y) - \sum_{j=1}^k \frac{\delta_j}{(\tau + \delta_j)^2}}{\text{tr} \Omega_{\tau\delta}^2} \\
&\Leftrightarrow \tau = \frac{\sum_{j=1}^k \frac{1}{(\tau + \delta_j)^2} (X\beta - y)^2 - \sum_{j=1}^k \frac{\delta_j}{(\tau + \delta_j)^2}}{\text{tr} \Omega_{\tau\delta}^2} \\
&\Leftrightarrow \tau = \frac{\sum_{j=1}^k \frac{1}{(\tau + \delta_j)^2} ((X\beta - y)^2 - \delta_j)}{\text{tr} \Omega_{\tau\delta}^2}
\end{aligned}$$

□

An important note for the implementation of Corollary 2.18: Never ever manually invert a matrix in any computation ever. Equation (2.51) is Equation (2.50) solved for β . A mathematician may usually be trained to think of β as the result of the linear operator $B_{\tau\delta}$ acting on y . Instead, think of β as a vector of coefficients that expands the vector $X^\top \Omega_{\tau\delta} y$ uniquely in the basis consisting of the columns of $(X^\top \Omega X)$. In other words, think of $V = (X^\top \Omega X)^{-1}$ as a matrix of basis transformation. Both approaches yield the same solution to the same problem. The former, namely By , may be more appealing, the latter, however, should always be used for any algorithmic computation. Evaluating By is computationally inefficient and numerically unstable. It is strongly suggested to always use a linear solver for linear equations instead. For example, when using the statistical software environment R, use

$$\beta := \text{solve}(V^{-1}, X^\top \Omega_{\tau\delta} y) \quad (2.53)$$

for calculating β .

Note that the solution of the above equations may yield negative values for τ . If this is the case, it should be checked if the likelihood function attains its maximum on the boundary $\{(\beta^\top, \tau)^\top : \tau = 0\}$. Restricting $l(\beta, \tau)$ to this boundary subspace makes the maximisation problem easier by losing one dimension: In this case, the solution for β is simply (2.51) and setting $\tau = 0$.

2.10 Likelihood function of a contrast and its derivatives

Section 2.6 introduced the notion of an error contrast. It was mentioned that inference on the variance components can be derived by studying the distribution

of a full rank error contrast applied to the response vector, here Y . In this section, the likelihood function of $K^\top Y$ will be discussed for a full rank error contrast matrix $K \in \mathbb{R}^{k \times k-p}$. Recall that a vector $u \in \mathbb{R}$ is called an error contrast if $\mathbb{E}(u^\top Y) = 0$. Hence, K is chosen such that any contrast u lies in the column space of K and the columns of K form a basis of the space of all error contrasts, namely the kernel of X^\top . Since $Y \sim \mathfrak{N}_k(X\beta, \Omega_{\tau\delta}^{-1})$, it follows that $K^\top Y \sim \mathfrak{N}_{k-p}(0, K^\top \Omega_{\tau\delta}^{-1} K)$. The density function of $U = K^\top Y$ is, thus,

$$f_{\tau\delta}(u) = \mathbf{t}^{-\frac{k-p}{2}} \det(K^\top \Omega_{\tau\delta}^{-1} K)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} u^\top (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} u\right). \quad (2.54)$$

In particular, by Theorem 2.11,

$$\begin{aligned} f_{\tau\delta}(K^\top y) &= \mathbf{t}^{-\frac{k-p}{2}} \det(K^\top \Omega_{\tau\delta}^{-1} K)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} y^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y\right) \\ &= \mathbf{t}^{-\frac{k-p}{2}} \det(K^\top \Omega_{\tau\delta}^{-1} K)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} q_{y\delta}(\tau)\right). \end{aligned}$$

Therefore,

$$l(\tau|y, \delta) = -\frac{1}{2} \left((k-p) \ln \mathbf{t} + \ln \det(K^\top \Omega_{\tau\delta}^{-1} K) + q_{y\delta}(\tau) \right).$$

Since the regression coefficients β are nuisance to the estimation of the heterogeneity τ , this likelihood can be used for the estimation of τ that is not confined by an estimation of β . Since $K^\top \Omega^{-1} K \in \mathbb{G}_{k-p}$ and $D_\tau \Omega^{-1} = I_k$,

$$\begin{aligned} D_\tau \det(K^\top \Omega^{-1} K) &= \det(K^\top \Omega^{-1} K)^{-1} \cdot \text{tr}(K^\top \Omega^{-1} K D_\tau (K^\top \Omega^{-1} K)) \\ &= \det(K^\top \Omega^{-1} K)^{-1} \cdot \text{tr}(K^\top \Omega^{-1} K K^\top D_\tau \Omega^{-1} K) \\ &= \det(K^\top \Omega^{-1} K)^{-1} \cdot \text{tr}(K^\top \Omega^{-1} K K^\top K) \\ &= \det(K^\top \Omega^{-1} K)^{-1} \cdot \text{tr}(\Omega^{-1} K K^\top K K^\top). \end{aligned}$$

Hence, by Lemma 2.12,

$$\begin{aligned} D_\tau l(\tau) &= D_\tau \ln \det(K^\top \Omega^{-1} K) + D_\tau q_y(\tau) \\ &= \frac{\det(K^\top \Omega^{-1} K)^{-1} \text{tr}(\Omega^{-1} K K^\top K K^\top)}{\det(K^\top \Omega^{-1} K)} - y^\top G y, \\ &= \frac{\text{tr}(\Omega^{-1} K K^\top K K^\top)}{\det^2(K^\top \Omega^{-1} K)} - y^\top G y, \end{aligned}$$

for the positive definite $G := A^\top A \in \mathbb{G}_k$ with $A := K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top$.

There is a discussion when, where, and whether to use a conditional likelihood approach for the estimation of variance components together with a strong pro argument in (Harville, 1977).

2.11 Independence of projected random vectors

For known heterogeneity τ and heteroscedasticity δ , the functions $y \mapsto \|E_{\tau\delta}y\|_{\tau\delta}^2$ and $y \mapsto \|\beta - B_{\tau\delta}y\|_{V_{\tau\delta}}^2$ are statistics of the random Y . The reformulation of the density function of the random effects meta regression model has already shown that these statistics play an important role here. They are both quadratic forms in y . The next section shall study the distributions of $\|E_{\tau\delta}Y\|_{\tau\delta}^2$ and $\|\beta - B_{\tau\delta}Y\|_{V_{\tau\delta}}^2$. The following lemma states a result concerning their independence. It shows that the estimators of $b_i^\top Y$ of each β_i are independent of the residual vector EY where b_i^\top denotes the i th row of B . Hence, $b_i^\top y = (By)_i$.

Lemma 2.19 If $Y \sim \mathfrak{F}_{\beta\tau\delta}$ follows the random effects meta regression model, then all linear forms in $B_{\tau\delta}Y$ are independent of $\|E_{\tau\delta}Y\|_{\tau\delta}^2$. In signs,

$$\forall i. (B_{\tau\delta}Y)_i \perp\!\!\!\perp \|E_{\tau\delta}Y\|_{\tau\delta}^2. \quad (2.55)$$

Proof. This is a consequence of Theorem 4.17 in (Graybill, 1961). \square

2.12 Distributions of residual squared lengths

Let Y denote a k -variate Gaussian distributed random vector with sample space \mathbb{R}^k . The previous sections studied the moments of quadratic forms $q_A(Y)$ for $A \in \mathbb{R}^{k \times k}$. This section shall study the distributions of $\|E_{\tau\delta}Y\|_{\tau\delta}^2$. The standard literature usually only considers the case in which the sample space is mapped to subspaces by orthogonal projections. If this is the case, theorems such as Theorem 2.20 from (Graybill, 1961) usually play important roles in the development of the theory of linear models and are used to deduce the distribution of the respected quadratic forms.

Theorem 2.20 (Distributions of orthogonal projections) Let $A \in \mathbb{R}^{k \times k}$ be symmetric and let $Y \sim \mathfrak{N}_k(\theta, V)$ for some vector $\theta \in \mathbb{R}^k$ and some positive definite unit $V \in \mathbb{G}_n$.

a) Let $\theta = 0$ and $V = I_k$. Then A is idempotent and of rank r if and only if

$$q_A(Y) \sim \mathfrak{X}^2_r. \quad (2.56)$$

b) Let $V = I_k$. Then A is idempotent and of rank r if and only if

$$q_A(Y) \sim \mathfrak{X}^2_r \left(\frac{1}{2} q_A(\theta) \right). \quad (2.57)$$

c) Let $V = \lambda I_k$ for some $\lambda > 0$. Then A is idempotent and of rank r if and only if

$$q_A(Y) \sim \mathfrak{X}^2_r \left(\frac{1}{2\lambda} q_A(\theta) \right). \quad (2.58)$$

Above, $\mathfrak{X}^2_r(\theta)$ shall denote the distribution of a non-centred chi-squared distributed random element with r degrees of freedom and non-centrality parameter θ .

Now, let Y follow the distribution of a random effects meta regression model. Then even though none of the above results can be applied to Y directly, a result about the distribution of $q_y(Y)$ can nevertheless be proven.

Corollary 2.21 Let $Y \sim \mathfrak{N}_k(X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and fixed $\tau \geq 0$ and $\delta > 0$. Then

- a) $\|\beta - B_{\tau\delta}Y\|_{V_{\tau\delta}}^2 \sim \mathfrak{X}^2_p$.
- b) $\|E_{\tau\delta}Y\|_{\tau\delta}^2 \sim \mathfrak{X}^2_{k-p}$.
- c) $\|\beta - B_{\tau\delta}Y\|_{V_{\tau\delta}}^2 + \|E_{\tau\delta}Y\|_{\tau\delta}^2 \sim \mathfrak{X}^2_k$.

Proof.

a) Since $V^{-\frac{1}{2}}(BY - \beta) \sim \mathfrak{N}_p(0, I_p)$, this yields

$$(BY - \beta)^\top V^{-1} (BY - \beta) \sim \mathfrak{X}^2_p. \quad (2.59)$$

b) By Theorem 2.11,

$$\|E_{\tau\delta}Y\|_{\tau\delta}^2 = q_y(\tau) = y^\top K(K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top y. \quad (2.60)$$

Equation (2.54) has shown that $K^\top Y \sim \mathfrak{N}_{k-p}(0, K^\top \Omega_{\tau\delta}^{-1} K)$. From this it follows directly that

$$Y^\top K (K^\top \Omega_{\tau\delta}^{-1} K)^{-1} K^\top Y \sim \mathfrak{X}_{k-p}^2. \quad (2.61)$$

- c) The model assumption is $Y \sim \mathfrak{N}_k(X\beta, \Omega_{\tau\delta}^{-1})$. Thus, $\Omega_{\tau\delta}^{\frac{1}{2}}(Y - X\beta) \sim \mathfrak{N}_k(0, I_k)$ and, therefore, $(Y - X\beta)' \Omega_{\tau\delta} (Y - X\beta) \sim \mathfrak{X}_k^2$. \square

Lemma 2.22 allows to construct confidence intervals for the location parameters β_i , $i = 1, \dots, p$. In Lemma 2.19, it was shown that each linear form in $B_{\tau\delta} Y$ is independent of $q_Y(\tau)$. Let \mathfrak{T}_d denote the centred Student's t-distribution with d degrees of freedom.

Lemma 2.22 Let $Y \sim \mathfrak{N}_k(X\beta, \Omega_{\tau\delta}^{-1})$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and fixed $\tau \geq 0$ and $\delta > 0$. Then,

$$\frac{b_i Y - \beta_i}{\sqrt{v_i \frac{q_Y(\tau)}{k-p}}} \sim \mathfrak{T}_{k-p}, \quad (2.62)$$

where v_i denotes the i th element of $(v_1, \dots, v_k) = \text{diag } V_{\tau\delta}$.

Proof. By Lemma 2.19, Lemma 2.21, and since $b_i Y \sim \mathfrak{N}(\beta_i, v_i)$. \square

2.13 *Extension:* Mean responses as summary statistics

In cases in which it is not possible to establish within-study variabilities trustworthy enough, one may not feel comfortable enough to assume the δ_j 's to be known. One nevertheless has to use study a specific estimator D_j for each δ_j . Usually, one assumes the D_j 's to be at least consistent for each δ_j and sets $\hat{\delta}_j := D_j$ for all $j = 1, \dots, k$ in the estimation process.

Consider the case in which each effect measure $Y_j | \theta_j$ denotes the mean response of study j with expected value θ_j and variance equal to

$$\delta_j := \frac{\sigma_j^2}{n_j}. \quad (2.63)$$

Here, σ_j^2 denotes the *population variance* and n_j the number of subjects in the j th study. Assuming again Gaussian responses, model (2.2)+(2.3) can be restated as

$$Y_j | \theta_j \sim \mathfrak{N} \left(\theta_j, \frac{\sigma_j^2}{n_j} \right), \quad (2.64)$$

$$\theta_j \sim \mathfrak{N} \left(x_j^\top \beta, \tau \right). \quad (2.65)$$

Let S_j denote the sum of squares of the j th study, in signs: $S_j = \sum_i (Y_{ji} - Y_j)^2$, where each Y_{jl} denotes the l th individual response in the j th study. Then $D_j := \frac{S_j}{n_j(n_j-1)}$ is an estimator for δ_j , since $\mathbb{E}(S_j) = \sigma_j^2(n_j - 1)$. Also,

$$\frac{S_j}{\sigma_j^2} \sim \mathfrak{X}^2_{n_j-1}, \quad \text{for } j = 1, \dots, k, \quad (2.66)$$

since \mathfrak{N} is infinitely divisible. From (2.66) follows

$$D_j = \frac{S_j}{n_j(n_j-1)} = \frac{\sigma_j^2}{n_j} \cdot \frac{1}{n_j-1} \cdot \frac{S_j}{\sigma_j^2} = \frac{\delta_j}{n_j-1} \cdot K_j, \quad (2.67)$$

where $K_j \sim \mathfrak{X}^2_{n_j-1}$. Thus, if the δ_j are truly assumed to be unknown, the underlying model is in fact

$$Y_j | \theta_j \sim \mathfrak{N}(\theta_j, D_j), \quad (2.68)$$

$$\theta_j \sim \mathfrak{N} \left(x_j^\top \beta, \tau \right), \quad (2.69)$$

$$K_j \sim \mathfrak{X}^2_{n_j-1}, \quad (2.70)$$

$$D_j = \frac{\delta_j}{n_j-1} \cdot K_j. \quad (2.71)$$

The (Y_j, D_j) , $j = 1, \dots, k$, denote the summary statistics of the j th study. This model, however, is no longer identifiable.

2.14 *Extension*: Binomial responses, risk ratios, and their limits

This extension to the random effects meta regression model is applied to *count data* that comes in the form of the example introduced in Section 1.2. Thompson and Sharp (1999) suggested a binomial-Gaussian hierarchical model description for modelling data such as the BCG vaccine efficacy data in a Bayesian framework. The modelling strategy of this section is loosely based on their idea.

The example introduced in Section 1.2 consisted of data of clinical trials in which group assignments were highly unbalanced. A measure of unbalance b was defined

when the data set was introduced. Let v denote the number of vaccinated subjects within a particular trial, and let s be its total study size. Let $\neg v = s - v$. Then the *measure of unbalance* was defined as

$$\begin{aligned} b &:= \frac{v - \neg v}{v + \neg v} = \frac{v - (s - v)}{s} = \frac{2v - s}{s} \\ \Leftrightarrow \quad 2v &= sb + s = s(b + 1) \\ \Leftrightarrow \quad v &= \frac{s(b + 1)}{2}. \end{aligned}$$

Given parameters $b = (b_1, \dots, b_k) \in (0, 1)^k$ and $s = (s_1, \dots, s_k) \in \mathbb{N}^k$, define

$$v_j = \left(\frac{s_j(b_j + 1)}{2} \right)_{\mathbb{N}}, \quad \neg v_j = s_j - v_j.$$

Here, $(c)_{\mathbb{N}}$ shall denote c rounded to the nearest integer. For a full rank design matrix $X \in \mathbb{R}^{k \times p}$ with rows x_j^\top , a vector $\beta \in \mathbb{R}^p$, and a parameter of heterogeneity $\tau > 0$, define the following hierarchical binomial-Gaussian model:

$$n_{j1} \sim \mathbf{Binom}(v_j, p_j), \quad n_{j2} = v_j - n_{j1}, \quad (2.72)$$

$$m_{j1} \sim \mathbf{Binom}(\neg v_j, q_j), \quad m_{j2} = \neg v_j - m_{j1}, \quad (2.73)$$

$$q_j = \text{some fixed number in } (0, 1), \quad (2.74)$$

$$p_j = q_j \cdot \exp(Y_j), \quad (2.75)$$

$$Y_j = x_j^\top \beta + e_j, \quad (2.76)$$

$$e_j \sim \mathfrak{N}(0, \tau). \quad (2.77)$$

Above, $\mathbf{Binom}(n, p)$ denotes the distribution of a binomial distributed random element of sample size n and success probability p . By (2.75), each p_j , q_j and Y_j stand in the relation

$$Y_j = \ln(p_j) - \ln(q_j). \quad (2.78)$$

It is well known that $\hat{p}_j = \frac{n_{j1}}{n_{j1} + n_{j2}}$ and $\hat{q}_j = \frac{m_{j1}}{m_{j1} + m_{j2}}$ are estimates of p_j and q_j , and

$$\begin{aligned} \sqrt{v_j} (\hat{p}_j - p_j) |_{p_j} &\rightsquigarrow \mathfrak{N}(0, p_j(1 - p_j)), \\ \sqrt{\neg v_j} (\hat{q}_j - q_j) |_{q_j} &\rightsquigarrow \mathfrak{N}(0, q_j(1 - q_j)). \end{aligned}$$

A straight forward application of the delta method yields that

$$\begin{aligned}\sqrt{v_j} (\ln(\hat{p}_j) - \ln(p_j)) |_{p_j} &\rightsquigarrow \mathfrak{N} \left(0, \frac{1-p_j}{p_j} \right), \\ \sqrt{-v_j} (\ln(\hat{q}_j) - \ln(q_j)) |_{q_j} &\rightsquigarrow \mathfrak{N} \left(0, \frac{1-q_j}{q_j} \right).\end{aligned}$$

This shows that the *logarithm of the relative risk*, namely $\ln(\hat{p}_j) - \ln(\hat{q}_j)$, has expectation $x_j^\top \beta$ and its variance is approximately

$$d_j = \frac{1-p_j}{v_j p_j} + \frac{1-q_j}{-v_j q_j}. \quad (2.79)$$

This suggests to define

$$y_j := \ln \left(\frac{n_{j1}}{n_{j1} + n_{j2}} \right) - \ln \left(\frac{m_{j1}}{m_{j1} + m_{j2}} \right) \quad (2.80)$$

as the effect size of the meta regression and to estimate its variance as

$$\begin{aligned}\hat{d}_j &:= \frac{1-\hat{p}_j}{v_j \hat{p}_j} + \frac{1-\hat{q}_j}{-v_j \hat{q}_j} \\ &= \frac{1-\hat{p}_j}{n_{j1}} + \frac{1-\hat{q}_j}{m_{j1}} \\ &= \frac{1}{n_{j1}} - \frac{1}{n_{j1} + n_{j2}} + \frac{1}{m_{j1}} - \frac{1}{m_{j1} + m_{j2}}.\end{aligned}$$

In the case of the BCG vaccine efficacy data, the question is whether the distance of a clinic to the equator has a linear influence on the logarithm of the relative risk. Regressing on *absolute latitude* results in setting $x_j = (1, a_j)^\top$ and setting $n_{j1}, n_{j2}, m_{j1}, m_{j2}$, and a_j to the values in Table 1.1 in the fashion:

Trial	Author	Year	Vaccinated		Not vaccinated		Absolute Latitude
			Diseased	Not diseased	Diseased	Not diseased	
j	n_{j1}	n_{j2}	m_{j1}	m_{j2}	a_j

3 Theoretical background about generalised inference

3.1 Introduction

One aim of this text is to study whether constructions based on generalised inference principles may improve the performance of statistical estimators for the random effects meta regression model for small number of studies.

Analogue to Bayesian methods, *generalised inference* has the goal of constructing confidence sets which are the result of exact probability statements. In contrast to Bayesian methods, however, generalised methods dismiss the use of priors. Of course, the presence of nuisance parameters, here τ and δ , complicates matters. Tsui and Weerahandi (1989) advertised an approach to carry out such constructions which they coined with the label *generalised inference*. The paper was followed by an extension of the principle to confidence intervals in (Weerahandi, 1993), and the book (Weerahandi, 1995). Hannig (2009) was able to show that the generalised inference principle is in fact related to the *fiducial argument* by Fisher (1935) and that both can be unified into, what he calls, the *generalised fiducial recipe*.

3.2 Basic concepts of generalised inference for confidence intervals

The definitions of this section are more or less directly taken from (Weerahandi, 1995). As this whole chapter is of more general nature and, thus, not particularly concerned with the random effects meta regression model itself, the notation will be slightly altered to reflect the changed emphasis: the random vector capturing the data generating processes will be denoted by X , the parameter of interest by θ , and the nuisance parameter by δ .

So, let us assume that a continuous random vector X is given which distribution belongs to a family of distributions F indexed by a parametrisation \mathfrak{F} with convex and closed index sets $\Theta \subseteq \mathbb{R}$ and $\Delta \subseteq \mathbb{R}^k$,

$$\mathfrak{F}: \Theta \times \Delta \rightarrow F, \quad (\theta, \delta) \mapsto \mathfrak{F}_{\theta\delta}. \quad (3.1)$$

Hence, $X \sim \mathfrak{F}_{\theta\delta}$ for some $(\theta, \delta) \in \Theta \times \Delta$. It is of interest to construct confidence intervals for θ in the presence of the nuisance parameter δ . The problem is tackled by defining a (generalised) pivotal quantity R for θ . Tsui and Weerahandi (1989) call the pivot R *generalised*, since the value of R is not only allowed to depend on the parameters (θ, δ) but also on an observed data point x of X .

Definition 3.1 Let $R = R(X, x, \theta, \delta)$ be a real valued function. The function R is called a *generalised pivotal quantity* if and only if

- a) $R(x, x, \theta, \delta)$ does not depend on the nuisance parameter δ . The value

$$r(x, \theta) := R(x, x, \theta, \delta) \quad (3.2)$$

is called the *observed value* of R .

- b) As a function in θ , the observed value $r(x, \cdot): \Theta \rightarrow \mathbb{R}$ is monotone and continuous.
 c) $R(X, x, \theta, \delta)$ as a function in X is a statistic of X for any given x, θ and δ . In case of $X \sim \mathfrak{F}_{\theta\delta}$, the distribution of $R(X, x, \theta, \delta)$ does not depend on θ nor δ .

The above axioms stipulated for a generalised pivotal quantity are of technical character. Property a) assures that $r(x, \theta)$ can be calculated for any given θ without the knowledge of the nuisance parameter δ , property c) assures that it is possible to calculate probabilities of the form $P(R \in C)$ for any $C \in \text{Bor}(\mathbb{R})$, whereas property b) assures that the inferred confidence set is a genuine interval in \mathbb{R} .

Let $\mathfrak{R}^{(x)}$ denote the distribution of $R(X, x, \theta, \delta)$ and let $\mathfrak{r}_\gamma^{(x)}$ be its γ -quantile for any $\gamma \in (0, 1)$. The superscript $^{(x)}$ shall stress that $\mathfrak{R}^{(x)}$ and $\mathfrak{r}^{(x)}$ depend on the observed data x . For $\gamma \in (0, \frac{1}{2})$ and any observed x , it is

$$P_{\theta\delta} \left(\mathfrak{r}_{\frac{\gamma}{2}}^{(x)} \leq R(X, x, \theta, \delta) \leq \mathfrak{r}_{1-\frac{\gamma}{2}}^{(x)} \right) = 1 - \gamma. \quad (3.3)$$

This is what is called an *extreme region* of a pivotal quantity R and should motivate the next definition.

Definition 3.2 Let $(\gamma, x) \mapsto C_\gamma(x)$ be such that each $C_\gamma(x)$ is a closed connected interval in \mathbb{R} for all observable x , and let R be a generalised pivotal quantity. The sequence $(C_\gamma(x))_{\gamma \in (0,1)}$ is called a *family of extreme regions* for R in x if and only if

$$P_{\theta\delta}(R(X, x, \theta, \delta) \in C_\gamma(x)) = 1 - \gamma \quad (3.4)$$

for all γ . If this is the case, the (genuine) interval

$$I_\gamma(x) := \{\theta : R(x, x, \theta, \delta) \in C_\gamma\} = \{\theta : r(x, \theta) \in C_\gamma\} \quad (3.5)$$

is called a *generalised $(1 - \gamma)$ -confidence interval* for θ .

If $R(X, x, \theta, \delta)$ is a generalised pivotal quantity, then the probability in (3.4) neither depends on θ nor δ . From (3.3) follows that

$$\left(\left[\mathbf{r}_{\frac{\gamma}{2}}^{(x)}, \mathbf{r}_{1-\frac{\gamma}{2}}^{(x)} \right] \right)_{\gamma \in (0,1)} \quad (3.6)$$

is a family of extreme regions for R . Hence,

$$I_\gamma(x) = \left\{ \theta : r(x, \theta) \in \left[\mathbf{r}_{\frac{\gamma}{2}}^{(x)}, \mathbf{r}_{1-\frac{\gamma}{2}}^{(x)} \right] \right\} \quad (3.7)$$

is a $(1 - \gamma)$ -confidence set for θ . Since $r(x, \theta)$ is monotone and continuous in θ , the set in (3.7) is a connected and closed genuine interval.

As an example, consider the case in which the observed value of R is simply equal to θ itself, in signs: $r(x, \theta) = \theta$. Then the extreme region

$$\left[\mathbf{r}_{\frac{\gamma}{2}}^{(x)}, \mathbf{r}_{1-\frac{\gamma}{2}}^{(x)} \right] \quad (3.8)$$

itself already defines a generalised $(1 - \gamma)$ -confidence interval for θ . In fact, without loss of generality, it can always be assumed that this is the case as otherwise one may replace $R(X, x, \theta, \delta)$ by $R(X, x, \theta, \delta) - R(x, x, \theta, \delta) + \theta$. It can also be assumed without loss of generality that $r(x, \cdot)$ is monotone increasing as otherwise one can always replace $R(X, x, \theta, \delta)$ by its negative: $-R(X, x, \theta, \delta)$.

3.3 Higher dimensions: generalising generalised inference

Definition 3.1 as defined by Weerahandi (1995) is sufficient to define generalised confidence intervals for each of the regression coefficients β_i , $i = 1, \dots, p$, of the random effects meta regression model. The construction is laid out in Chapter 5, and a study of the performance of these estimators is studied in Section 5.6. If it is possible to construct confidence intervals for each single β_i , it is straight forward to construct confidence intervals for all β_i simultaneously meeting a given overall confidence: setting each individual confidence level to $1 - \frac{\gamma}{p}$, the overall confidence will be equal to *or greater* than $1 - \gamma$. This is known as the Bonferroni correction. Doing so, one has in fact constructed a confidence cuboid for β in \mathbb{R}^p . Sadly, this higher dimensional confidence set is very conservative and, when p is large, of little practical use.

In this section, let X be a continuous random vector which distribution belongs to a family of distributions F indexed by a parametrisation \mathfrak{F} with convex and closed index sets $\Theta \subseteq \mathbb{R}^p$ and $\Delta \subseteq \mathbb{R}^k$,

$$\mathfrak{F}: \Theta \times \Delta \rightarrow F, \quad (\theta, \delta) \mapsto \mathfrak{F}_{\theta\delta}. \quad (3.9)$$

The set-up is, thus, the same as in the previous section, only that θ now lives in a higher dimensional space. Hence, $X \sim \mathfrak{F}_{\theta\delta}$ for some $(\theta, \delta) \in \Theta \times \Delta$.

The goal, now, is to construct (generalised) confidence sets in \mathbb{R}^p . However, property c) of Definition 3.1 is a problem as there exists no canonical order on \mathbb{R}^p for $p > 1$. Hence, *monotonicity* as used in Definition 3.1 makes little sense in higher dimensions. When it comes to confidence sets, the adequate generalisation of (one dimensional) intervals living \mathbb{R} are ellipses in \mathbb{R}^2 or, more general, ellipsoids in \mathbb{R}^p , $p > 2$. In classical statistical theory, ellipsoids occur naturally when constructing confidence sets for parameter vectors in higher dimensions as they are of minimal Lebesgue measure in most applications.

On the basis of ellipsoids lie quadratic forms, as they have already been discussed, or, more generally, polynomials of degree 2. As no ambiguity can occur, no difference will be made notationally between a polynomial $q \in \mathbb{R}[\theta]$ and its induced polynomial map $q: \mathbb{R}^p \rightarrow \mathbb{R}$. Thus, the induced quadratic form of any *homogeneous degree 2 polynomial* q will also be denoted as q , just as it was the case in previous chapters.

Let q be a real polynomial in $\theta = (\theta_1, \dots, \theta_p)$, in signs: $q \in \mathbb{R}[\theta]$. Every degree 2 polynomial q can be written in the form

$$q(\theta) = \theta^\top A \theta + b \theta + r \quad (3.10)$$

where $A = (a_{ij})$ is a $p \times p$ matrix, $b \in \mathbb{R}^p$ and $r \in \mathbb{R}$. In (3.10), the a_{ij} are the coefficients of the monomials $\theta_i \theta_j$ of degree 2, the b_i are the coefficients of the monomials θ_i of degree 1, and r is the coefficient of the unique monomial of degree 0. Note that A in (3.10) can be assumed to be symmetric as otherwise the matrix can be replaced by $\frac{1}{2}(A + A^\top)$, which yields the same polynomial. Let us, therefore, assume without loss of generality that A is symmetric.

Interest lies particularly in polynomials in which A is not only symmetric but also positive definite. Assume that q can be written in the form

$$\begin{aligned} q(\theta) &= (\theta - c)^\top A (\theta - c) \\ &= \theta^\top A \theta - \theta^\top A c - c^\top A \theta + c^\top A c \\ &= \theta^\top A \theta - 2c^\top A \theta + c^\top A c \end{aligned} \quad (3.11)$$

for some fixed $c \in \mathbb{R}^p$. In case $c = 0$, then q is *homogeneous*, i.e., all terms in q are monomials of the same degree. The induced polynomial map of a homogeneous polynomial of degree 2 is a *quadratic form*. If q is of the form (3.11) with positive definite A , then q itself is called *positive definite*. Conveniently, every positive definite q defines a genuine ellipsoid in \mathbb{R}^p as for any $l \in \mathbb{R}$, $l > 0$ the preimage

$$q^{-1}(l) = \{\theta : q(\theta) = l\} \quad (3.12)$$

is a non-empty elliptic surface in \mathbb{R}^p with centre c . Let v_1, \dots, v_p and $\lambda_1, \dots, \lambda_p$ be the eigenvectors and eigenvalues of A respectively. Then the elliptic surface $q^{-1}(l)$ has the principle components v_1, \dots, v_p and the equatorial radii $\sqrt{\frac{l}{\lambda_1}}, \dots, \sqrt{\frac{l}{\lambda_p}}$.

If $r, l \in \mathbb{R}$ and $r < l$, then $q^{-1}(r)$ lies entirely within $q^{-1}(l)$ having the same centre and principle components. In particular, $q^{-1}([0, l])$ is the closed solid ellipsoid having $q^{-1}(l)$ as its surface.

If q is a homogeneous positive definite polynomial of degree 2, then

$$b(\theta, \iota) := \frac{1}{2}(q(\theta + \iota) - q(\theta) - q(\iota)), \quad (3.13)$$

as a map from $(\mathbb{R}^p)^2$ to \mathbb{R} , defines a symmetric positive definite bilinear form on \mathbb{R}^p as the induced map of q is a positive definite quadratic form. If $q(\theta) = \theta^\top A \theta$, then $b(\theta, \iota) = \theta^\top A \iota$. In other words, b is an inner product making (\mathbb{R}^p, b) into an inner product space.

Definition 3.3 Let $p > 1$ and $R = R(X, x, \theta, \delta)$ be a real valued function.

The function R is called a *generalised pivotal quantity* for θ if and only if

- a) the observed value $R(x, x, \theta, \delta)$ does not depend on the nuisance parameter δ . For any δ , let $r(x, \theta) := R(x, x, \theta, \delta)$.
- b) For all x , the observed value $r(x, \theta)$ as a function in θ can be written as a positive definite quadratic form $q_x \in \mathbb{R}[\theta]$, in signs: $q_x(\theta) = r(x, \theta)$.
- c) If $X \sim \mathfrak{F}_{\theta\delta}$, then $R(X, x, \theta, \delta)$ as a function in X is a random variable which distribution is free of θ and δ .

Since q is positive definite, all observed values $r(x, \theta)$ lie in $\mathbb{R}_{\geq 0}$. Again, let $\mathfrak{R}^{(x)}$ denote the distribution of $R(X, x, \theta, \delta)$ and let $\mathbf{r}_\gamma^{(x)}$ denote its γ -quantile. Then for all x and $\gamma \in (0, 1)$

$$P\left(R(X, x, \theta, \delta) \leq \mathbf{r}_{1-\gamma}^{(x)}\right) = 1 - \gamma. \quad (3.14)$$

Hence,

$$I_\gamma(x) := \left(\left(-\infty, \mathbf{r}_{1-\gamma}^{(x)} \right] \right)_{\gamma \in (0,1)} \quad (3.15)$$

is a family of extreme regions of R in the sense of Definition 3.2. Thus, (3.15) defines a generalised $(1 - \gamma)$ -confidence set $I_\gamma(x)$ for θ . Since $r(x, \theta) \geq 0$ for all θ ,

$$I_\gamma(x) = \left\{ \theta : r(x, \theta) \in \left[0, \mathbf{r}_{1-\gamma}^{(x)} \right] \right\} = q_x^{-1} \left(\left[0, \mathbf{r}_{1-\gamma}^{(x)} \right] \right). \quad (3.16)$$

Since q_x is a positive definite quadratic form, this preimage $I_\gamma(x)$ is a solid closed ellipsoid in \mathbb{R}^p .

3.4 Discussion

As pointed out at the beginning of this chapter, the generalised inference principle has certain analogies to inference in the Bayesian framework. Both frameworks result in a distribution on the parameter space. In the Bayesian framework, this distribution is usually called the *posterior* distribution. As Cameron and Trivedi (2005) point out, Bayesian methods can produce entire estimates of this distribution. This, too, is the case for generalised inference methods. As a result, the choice of which moment or which quantile of the parameter distribution should be reported is left until to the end of the analysis. Confidence intervals for the parameters of interest are only one choice. In particular, there is no need for separate estimators for the mean, the variance, the mode, the median, or other moments and quantiles of the distribution.

In the upcoming simulation studies, point estimators based on the median of the constructed generalised pivotal distributions are tested against the performance of other more commonly used point estimators for the heterogeneity and the regression coefficients in the random effects meta regression model. Other choices are obviously also possible such as the mode or the mean of the respected distributions.

4 Inference on the heterogeneity parameter

4.1 Introduction

Estimating the variance components of a model is an integral part in any statistical inference. In random effects meta regression, accurate estimation of the heterogeneity parameter is of particular importance. For example, the decision whether to analyse data using a meta regression model, instead of a meta analysis model, may be based on how much the magnitude of the respected heterogeneity parameters can be reduced by using the former model. In the meta regression context this is of particular importance. One may say, the goal of a meta analysis is the question whether there *exists* an effect. The goal in performing a meta regression is to study *how* a covariate influences the measured effect. The question is not *if* but *how*. If the heterogeneity component τ can greatly be reduced by introducing a covariate to the meta regression model, this can be interpreted as a sign that the covariate should indeed be considered for data analysis.

As there exists no non-negative unbiased estimator for τ , (LaMotte, 1973), a variety of different choices for estimating the heterogeneity have been proposed in the literature. Six of these estimators will be discussed in the following. Most of these were originally proposed for estimating the heterogeneity in the random effects meta analysis model (where $p = 1$ and $\beta \in \mathbb{R}$), namely the Hedges estimator, the DerSimonian–Laird estimator, the Sidik–Jonkman estimator, and the Mandel–Paule estimator. Also, the maximum likelihood estimator and the restricted maximum likelihood estimator are presented. Additionally, iterated versions of the DerSimonian–Laird and Sidik–Jonkman estimators are discussed. For the conventional random effects meta analysis model, some variance and bias estimates for the heterogeneity estimators have been established in (Viechtbauer, 2005).

4.2 Methods of moments

The Hedges, the DerSimonian–Laird, the Mandel–Paule, and Sidik–Jonkman estimators are method of moments estimators. Thus, each of these are essentially based on the same idea. The goal in this section is to put each of these estimators into a common context and to integrate each in the general theory. Fortunately, Corollary 2.8 will be doing all the hard work for us in this section.

Assume that $Y \sim \mathfrak{F}_{\beta\tau\delta}$ follows a random meta regression model. For given $\tau' > 0$ and $\delta' > 0$, let $E' = E_{\tau'\delta'}$. Then first note that the residual projection $E'Y$ is a consistent estimator for $\mathbb{E}(E'Y)$, and $E'y$ is an estimate of $\mathbb{E}(E'Y)$. Since the

squared norms $\|\cdot\|^2$ and $\|\cdot\|_{\tau'\delta'}^2$ are continuous mappings, it follows by the continuous mapping theorem that also $\|E'Y\|^2$ and $\|E'Y\|_{\tau'\delta'}^2$ are consistent estimators for $\mathbb{E}(\|E'Y\|^2)$ and $\mathbb{E}(\|E'Y\|_{\tau'\delta'}^2)$ respectively.

Suppose δ is known or that at least a trustworthy estimate, say $\hat{\delta}$, is given for δ . Then each part of Corollary 2.8 consists of an equation in the unknown true heterogeneity parameter τ . Taking a ‘‘good guess’’, say τ' , for τ , one defines $E' = E_{\tau'\hat{\delta}}$. Then using this estimate (namely E' for E), one uses the equations in Corollary 2.8 and solves for τ . This yields a new guess for $\hat{\tau}$. One could iterate this procedure indefinitely till the sequence of $\hat{\tau}$'s converges or stop at a predefined point. The literature usually stops after one step of the iteration process.

Note that when studying the following methods of moments estimators, no parametric assumptions are made about the distribution of the study effects Y . In this section, it will only be assumed that the first two moments of Y follow the structure of a random effect meta regression model as defined in Definition 2.1, which is given by its mean and variance-covariance structure:

$$Y \sim (X\beta, [\tau]_k + [\delta]). \quad (4.1)$$

Hedges type estimators

As mentioned before, Corollary 2.8 will be doing the hard work in this section. The Hedges type estimator is based on Corollary 2.8a). Let $E' = E_{\tau'\delta'}$ be a residual projection for some given τ' and δ' . Solving 2.8a) for τ suggests the method of moments estimator

$$\hat{\tau} = \left(\frac{\|E'y\|^2 - \text{tr}(E'[\hat{\delta}])}{k - p} \right)_{\geq}. \quad (4.2)$$

Here, $(c)_{\geq}$ shall denote c truncated at 0 from below: $(c)_{\geq} = \max(0, c)$. The truncation is necessary as otherwise the above estimate may yield negative values in any practical application. This has the side effect of $\hat{\tau}$ becoming biased. The Hedges type estimator follows from (4.2) by applying the orthogonal projection

$$E_{01} = I_k - X(X^\top X)^{-1}X^\top \quad (4.3)$$

for E' .

Definition 4.1 (Hedges type estimator) Let Y follow the random effects meta regression model with heterogeneity τ and heteroscedasticity δ . Let $\hat{\delta}$ be an estimate of δ and let $E' = E_{01}$. The method of moments estimator

$$\hat{\tau}_{HE} = \left(\frac{\|E_{01}y\|^2 - \text{tr}(E_{01}[\hat{\delta}])}{k - p} \right)_{\geq} \quad (4.4)$$

is called a *Hedges type estimator* for the heterogeneity τ .

If $p = 1$ and $X = 1_k$, then the random effects meta regression model reduces to the random effects meta analysis model, in which case $\hat{\tau}_{HE}$ reduces to the heterogeneity estimate introduced by Hedges (1983).

DerSimonian–Laird type estimators

The positive definite variance-covariance matrix of a non-degenerate multivariate Gaussian distribution defines an inner product on the sample space. In the case of the random effects meta regression model, the product is given by

$$\langle x, y \rangle_{\tau\delta} = x^\top ([\tau]_k + [\delta])^{-1} y = \frac{1}{\tau} \langle x, y \rangle^2 + \sum_j \frac{x_j y_j}{\delta_j}. \quad (4.5)$$

The latter summand being a weighted sum. Assume that δ is known and the hypothesis $\tau = 0$ is supposed to be tested. This would suggest to work in the Euclidean space $(\mathbb{R}^k, b_{[\delta]})$, in which

$$b_{[\delta]^{-1}}(x, y) = x^\top [\delta]^{-1} y = \sum_j \frac{x_j y_j}{\delta_j} \quad (4.6)$$

defines the inner product between $x, y \in \mathbb{R}^k$. In particular, $\|\cdot\|_{0\delta}$ defines the induced norm on \mathbb{R}^k .

Lemma 4.2 Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and fixed $\tau \geq 0$ and $\delta > 0$. Let $E' = E_{0\delta}$. Then

$$\mathbb{E}_{\tau\delta} (\|E'Y\|_{0\delta}^2) = \tau \cdot \text{tr}(E'^\top [\delta]^{-1}) + k - p. \quad (4.7)$$

Proof. Setting $\tau' = 0$ in Corollary 2.8c) yields

$$\begin{aligned}
\mathbb{E}_{\tau\delta} (\|E'Y\|_{0\delta}^2) &= \tau \operatorname{tr} (E'^{\top} [\delta]^{-1}) + \operatorname{tr} (E'[\delta]^{-1}[\delta]) \\
&= \tau \operatorname{tr} (E'^{\top} [\delta]^{-1}) + \operatorname{tr} (E'^{\top}) \\
&= \tau \operatorname{tr} (E'^{\top} [\delta]^{-1}) + k - p
\end{aligned}$$

□

The DerSimonian–Laird type estimator follows from applying the projection

$$E_{0\hat{\delta}} = I_k - X(X^{\top}[\hat{\delta}]^{-1}X)^{-1}X^{\top}[\hat{\delta}]^{-1} \quad (4.8)$$

to E' for a heteroscedasticity estimate $\hat{\delta}$ and solving (4.7) for the unknown τ . Again, this suggests a method of moments estimator.

Definition 4.3 (DerSimonian–Laird type estimator) Let Y follow the random effects meta regression model, i.e., $Y \sim (X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$ and fixed $\tau \geq 0$ and $\delta > 0$. Let $\hat{\delta}$ be an estimate of δ . The method of moments estimator

$$\hat{\tau}_{DL} = \left(\frac{\|E_{0\hat{\delta}}y\|^2 - k + p}{\operatorname{tr} (E_{0\hat{\delta}}^{\top}[\hat{\delta}]^{-1})} \right)_{\geq} \quad (4.9)$$

is called a *DerSimonian–Laird type estimator* for the heterogeneity τ .

Again, the truncation in (4.9) is necessary as otherwise the estimate might yield negative values. Thus, also $\hat{\tau}_{DL}$ is biased. If $p = 1$ and $X = \mathbf{1}_k$, the estimate $\hat{\tau}_{DL}$ reduces to the estimate in (DerSimonian and Laird, 1986) for the random effects meta analysis model.

Replacing $E_{0\hat{\delta}}$ with $E_{\hat{\tau}_{DL}\hat{\delta}}$ in (4.9) allows the iteration of the above method. The convergence of this sequence of τ -estimates might be worth studying, and, if its limit exists, the properties of its resulting estimator.

Mandel–Paule type estimators

Section 2.6 defined the function

$$\tau \mapsto q_y(\tau), \quad q_y(\tau) := \|E_{\tau\delta}y\|_{\tau\delta}^2,$$

and it was shown in Corollary 2.13 that this function is strictly monotone decreasing with limit point 0. By Corollary 2.8b),

$$\mathbb{E} (\|E_{\delta\tau}Y\|_{\tau\delta}^2) = k - p. \quad (4.10)$$

The Mandel–Paule type estimator is the unique solution to the equation $\|E_{\tau\delta}y\|_{\tau\delta}^2 = k - p$. The inverse function of q_y was defined as

$$p_y(\eta) = \begin{cases} q_y^{-1}(\eta) & : 0 < \eta < q_y(0) \\ 0 & : \text{otherwise} \end{cases}$$

for $\eta \in \mathbb{R}$.

Definition 4.4 (Mandel–Paule type estimator) Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and fixed $\tau \geq 0$ and $\delta > 0$. Let $\hat{\delta}$ be an estimate of δ . The method of moments estimator

$$\hat{\tau}_{MP} = p_y(k - p) \quad (4.11)$$

is called a *Mandel–Paule type estimator* for τ .

In the reduced case $p = 1$, $\beta \in \mathbb{R}$, and $X = 1_k$, this estimator was first suggested for a special case of the random effects meta analysis model in (Mandel and Paule, 1970) and then in its general form in (Paule and Mandel, 1982).

When $q_y(0) \leq k - p$, which can happen in any practical application, then it follows that $p_y(k - p) = 0$ by definition. Hence, in this case the Mandel–Paule type estimator will yield $\hat{\tau}_{MP} = 0$. In a certain way, this resembles the truncation by 0 as in the previous method of moments estimators: it represents a discontinuity in the estimation.

Sidik–Jonkman type estimators

The Sidik–Jonkman type estimator is based on a different parametrisation of the meta regression model. Define $\xi := \frac{1}{\tau}\delta = (\frac{\delta_1}{\tau}, \dots, \frac{\delta_k}{\tau})^\top$. Then

$$[\tau]_k + [\delta] = \tau \left(I_k + \frac{1}{\tau} [\delta] \right) = \tau (I_k + [\xi]) = \tau \cdot [1_k + \xi]. \quad (4.12)$$

Here, I_k shall denote the unity in \mathbb{G}_k , and 1_k the k -vector of 1's, i.e., $1_k \in \mathbb{R}^k$ and $1_k = (1, \dots, 1)$. In the Sidik–Jonkman approach the variance components of the random effects meta regression model are parametrised by (τ, ξ) instead of (τ, δ) . This still makes $\xi > 0$ a nuisance parameter. Assuming ξ known, this suggests to consider the inner product induced by $b_{[1_k + \xi]^{-1}}$ on \mathbb{R}^k . Note that $[1_k + \xi] = \Omega_{1\xi}^{-1}$. Thus, in the notation for quadratic forms and norms adopted so far, $\langle \cdot, \cdot \rangle_{1\xi}$ and $\|\cdot\|_{1\xi}$ define the corresponding inner product and induced norm on \mathbb{R}^k .

Lemma 4.5 Let $Y \sim (X\beta, \tau[1 + \xi])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and $\tau \geq 0$ and $\xi > 0$. Then

$$\mathbb{E}_{\tau\delta} \left(\|E_{1\xi} Y\|_{1\xi}^2 \right) = \tau(k - p). \quad (4.13)$$

Proof. From Theorem 2.7 follows that

$$\begin{aligned} \mathbb{E}_{\tau\delta} \left(\|E_{1\xi} Y\|_{1\xi}^2 \right) &= \text{tr} \left(E_{1\xi}^\top [1 + \xi]^{-1} E_{1\xi} \tau [1 + \xi] \right) \\ &= \text{tr} \left(\tau E_{1\xi}^\top [1 + \xi]^{-1} E_{1\xi} [1 + \xi] \right) \\ &= \text{tr} \tau \left(E_{1\xi}^\top \right) \\ &= \tau(k - p). \end{aligned}$$

□

Assuming δ known, $\xi = \xi(\tau)$ can be understood as a function of τ . For ξ evaluated at some “good guess”, say τ' , and solving (4.13) for τ yields the basis of a method of moments estimator for τ . Sidik and Jonkman (2005) suggested to use $\tau' = \text{var}(y)$ as a starting value. It might be argued that this arguably very large “guess” is not the most reasonable starting point when performing a meta regression.

Definition 4.6 (Sidik–Jonkman type estimator) Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$, and fixed $\tau \geq 0$ and $\delta > 0$. Let $\hat{\delta}$ be an estimate of δ . Define $\xi(\tau) = \frac{\hat{\delta}}{\tau}$. The method of moments estimator

$$\hat{\tau}_{SJ} = \frac{\|E_{1\xi(\tau')} y\|_{1\xi(\tau')}^2}{k - p} \quad (4.14)$$

is called a *Sidik–Jonkman type estimator* in τ' .

Note that the estimator defined by (4.14) yields positive values as $\|\cdot\|_{1\xi}$ is a positive function. Thus, in particular $P(\hat{\tau}_{SJ} \geq 0) = 1$. If $p = 1$ and $X = 1_k$, the estimate $\hat{\tau}_{DL}$ reduces to the estimate in (Sidik and Jonkman, 2005). As with the other methods of moments estimators defined so far, (4.14) can be iterated by replacing τ' with $\hat{\tau}_{SJ}$ in (4.14).

If the sequence is converging, the limit point $\hat{\tau}$ for which

$$\hat{\tau} = \frac{\|E_{1\xi(\hat{\tau})}y\|_{1\xi(\hat{\tau})}^2}{k-p} \quad (4.15)$$

holds is called the *iterated Sidik–Jonkman type estimator*.

4.3 Interval estimation for Gaussian responses

When parametric assumptions are made about the form of the distribution of Y , then also confidence intervals for the heterogeneity parameter can be constructed. So, let us assume that Y follows Definition 2.15 of a random effects meta regression model, namely: $Y \sim \mathfrak{N}_k(X\beta, [\tau]_k + [\delta])$ for some $\beta \in \mathbb{R}^p$, some full rank design matrix $X \in \mathbb{R}^{k \times p}$ and fixed $\tau \geq 0$ and $\delta > 0$.

Lemma 2.21 shows that the residual vector of the random effects meta regression model follows a chi-squared distribution:

$$\|E_{\delta\tau}Y\|_{\tau\delta}^2 \sim \chi^2_{k-p}. \quad (4.16)$$

Also, recall the function $q_y(\tau) = \|E_{\delta\tau}Y\|_{\tau\delta}^2$ as a function of τ , which has been shown to be strictly decreasing and having the inverse

$$p_y(\eta) = \begin{cases} q_y^{-1}(\eta) & : 0 < \eta < q_y(0), \\ 0 & : \text{otherwise.} \end{cases}$$

Then also p_y is strictly decreasing and $\eta < \eta'$ implies $p_y(\eta) > p_y(\eta')$. Hence,

$$\left\{ \tau : \mathfrak{x}^2_{k-p, \frac{\gamma}{2}} \leq q_y(\tau) \leq \mathfrak{x}^2_{k-p, 1-\frac{\gamma}{2}} \right\} = \left[p_y \left(\mathfrak{x}^2_{k-p, 1-\frac{\gamma}{2}} \right), p_y \left(\mathfrak{x}^2_{k-p, \frac{\gamma}{2}} \right) \right] \quad (4.17)$$

is a genuine $(1-\gamma)$ -confidence interval for τ for $\gamma \in (0, \frac{1}{2})$. Here, $\mathfrak{x}^2_{k-p, \gamma}$ denotes the γ -quantile of a chi-squared distributed random variable. An analogue construction was used by Hartung and Knapp (2005) and Knapp et al. (2006) for constructing interval estimates for τ in the one-way ANOVA and the random effects meta analysis model.

When replacing the unknown δ by within-study variance estimates $\hat{\delta}$, the estimate (4.17) is called the *q-profiling* interval estimate for τ .

Similar to the argument when introducing the Mandel–Paule type estimator for τ , note that the lower bound of (4.17) will be zero when

$$q_y(0) \leq \mathfrak{x}^2_{k-p, 1-\frac{\gamma}{2}}. \quad (4.18)$$

Note that the upper bound of this interval may be 0, too, in case

$$q_y(0) \leq \mathfrak{x}^2_{k-p, \frac{\gamma}{2}}. \quad (4.19)$$

It is, thus, possible for this confidence interval to only consist of a singleton.

4.4 Methods based on maximum likelihood

Maximum likelihood estimators

In Section 2.9, it was shown that solutions $\hat{\tau}$ and $\hat{\beta}$ to the system of equations

$$\beta = B_{\tau\delta}y, \quad (4.20)$$

$$\tau = \frac{\sum_{j=1}^k \frac{1}{(\tau + \delta_j)^2} ((X\beta - y)^2 - \delta_j)}{\text{tr } \Omega_{\tau\delta}^2} \quad (4.21)$$

maximise the likelihood function of the random effects meta regression model. This system of equations can be solved numerically by iterating over τ and β . Either the iteration is started by an initial value β_0 for β or by an initial value τ_0 for τ . Software packages commonly start with an initial value for τ . An example is the implementation by Viechtbauer (2010) which, by default, starts the iteration at the Hedges type estimator.

Algorithm 4.7 Let $\tau = (\tau_l)_{l \in \mathbb{N}}$ and $\beta = (\beta_l)_{l \in \mathbb{N}}$ be sequences in $\mathbb{R}_{\geq 0}$ and \mathbb{R}^k respectively, such that

$$\begin{aligned} \beta_{l+1} &= B_{\tau_l\delta}y, \\ \tau_{l+1} &= \frac{\sum_{j=1}^k \frac{1}{(\tau_l + \delta_j)^2} ((X\beta_{l+1} - y)^2 - \delta_j)}{\text{tr } \Omega_{\tau_l\delta}^2}, \end{aligned}$$

for all $l \in \mathbb{N}$. The sequence $(\beta_l, \tau_l)_{l \in \mathbb{N}}$ is converging, and

$$\hat{\beta} = \lim_l \beta_l, \quad \hat{\tau} = \lim_l \tau_l$$

are a solution to (4.20) and (4.21).

The reason that the above sequence converges is essentially a consequence of the Banach fix point theorem and of the above function being a contraction, i.e., Lipschitz continuous with Lipschitz constant strictly less than 1. If a limit point $(\hat{\beta}, \hat{\tau})$ is found by the above algorithm, it needs to be checked if the algorithm yielded a positive result for $\hat{\tau}$. It is not guaranteed that the global maximum of the likelihood

function lies within $\mathbb{R}^p \times \mathbb{R}_{\geq 0}$. If $\hat{\tau} < 0$, it can be truncated by 0 from below, and it should be checked, whether

$$\det \text{Hess } l(\hat{\beta}, (\hat{\tau})^+) \quad (4.22)$$

is indeed negative on this boundary set. A very important note concerning the implementation of the above algorithm is that iterating Algorithm 4.7 under the constraint of τ being non negative will unlikely yield a correct result. For the algorithm to converge, the marginal sequence of $(\tau_l)_{l \in \mathbb{N}}$ needs to be allowed to run through negative values, too. Otherwise, such an implementation may terminate the algorithm to prematurely.

Viechtbauer (2005) states that convergence is usually achieved within less than ten iterations. In a different implementation by Erez et al. (1996), the authors claim that convergence was achieved in less than seven iterations starting the iteration at $\tau_0 = 0$. A ready-to-use implementation can be found in (Viechtbauer, 2010).

Restricted maximum likelihood estimators

The conditional restricted maximum likelihood (REML) estimate for τ needs to be found by an iterative procedure, too, as it was the case of the maximum likelihood estimator. Formulae for the conditional maximum likelihood estimator are not unambiguously in the literature. As Viechtbauer (2005) points out, different formulae for the REML estimator can be found, some of which may have originated from an approximation to the REML estimator that is due to Morris (1983). This approximation, however, does not hold true in general. Also, as Harbord and Higgins (2008) point out, the iterative algorithm defined in (Morris, 1983) is not guaranteed to converge.

As there exists no closed form expression for this estimator, the evaluation of the restricted maximum likelihood estimator relies on numerical approximations, (Viechtbauer, 2005). The `metagen` package does not re-implement the necessary algorithms for performing this task but falls back to the already tested and optimised implementation of the `metafor` package by Viechtbauer (2010).

4.5 Methods based on generalised inference

In this section, a generalised pivotal quantity will be constructed for the heterogeneity parameter τ . An introduction to the principle of generalised inference was given Chapter 3. The generalised pivotal quantity for the regression coefficient β will be based on the construction of this section. As it will turn out, the generalised pivotal for τ can be based on a generalised pivotal for the heteroscedasticity δ , too. This may allow for incorporating the uncertainty that comes with the estimation

of δ into the interval estimation of τ and β . The functions q_y and p_y will play a prominent role in this section.

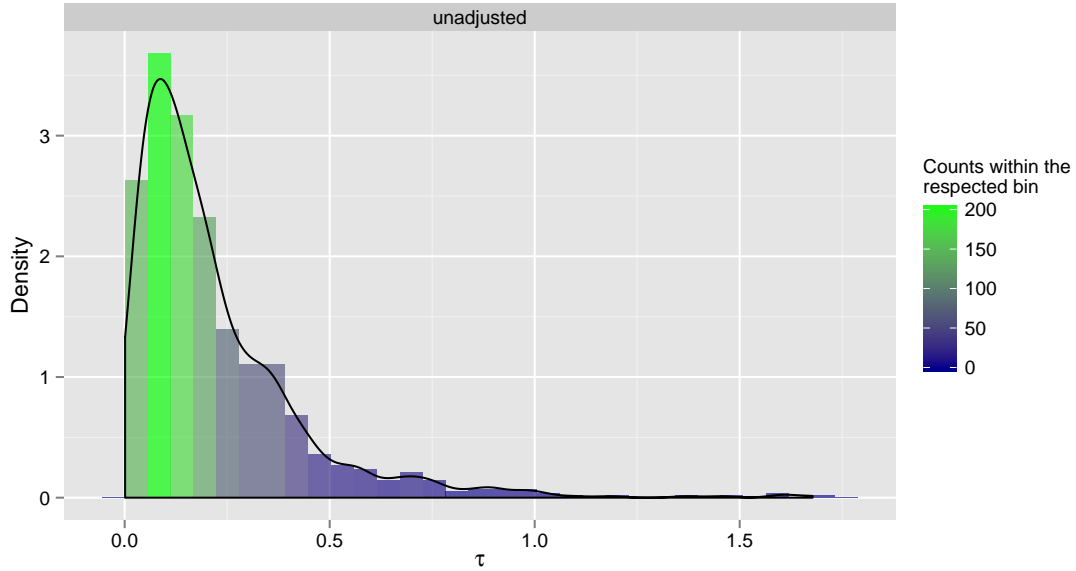


Figure 4.1 Density histogram and kernel density estimate of the density of the generalised pivotal quantity \tilde{T} for the heterogeneity parameter τ in the BCG vaccine efficacy example. The estimates are based on $m = 1000$ independent draws from the pivotal distribution using Algorithm 4.8.

For now, it is assumed that the within study variances are well-established and it is safe to assume that δ is known. The construction is based on (Iyer et al., 2004). Lemma 2.21 has shown that $q_Y(\tau) \sim \mathcal{X}^2_{k-p}$. Define

$$\tilde{T} := p_y(q_Y(\tau)) = p_y(Q) \quad (4.23)$$

for $Q \sim \mathcal{X}^2_{k-p}$. Thus, the distribution of \tilde{T} is free of unknowns. Since q_y is strictly monotone decreasing and $\tau > 0$, the observed value \tilde{t} of \tilde{T} is equal to

$$\tilde{t} = p_y(q_y(\tau)) = q_y^{-1}(q_y(\tau)) = \tau, \quad (4.24)$$

which is, well, monotone in τ . Hence, \tilde{T} is a generalised pivotal quantity for τ . Since the observed value of the pivot \tilde{T} is τ itself, the quantiles of its distribution can directly be used for the construction of confidence intervals. See the example after Definition 3.2. Let \mathfrak{T} denote the distribution of \tilde{T} and let \mathbf{t}_γ denote its γ -quantile. Then

$$\left[\mathbf{t}_{\frac{\gamma}{2}}, \mathbf{t}_{1-\frac{\gamma}{2}} \right] \quad (4.25)$$

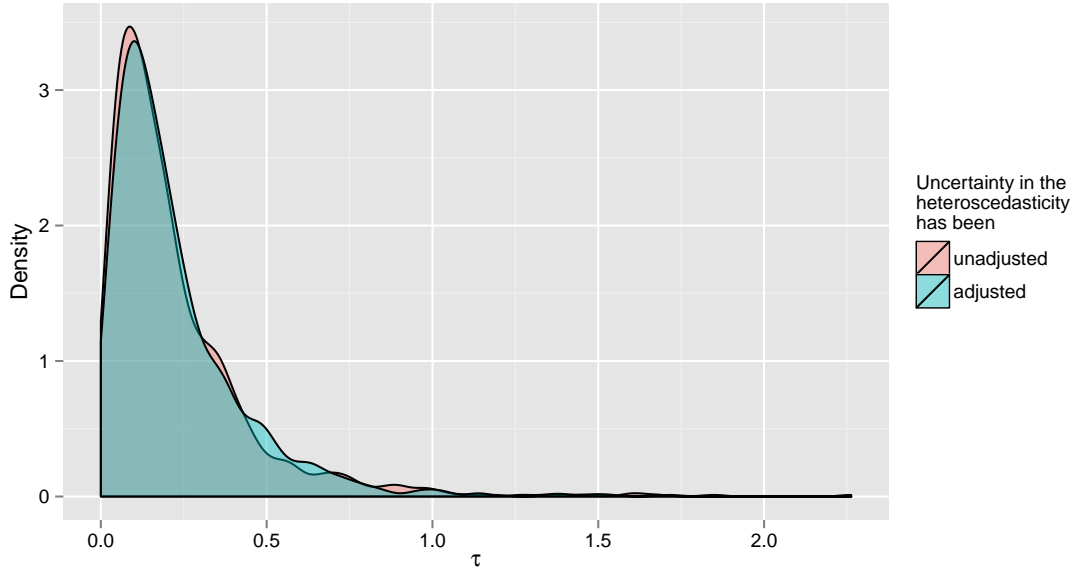


Figure 4.2 Two kernel density estimates of the densities of generalised pivotal quantities for the heterogeneity parameter τ , one based on Algorithm 4.8, the other additionally adjusted for uncertainty in the heteroscedasticity estimate.

is a generalised $(1 - \gamma)$ -confidence interval for τ . The quantiles \mathbf{t}_γ can be estimated numerically by generating random quantities $Q \sim \mathfrak{X}_{k-p}^2$, calculating their inverses with respect to q_y^{-1} , and obtaining the empirical quantiles of these inverses. The procedure is summarised in Algorithm 4.8.

The algorithm involves the evaluation of $p_y(\eta)$ many times for different values of η which may be computationally expensive if done naively. For this reason, some issues related to the implementation of p_y are discussed. Theorem 2.11 has shown how to rewrite q_y in a numerically stable form. Define $q_0 := q_y(0)$. If $\eta \geq q_0$, then $p_y(\eta) = 0$ by the definition of p_y . If $\eta < q_0$, then

$$p_y(\eta) = \tau \Leftrightarrow q_y(\tau) = \eta \Leftrightarrow q_y(\tau) - \eta = 0. \quad (4.26)$$

Hence, obtaining the inverse $q_y^{-1}(\eta)$ is the same as finding the root of the function $f(\tau) := q_y(\tau) - \eta$. This is a smooth function. It is possible to take advantage of this property when deciding on the proper root finding algorithm, for example, when using the well known Newton's method. When converging, Newton's method is converging fast. However, when the starting value for the algorithm lies too far from the true value, the algorithm may not converge at all.

An alternative might be to use algorithms such as Brent's method, (Brent, 2002). This method does not rely on f being continuously differential. It has the advantage of being robust, numerically stable and quite popular. The latter is not to be

neglected as the good implementation of an algorithm is usually half the battle when relying on numerically stable results. When using \mathbf{R} , e.g., it is the default root finding method which is available via the function `uniroot()`.

The obvious disadvantage of Brent's method is that it does not take full advantage of the properties of q_y , e.g., that q_y is smooth. Another is that two initial boundary values ε, ζ need to be given to the algorithm a-priori with the property that f has a unique root within (ε, ζ) . A sufficient condition for this is that $f(\varepsilon)$ and $f(\zeta)$ are of opposite signs. An obvious choice for $f(\tau) = q_y(\tau) - \eta$ is $\varepsilon = 0$. In any worst case scenario, though, ζ might need to be chosen arbitrarily large. Thus, hard-coding the upper bound ζ , however large, is generally a bad idea. In praxis, it is advisable to define ζ as $\zeta = e^n$ for $n = \operatorname{argmin}_n \{n : f(e^n) < 0\}$.

Algorithm 4.8 For $\gamma \in (0, 1)^s$, $s \in \mathbb{N}$, this algorithm returns a sequence of s -tuples $(t_{\gamma, m})_{m \in \mathbb{N}}$ of elements in $\mathbb{R}_{\geq 0}^s$, such that

$$t_{\gamma, m} \xrightarrow{P} \mathbf{t}_\gamma, \quad m \rightarrow \infty. \quad (4.27)$$

- 1) Generate a sequence $(Q_l)_{l \in \mathbb{N}}$ with $Q_l \sim \mathfrak{X}_{k-p}^2$.
- 2) Find a single value decomposition UDV^\top of X^\top where X is the design matrix of the model. Define $K := (v_{p+1}, \dots, v_k)$ where v_1, \dots, v_k denote the columns of the matrix V . Given the observed response vector y , the design matrix X , and the fixed response variances δ define q_y as in (2.28) and Dq_y as in (2.29).
- 3) For $l \in \mathbb{N}$, use a root finding algorithm to find the roots τ_l of each function $q_y - Q_l$. Here, Newton's or Brent's method should come to mind. This yields a sequence $(\tau_l)_{l \in \mathbb{N}}$ in $\mathbb{R}_{\geq 0}$.
- 4) For $r = 1, \dots, s$ and any given $m \in \mathbb{N}$, let $t_{\gamma_r, m}$ denote the empirical γ_r -quantile of $(\tau_l)_{l \leq m}$.
- 5) Define $t_{\gamma, m} = (t_{\gamma_1, m}, \dots, t_{\gamma_s, m})$.

Returning again to the example of the BCG vaccine efficacy data, Figure 4.1 shows density estimates of \mathfrak{X} based on the above algorithm for $m = 1000$. The plot shows a right skewed density similar in shape to a chi-squared distribution. Figure 4.2 shows the same density estimate overlaid by a version of \tilde{T} that adjusts for the uncertainty in the heteroscedasticity estimate δ in the BCG vaccine efficient data. This adjustment is discussed in Section 5.5.

When applying Algorithm 4.8 for $(\frac{\gamma}{2}, 1 - \frac{\gamma}{2}) \in (0, 1)^2$, consistent estimators for $\mathbf{t}_{\frac{\gamma}{2}}$ and $\mathbf{t}_{1 - \frac{\gamma}{2}}$ are obtained that can be used as an approximation for the quantiles

in (4.25). Note that m in the above approximation can be chosen arbitrarily large and is only limited by available computation time.

The median of the above random sequence τ_l , i.e, when the algorithm is applied for $\gamma = \frac{1}{2}$, yields a point estimate for τ . This point estimate will also be discussed in the performance study in Section 4.6. Obviously, other properties of \mathfrak{T} may also be used as point estimates for τ such as its mean or mode.

Excursion 4.9 (Independent draws from the distribution of \tilde{T}) Equation (4.23) defines a generalised pivotal quantity \tilde{T} for the heterogeneity parameter τ . Instructions on how to generate independent draws from its distribution are given in Algorithm 4.8. The algorithm is implemented in the function `pivotalStream` in the `metagen`-package. The following code shows an example on how to call this function.

```
library(metagen)
bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
s <- bcg$size
x <- cbind(1,bcg$x)

set.seed(865287113)
piv <- pivotalStream(1000, y=y, d=d, x=x, adjusted=FALSE)
```

This will in fact produce pivotal quantities for the regression coefficient vector β , too, as it will be discussed in Algorithm 5.3 and Algorithm 5.5. The draws from the distribution of \tilde{T} are contained in the first row of the object `piv`. For convenience, the `metagen` package contains functions for extracting the pivotal draws from `piv` and summarises them in a data frame for further processing.

```
pivh <- joinPivotalHeterogeneity(piv)
```

The plot in Figure 4.1 can be produced by invoking:

```
plotDensityH2(pivh)
```

In Section 5.5, an adjustment to the generalised pivotal quantity (4.23) will be discussed that allows for incorporating the unknown uncertainty of the heteroscedasticity estimate δ into the pivot. This modifies Algorithm 4.8 as it will be discussed later in the text. To generate independent draws from this adjusted distribution, run:

```
set.seed(865287113)
pivAdjusted <- pivotalStream(1000, y=y, d=d, x=x, s=s, adjusted=TRUE)
```

The following code produces plots of density estimates of both versions next to each other.

```
pivh <- joinPivotalHeterogeneity(piv, pivAdjusted)
plotDensityH(pivh)

library(ggplot2)
plotDensityH2(pivh) + facet_wrap(~type)
```

4.6 Performance study of inference methods about the heterogeneity

General set up of the simulation and experimental design

Weerahandi (1995) calls methods based on the generalised inference principle with the somewhat misleading term *exact*. The term is misleading, since these methods are not exact in a strict frequentist sense: The coverage probability of a $(1 - \gamma)$ -confidence interval should, by definition, be $1 - \gamma$. Rather, Weerahandi speaks of exact methods when they are the result of exact *probability statements*. This is similar to the understanding of the term in Bayesian statistics. Thus, even though these methods are *exact* in this sense, the true coverage probabilities of generalised confidence intervals need to be evaluated, for example, by simulation studies. In this section, the results of such a study are summarised which was targeted to estimate the coverage probabilities and average lengths of the presented point and interval estimates for the heterogeneity parameter τ of this section.

In the past sections, different point and interval estimates for the heterogeneity parameter in the random effects meta regression model have been developed. Four point estimates have been discussed that were based on methods of moments, namely the Hedges type estimator, the DerSimonian–Laird type estimator, the Sidik–Jonkman type estimator, and the Mandel–Paule type estimator. Two point estimates based on likelihood methods have been discussed, the maximum likelihood estimator, and the conditional maximum likelihood estimator (REML) for the heterogeneity. A new point estimator based on generalised inference principles was defined as the median of the generalised pivotal distribution \mathfrak{Z} . In this section, the

performance of these point estimators is evaluated by an extensive simulation study. The performance evaluation is based on the following features: the mean squared error, the bias, and the standard deviation of the estimators.

The past sections have also introduced two different constructions for intervals estimators for the heterogeneity parameter. One based on a pivotal quantity for τ , the other based on a generalised pivotal quantity for τ . The performance of these two interval estimates is also compared by the same simulation study. The features for the performance evaluation are their coverage and their average width.

In Chapter 2, different modelling strategies for the random effects meta regression model have been discussed. First the canonical model was introduced in Definition 2.15. Then a non-identifiable extensions to the canonical model was discussed in Section 2.13. For the simulation study, data has been generated using each of these two models, once for $k = 7$ and once for $k = 13$ number of studies. The number of regression parameters was fixed at $p = 2$, allowing the model to have an intercept and a slope. The design matrix in each scenario was set to

$$X = \begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & k \end{pmatrix}. \quad (4.28)$$

The simulation studies' main focus of attention was the comparison and influence of different relations of the heteroscedasticity parameter vector and the heterogeneity parameter of the model on the above performance features. To study this relation, the heterogeneity τ was varied within $[0, 1]$ and heteroscedasticity vector δ within $[0.01, 2]^k$. The experimental design of the computer experiment was defined via a Latin hypercube design of 150 within

$$[0, 1] \times [0.01, 2]^k \quad (4.29)$$

for $k = 7$ and $k = 13$ each. Thus, 150 different parameter configurations for each sample size scenario were considered in the simulation study.

When the mean response extension model was used for data generation, the experimental design for the study size vector s was also defined by a Latin hypercube design with 150 points in

$$[200, 2000]^k. \quad (4.30)$$

These 150 size configurations were randomly allocated to each pair (τ, δ) from the first design. Thus, also for the mean response extension, 150 parameter configurations were used for the simulation study for $k = 7$ and $k = 13$ each. For each parameter configuration (τ, δ) or (τ, δ, s) , the respected performance feature was estimated by drawing 1000 independent data sets from either the canonical model or

its mean response extension. These shall be called the two modelling scenarios of the simulation study.

The respected designs were created by using the `maximinLHS`-function of the R-package `lhs`, (Carnell, 2012). This function aims to maximise the minimal distance between the points of the design and, thus, to spread out the design points as even as possible throughout the parameter space, (Stein, 1987), (Santner et al., 2003). Such a simulation study has not been performed in the literature in the context of estimating performance features of point and interval estimators for a parametric model such as the random effects meta regression model, yet.

When the canonical model was used for data generation, a simulated data set consists of a tuple (y, δ) where $y \in \mathbb{R}^k$ is simulated according to the canonical model and δ equals the true heteroscedasticity. In the canonical scenario, the methods for point and interval estimation are, therefore, fed with the true heteroscedasticity parameter δ that has been used for the data simulation. This scenario corresponds to the underlying model assumption of all presented methods, namely: *δ is known*.

In contrast, when the mean response extension is used for the data generation, a simulated data set consists of a tuple (y, d) where both y and the heteroscedasticity d are simulated. Thus, in the second scenario, the methods only see an estimate for the heterogeneity: namely $\hat{\delta} = d$ where d is a simulated element D following the distribution of (2.67). As the underlying study sizes are known in this second scenario, also the *adjustments* to the generalised methods can be evaluated in this scenario.

Excursion 4.10 (Running a computer experiment) Essentially, a computer experiment was performed for each combination of method and scenario together with an experimental design that defines the different parameter configurations for τ and δ . The `metagen` package provides a function that allows the easy set up of the described simulation studies.

```
library(metagen)
library(BatchExperiments)
```

The following produces the set-up of the described computer experiment for $k = 7$:

```
set.seed(8759179L)
reg07 <- setupExperiment( name="k07regression", seed=5311151L, n=1000L
  , resolution=150L
  , bounds=list(h=c(0,1), d=c(0.01, 2), s=c(200L,2000L))
  , x=cbind(1,1:7), b=c(0.05, -0.5)
  , sgnf=c(0.01, 0.025, 0.05, 0.1)
  , piv_draws=1000L
  , file.dir="k07")
```


And for $k = 13$:

```
set.seed(5731287L)
reg13 <- setupExperiment( name="k13regression", seed=5311151L, n=1000L
  , resolution=150L
  , bounds=list(h=c(0,1), d=c(0.01, 2), s=c(200L,2000L))
  , x=cbind(1,1:13), b=c(0.05, -0.5)
  , sgnf=c(0.01, 0.025, 0.05, 0.1)
  , piv_draws=1000L
  , file.dir="k13")
```

This will generate two registries with $2 \cdot 150 = 300$ jobs each. One job corresponds to a single parameter configuration either for the *canonical model* or for the *mean response extension*.

Together, the two computer experiments consist of 600 such jobs in total. Each job has an approximate run time of around two to two and a half hours on one single CPU node on a high performance cluster.

To submit the generated jobs to the batch system execute:

```
submitJobs(reg07)
submitJobs(reg13)
```

The results of the finished jobs can be extracted by the function `collectAllExperiments` which scans the finished jobs of the registry, collects the results, and summarises them in an accessible data frame for further preprocessing in R.

```
results07 <- collectAllExperiments(reg07)
results13 <- collectAllExperiments(reg13)
```

At the backbone for the efficient batch distribution of the mentioned jobs, `metagen` is using the package `BatchJobs` by (Bischi et al., 2012). The functionality of this package allows to add additional jobs to the above registry based on methods that are not a-priori implemented in the `metagen` package. As new jobs added to the same registry can share experimental designs, here defined via the `metagen` package, this allows other researchers to add their own methods to the simulation study.

Evaluating point estimators for the heterogeneity

Different boxplots of the estimated mean squared errors, the estimated bias, and the estimated standard deviation for different point estimates for τ are presented in Figures 4.3 to 4.5. Each boxplot is based on 150 points corresponding to the different parameter configurations which were discussed above. Thus, each box plot in Figures 4.3 is based on 150 estimated mean squared errors corresponding to different model parameter configurations. In Figures 4.4 and 4.5 respectively, the box plots are based on 150 estimates of the bias and the standard deviations of these different estimators. Each of these point estimates of mean squared error, bias and standard deviation are based on 1000 simulated studies following the respected parameter configuration and the respected modelling scenario. For the canonical simulation scenario, no performance measures for the adjusted generalised methods are plotted in Figures 4.3 to Figure 4.5 as these methods do not apply to these settings.

As Figure 4.3 shows, the maximum likelihood has the lowest mean squared error in comparison with all other point estimators of τ . For $k = 7$, the largest mean squared errors are shown by the new point estimators based on the generalised pivotal quantities. For $k = 13$, the Sidik–Jonkman type estimator shows the largest mean squared error. The high reduction in mean squared error of the generalised estimators is interesting to acknowledge. For $k = 13$, the mean squared error of the generalised point estimators are on a comparable level with the Hedges type estimator. As expected, differences between the adjusted and unadjusted generalised methods are hardly visible. Also as expected, the mean squared error of the all point estimate is reduced when sample size is increased from $k = 7$ to $k = 13$. This reduction is particular visual when the mean response extension is the underlying data generating process.

Of the point estimators based on the method of moments approach, i.e., the Mandel–Paule, Sidik–Jonkman, DerSimonian–Laird, and Hedges type estimators, the DerSimonian–Laird type estimator shows the least mean squared error. The Sidik–Jonkman type estimator shows the largest average mean squared error.

In Figure 4.4, box plots of the estimated bias of the different point estimators for τ are summarised. The plots show that the low mean squared error of the maximum likelihood estimator that one could see in Figure 4.3 was paid for by a very high bias: the maximum likelihood estimator will on average underestimate the true heterogeneity τ of the underlying model. In comparison, the restricted maximum likelihood estimator tends to slightly overestimate the true heterogeneity.

Among the method of moments estimators, the Sidik–Jonkman type estimator shows the largest bias, overestimating the true heterogeneity even more severe as the maximum likelihood estimator is underestimating it. Within the group of method of moment estimators, the best performing method with respect to bias

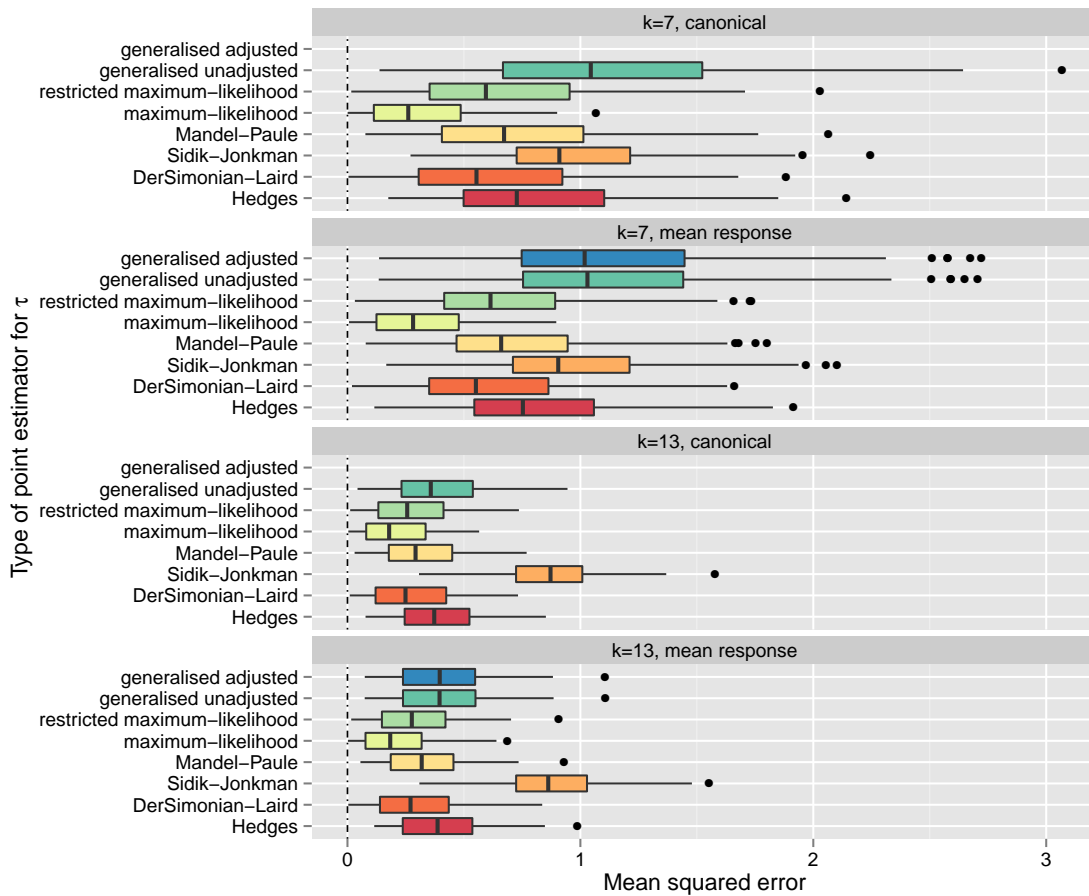


Figure 4.3 Box plots of the estimated mean squared errors of different point estimators for the heterogeneity parameter τ .

is the DerSimonian–Laird type estimator. The new methods based on generalised inference are on average overestimating the true heterogeneity, however, not as severe as the Sidik–Jonkman type estimator. The DerSimonian–Laird type and the restricted maximum likelihood estimator show the least bias of all estimators. The Mandel–Paule and the Hedges type estimator show similar bias.

The standard deviations for the different point estimators are shown in Figure 4.5. Note that this plot does not make any claim on how well an estimator estimates the true value. It rather shows how much the estimator varies for different parameter configurations. A low standard deviation does not necessarily mean a good estimator but it may give a hint on how reproducible an estimate is when an experiment is repeated. This is shown exemplary by the standard deviation of the maximum likelihood estimator. As Figure 4.4 has shown, the maximum likelihood estimator has very high negative bias marking this estimator as not very reliable. The low

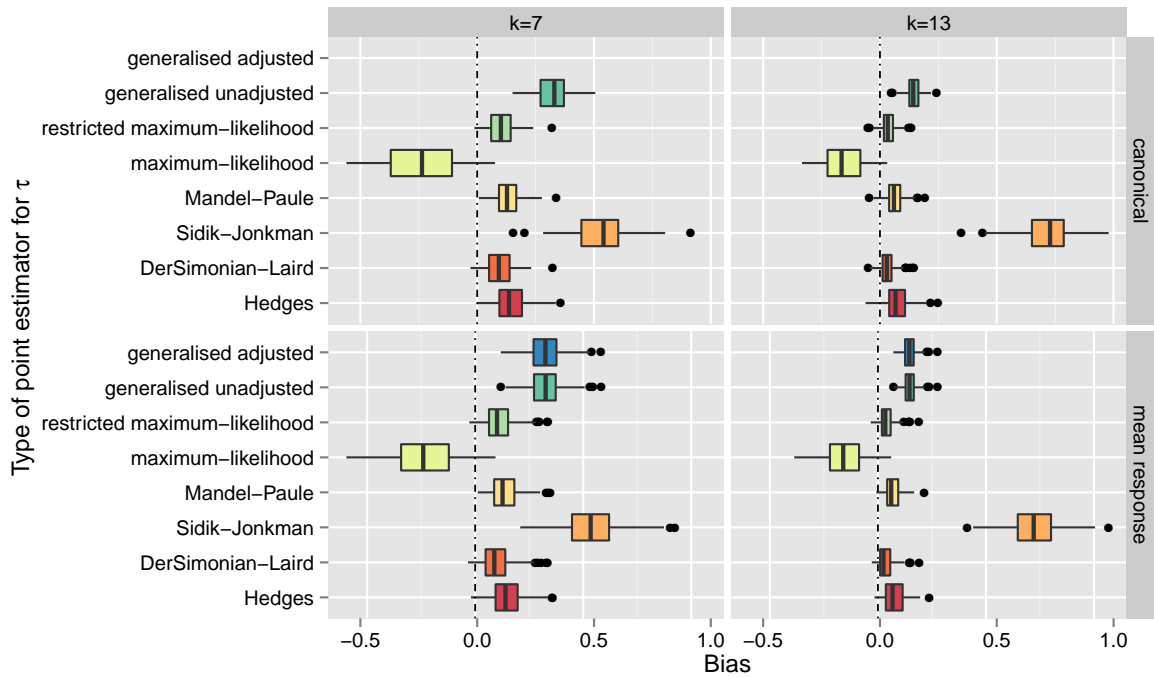


Figure 4.4 Box plots of the estimated bias of different point estimators for the heterogeneity parameter τ .

average mean squared error in Figure 4.3 is a consequence of the low standard deviation shown in Figure 4.5.

The methods based on the generalised inference principle show the highest variability, though not utterly severe. As expected, the estimated standard deviations of all estimators are smaller, when the number of available studies increases. The biggest decrease in standard deviation are shown by the generalised methods.

Figures 4.6 to 4.14 show scatter plots of all discussed point estimators with respect to the true underlying heterogeneity τ used in the respected parameter configuration. Figures 4.6 to 4.8 show scatter plots between the true τ and the estimated mean squared error for different heteroscedasticity parameters δ . As one may expect, the mean squared error of all estimators increases when the heterogeneity between the study increases. The plots, however, show that this increase is happening with different magnitudes for the different point estimators. This effect is less severe if the number of studies in the analysis gets larger. A linear model has been fitted to the data in each plot, and the estimated regression line has been overlaid.

In Figure 4.6, it is interesting to see that the Sidik–Jonkman type estimator for the heterogeneity starts off (small values of τ) with having the largest mean squared error among the method of moments estimators. For $k = 7$ though, the

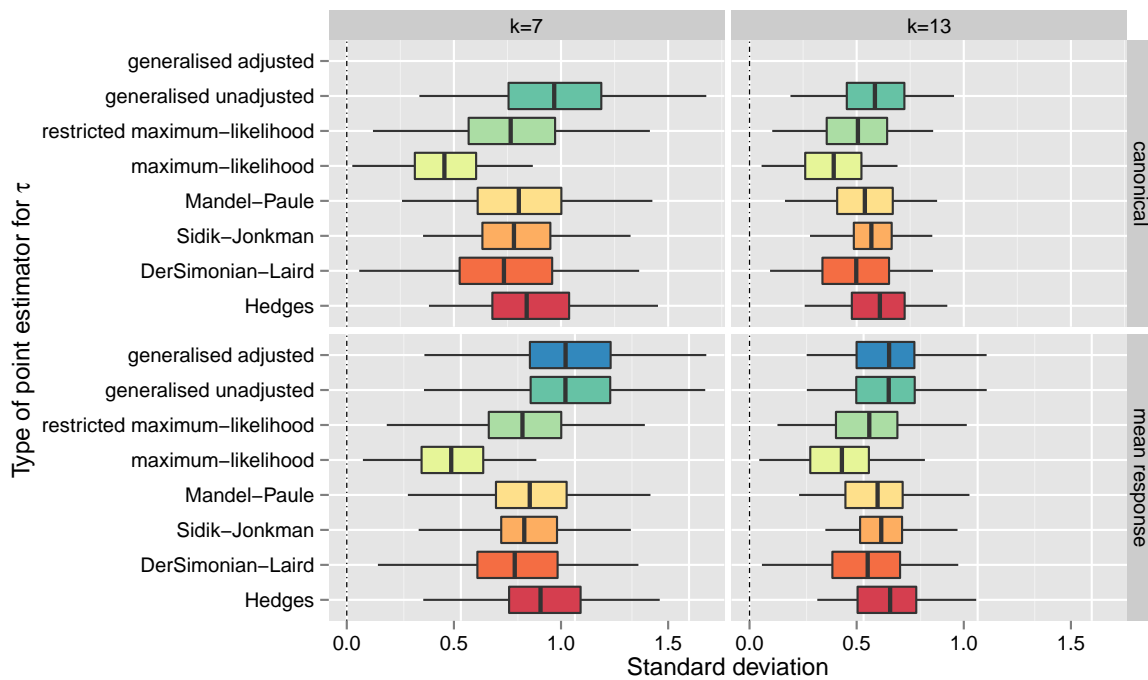


Figure 4.5 Box plots of the estimated standard deviation of different point estimators for the heterogeneity parameter τ .

mean squared error of this estimator falls below the Hedges type estimator with increasing τ .

The scatter plots in Figures 4.9 to 4.11 show the true heterogeneity and the estimated bias for the different heterogeneity vectors in the simulation design. Interesting in these figures is the large positive bias of the Sidik–Jonkman type estimator for τ in Figure 4.9 and the large negative bias of the maximum likelihood estimator in Figure 4.10. The large magnitudes of these biases will have an impact on the performance of the interval estimates for the regression coefficients which are based on these estimates for τ yielding quite surprising results for the Sidik–Jonkman type estimator.

The scatter plots showing τ and the estimated standard deviations of the estimators, see Figures 4.12 to 4.14, reflect the picture drawn by the previous plots.

It is advisable to choose a point estimator for τ which shows low bias and a low average mean squared error. Unfortunately, the new methods based on generalised inference principle were not able to shine in the performance study.

When bias is an issue and the expected heterogeneity of the studies is high, Figure 4.9 and Figure 4.10 suggest to either use the DerSimonian–Laird type estimator or the restricted maximum likelihood estimator for the point estimation.

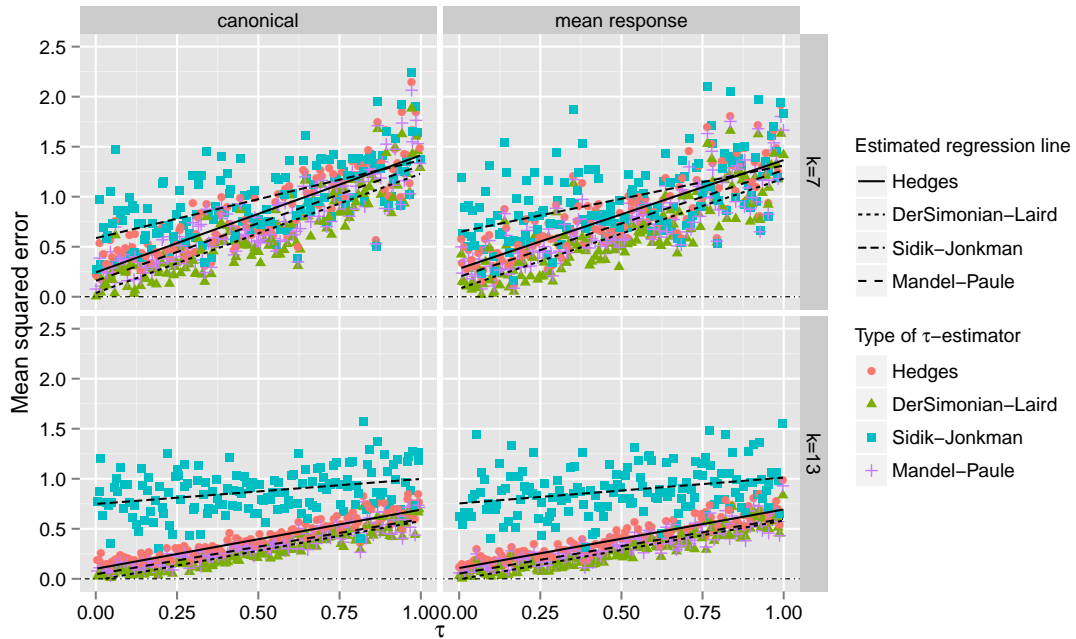


Figure 4.6 Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ which are based on method of moments.

The restricted maximum likelihood method is slightly outperforming the DerSimonian–Laird estimator. Note that the restricted maximum likelihood estimator relies on non-trivial numerical optimisation methods that may not be implemented in all software packages. On the other hand, the DerSimonian–Laird type estimator can be calculated on every standard office calculator if necessary.

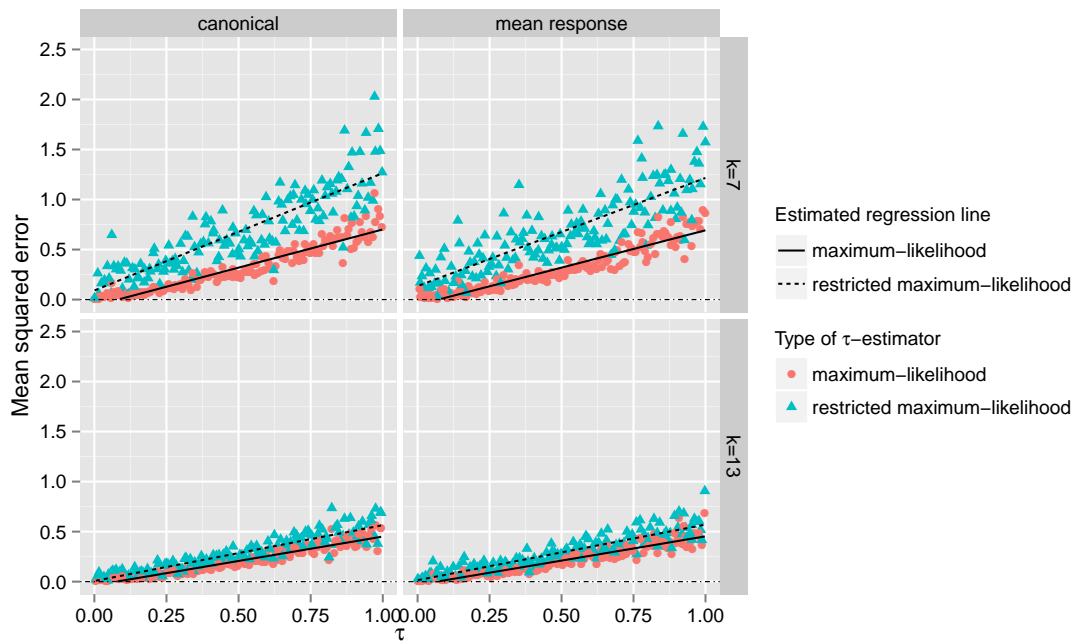


Figure 4.7 Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ which are based on likelihood methods.

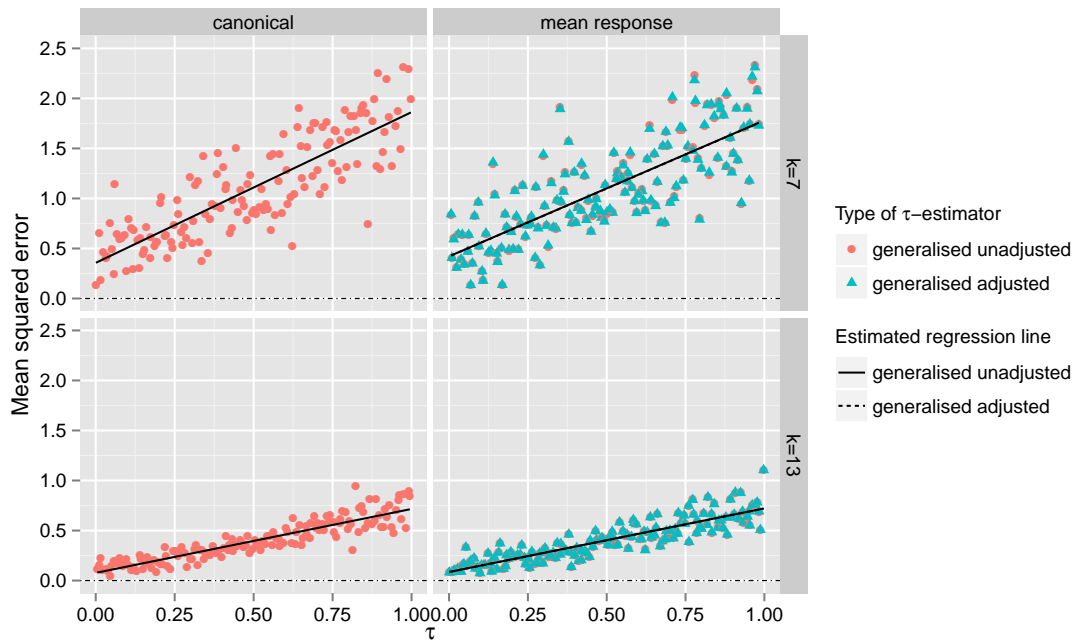


Figure 4.8 Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ based on generalised inference principles.

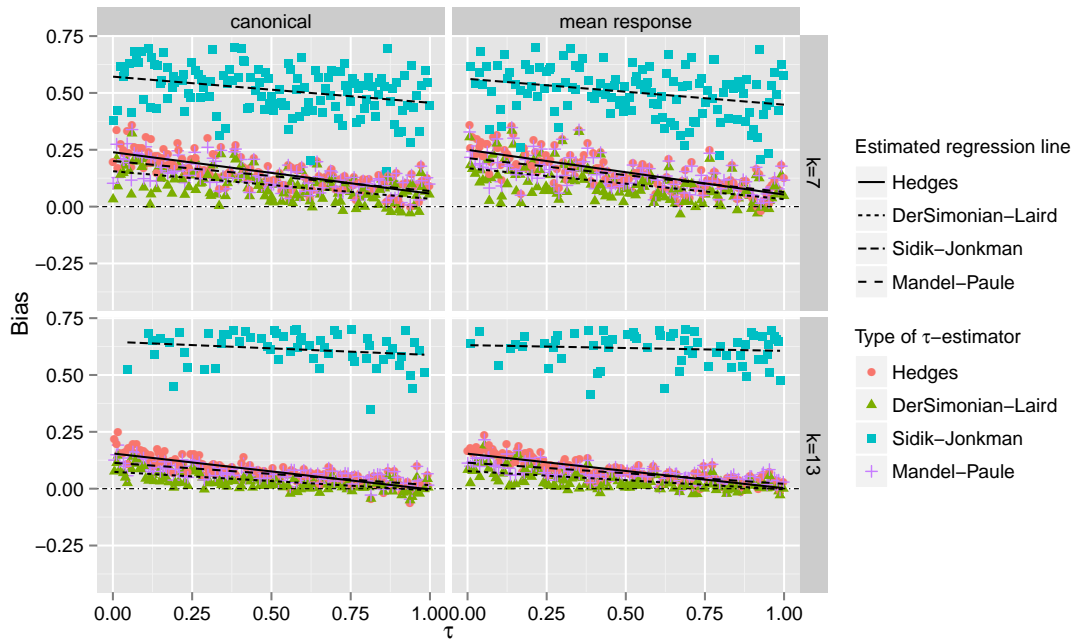


Figure 4.9 Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on method of moments.

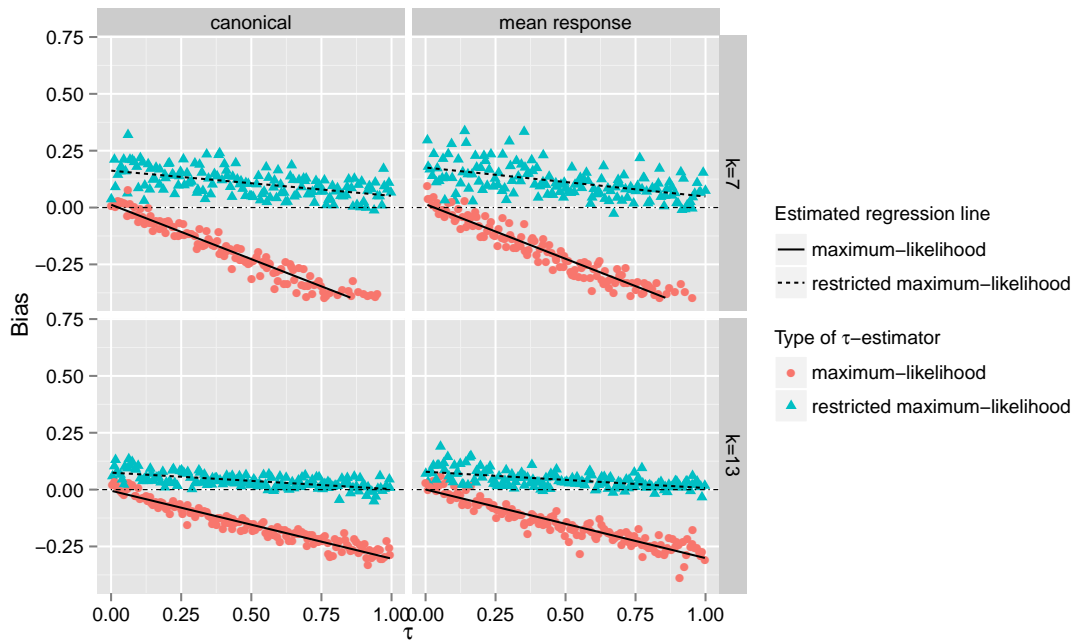


Figure 4.10 Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on likelihood methods.

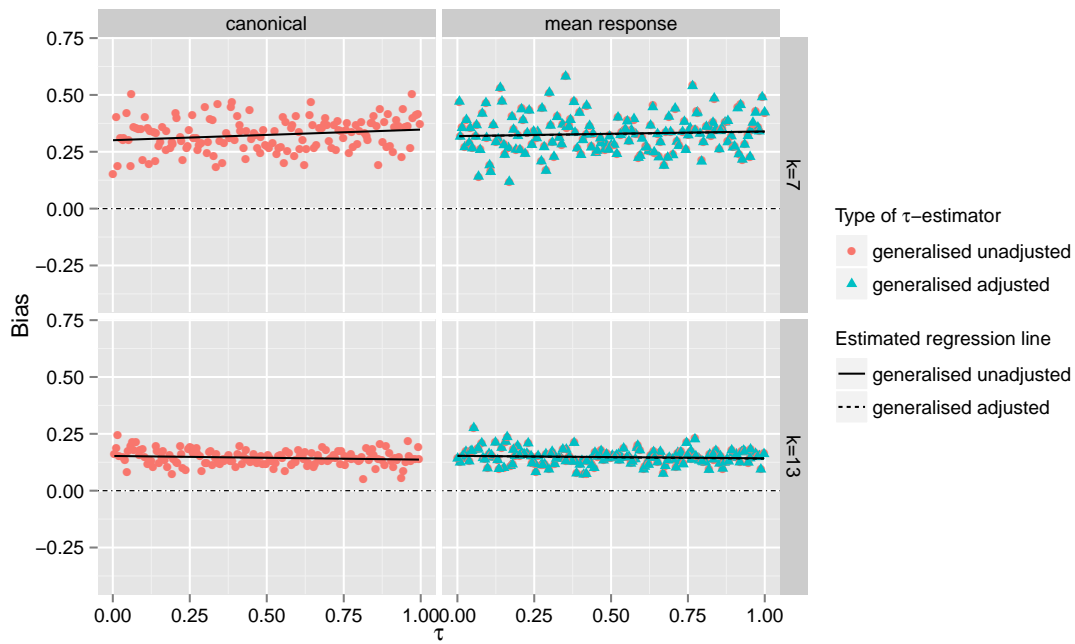


Figure 4.11 Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on generalised inference principles.

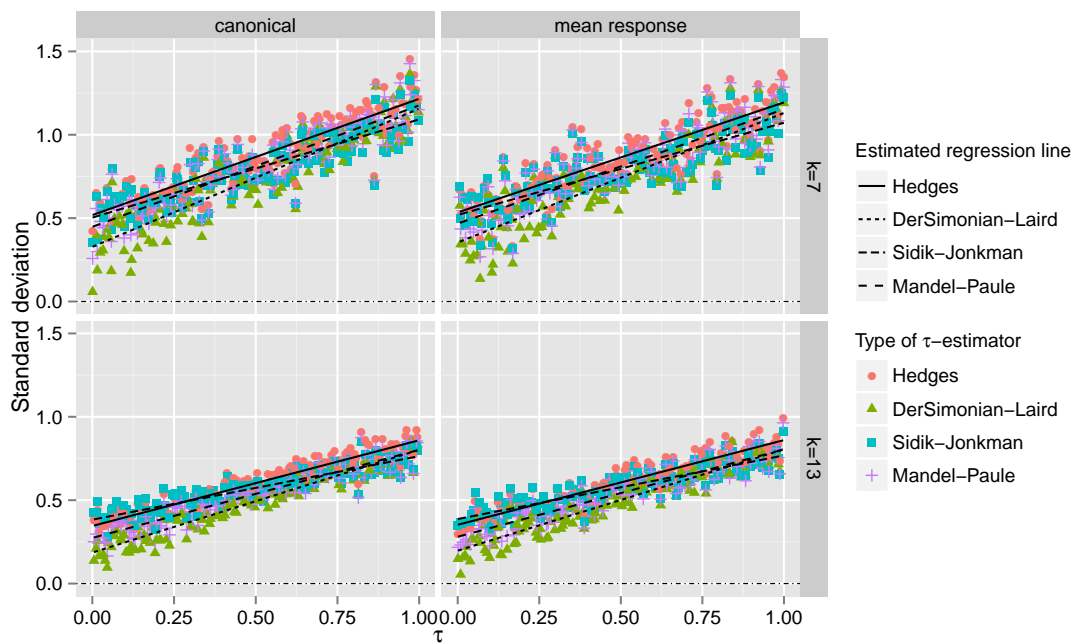


Figure 4.12 Scatter plots of the estimated standard deviations of point estimators for the heterogeneity parameter τ which are based on method of moments.

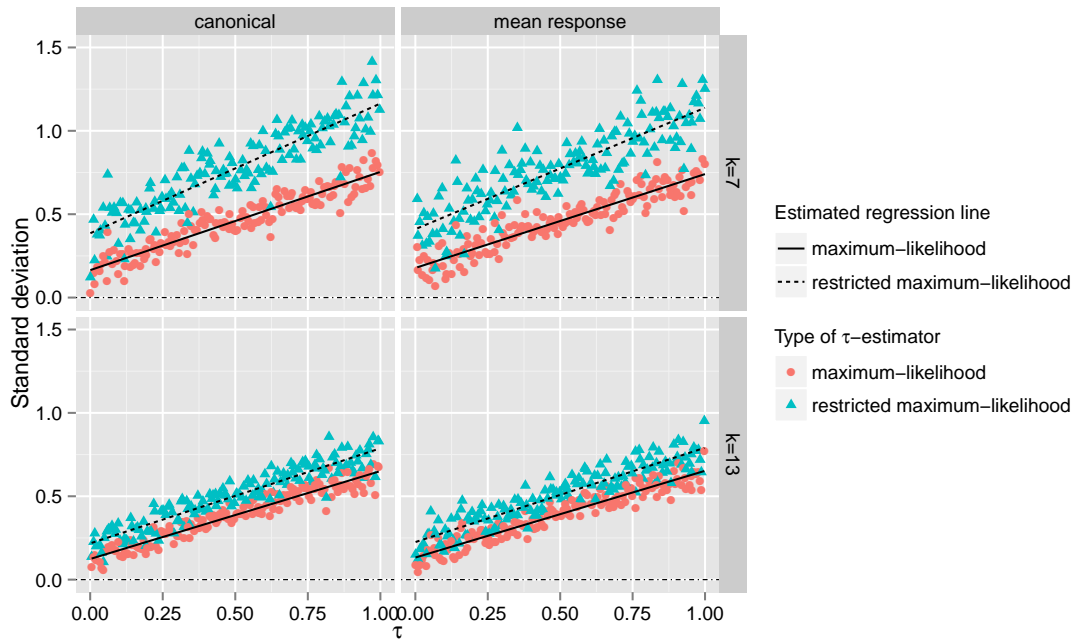


Figure 4.13 Scatter plots of the estimated standard deviation of point estimators for the heterogeneity parameter τ which are based on likelihood methods.

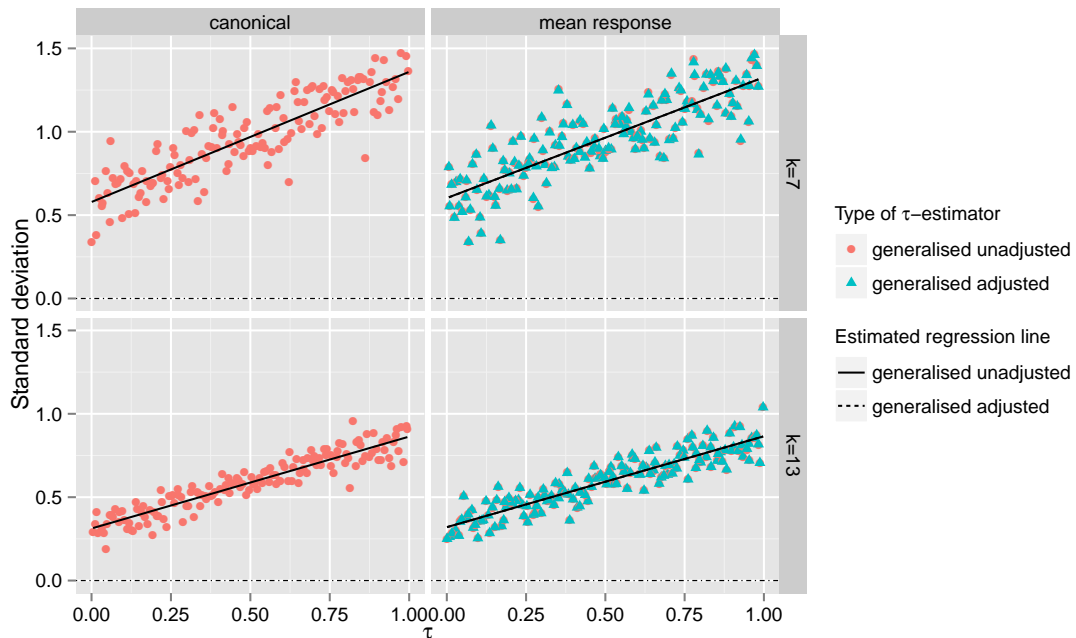


Figure 4.14 Scatter plots of the estimated standard deviations of point estimators for the heterogeneity parameter τ which are based on generalised inference principles.

Evaluating interval estimators for the heterogeneity

Two different interval estimates were discussed in the text. The first is the q-profiling interval estimator (4.17), the other one was based on generalised inference principles, namely (4.25). The performances of both interval estimates were evaluated with respect to their coverage and mean width.

For $i = 1, \dots, p$ and $\gamma \in (0, \frac{1}{2})$, the *coverage* of an interval estimator $I^i(\gamma)$ for β_i is the expectation of $I^i(\gamma)$ to contain β_i , in signs

$$\mathbb{E}(\beta_i \in I^i(\gamma)). \quad (4.31)$$

For a well behaved interval estimate $I^i(\gamma)$, the above expectation should yield a value close to $1 - \gamma$. If

$$\mathbb{E}(\beta_i \in I^i(\gamma)) > 1 - \gamma, \quad (4.32)$$

then I^i is *overconfident*. Clinical settings usually aim to reduce the probability of false positive outcomes. For this reason, if the performance of

$$\mathbb{E}(\beta_i \in I^i(\gamma)) - (1 - \gamma) \quad (4.33)$$

results in values below zero, they are regarded as less severe than the other way around.

The coverage of the discussed interval estimate was evaluated for

$$\gamma = 0.01, 0.025, 0.05, 0.1 \quad (4.34)$$

in each of the parameter configurations, in each of the modelling, and in each of the sample size scenario of the previous section. Figure 4.15 shows box plots of the coverage distributions of the interval estimators. As mentioned, it is desirable for a method to be close to the zero line but preferable to be below rather than above this line as this corresponds to a method being more conservative rather than being too liberal in comparison to its aspired confidence level. The figure also includes box plots of an adjustment of the generalised method that will be discussed in Section 5.5.

The plots show that the q-profiling method performs on average better than the methods based on generalised principles, though not severely. As expected, the adjusted and unadjusted generalised methods perform visually undistinguishable in the mean response scenario.

Figure 4.16 shows a scatter plot of the true τ and the coverage of the interval estimators. No big variations can be seen with respect to varying heterogeneities. Figure 4.17 shows that the average width of the generalised interval estimate for τ tends to be slightly shorter than the q-profiling method, though only minimal.

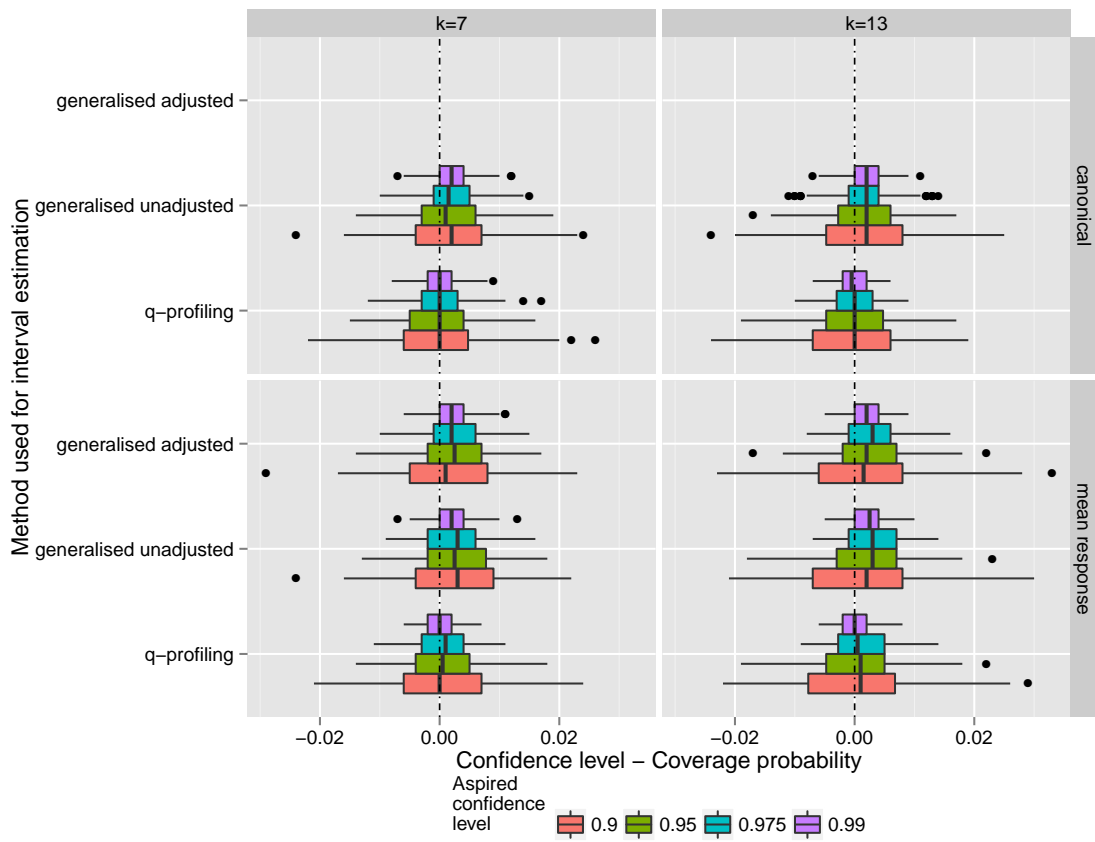


Figure 4.15 Box plots of the estimated coverage of the interval estimators for the heterogeneity parameter τ .

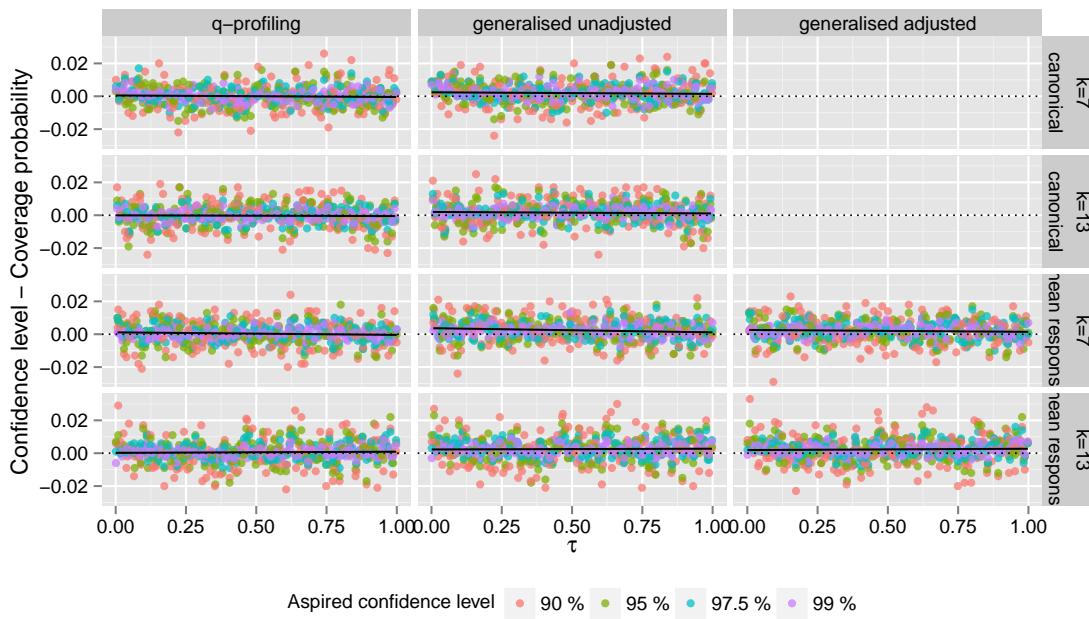


Figure 4.16 Scatter plots of the estimated coverage of the interval estimators for the heterogeneity parameter τ .

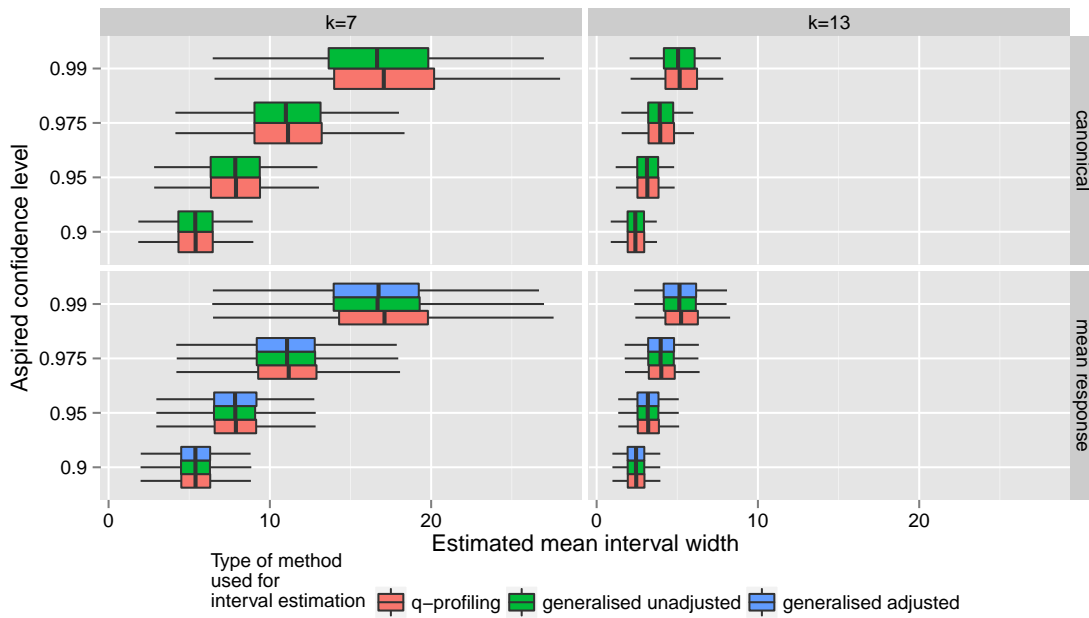


Figure 4.17 Box plots of the estimated mean width of the discussed interval estimators for the heterogeneity parameter τ .

4.7 *Example:* Heterogeneity of clinical trials studying BCG vaccine efficacy

The above methods for point and interval estimation of the heterogeneity parameter τ in the random effects meta regression model are now applied to the BCG vaccine efficacy data discussed in Section 1.2.

Point estimates for the	heterogeneity
Hedges	0.2090
DerSimonian–Laird	0.0633
Sidik–Jonkman	0.2318
Mandel–Paule	0.1421
maximum likelihood	0.0344
restricted maximum likelihood	0.0764
generalised unadjusted	0.1579
generalised adjusted	0.1678

Table 4.1 Point estimates of the heterogeneity parameters in the BCG vaccine efficacy example.

Method	Confidence	Lower	Upper
q-profiling	0.900	0.02688	0.615
generalised unadjusted	0.900	0.02471	0.624
generalised adjusted	0.900	0.02732	0.612
q-profiling	0.950	0.01667	0.785
generalised unadjusted	0.950	0.01612	0.834
generalised adjusted	0.950	0.01575	0.737
q-profiling	0.975	0.01012	0.976
generalised unadjusted	0.975	0.00756	0.991
generalised adjusted	0.975	0.00952	0.999
q-profiling	0.990	0.00465	1.268
generalised unadjusted	0.990	0.00333	1.179
generalised adjusted	0.990	0.00454	1.358

Table 4.2 Interval estimates of the heterogeneity parameter in the BCG vaccine efficacy example.

Applying the methods for point estimation to the data yield the heterogeneity estimates in Table 4.1. The estimates have a rather large dynamical range from 0.03 (maximum likelihood) up to 0.23 (Sidik–Jonkman). The performance study has shown that the Sidik–Jonkman type estimator tends to overestimate the true heterogeneity, so it was to be expected that this estimator would yield a large estimate of the heterogeneity. The performance study suggests to trust the DerSimonian–Laird type estimator to be close to the truth.

The bounds of the different interval estimates for the heterogeneity τ are summarised in Table 4.2. The interval estimates based on the q-profiling method yields slightly sharper confidence intervals than the generalised constructions in this example. The unadjusted generalised method results in slightly larger estimates than the adjusted generalised method.

The following excursion shows some code examples on how to reproduce the above results.

Excursion 4.11 (Between-study variance of BCG vaccine efficacy trials)

Using the `metagen` package, the analysis of the BCG vaccine efficacy data can be performed by the following code:

```
library(metagen)
sgnf <- c(0.01, 0.025, 0.05, 0.1)

bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
s <- bcg$size
x <- cbind(1, bcg$x)

set.seed(531553313)
estimation <- metagen( y=y, d=d, x=x
                      , sgnf=sgnf, s=s
                      , n=1000, adjusted=TRUE)
```

The `estimation` object is a list that contains point and interval estimates of the heterogeneity and the regression coefficient of the regression model. The point and interval estimates of the heterogeneity are found in

```
estimation$pointh
estimation$confh
```

respectively. The list of confidence intervals can easily be ordered by the aspired confidence level via

```
with(estimation, with(confh, confh[order(confidence),]))
```

A graphical representation of the interval estimates can be produced by

```
plotHeterogeneityInterval(estimation$confh)
```


5 Inference on the regression coefficients

5.1 Introduction

Having developed different point and interval estimates for the variance components in the random effects meta regression model in Chapter 4, the focus will now shift to the location parameters of the model. Different constructions of point and interval estimates for the regression parameters β_i will be constructed, including the current state-of-the-art interval estimate developed by Knapp and Hartung (2003). The key construction of this chapter will be the development of a new, genuine interval estimate for the β_i 's based on generalised inference principles. As before, the construction will also yield new, genuine points estimate for these parameters.

5.2 Theoretical considerations

Let Y follow the distribution of a random effects meta regression model with design matrix X , heterogeneity τ , and heteroscedasticity δ . Conditional on an observed data y of Y , the aim is to construct point and interval estimates for the regression coefficients $\beta = (\beta_1, \dots, \beta_p)^\top$. A common approach is to minimise the distance between the expected responses $\mathbb{E}(Y) = X\beta$ and the observed response vector y with respect to β in some fixed Hilbert space. This approach is commonly referred to as the *method of minimal least squares* or *method of weighted minimal least squares*. Here, the choice of the Hilbert space is equivalent to the choice of an inner product for \mathbb{R}^k which has to be defined a-priori. Fortunately, there are some canonical options for how to choose a suitable inner product or norm. Choices are the canonical inner product on \mathbb{R}^k , which would correspond to the Euclidean norm, or the inner product $\langle \cdot, \cdot \rangle_{\tau\delta}$ on \mathbb{R}^k , which has been introduced earlier.

The variance components τ and δ are obviously not known prior to the analysis of the model. Loosely speaking, when applying inference methods to the observed y , somewhat appropriate values for either τ , δ or both need to be chosen. In this chapter, whenever the variables τ and δ are used in formulae, their values shall correspond to the true variance components of the model. In contrast, $\hat{\tau}$ and $\hat{\delta}$ shall correspond to any choice of τ and δ in the respected parameter spaces $\mathbb{R}_{\geq 0}$ and $\mathbb{R}_{> 0}^k$ and not necessarily being equal to the truth. The notation shall obviously suggest that $\hat{\tau}$ and $\hat{\delta}$ shall later be treated as consistent estimators for τ and δ respectively. For now, though, $\hat{\tau}$ and $\hat{\delta}$ can be arbitrarily chosen and they are treated as fixed.

Different linear operators on the sample space have been defined on page 11. Let $\hat{\Omega} = \Omega_{\hat{\tau}\hat{\delta}}$, $\hat{V} = V_{\hat{\tau}\hat{\delta}}$, $\hat{B} = B_{\hat{\tau}\hat{\delta}}$, $\hat{H} = H_{\hat{\tau}\hat{\delta}}$, and $\hat{E} = E_{\hat{\tau}\hat{\delta}}$. For any arbitrarily

but fixed choice of $\hat{\tau}$ and $\hat{\delta}$, there exists a minimal distance estimator with respect to $\langle \cdot, \cdot \rangle_{\hat{\tau}\hat{\delta}}$, or, consequently, with respect to $\|\cdot\|_{\hat{\tau}\hat{\delta}}$.

Theorem 5.1 Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some fixed τ and δ . Then for any $\hat{\tau}, \hat{\delta}$ the minimisation problem

$$\operatorname{argmin}_{\beta} \|y - X\beta\|_{\hat{\tau}\hat{\delta}} \quad (5.1)$$

has the unique solution $\hat{B}Y$ with expectation and variance-covariance structure

$$\hat{B}Y \sim \left(\beta, \hat{B}\Omega_{\tau\delta}^{-1}\hat{B}^{\top} \right). \quad (5.2)$$

In particular, $\hat{B}Y$ is an unbiased linear estimator of β .

Proof. It needs to be proven that $\operatorname{argmin}_{\beta} \|y - X\beta\|_{\tau\delta}$ is a singleton. For this it is sufficient to show that the smooth $\beta \mapsto \|y - X\beta\|_{\tau\delta}$ has a unique minimum. This is standard calculus,

$$\begin{aligned} \|y - X\beta\|_{\hat{\tau}\hat{\delta}} &= (y - X\beta)^{\top} \Omega_{\hat{\tau}\hat{\delta}} (y - X\beta) \\ &= y^{\top} \Omega_{\hat{\tau}\hat{\delta}} y - 2y^{\top} \Omega_{\hat{\tau}\hat{\delta}} X\beta + \beta^{\top} X^{\top} \Omega_{\hat{\tau}\hat{\delta}} X\beta, \\ D_{\beta} \|y - X\beta\|_{\hat{\tau}\hat{\delta}} &= -2X^{\top} \Omega_{\hat{\tau}\hat{\delta}} y + 2X^{\top} \Omega_{\hat{\tau}\hat{\delta}} X\beta, \\ D_{\beta} \|y - X\beta\|_{\hat{\tau}\hat{\delta}} = 0 &\Leftrightarrow \beta = (X^{\top} \Omega_{\hat{\tau}\hat{\delta}} X)^{-1} X^{\top} \Omega_{\hat{\tau}\hat{\delta}} y = B_{\hat{\tau}\hat{\delta}} y. \end{aligned}$$

The expectation of $\hat{B}Y$ is $\mathbb{E}(B_{\hat{\tau}\hat{\delta}}Y) = B_{\hat{\tau}\hat{\delta}}\mathbb{E}(Y) = B_{\hat{\tau}\hat{\delta}}X\beta = \beta$ and its variance-covariance structure is $\mathbb{V}(B_{\hat{\tau}\hat{\delta}}Y) = B_{\hat{\tau}\hat{\delta}}\mathbb{V}(Y)B_{\hat{\tau}\hat{\delta}}^{\top} = B_{\hat{\tau}\hat{\delta}}\Omega_{\tau\delta}^{-1}B_{\hat{\tau}\hat{\delta}}^{\top}$. \square

This is quite an important result as it shows that any $\hat{B}Y$ is in particular consistent for any choice of $\hat{\tau}$ and $\hat{\delta}$. The essential message of the above corollary is that a wrong choice of $\hat{\tau}$ and $\hat{\delta}$ might be levelled out by sufficient amount of data. Unfortunately, the amount of data in a meta analysis is typically rather limited. This makes the choices of good estimates for τ and δ important again but also to use statistical methods for inference on β with a sufficient high rate of convergence.

In the special case in which $\hat{\tau} = 0$ and $\hat{\delta} = 1_k$, the above theorem defines a rather commonly chosen estimator for β , namely the *minimal least squares estimator*

$$B_{01}Y = (X^{\top}X)^{-1}X^{\top}Y. \quad (5.3)$$

In case there is no heteroscedasticity present in the data, only a single variance component remains in the model. Without loss of generality, $\delta > 0$ can assumed to

be constant and $\tau = 0$. The reason that $\delta = 1_k$ can deliberately be chosen as *one* is essentially a consequence of $(\delta \in \mathbb{R}_{>0})$

$$\begin{aligned}\mathbb{V}_{0\delta_k}(B_{01}Y) &= \delta \cdot V_{01}, \\ \mathbb{E}_{0\delta_k}(\|E_{01}Y\|) &= \delta \cdot \|E_{01}Y\|_{01},\end{aligned}$$

which holds since $\mathbb{V}_{0\delta_k}(B_{01}Y) = B_{01}\Omega_{0\delta_k}^{-1}B_{01}^\top = \delta B_{01}B_{01}^\top = \delta(X^\top X)^{-1} = \delta V_{01}$. Hence, the variance-covariance matrix and the norm of the residuals are essentially multiples of the heterogeneity/heteroscedasticity parameter δ , which, in turn, has no influence on the minimisation problem in Theorem 5.1.

The other interesting special case occurs when the true values of τ and δ are known. This case is interesting as, when τ and δ are replaced by consistent estimators, the delta method produces many fruitful estimates for the β_i since

$$(\tau, \delta) \mapsto B_{\tau\delta}Y, \quad (5.4)$$

as a function of τ and δ , is jointly continuous and, in fact, smooth in $(\tau, \delta_1, \dots, \delta_k)$. Thus, consistency of $\hat{\tau}$ and $\hat{\delta}$ yields consistency in $\hat{B}Y$.

The bias of a multivariate estimator, say $\hat{\beta}$ of β , is defined as $\mathbb{B}_{\beta\tau\delta}(\hat{\beta}) := \mathbb{E}_{\beta\tau\delta}(\hat{\beta}) - \beta$. The mean squared error of $\hat{\beta}$ is defined as

$$\mathbb{E}_{\beta\tau\delta}(\|\hat{\beta} - \beta\|^2) = \text{tr } \mathbb{V}_{\tau\delta}(\hat{\beta}) + \|\mathbb{B}_{\beta\tau\delta}(\hat{\beta})\|^2. \quad (5.5)$$

In particular, the mean squared error of an unbiased estimator is equal to the trace of its variance-covariance matrix. The following corollary summarises some basic properties of the estimator $\hat{B}Y$ such as bias, consistency, and its mean squared error.

Corollary 5.2 (Mean squared error) Let $Y \sim (X\beta, [\tau]_k + [\delta])$ for some fixed τ and δ . Let $\hat{\tau} \geq 0$ and $\hat{\delta} > 0$. Then $\hat{B}Y$ is unbiased and its mean squared error is equal to

$$\mathbb{E}_{\beta\tau\delta}(\|\hat{B}Y - \beta\|^2) = \text{tr } \hat{B}^\top \hat{B} \Omega_{\tau\delta}^{-1}. \quad (5.6)$$

In particular, when $\hat{\tau} = \tau$ and $\hat{\delta} = \delta$ are equal to the truth, then

$$\mathbb{E}_{\beta\tau\delta}(\|B_{\tau\delta}Y - \beta\|^2) = \text{tr } V_{\tau\delta}. \quad (5.7)$$

Proof. First note that $\mathbb{E}_{\tau\delta} \left(\|\hat{B}Y - \beta\|^2 \right) = \text{tr} \mathbb{V}_{\tau\delta} (\hat{B}Y) = \text{tr} \hat{B}^\top \hat{B} \Omega_{\tau\delta}^{-1}$. Then in particular, $\mathbb{V}(BY) = B\Omega^{-1}B^\top = (X^\top \Omega X)^{-1} X^\top \Omega \Omega^{-1} \Omega X (X^\top \Omega X)^{-1} = (X^\top \Omega X)^{-1}$. \square

From now on, $\hat{\tau}$ and $\hat{\delta}$ shall be treated as consistent estimators for τ and δ respectively.

5.3 Point estimation of the regression coefficients

We have just seen that $\mathbb{E}_{\beta\tau\delta}(\hat{B}Y) = \beta$. Thus, we could say that the (weighted) minimal least squares estimator $\hat{B}Y$ is a method of moments estimator for the regression coefficient vector β . When τ and δ were known, BY is the maximum likelihood estimator of β . If τ and δ are replaced by consistent estimators, say $\hat{\tau}$ and $\hat{\delta}$, the estimator $B_{\hat{\tau}\hat{\delta}}Y$ and its estimate $B_{\hat{\tau}\hat{\delta}}y$ are usually *both* denoted by $\hat{\beta}$. This convention shall be adapted here as well. The ambiguous notation usually proves to be unproblematic, since it is mostly clear from the context if the estimator or the estimate is meant in any given formula.

Formulae for generalised test statistics and generalised pivotal quantities, however, can contain both quantities: the observed data y and its random counterpart Y . If no ambiguity can occur, then $\hat{\beta}$ will be used to denote either estimator or estimate depending on the context. This will make formulae easier on the eye. If a formula contains both, observed data and the underlying random element, the non-ambiguous BY - and By -notation will be used instead.

If $\hat{\tau}$ and $\hat{\delta}$ are consistent estimators for τ and δ , then $\hat{B}y$ is a consistent estimator for β by the delta method.

5.4 Interval estimation for Gaussian responses

Two-step maximum likelihood approach

Let us now assume that Y follows a multivariate Gaussian distributed random effects meta regression model

$$Y \sim \mathfrak{N}_k(X\beta, [\tau]_k + [\delta]), \quad (5.8)$$

as it was defined in Definition 2.15 with design matrix $X \in \mathbb{R}^{k \times p}$, heterogeneity τ , and heteroscedasticity δ . A classic interval construction for β is based on the distribution of $\hat{B}Y$. Since B is linear, BY is multivariate Gaussian distributed, and it follows from Theorem 5.1 that

$$BY \sim \mathfrak{N}_p(\beta, V). \quad (5.9)$$

Let $(v_1, \dots, v_p) := \text{diag } V$. Since $\hat{\beta}_i \rightsquigarrow \mathfrak{N}(\beta_i, v_i)$, when replacing τ and δ by consistent estimators $\hat{\tau}$ and $\hat{\delta}$, this yields

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{v}_i}} \rightsquigarrow \mathfrak{N}(0, 1). \quad (5.10)$$

Here, \rightsquigarrow denotes convergence in distribution, and \hat{v}_i denotes the i th element of the tuple $(\hat{v}_1, \dots, \hat{v}_p) = \text{diag } \hat{V}$. For any $\gamma \in (0, \frac{1}{2})$, this yields the approximate $(1 - \gamma)$ -confidence interval for β_i

$$\left[\hat{\beta}_i - \mathfrak{z}_{1-\frac{\gamma}{2}} \cdot \sqrt{\hat{v}_i}, \hat{\beta}_i + \mathfrak{z}_{1-\frac{\gamma}{2}} \cdot \sqrt{\hat{v}_i} \right]. \quad (5.11)$$

Here, \mathfrak{z}_γ denotes the γ -quantile of the standard Gaussian distribution. The confidence interval in (5.11) is probably the most straightforward method for making inference on the regression coefficients in the random effects meta regression model. It is implemented in many statistical software packages as the default methods for parameter estimation, see for example (Viechtbauer, 2010).

The Knapp–Hartung-Adjustment

As the simulation studies will show, the interval estimate in (5.11) is generally too short and tends to be overconfident, i.e., it is too liberal and its actual confidence is lower than its aspired level. For this very reason, Knapp and Hartung (2003) suggested an adjustment to (5.11) based on the function $q_y(\tau)$ defined in (2.24), which was used multiple times for different constructions for point and interval estimates for τ in Chapter 4. The adjustment takes advantage of properties of q_y , which were discussed thoroughly in Chapter 2.

So, let

$$\tau \mapsto q_y(\tau), \quad q_y(\tau) = \|E_{\tau\delta} y\|_{\tau\delta}.$$

It was shown that $q_Y(\tau) \sim \mathfrak{X}_{k-p}^2$ and, in particular, that $\mathbb{E}(q_Y(\tau)) = k - p$. Also, $q_y(\tau)$ is strictly monotone decreasing, injective, smooth and converging to 0 as τ goes to infinity. The dependence of $q_y(\tau)$ on the parameter δ will not be written explicitly. In Lemma 2.22 it was shown that

$$\frac{b_i Y - \beta_i}{\sqrt{v_i \frac{q_Y(\tau)}{k-p}}} \sim \mathfrak{T}_{k-p}, \quad (5.12)$$

where \mathfrak{T}_{k-p} denotes the centred Student's t-distribution with $k - p$ degrees of freedom. If the nuisance parameters τ and δ are known, (5.12) would yield an exact

confidence interval for β_i . Exact, here, shall not only mean that the corresponding interval estimate is the result of an exact probability statement but shall also be understood in the frequentist sense: for any $\gamma \in (0, 1)$

$$P \left(\mathbf{t}_{k-p, \frac{\gamma}{2}} \leq \frac{b_i Y - \beta_i}{\sqrt{v_i \frac{q_Y(\tau)}{k-p}}} \leq \mathbf{t}_{k-p, 1-\frac{\gamma}{2}} \right) = 1 - \gamma. \quad (5.13)$$

For any consistent estimators $\hat{\tau}$ and $\hat{\delta}$, it follows by the delta method and the joint continuity of $q_y(\tau)$ in (τ, δ) that

$$q_y(\hat{\tau}) \xrightarrow{P} q_y(\tau). \quad (5.14)$$

Again, the delta method can be applied to yield

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{v}_i \frac{q_Y(\hat{\tau})}{k-p}}} \rightsquigarrow \mathfrak{T}_{k-p}. \quad (5.15)$$

This results in the approximative $(1 - \gamma)$ -confidence interval

$$\left[\hat{\beta}_i - \mathbf{t}_{k-p, 1-\frac{\gamma}{2}} \cdot \sqrt{\hat{v}_i \frac{q_y(\hat{\tau})}{k-p}}, \hat{\beta}_i + \mathbf{t}_{k-p, 1-\frac{\gamma}{2}} \cdot \sqrt{\hat{v}_i \frac{q_y(\hat{\tau})}{k-p}} \right]. \quad (5.16)$$

Here, $\mathbf{t}_{k-p, \gamma}$ denotes the γ -quantile of the centred Student's t-distribution \mathfrak{T}_{k-p} with $k - p$ degrees of freedom.

If $q_y(\hat{\tau}) < k - p$, the interval estimate in (5.16) will turn out tighter than the interval in (5.11). As it was desired to construct an interval estimate that is in any case more conservative than (5.11), Knapp and Hartung (2003) suggest to truncate the factor $\frac{q_\delta(\hat{\tau})}{k-p}$ by 1 below as an additional *ad hoc* improvement to (5.16). Let

$$\check{v}_i := \hat{v}_i \cdot \max \left(1, \frac{q_\delta(\hat{\tau})}{k-p} \right) \quad (5.17)$$

denote the result of this truncation. Then for each β_i , $i = 1, \dots, p$ and aspired significance $\gamma \in (0, \frac{1}{2})$, this yields the approximate $(1 - \gamma)$ -confidence interval

$$\left[\hat{\beta}_i - \mathbf{t}_{k-p, 1-\frac{\gamma}{2}} \cdot \sqrt{\check{v}_i}, \hat{\beta}_i + \mathbf{t}_{k-p, 1-\frac{\gamma}{2}} \cdot \sqrt{\check{v}_i} \right]. \quad (5.18)$$

The interval estimate in (5.18) is probably the current state-of-the-art for inference on the regression coefficients in the random effects meta regression model.

The interval estimates in (5.11) and (5.18) are implemented in the R-package `metafor` by Viechtbauer (2010), and in the STATA-package `metareg` by Harbord and

Higgins (2008). All three intervals, (5.11), (5.16), and (5.18), are implemented in the R-package `metagen` by Möbius (2014). A comparison study of the two adjusted interval estimates, with and without truncation, is missing in the literature so far and is included in the performance study of Section 5.6, which yields quite surprising results.

Inference on the full coefficient vector

The pivotal argument of the previous section can be extended to higher dimensions. Recall that

$$BY \sim \mathfrak{N}_p(\beta, V). \quad (5.19)$$

The observed value of BY is By . If τ, δ are assumed to be known, BY is the maximum likelihood estimator of β . The model assumptions are

$$Y \sim \mathfrak{N}_k(X\beta, \Omega_{\tau\delta}^{-1}). \quad (5.20)$$

Therefore, $\Omega_{\tau\delta}^{\frac{1}{2}}(Y - X\beta) \sim \mathfrak{N}_k(0, I_k)$ and thus, $(Y - X\beta)^\top \Omega_{\tau\delta}(Y - X\beta) \sim \mathfrak{X}_k^2$. Replacing τ and δ with consistent estimators, say $\hat{\tau}$ and $\hat{\delta}$, yields the approximate $(1 - \gamma)$ -confidence ellipsoid

$$A_1 := \left\{ \beta \in \mathbb{R}^p : (y - X\beta)^\top \hat{\Omega} (y - X\beta) \leq \mathfrak{r}_{k,1-\gamma}^2 \right\} \quad (5.21)$$

for β and $\gamma \in (0, 1)$. Here, $\mathfrak{r}_{k,\gamma}^2$ denotes the γ -quantile of the chi-square distribution \mathfrak{X}_k^2 with k degrees of freedom. A different approach follows from (5.19): from $V^{-\frac{1}{2}}(BY - \beta) \sim \mathfrak{N}_p(0, I_p)$ follows that $(BY - \beta)^\top V^{-1}(BY - \beta) \sim \mathfrak{X}_p^2$. Replacing τ and δ with consistent estimators yields

$$A_2 := \left\{ \beta \in \mathbb{R}^p : \left(\hat{B}y - \beta \right)^\top \hat{V}^{-1} \left(\hat{B}y - \beta \right) \leq \mathfrak{r}_{p,1-\gamma}^2 \right\}, \quad (5.22)$$

as an approximate $(1 - \gamma)$ -confidence ellipsoid for β and $\gamma \in (0, 1)$. The confidence set in (5.22) reassembles familiarity: it is based on an unbiased point estimator of the parameter of interest, namely $\hat{B}Y$ of β , and the form of its inherited inequality allows to immediately read off basic properties of its defining ellipsoid; the centre of the ellipsoid is $\hat{B}y$ and the principle components of the ellipsoid are the eigenvectors of \hat{V}^{-1} .

In contrast, the confidence set in (5.21) appears peculiar and naive. Their respected ellipsoids, though, have the same centre and share the same principle components. They only differ in their respected equatorial radii. Recall that

$$\|X\beta - y\|_{\tau\delta}^2 = \|\beta - By\|_{V_{\tau\delta}}^2 + \|E_{\tau\delta}y\|_{\tau\delta}^2. \quad (5.23)$$

Define the following function.

$$p_y(\beta, \tau) := (\beta - By)^\top V^{-1} (\beta - By). \quad (5.24)$$

In terms of $p_y(\beta, \tau)$ and $q_y(\tau)$, the ellipsoids (5.21) and (5.22) can be rewritten as

$$\{\beta \in \mathbb{R}^p : p_y(\beta, \tau) \leq \mathfrak{r}_{p,1-\gamma}^2\} \text{ vs. } \{\beta \in \mathbb{R}^p : p_y(\beta, \tau) \leq \mathfrak{r}_{k,1-\gamma}^2 - q_y(\hat{\tau})\}. \quad (5.25)$$

This shows that the ellipsoid defined in (5.21) has the same centre and the same principle components as (5.22). To summarise previous results in the new notation, note that

$$\begin{aligned} p_Y(\beta, \tau) &\sim \mathfrak{X}_p^2, \\ q_Y(\tau) &\sim \mathfrak{X}_{k-p}^2, \\ p_Y(\beta, \tau) + q_Y(\tau) &\sim \mathfrak{X}_k^2. \end{aligned}$$

Thus, in order to understand the difference between the confidence ellipsoids in (5.25), it needs to be understood what the term $q_y(\tau)$ is quantifying. But this has already been established in Chapter 4. As a diagonal matrix with positive eigenvalues, $\Omega_{\tau\delta}$ is positive definite. In particular, $\Omega_{\tau\delta}$ defines an inner product on \mathbb{R}^k

$$\langle a, b \rangle_{\tau\delta} := a^\top \Omega_{\tau\delta} b \quad (5.26)$$

for $a, b \in \mathbb{R}^k$, and, therefore, a norm on \mathbb{R}^k which is denoted by $\|a\|_{\tau\delta} := \sqrt{\langle a, a \rangle_{\tau\delta}}$. This shows that the difference between the two ellipsoids lies in the equatorial radii of the principal components, namely for $i = 1, \dots, p$

$$\sqrt{\frac{\mathfrak{r}_{p,1-\gamma}^2}{\lambda_i}} \text{ vs. } \sqrt{\frac{|\mathfrak{r}_{k,1-\gamma}^2 - \|Ey\|_{\hat{\tau}\delta}^2|}{\lambda_i}}. \quad (5.27)$$

Let $\gamma_0 \in (0, 1)$. If

$$\mathfrak{r}_{k,1-\gamma_0}^2 < \|Ey\|_{\hat{\tau}\delta}^2, \quad (5.28)$$

then the confidence set in (5.21) evaluates to the empty set for all $\gamma \geq \gamma_0$, since $\mathfrak{r}_{k,1-\gamma}^2$ is strictly monotone decreasing in γ . However, since

$$\lim_{\gamma \rightarrow 0} \mathfrak{r}_{k,1-\gamma}^2 = \infty \quad \text{and} \quad \lim_{\gamma \rightarrow 1} \mathfrak{r}_{k,1-\gamma}^2 = 0, \quad (5.29)$$

there always exists some $\gamma_0 \in (0, 1)$ such that $\mathfrak{r}_{k,1-\gamma}^2 \geq q_y(\hat{\tau})$ for all $\gamma \leq \gamma_0$ and for which (5.21) is not empty. We can even compute γ_0 explicitly:

$$\gamma_0 := \mathfrak{X}_k^2 \left(\|Ey\|_{\hat{\tau}\delta}^2 \right). \quad (5.30)$$

On the other hand, the probability that (5.21) is empty can be calculated for any fixed $\gamma \in (0, 1)$. Recall that $q_Y(\tau) = \|EY\|_{\tau\delta}^2 \sim \mathfrak{X}_{k-p}^2$. If τ and δ were known, then

$$P(A_1 = \emptyset) = P(\mathfrak{x}_{k,1-\gamma}^2 < \|EY\|_{\tau\delta}^2) = 1 - \mathfrak{X}_{k-p}^2(\mathfrak{x}_{k,1-\gamma}^2). \quad (5.31)$$

It might, nevertheless, be advisable to use (5.22) for the construction of confidence ellipsoids as this estimator is guaranteed to yield non-empty sets.

Methods based on a full likelihood approach

The maximum likelihood estimators for τ and β have already been established in Section 4.3. Necessary and sufficient conditions have been established in Section 2.9. For details see the discussion after Algorithm 4.7.

5.5 Methods based on generalised inference

First construction based on generalised inference principles

From a theoretical perspective, the following construction shows how to construct new generalised pivotal quantities from old ones. In Section 4.5, a generalised pivotal quantity for the heterogeneity parameter τ was constructed. This quantity will now be used to build a generalised pivotal quantity for the regression coefficients β_i , $i = 1, \dots, p$.

So, let \tilde{T} be the generalised pivotal quantity for τ defined in (4.23), namely

$$\tilde{T} := p_\delta(Q_\delta(\tau)) = p_\delta(Q) \quad (5.32)$$

for $Q \sim \mathfrak{X}_{k-p}^2$. Define $(\tilde{v}_1, \dots, \tilde{v}_p) := \text{diag } V_{\tilde{T}\delta}$ and let \tilde{b}_i denote the i th row of $B_{\tilde{T}\delta}$, hence $\tilde{v}_i = (V_{\tilde{T}\delta})_{ii}$ and $\tilde{b}_i = (V_{\tilde{T}\delta})_i X^\top \Omega_{\tilde{T}\delta}$. The observed values of \tilde{v}_i and \tilde{b}_i are v_i and b_i respectively. As \tilde{T} is a generalised pivotal quantity, the respected (multivariate) distributions of \tilde{v}_i and \tilde{b}_i can explicitly be calculated:

$$\tilde{v}_i = \left(\left(X^\top ([\tilde{T}]_k + [\delta])^{-1} X \right)^{-1} \right)_{ii}, \quad (5.33)$$

$$\tilde{b}_i = \left(\left(X^\top ([\tilde{T}]_k + [\delta])^{-1} X \right)^{-1} \right)_i X^\top ([\tilde{T}]_k + [\delta])^{-1}. \quad (5.34)$$

Since the distribution of \tilde{T} is free of unknown parameters, so are the distributions of \tilde{v}_i and \tilde{b}_i free of unknowns. Define

$$L_i := \tilde{b}_i y - \frac{b_i Y - \beta_i}{\sqrt{v_i}} \cdot \sqrt{\tilde{v}_i} = \tilde{b}_i y - N \cdot \sqrt{\tilde{v}_i} \quad (5.35)$$

where $N \sim \mathfrak{N}(0, 1)$. By (5.33), (5.34), and (5.35), the distribution of L_i is free of unknowns, though, it depends on the data via the observed response vector y . The observed value of L_i is β_i . Hence, L_i is a generalised pivotal quantity for β_i . Let \mathfrak{L}_i denote the distribution of L_i and let $\mathfrak{l}_{i\gamma}$ denote its γ -quantile for any $\gamma \in (0, 1)$. Then for any fixed $\gamma \in (0, \frac{1}{2})$ and fixed $i = 1, \dots, p$,

$$\left[\mathfrak{l}_{i\frac{\gamma}{2}}, \mathfrak{l}_{i1-\frac{\gamma}{2}} \right] \quad (5.36)$$

is a generalised $(1 - \gamma)$ -confidence interval for β_i .

The following algorithm can be used to estimate the quantiles $\mathfrak{l}_{i\gamma}$ for any $\gamma \in (0, 1)$. If $\gamma \in (0, 1)^s$, let $\mathfrak{l}_\gamma = (\mathfrak{l}_{i\gamma_1}, \dots, \mathfrak{l}_{i\gamma_s})$ denote the s -tuple of γ_r -quantiles of \mathfrak{L}_i , $r = 1, \dots, s$.

Algorithm 5.3 Let $i = 1, \dots, p$ be fixed. For $\gamma \in (0, 1)^s$, this algorithm returns a sequence $(l_{i,\gamma,m})_{m \in \mathbb{N}}$ of random elements in \mathbb{R}^s , such that

$$l_{i,\gamma,m} \xrightarrow{P} \mathfrak{l}_{i\gamma}, \quad m \rightarrow \infty. \quad (5.37)$$

- 1) Let $(\tau_l)_{l \in \mathbb{N}}$ be a stream of generalised pivotal quantities generated, for example, using Algorithm 4.8.
- 2) Let $(N_l)_{l \in \mathbb{N}}$ be a stream of random elements with $N_l \sim \mathfrak{N}(0, 1)$.
- 3) For each $l \in \mathbb{N}$, solve the system of linear equations

$$(X^\top \Omega_{\tau_l \delta} X) \beta_l = X^\top \Omega_{\tau_l \delta} y, \quad (5.38)$$

which yields a sequence of $(\beta_l)_l$.

- 4) Let v_{il} , $i = 1, \dots, p$ denote the diagonal elements of $(X^\top \Omega X)^{-1}$. This yields a sequence of $(v_{il})_{l \in \mathbb{N}}$.
- 5) Take the random sequences $(\tilde{v}_{il})_{l \in \mathbb{N}}$, $(\tilde{b}_{il})_{l \in \mathbb{N}}$, and $(N_l)_{l \in \mathbb{N}}$ and zip them together using (5.35). This yields a random sequence $(L_{il})_{l \in \mathbb{N}}$. Then each L_{il} has distribution \mathfrak{L}_i .

Finally, for any $m \in \mathbb{N}$, define $l_{i,\gamma_r,m}$ as the empirical γ_r -quantile of $(L_{il})_{l \leq m}$. This yields

$$l_{i,\gamma,m} := (l_{i,\gamma_1,m}, \dots, l_{i,\gamma_s,m}). \quad (5.39)$$

Instead of obtaining confidence intervals for β_i which are based on stochastic approximation, such as (5.11), (5.16), and (5.18), the above construction produces confidence intervals which are the result of exact probability statements.

The pivotal distribution in (5.35), the resulting interval estimate in (5.36), and a short version of Algorithm 5.3 have already been successfully published in (Friedrich and Knapp, 2013)

Note that step 4) forces to invert the matrix $(X^\top \Omega X)$. This is computationally expensive as the algorithm is potentially evaluated around 1000 times or more. It will be discussed later how to modify this algorithm such that $(X^\top \Omega X)^{-1}$ does not need to be computed.

Excursion 5.4 The evaluation of the generalised pivotal quantity L_i is implemented in the function `formulaL` in the `metagen` package.

```
library(metagen)
```

First the summary statistics of the BCG vaccine efficacy data are collected.

```
bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
x <- cbind(1, bcg$x)

k <- dim(x)[1]
p <- dim(x)[2]
```

Points of the independent random streams of step 1) and 2) are generated by

```
set.seed(8572178)
h <- pfunc(y, d, x)(rchisq(1, k-p))
g <- rnorm(1)
```

Then the following call will produce a random pair from the generalised pivotal distributions \mathfrak{L}_i , $i = 1, 2$.

```
formulaL(y, d, h, g, x)
# [1] 0.42897269 -0.02472352
```

Second construction based on generalised inference principles

Recall that $V^{-\frac{1}{2}}(BY - \beta) \sim \mathfrak{N}_p(0, I_p)$. This distribution is free of unknowns. Let $\tilde{V} := V_{\tilde{T}\delta}$ and $\tilde{B} = B_{\tilde{T}\delta}$ where \tilde{T} is a pivotal quantity of τ . Then define

$$\begin{aligned} R &:= \tilde{B}y - \tilde{V}^{\frac{1}{2}} \cdot V^{-\frac{1}{2}}(BY - \beta) \\ &= \tilde{B}y - \tilde{V}^{\frac{1}{2}} \cdot N, \end{aligned} \quad (5.40)$$

for some $N \sim \mathfrak{N}_k(0, 1)$. The latter formula already shows that the distribution of this quantity is free of unknowns. The observed values of \tilde{V} and \tilde{B} are V and B respectively. Hence, the observed value of R is simply $r = \beta$. The distribution of R is given in detail by the equation

$$\begin{aligned} R &= B_{p_\delta(Q)\delta} \cdot y - V_{p_\delta(Q)\delta}^{\frac{1}{2}} \cdot N \\ &= V_{p_\delta(Q)\delta} X^\top \Omega_{p_\delta(Q)\delta} \cdot y - V_{p_\delta(Q)\delta}^{\frac{1}{2}} \cdot N \\ &= (X^\top ([p_\delta(Q)]_k + [\delta])^{-1} X)^{-1} X^\top ([p_\delta(Q)]_k + [\delta])^{-1} y \\ &\quad - (X^\top ([p_\delta(Q)]_k + [\delta])^{-1} X)^{-1} \cdot N \end{aligned}$$

where $N \sim \mathfrak{N}_p(0, 1)$ and $Q \sim \mathfrak{X}_{k-p}^2$. Each marginal projection R_i of R is a generalised pivotal quantity for β_i in the sense of Definition 3.1. Let \mathfrak{R} denote the p -variate distribution of R and \mathfrak{r}_γ denote its (p -variate) γ -quantile for any $\gamma \in (0, 1)$. This suggests an alternative route on how to obtain a generalised pivotal quantity for each β_i , $i = 1, \dots, p$. If \mathfrak{R}_i denotes the i th marginal distribution of \mathfrak{R} and $\mathfrak{r}_{i\gamma}$ its γ -quantile, then each

$$\left[\mathfrak{r}_{i\frac{\gamma}{2}}, \mathfrak{r}_{i1-\frac{\gamma}{2}} \right] \quad (5.41)$$

is a generalised $(1 - \gamma)$ -confidence interval for β_i .

Computationally, R is a solution to the system of linear equations

$$V_{p_\delta(Q)\delta}^{-\frac{1}{2}} \cdot R = V_{p_\delta(Q)\delta}^{\frac{1}{2}} X^\top \Omega_{p_\delta(Q)\delta} \cdot y - N. \quad (5.42)$$

In different notation, this is equal to

$$\begin{aligned} \tilde{V}^{-\frac{1}{2}} \cdot R &= \tilde{V}^{-\frac{1}{2}} \tilde{B}y - N \\ &= \tilde{V}^{\frac{1}{2}} X \tilde{\Omega} y - N, \end{aligned}$$

since $\tilde{V}^{-\frac{1}{2}} \tilde{B} = \tilde{V}^{-\frac{1}{2}} \tilde{V} X^\top \tilde{\Omega} = \tilde{V}^{\frac{1}{2}} X^\top \tilde{\Omega}$. Hence, $x := \tilde{V}^{-\frac{1}{2}} \tilde{B}y$ is the solution to the system of linear equations

$$(X^\top \Omega_{\tilde{T}_\delta} X)^{\frac{1}{2}} x = X^\top \Omega_{\tilde{T}_\delta} y. \quad (5.43)$$

Thus, the generalised pivotal quantity R can be computationally calculated by solving sequentially two systems of linear equations. This is numerically more efficient than inverting a single matrix. The above argument yields Algorithm 5.5.

Algorithm 5.5 Let $i = 1, \dots, p$ be fixed. For $\gamma \in (0, 1)^s$, this algorithm returns a sequence $(r_{i,\gamma,m})_{m \in \mathbb{N}}$ of elements in \mathbb{R}^s such that

$$r_{i,\gamma,m} \xrightarrow{P} \mathbf{r}_{i\gamma}, \quad m \rightarrow \infty. \quad (5.44)$$

- 1) Let $(\tilde{T}_l)_{l \in \mathbb{N}}$ be a stream of independent generalised pivotal quantities generated, for example, using Algorithm 4.8.
- 2) Let $(N_l)_{l \in \mathbb{N}}$ be a stream of independent random elements such that each $N_l \sim \mathfrak{N}(0, 1)$.
- 3) For each $l \in \mathbb{N}$, calculate the principle root $\tilde{V}_l^{-\frac{1}{2}}$ of

$$\tilde{V}_l^{-1} = \left(X^\top \Omega_{\tilde{T}_l \delta} X \right), \quad \text{yielding} \quad \left(\tilde{V}_l^{-\frac{1}{2}} \right)_{l \in \mathbb{N}}, \quad (5.45)$$

a sequence of matrices in $\mathbb{R}^{p \times p}$. Even though the notation might suggest differently, note that no matrix needs to be inverted during this step.

- 4) For each $l \in \mathbb{N}$, solve the system of linear equations

$$\tilde{V}_l^{-\frac{1}{2}} x_l = X^\top \Omega_{\tilde{T}_l \delta} y \quad (5.46)$$

for x_l , which yields a sequence of $(x_l)_{l \in \mathbb{N}} \in \mathbb{N}$.

- 5) For each $l \in \mathbb{N}$, solve the system of linear equations

$$\tilde{V}_l^{-\frac{1}{2}} \cdot R_l = x_l - N \quad (5.47)$$

for R_l . This yields a sequence of $(R_l)_{l \in \mathbb{N}}$. Then each R_l has the distribution \mathfrak{R} .

For any $m \in \mathbb{N}$, let $r_{\gamma_r, m}$ denote the empirical γ_r -quantile of $(R_l)_{l \leq m}$. Define

$$r_{\gamma, m} := (r_{\gamma_1, m}, \dots, r_{\gamma_s, m}), \quad r_{i, \gamma, m} = (r_{\gamma, m})_i.$$

Algorithm 5.5 is computationally more efficient and numerically more stable than Algorithm 5.3. The construction in (5.40), the generalised pivotal quantity \mathfrak{R}_i , its resulting interval estimate (5.41), and Algorithm 5.5 are genuine and have not been published prior to this dissertation.

Excursion 5.6 (Continuation of Excursion 5.4) The generalised pivotal quantity R_i is implemented in `formulaR` in an analogue fashion to L_i in the function `formulaL`. For the BCG vaccine efficacy data, the following call will produce a random draw from the distributions \mathfrak{R} .

```
set.seed(8572178)
h <- pfunc(y, d, x)(rchisq(1, k-p))
g <- rnorm(1)

formulaR(y, d, h, g, x)
# [1] 0.42479083 -0.03136872
```

Adjusting for uncertainties in the heteroscedasticity estimate

In Section 2.13, it was argued that most practical applications are more closely modelled by an extension to the random effects meta regression model that is, however, not identifiable. Using the generalised inference principle, the modelling assumptions may nevertheless be used to compensate for the added uncertainty of the heteroscedasticity estimates in a bottom-up approach. This is analogue to the use of the generalised pivotal quantity \tilde{T} in the construction of the generalised pivot quantities L_i and R_i .

It was argued that, in case the δ_j 's are truly assumed to be unknown, the model with whom one is actually working is

$$\begin{aligned} Y_j | \theta_j &\sim \mathfrak{N}(\theta_j, D_j), \\ \theta_j &\sim \mathfrak{N}\left(x_j^\top \beta, \tau\right), \\ K_j &\sim \mathfrak{X}_{n_j-1}^2, \\ D_j &= \frac{\delta_j}{n_j - 1} \cdot K_j, \end{aligned}$$

where (Y_j, D_j) denotes the summary statistics of the j th study.

Define

$$\tilde{D}_j := \delta_j \cdot \frac{d_j}{D_j} = \delta_j \cdot d_j \cdot \frac{n_j - 1}{\delta_j} \cdot K_j^{-1} = d_j(n_j - 1)K_j^{-1}. \quad (5.48)$$

As the distribution of \tilde{D}_j is independent of unknowns and its observed value is $\tilde{d}_j = \delta_j$, it is obvious that \tilde{D}_j is a generalised pivotal quantity for δ_j . If the unknown δ_j is replaced by the generalised pivotal quantity \tilde{D}_j in the definitions of \tilde{v}_i and \tilde{b}_i

in (5.33) and (5.34) respectively, the distribution of L_j in (5.35) would still stay free of unknowns. Also, its observed value would still be β_i . Let us denote these new random quantities by \bar{v}_i , \bar{b}_i and \bar{L}_j respectively.

For independent (K_1, \dots, K_k, Q, N) with $K_j \sim \mathfrak{X}^2_{n_j-1}$, with $Q \sim \mathfrak{X}^2_{k-p}$, and with $N \sim \mathfrak{N}(0, 1)$, the uni- and multivariate distributions are described via the formulae

$$\tilde{D}_j = d_j(n_j - 1)K_j^{-1}, \quad (5.49)$$

$$\bar{v}_i = \left(\left(X^\top \left([p_{\tilde{D}_j}(Q)]_k + [\tilde{D}_j] \right)^{-1} X \right)^{-1} \right)_{ii}, \quad (5.50)$$

$$\bar{b}_i = \left(\left(X^\top \left([p_{\tilde{D}_j}(Q)]_k + [\tilde{D}_j] \right)^{-1} X \right)^{-1} \right)_i X^\top \left([p_{\tilde{D}_j}(Q)]_k + [\tilde{D}_j] \right)^{-1}, \quad (5.51)$$

$$\bar{L}_i = \bar{b}_i y - N \cdot \sqrt{\bar{v}_i}. \quad (5.52)$$

If $\bar{\mathcal{L}}_i$ denotes the distribution of \bar{L}_i in (5.52) and $\bar{l}_{i,\gamma}$ its γ -quantile, then

$$\left[\bar{l}_{i,\frac{\gamma}{2}}, \bar{l}_{i,1-\frac{\gamma}{2}} \right] \quad (5.53)$$

is a generalised $(1 - \gamma)$ -confidence interval for β_i for $\gamma \in (0, \frac{1}{2})$ and fixed $i = 1, \dots, p$.

Note again that this confidence set is the result of an exact probability statement. The quantiles $\bar{l}_{i,\gamma}$ are obtained by numerical simulations by Algorithm 5.7 which is a modified version of Algorithm 5.3.

Algorithm 5.7 Let $i = 1, \dots, p$ be fixed. For $\gamma \in (0, 1)^s$, this algorithm returns a sequence $(\bar{l}_{i,\gamma,m})_{m \in \mathbb{N}}$ of random elements in \mathbb{R}^s , such that

$$\bar{l}_{i,\gamma,m} \xrightarrow{P} \bar{l}_{i,\gamma}, \quad m \rightarrow \infty. \quad (5.54)$$

- 1) For each $j = 1, \dots, k$, let $K_j = (K_{jl})_{l \in \mathbb{N}}$ be an independent sequence such that each $K_{jl} \sim \mathfrak{X}_{n_j-1}^2$.
- 2) Let $\tilde{D}_l = (\tilde{D}_{1l}, \dots, \tilde{D}_{kl})$ with $\tilde{D}_{jl} = d_j(n_j - 1)K_{jl}^{-1}$.
- 3) Let $(\tau_l)_{l \in \mathbb{N}}$ be a stream of generalised pivotal quantities that has been generated by applying Algorithm 4.8 but setting $\delta := \tilde{D}_l$ in the algorithm itself.
- 4) For each $l \in \mathbb{N}$, solve the system of linear equations

$$(X^\top \Omega_{\tau_l D_l} X) \beta_l = X^\top \Omega_{\tau_l D_l} y, \quad (5.55)$$

which yields a sequence of $(\beta_l)_l$, and define $v_{il} = ((X^\top \Omega_{\tau_l D_l} X)^{-1})_{ii}$.

- 5) Let $(N_l)_{l \in \mathbb{N}}$ be a stream of random elements with $N_l \sim \mathfrak{N}(0, 1)$.
- 6) Take the random sequences $(v_{il})_{l \in \mathbb{N}}$, $(b_{il})_{l \in \mathbb{N}}$, and $(N_l)_{l \in \mathbb{N}}$ and zip them together using (5.52), which yields a random sequence $(L_{il})_{l \in \mathbb{N}}$. Then each L_{il} has distribution $\bar{\mathfrak{L}}_i$.

For any $m \in \mathbb{N}$, let $l_{i,\gamma_r,m}$ denote the empirical γ_r -quantile of $(L_{il})_{l \leq m}$. Define

$$l_{i,\gamma,m} := (l_{i,\gamma_1}, \dots, l_{i,\gamma_s}). \quad (5.56)$$

The key in going from Algorithm 5.3 to Algorithm 5.7 is happening in steps 3) and 4) where the unknown inner-study variances δ_j are replaced with generalised pivots which are in turn based on the observed summary statistics $d_j = \hat{\delta}_j$ of the respected studies. This will increase the variance of the resulting generalised pivotal quantity and should adjust for the uncertainty in the estimation of the δ_j 's.

Note that replacing the δ_j 's by the generalised pivots \tilde{D}_j in Algorithm 4.8 already defines a new adjusted pivotal quantity for the heterogeneity parameter τ . An adjusted version of Algorithm 5.5 can also be constructed by modifying its steps 1) and 3) accordingly. This yields another adjusted pivotal quantity for the β_i 's.

Unfortunately, the above procedure has its limitations. Consider the case, when the effect measures of the studies are risk ratios as in the BCG example of Section 1.2. Then model (2.2)+(2.3) is used with the logarithms of the relative risks as response values, since these are approximately Gaussian distributed. In this case, however, the estimated within-study variances are no longer independent of the responses. Expressed in the above notation, each (Y_j, D_j) , $j = 1, \dots, k$, is no

longer independent. The derivation of the distribution of \tilde{D}_j , though, relies on this independence in Equations (2.66) and (2.67).

Example: Generalised pivotal distributions in the BCG vaccine efficacy data

The previous sections have shown that inference on the regression coefficients may either be based on the distributions L_i , $i = 1, \dots, p$, or on the marginal distributions of \mathfrak{R} . The obvious question is whether differences are to be expected between the pivotal distributions \mathfrak{L}_i and \mathfrak{R}_i . In case the underlying model follows the mean response extension, the question is whether it is worth the effort to apply the discussed adjustment of the last section.

Different density estimates of the respected distributions can be seen in Figure 5.1 and Figure 5.2. As one can see, no big differences are to be expected when using either \mathfrak{L}_2 or \mathfrak{R}_2 for making inference on the slope β_2 . One may argue that \mathfrak{R}_2 shows a sharper peak in the middle of the distribution and less heavy tails than \mathfrak{L}_2 . This suggests to base inference on \mathfrak{R} as the implementation of this generalised pivotal quantity allows a higher degree of numerical optimisation.

Differences between the adjusted and unadjusted versions of \mathfrak{L}_2 and \mathfrak{R}_2 are hardly visible Figure 5.2. One may think that this is due to the large sample sizes in some of the clinical trials of the BCG vaccine efficacy data. This supposition, however, could not be supported by artificially altering the sample sizes.

Excursion 5.8 Algorithm 4.8, Algorithm 5.3, Algorithm 5.5, and their respected adjustments analogue to Algorithm 5.7 can all be accessed by the function `pivotalStream` of the `metagen` package.

```
library(metagen)
bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
s <- bcg$size
x <- cbind(1, bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(1000, y=y, d=d, x=x, adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(1000, y=y, d=d, x=x, s=s, adjusted=TRUE)
```

The form of the response of this function, namely `pivUn` and `pivAd` above, is a matrix in which each row corresponds to the different pivotal distributions \mathfrak{T} , $\mathfrak{L}_1, \dots, \mathfrak{L}_p$, and $\mathfrak{R}_1, \dots, \mathfrak{R}_p$ where \mathfrak{T} is the pivotal distribution of the heterogeneity.

There exist, however, functions which join and extract the respected pivotal draws from this matrix and attach human readable labels to each of them.

```
pivh <- joinPivotalHeterogeneity(pivUn, pivAd)
pivr <- joinPivotalCoefficients(pivUn, pivAd)
```

To produce plots as shown in Figure 5.1 and Figure 5.2 try one of the following functions:

```
plotDensityH(pivh)
plotDensityH2(subset(pivh, variable=="unadjusted"))
plotDensityH2(subset(pivh, variable=="adjusted"))
plotDensitySlope(pivr)
plotDensitySlope2(pivr)
plotDensityIntercept(pivr)
plotDensityIntercept2(pivr)
```

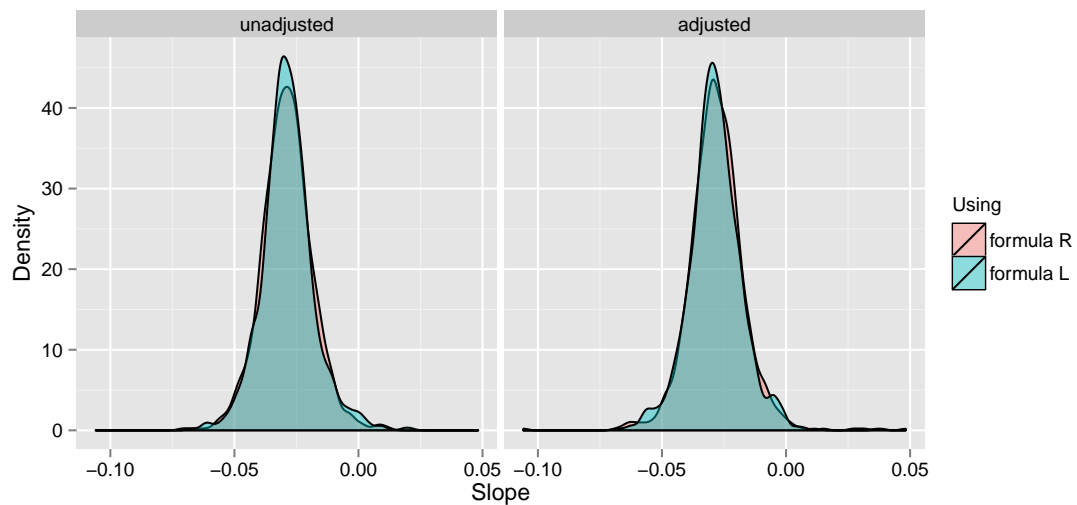


Figure 5.1 Density estimates of the distributions \mathcal{L}_2 and \mathfrak{R}_2 for the slope in the BCG vaccine efficacy example.

Generalised inference in higher dimensions

In the last sections, different point and interval estimates were defined for the individual regression coefficients β_i of the regression model. Here, the construction of

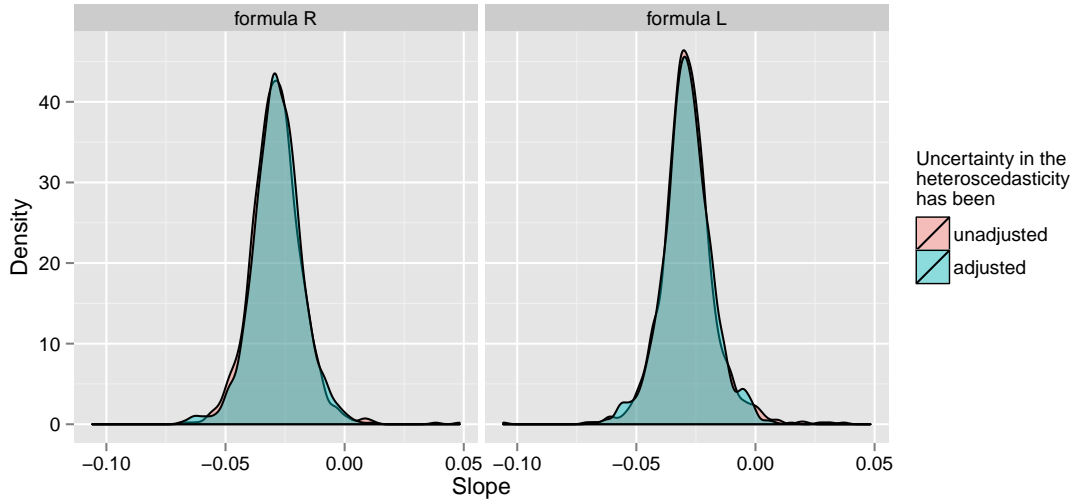


Figure 5.2 Density estimates of the distributions \mathcal{L}_2 and \mathcal{R}_2 for the slope in the BCG vaccine efficacy example.

generalised confidence ellipsoids for the vector $\beta = (\beta_1, \dots, \beta_p)$ shall be discussed. The goal is to construct a generalised pivotal quantity in the sense of Definition 3.3.

One might be tempted to define the cube

$$\left[\mathbf{r}_{\frac{\gamma}{2}}, \mathbf{r}_{1-\frac{\gamma}{2}} \right] \quad (5.57)$$

as a potential set estimate for β . However, the $(1-\gamma)$ -confidence of the cube, namely

$$\mu_{\mathfrak{R}} \left(\left[\mathbf{r}_{\frac{\gamma}{2}}, \mathbf{r}_{1-\frac{\gamma}{2}} \right] \right), \quad (5.58)$$

will usually be much smaller than $1-\gamma$. Applying a Bonferroni correction suggests to replace γ by $\frac{\gamma}{p}$ which yields p simultaneous $(1-\gamma)$ -confidence intervals,

$$\left[\mathbf{r}_{i\frac{\gamma}{2p}}, \mathbf{r}_{i1-\frac{\gamma}{2p}} \right] \quad (5.59)$$

for all β_i .

Instead of constructing simultaneous interval estimates for each β_i , or, equivalently, cube shaped confidence sets in \mathbb{R}^p , one could also aim for finding a convex area A_γ under the density $D\mathfrak{R}$ of \mathfrak{R} of minimal extent, or *measure*, such that $\mu_{\mathfrak{R}}(A_\gamma) \leq 1-\gamma$. This might be achieved by using numerical methods. Theoretical considerations suggest to define suitable quadratic forms instead. For example, define

$$R^\top R \quad (5.60)$$

as a test statistic. Let \mathfrak{R}^2 be the distribution of $R^\top R$, and \mathfrak{r}_γ^2 be its γ -quantile. Then indeed

$$\{\beta : \|\beta\|^2 < \mathfrak{r}_\gamma^2\} \quad (5.61)$$

is an ellipsoid in \mathbb{R}^p and a set estimate for β .

In case one has a (classical) Gaussian distributed pivotal quantity, such as $BY \sim \mathfrak{N}_p(\beta, V)$ for β , the usual course of action is to base further inference on quadratic forms such as

$$(BY - \beta)^\top V^{-1}(BY - \beta) \sim \mathfrak{X}_p^2. \quad (5.62)$$

It is possible to go an analogue path when constructing an higher dimensional generalised pivotal quantity in the sense of Definition 3.3. Let us assume that we are given a multivariate statistic for which each marginal component is a generalised pivotal quantity for each β_i respectively and which variance-covariance structure is free of unknowns. This holds true for both R , defined in (5.40), and $L = (L_1, \dots, L_p)$, defined in (5.35). Say, we are given $R \sim \mathfrak{R}$ and let $U := \mathbb{V}(R)$. Then the distribution of

$$T := R^\top U^{-1}R \quad (5.63)$$

is also free of unknowns. But note that the observed value T does also not depend on any nuisance parameters and is a quadratic form in β . Let \mathfrak{T} denote the distribution of T , and let \mathfrak{t}_γ denote its γ -quantile. Then

$$I_\gamma = \{\beta \in \mathbb{R}^p : \beta^\top U^{-1}\beta \leq \mathfrak{t}_\gamma\} \quad (5.64)$$

is a $(1 - \gamma)$ -confidence ellipsoid for β in \mathbb{R}^p . Given an algorithm that produces a stream of random draws from \mathfrak{R} , Algorithm 5.9 can be used to estimate its variance-covariance structure and the quantiles of the distribution \mathfrak{T} .

Algorithm 5.9 Let $i = 1, \dots, p$ be fixed. For $\gamma \in (0, 1)^s$, this algorithm returns a sequence $(t_{i,\gamma,m})_{m \in \mathbb{N}}$ of random elements in \mathbb{R}^s , such that

$$t_{\gamma,m} \xrightarrow{P} \mathbf{t}_\gamma, \quad m \rightarrow \infty. \quad (5.65)$$

- 1) Let $(R_l)_{l \in \mathbb{N}}$ be a stream of independent draws from the distribution \mathfrak{R} , for example, being produced by Algorithm 5.3.
- 2) Let $\bar{R}_m := \frac{1}{m} \sum_{l=1}^m R_l$. Then for each $i, j = 1, \dots, p$ let

$$c_{mij} := \frac{1}{m} \sum_{l=0}^m (R_{li} - \bar{R}_m) (R_{lj} - \bar{R}_m),$$

$$C_m := (c_{mij})_{ij}.$$

Then each $C_m, m \rightarrow \infty$ is an unbiased, consistent estimator of $\mathbb{V}(R)$.

- 3) Let

$$T_{lm} := R_l^\top C_m^{-1} R_l. \quad (5.66)$$

Then for each $m \in \mathbb{N}$, this will produce a list $(T_{ml})_{l \leq m}$ of length m . For any $m \in \mathbb{N}$, let $t_{\gamma_r, m}$ denote the empirical γ_r -quantile of $(T_{lm})_{l \leq m}$. Define

$$t_{\gamma, m} := (t_{\gamma_1}, \dots, t_{\gamma_s}). \quad (5.67)$$

The larger m the better the variance-covariance structure of R can be estimated by C_m .

5.6 Performance study of inference methods about the regression coefficients

Evaluating point estimators for the regression coefficients

As it has been discussed at the beginning of Section 4.6, the performance of point and interval estimators have to be evaluated, for example, by simulation studies. This holds true whether these methods are based on asymptotic statistical theory or if they are based on generalised inference principles, and it is particularly important if sample sizes are small.

As in the performance study of the heterogeneity estimators, the performance for the regression coefficient estimators will be evaluated with respect to the canonical

model and its mean response extension. The canonical model assumes the within-study variances δ to be known, i.e., the true δ -vector is fed to each of the methods. This corresponds to model (2.2)+(2.3). The mean response extension corresponds to model (2.68)–(2.70) in which the methods only see estimates $\hat{\delta}$ of the heteroscedasticity based on a vector of known study sizes.

When setting up the experimental design of the simulation study, the goal was to get a comprehensive overview on how the performance of the respected interval estimators changes due to differences in heterogeneity and heteroscedasticity. The same parameter settings and model strategies were used here as in Section 4.6.

The past sections have developed different point and interval estimators for the regression coefficients in the random effects meta regression model. As point estimators for the regression coefficients, essentially two different estimators were suggested. One based on asymptotic theory, namely $\hat{B}y$, the other based on generalised inference principles, either using the univariate distribution \mathfrak{L}_i or the i th marginal distribution of \mathfrak{R} .

Six possible point estimators for the heterogeneity τ in $\hat{B}y = B_{\tau\hat{\delta}}$ were discussed in Chapter 4. These yield six different point estimators for the regression coefficients each. It is interesting to see which of these heterogeneity estimators will result in the best performing point estimators for β . Recall that Section 4.6 concluded to either use the DerSimonian–Laird type estimator or the restricted maximum likelihood estimator for the point estimation of τ .

On the generalised inference side, inference can either be based on \mathfrak{L}_i or \mathfrak{R}_i . The plots in Figure 5.1 and Figure 5.2 suggested that not too much difference between these estimators is to be expected. In case the underlying model follows the mean response extension, also the adjusted versions $\bar{\mathfrak{L}}_i$ and $\bar{\mathfrak{R}}_i$ can be used as the base for inference. Thus, 10 different point estimators for the regression coefficients have been suggested in total, which will be discussed respectively.

This section will show multiple plots of different performance measurements. Just as in the case of the BCG vaccine efficacy example, most interest usually lies in the estimation of the slope parameters in a meta regression analysis. As performance results for the slope and the intercept have been very similar, performance measures will only be shown for the slope parameter β_2 .

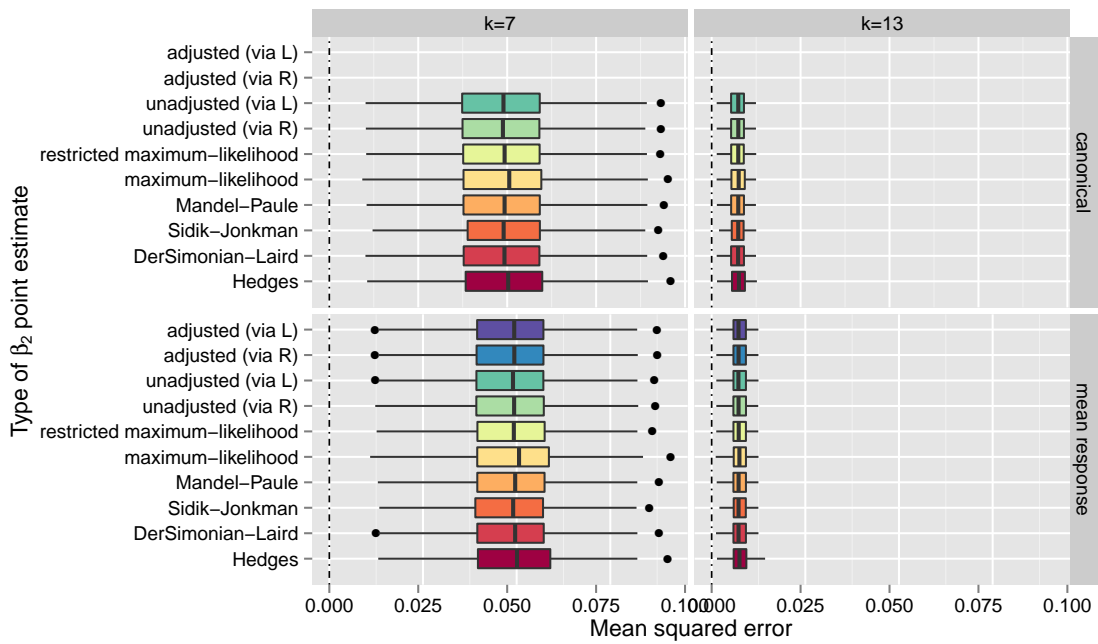


Figure 5.3 Box plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.

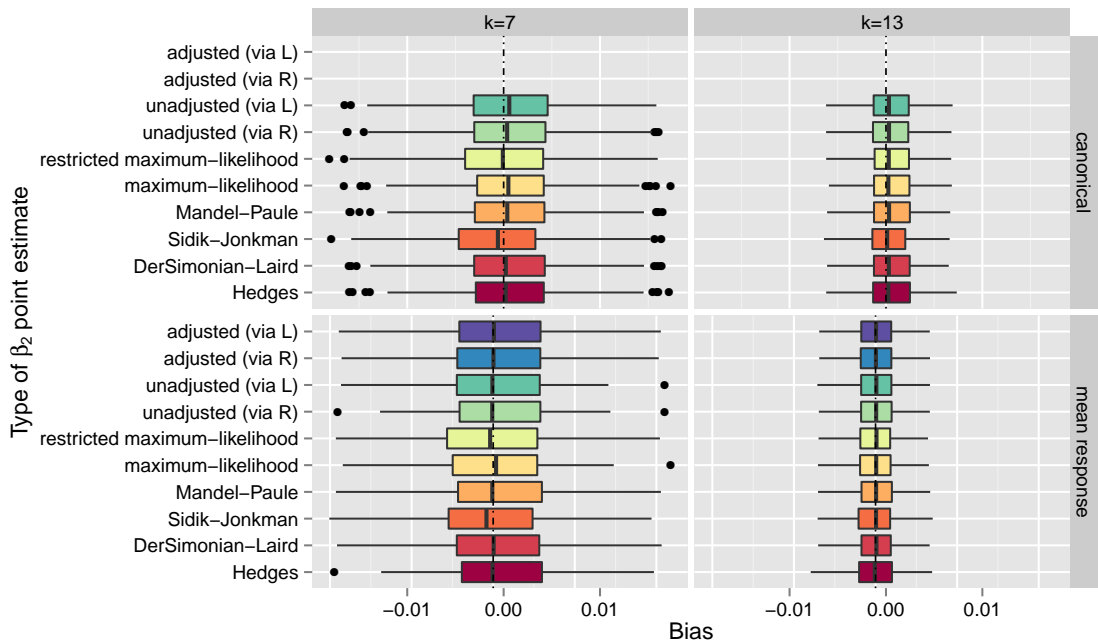


Figure 5.4 Box plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.

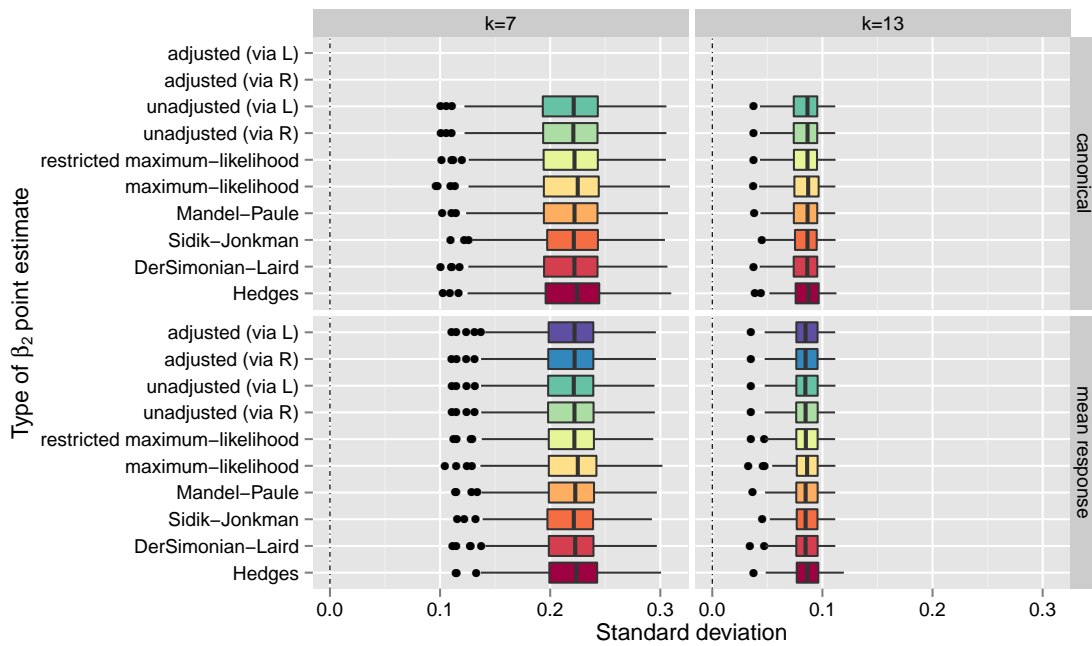


Figure 5.5 Box plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.

Figure 5.3, Figure 5.4, and Figure 5.5 show box plots of the estimated mean squared error, the bias, and the standard deviation of each of the 10 point estimators for the slope parameter. As it can be seen, the point estimators do not vary much in their respected mean squared errors. The minimal observed mean squared error for the different parameter configurations is slightly higher in case of the mean response extension, which is to be expected. Obviously, when sample sizes increase, the mean squared errors of the estimators are reduced. The methods based on generalised inference principles are very much in line with the methods based on more traditional approaches.

A more diverse picture is shown in Figure 5.4, which shows the estimated bias of the point estimators. Differences are not severe but one may argue that $\hat{B}y$ in conjunction with the Hedges type or Mandel–Paule type estimator shows a slight tighter distribution around 0 in comparison to others methods. For $k = 13$ and the mean response extension, the unadjusted generalised method based on \mathfrak{R}_2 shows the tightest bias distribution around 0. Figure 5.5 shows a very similar picture to Figure 5.3.

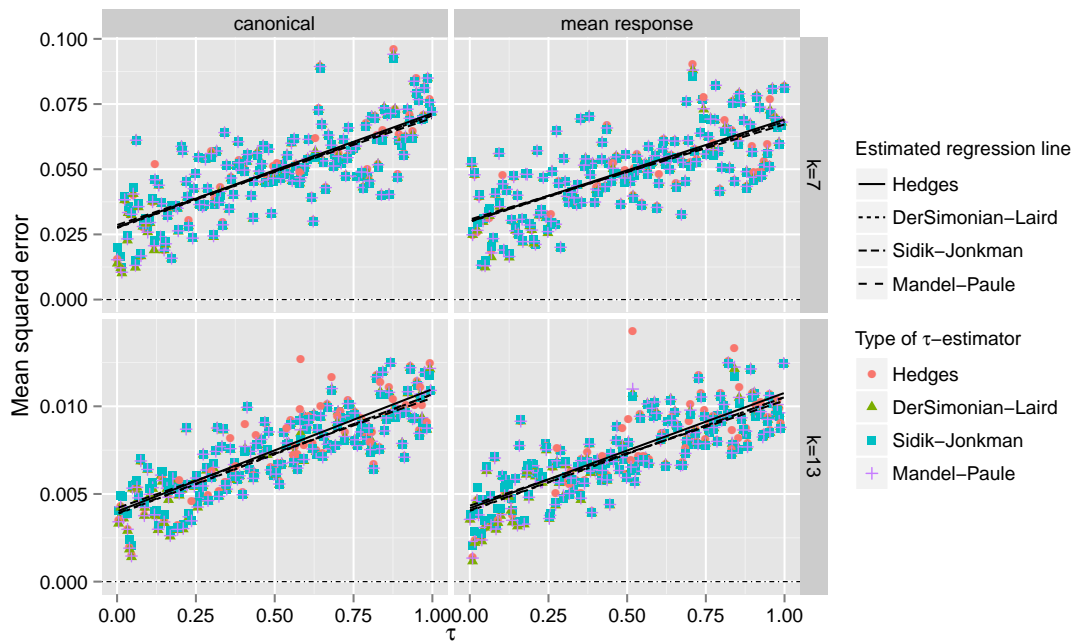


Figure 5.6 Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.

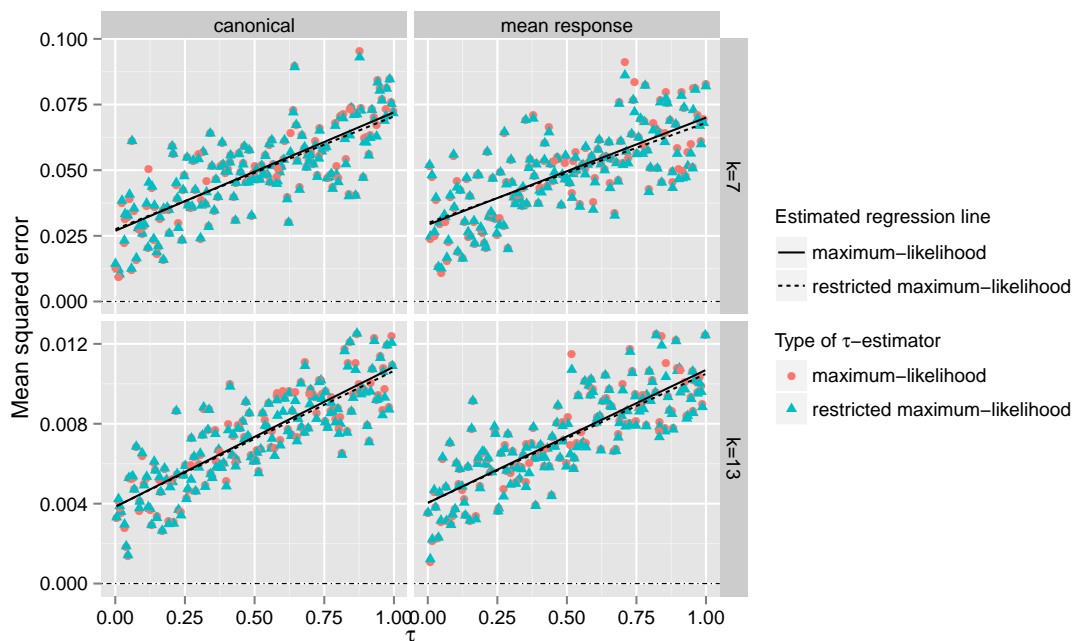


Figure 5.7 Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.

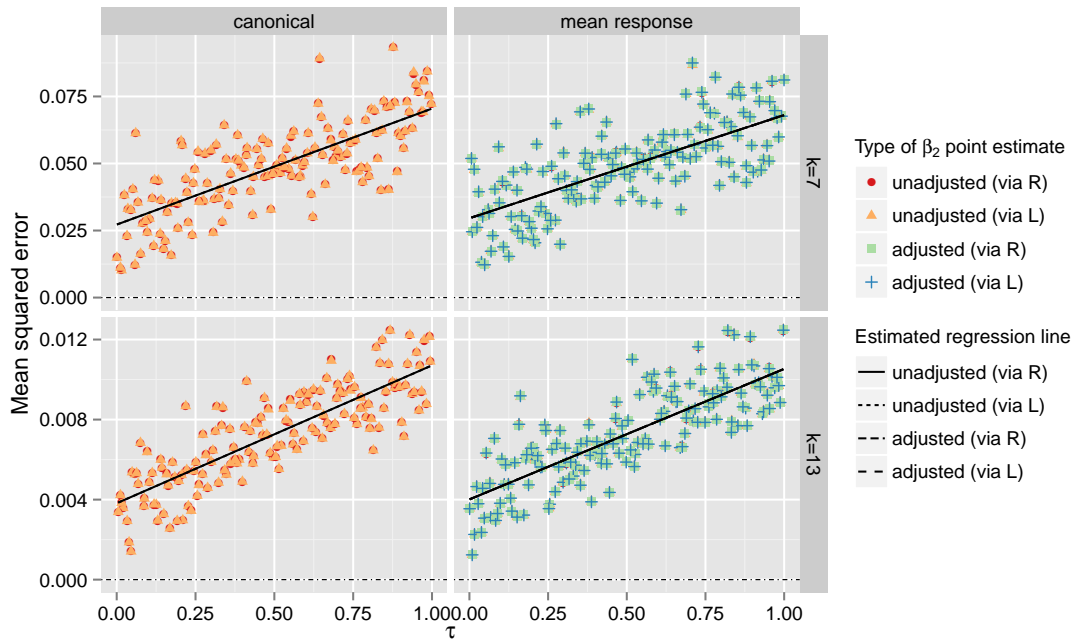


Figure 5.8 Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.

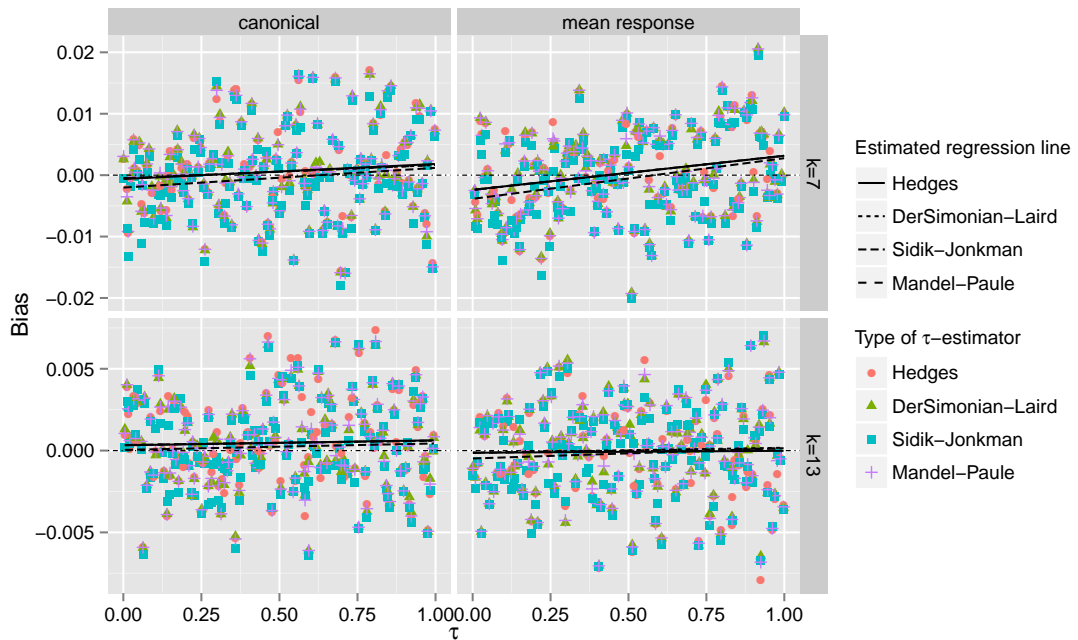


Figure 5.9 Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.

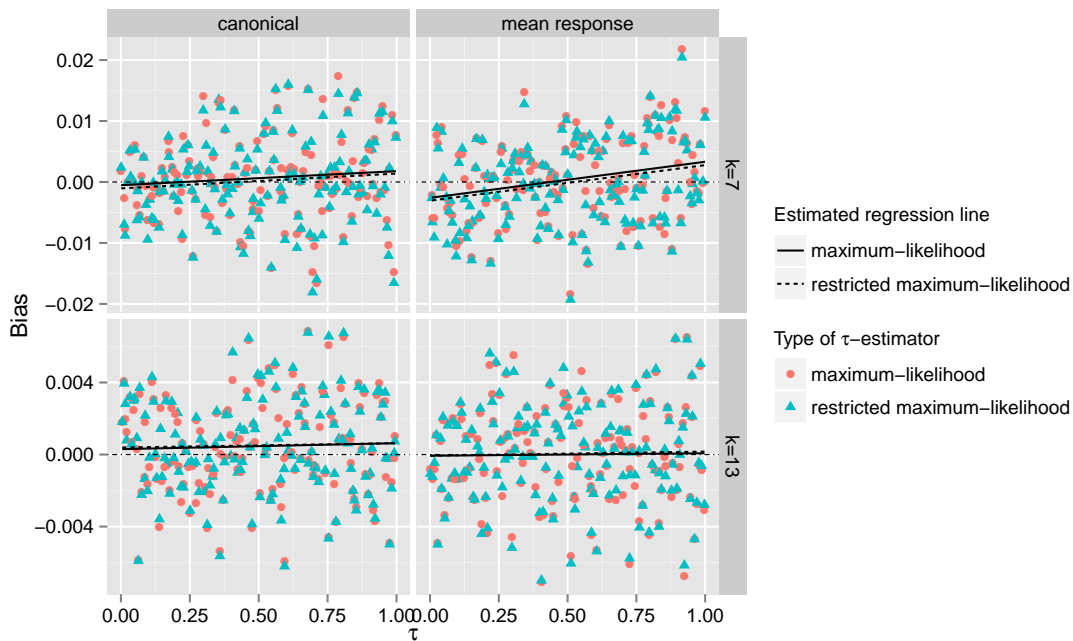


Figure 5.10 Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.

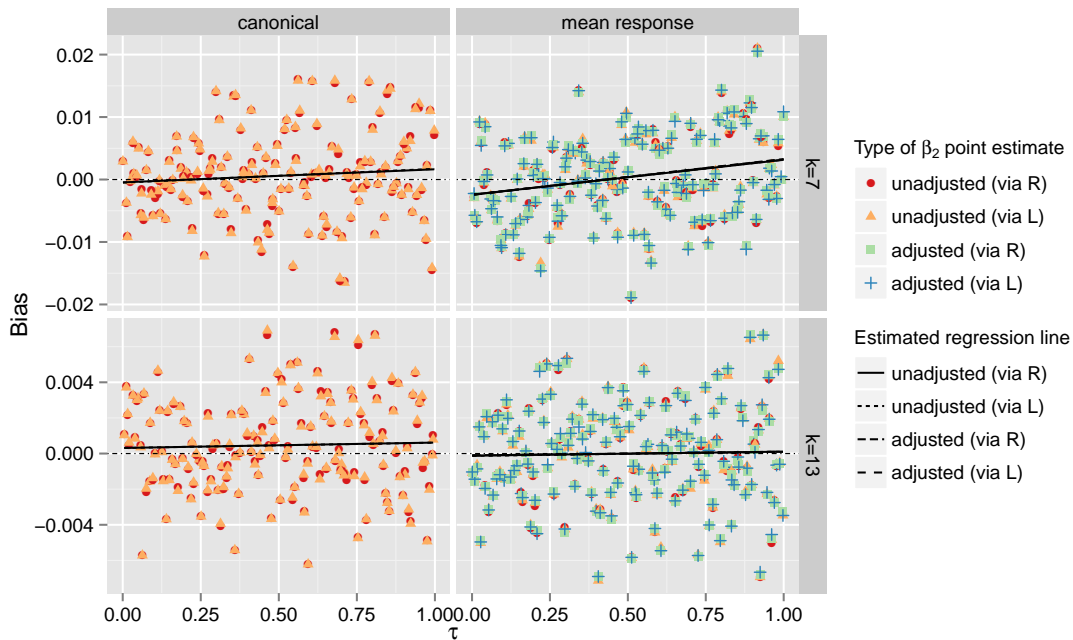


Figure 5.11 Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.

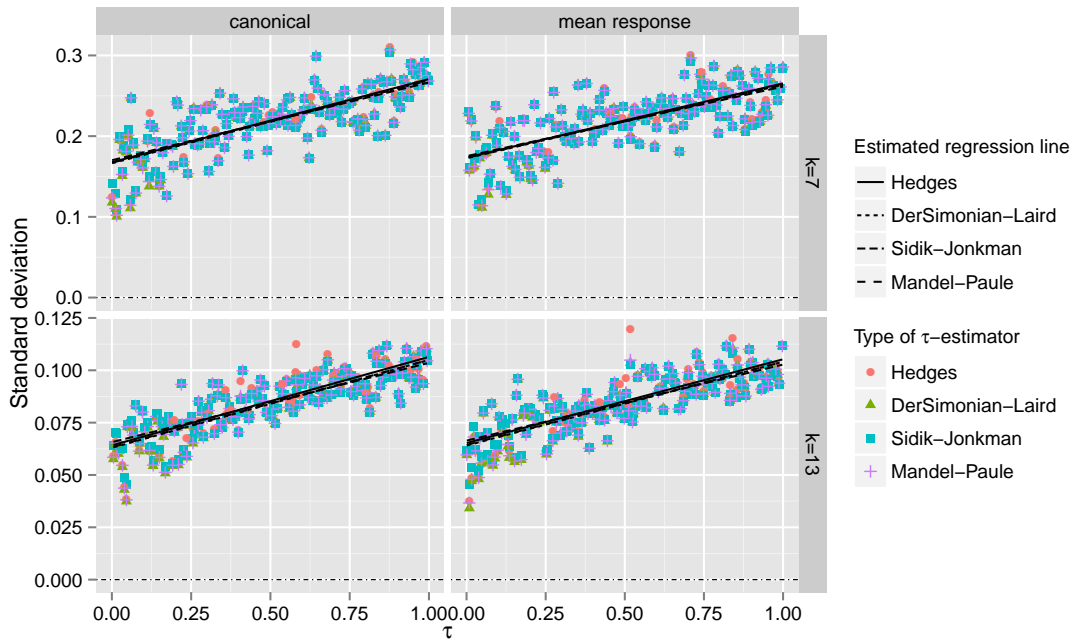


Figure 5.12 Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.

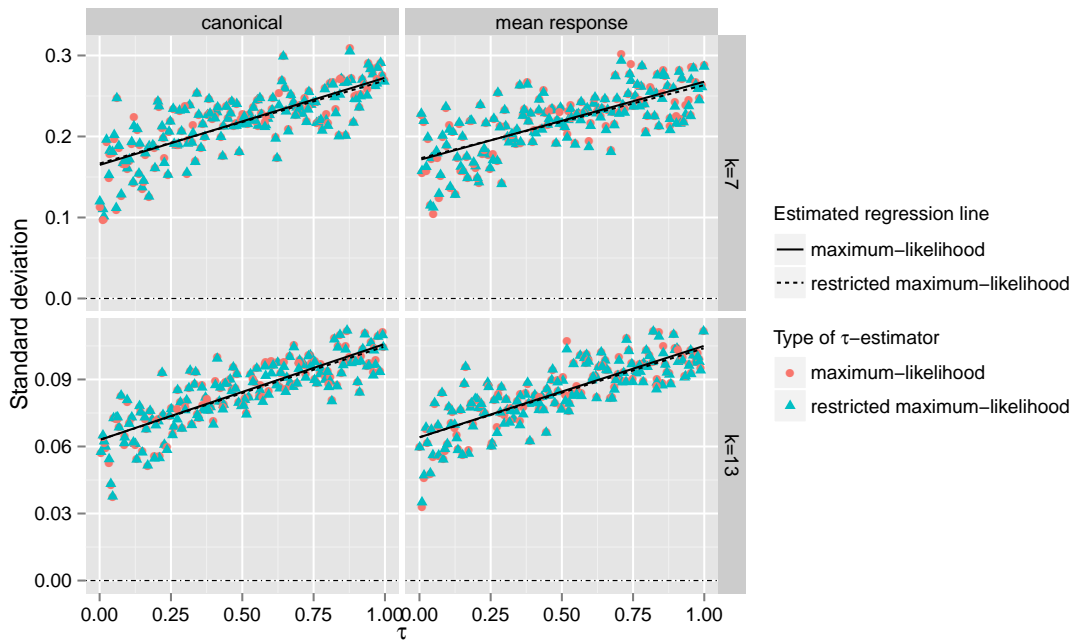


Figure 5.13 Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.

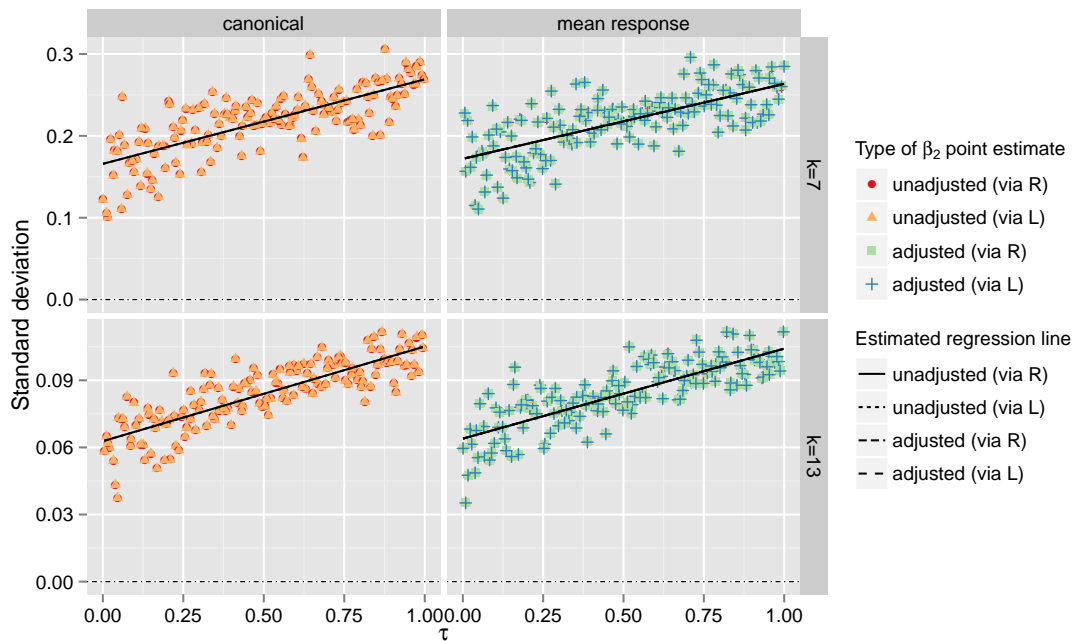


Figure 5.14 Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.

The scatter plots of the true underlying heterogeneity τ and estimated mean squared error of Figure 5.6, Figure 5.7, and Figure 5.8 show hardly any differences. The scatter plots of τ and the estimated bias are shown in Figure 5.9, Figure 5.10, and Figure 5.11. As Figure 5.9 suggests, the bias of the estimator $\hat{B}y$ in conjunction with the Mandel–Paule type estimator consistently lies below the bias of the other methods. Also note the difference in the regression lines for the likelihood methods in Figure 5.10 for $k = 7$ and $k = 13$. No mentionable differences are seen in the scatter plots of Figure 5.12, Figure 5.13, and Figure 5.14 of τ and the estimated standard deviation.

This suggests no clear recommendation of which point estimator to use for estimating the regression coefficients. One may argue in favour of $\hat{B}y$ in conjunction with the Hedges type estimator or the new generalised estimate based on the median of the unadjusted distribution \mathfrak{R} . Worth mentioning is that the point estimators based on the distributions \mathfrak{L}_i , \mathfrak{R}_i , $\bar{\mathfrak{L}}_i$, and $\bar{\mathfrak{R}}_i$ are very much in line with the performance of all other presented methods.

Evaluating interval estimators for the regression coefficients

This chapter has developed different interval estimators for the regression coefficients of the random effect meta regression model. Three of these were based on asymptotic

theory. The first being (5.11), the Knapp–Hartung adjustment (5.16), as well as an additional ad hoc improvement to this adjustment, the Knapp–Hartung ad hoc improvement (5.18), which is considered to be the current state-of-the-art. Each of these three interval estimators rely on a consistent point estimate for τ . In conjunction with the six estimators for τ which are based on either methods of moments or maximum likelihood, this results in 18 different interval estimators for each β_i .

Additionally, two generalised pivotal distributions, \mathfrak{L}_i and the i th marginal distribution of \mathfrak{R} , were developed which quantiles led to the interval estimators (5.53) and (5.41) respectively. In case of the mean response extension, these interval estimators may also be based on $\bar{\mathfrak{L}}_i$ and $\bar{\mathfrak{R}}_i$ instead. This section will, therefore, compare 22 different interval estimators for β_i . Only the performance with respect to β_2 will be discussed and their respected performance measures plotted since performance with respect to β_1 is comparable and β_2 is the parameter of practical importance in most applications.

For each of the 22 interval estimators, Figure 5.15 shows box plots of the aspired confidence level minus the estimated coverage. For a good performing interval estimate **confidence level - coverage probability** is expected to be close to 0. For a conservative method this value is preferably below 0 rather than above.

In conjunction with the Sidik–Jonkman type estimator for τ , the three interval estimators based on asymptotic theory show the most conservative performance. Among these, the Knapp–Hartung ad hoc improvement used in conjunction with the Sidik–Jonkman estimator is the most conservative of all discussed interval estimators, and it is, in fact, the only one which holds true to its aspired confidence level in all simulated parameter scenarios. In conjunction with the maximum likelihood estimator for τ , each of these interval estimators show the most liberal performance. The worst performing interval estimate is (5.11) in conjunction with the unconditional maximum likelihood estimator for the heterogeneity. Results based on this method are almost bound to overestimate their statistical significance.

Interesting to see is the wide spread of the estimated coverage that some of these interval estimators show with respect to different parameter configurations. This is especially severe in conjunction with the unconditional maximum likelihood method for τ . Recall the high bias of this estimator in Chapter 4. But also in conjunction with Hedges type estimator, the estimated coverage probability of the corresponding interval estimators for β_2 is quite diverse. This makes it difficult to predict the actual coverage of these interval estimate independent of any underlying (unknown) parameter configuration. Most accurately, the coverage can be predicted in conjunction with the Sidik–Jonkman type estimator. In conjunction with this estimator, the Knapp–Hartung adjustment without the ad hoc truncation shows a good mix between accuracy and precision.

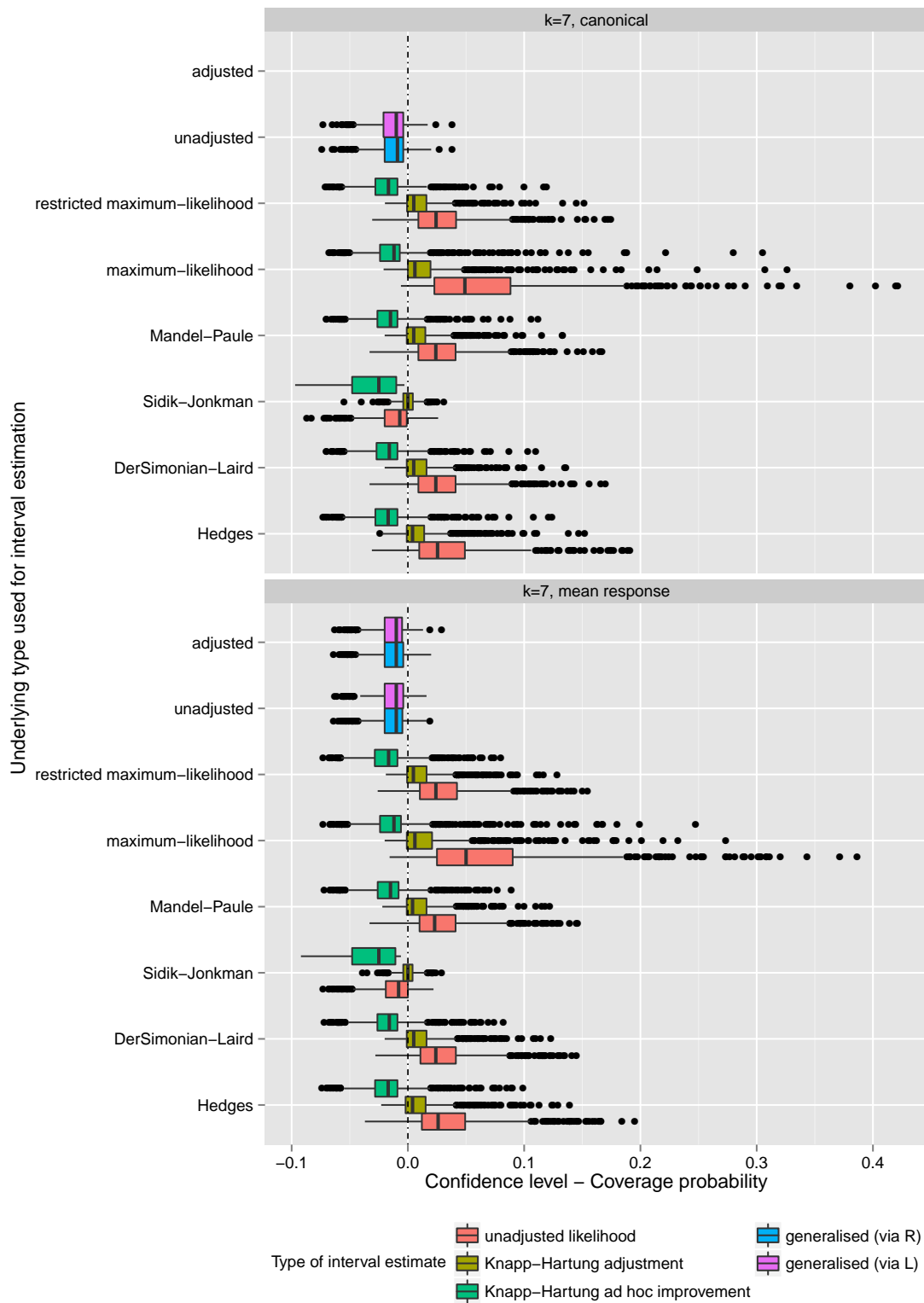


Figure 5.15 Boxplots of confidence level - coverage probability for different model parameters for $k = 7$ number of studies and different inferential methods shown for the slope parameter. The optimal line is the black vertical line at zero in each of the plots. Being to the left of this line is preferable.

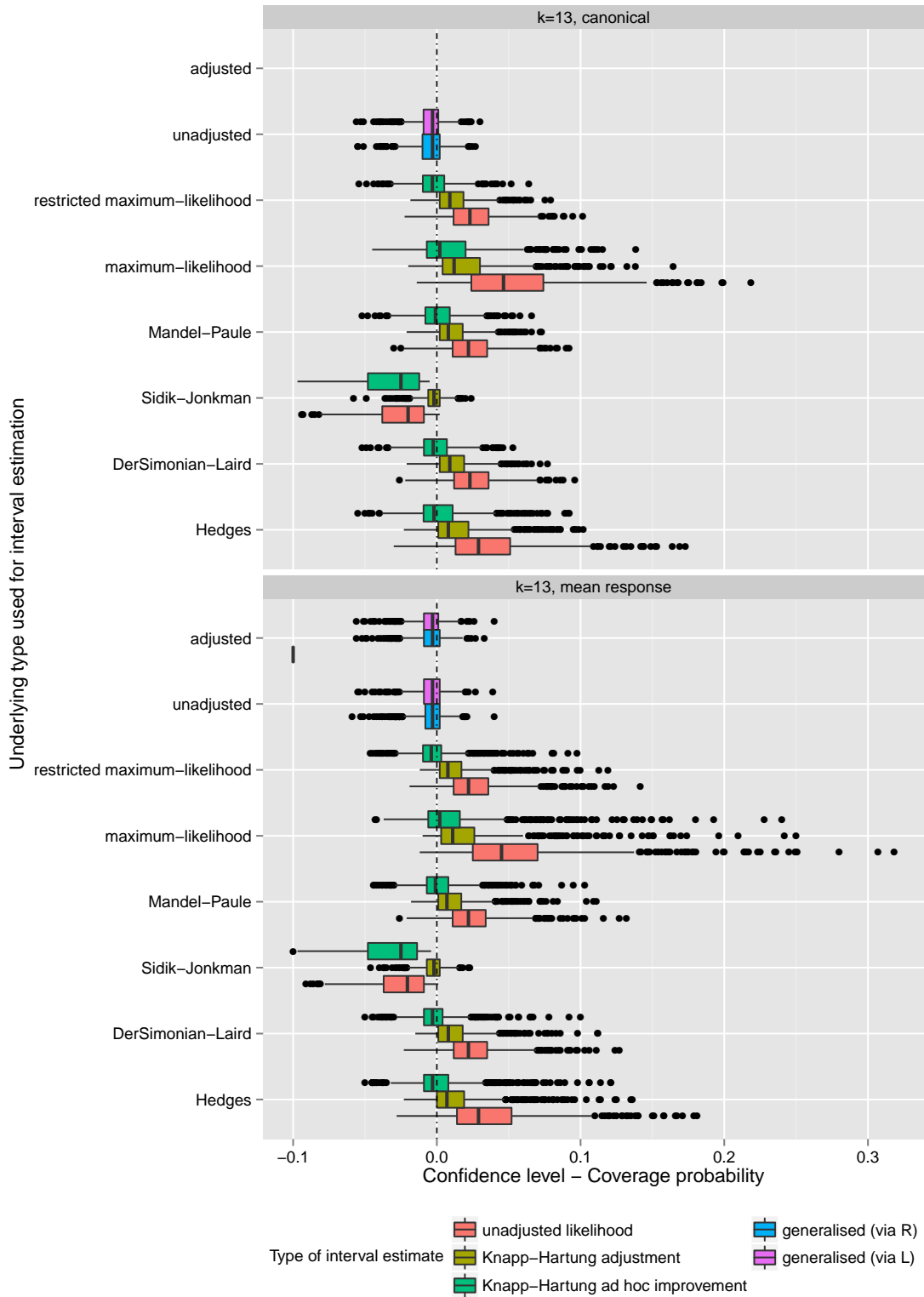


Figure 5.16 Boxplots of confidence level - coverage probability for different model parameters for $k = 13$ number of studies and different inferential methods shown for the slope parameter. The optimal line is the black vertical line at zero in each of the plots. Being to the left of this line is preferable.

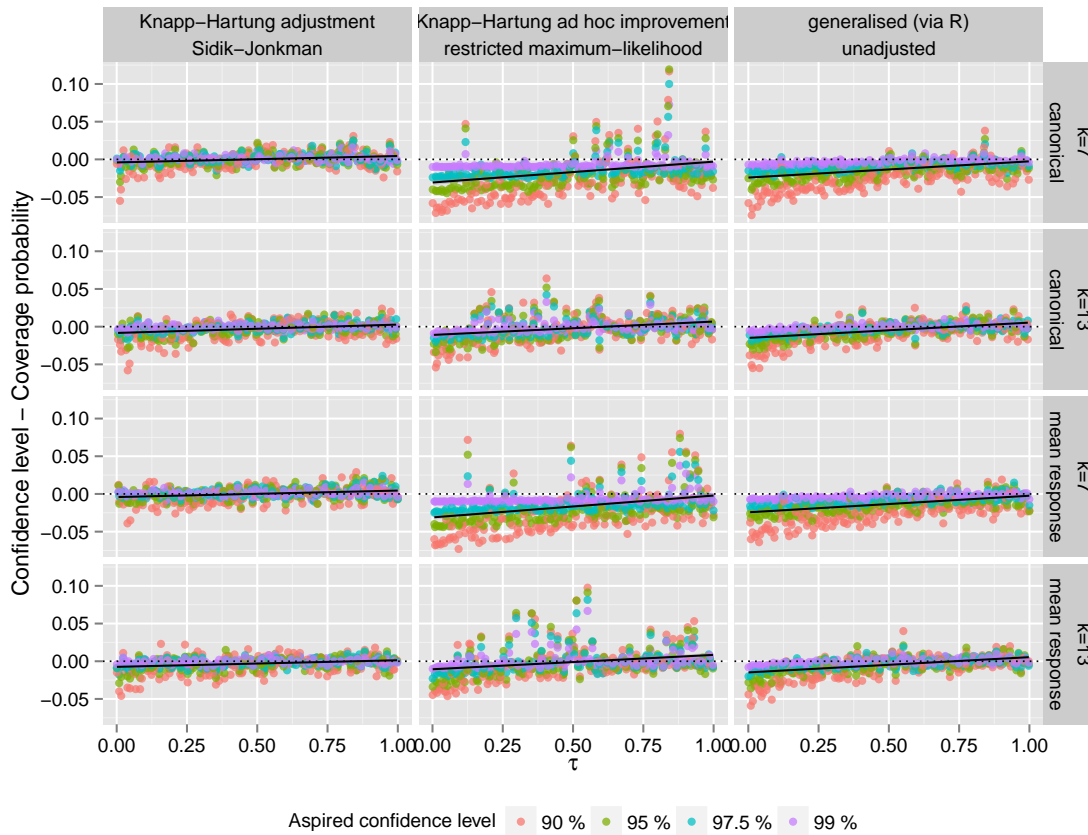


Figure 5.17 Scatter plots of estimated coverage of different interval estimators for the slope parameter for different model parameter configurations.

For the interval estimators based on generalised inferential principles, difference between the performance of the different pivotal distributions are not visible. As mentioned before, this was to be expected and is in line with Figure 5.1 and Figure 5.2. This suggests to use \mathfrak{R}_i when inference is based on generalised inference principles.

The performance of the interval estimate based on \mathfrak{R}_i is comparable with the performance of the Sidik–Jonkman–Knapp–Hartung adjusted interval estimate. This is also visible in the scatter plot of Figure 5.17 that plots the true heterogeneity τ and the performance measure **confidence-coverage**. For small magnitudes of heterogeneity, the generalised confidence interval shows more conservative behaviour. A direct comparison of these three interval estimators is shown in Figure 5.18 which shows kernel density estimates of the respected coverage distributions for the same interval estimators as in Figure 5.17. The density estimates suggest that the coverage of Sidik–Jonkman–Knapp–Hartung adjusted interval estimate lies tighter around its

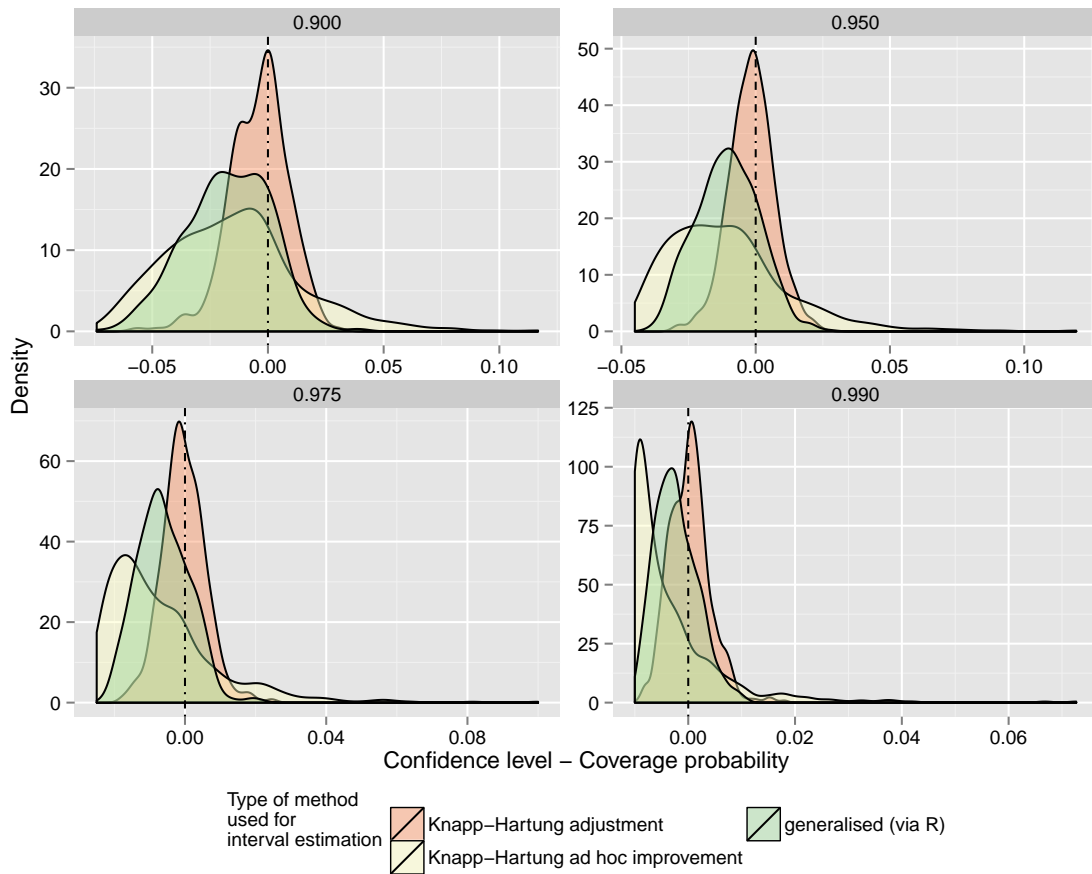


Figure 5.18 Density estimates of estimated coverage of different interval estimators for the slope parameter for different model parameter configurations.

aspired confidence level than the \mathfrak{R}_i -method. Both perform better than the so far believed state-of-the-art REML-Knapp–Hartung estimator. If it were not for the Sidik–Jonkman-Knapp–Hartung estimator, the new method based on generalised inference principles would have been the best performing interval estimate of all. The plots have also shown that the ad hoc truncation of

$$\frac{q_y(\hat{\tau})}{k - p}, \quad (5.68)$$

which was suggested by Knapp and Hartung (2003), seems to be making their original interval estimator unnecessarily conservative.

Last but not least, Figure 5.19 shows box plots of the estimated mean interval width for the same interval estimators as in Figures 5.17 and Figure 5.18. The mean interval widths are on comparable scales. Interestingly, even though

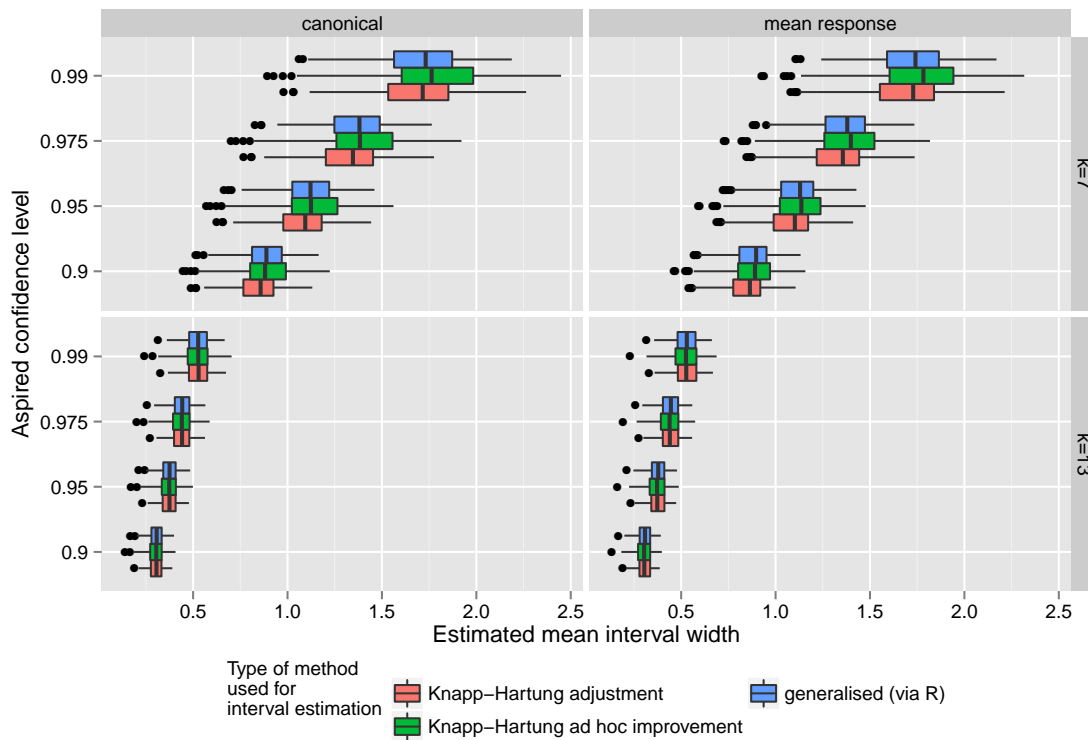


Figure 5.19 Box plots of estimated average width of different interval estimators for the slope parameter for different model parameter configurations.

the Knapp–Hartung ad hoc improved interval estimate in conjunction with the restricted maximum likelihood estimator for τ shows less conservative behaviour than the Sidik–Jonkman–Knapp–Hartung adjusted and the generalised- \mathfrak{R}_i interval estimate, its estimated mean interval widths lie above the latter two.

When using the classic and non-adjusted maximum likelihood approach for inference on the regression coefficients in the random effect meta regression model, e.g., in case no other methods are implemented in the software package used for the statistical analysis, the performance study suggests to use this estimator in conjunction with the Sidik–Jonkman type estimator for the heterogeneity parameter. If adjustments such as the Knapp–Hartung adjustment are available, it is strongly suggested to use this method. It is also suggested to use this method in conjunction with the Sidik–Jonkman type estimator as a *default* method. The Sidik–Jonkman type estimator will lead to comparably high accuracy and precision of the aspired coverage probability. When the Knapp–Hartung ad hoc improvement is implemented in the statistical software, the simulation study suggests to use the conditional restricted maximum likelihood method in conjunction with this estimator.

The surprise of the performance study is that the method that was so-far believed to be the current state-of-the-art, namely the Knapp–Hartung ad hoc improvement, is outperformed by its counterpart without the ad hoc truncation. Thus, two messages can be concluded from the performance study. The new presented method based on generalised inference principles was able to outperform this current state-of-the-art. Second, the performance study gives a clear suggestion to the actual state-of-the-art method, namely the Sidik–Jonkman–Knapp–Hartung adjusted interval estimator without truncation. This is new and unexpected.

5.7 *Example:* Regressing BCG vaccine efficacy on geographic location

In the BCG vaccine efficacy example of Section 1.2, the logarithm of the relative risk is used as the effect size in the statistical analysis in order to study whether the geographic distance of a clinic to the equator has a linear influence on this effect. The absolute latitude of the geographic location of a clinical trial is used as a surrogate for this distance. The effect size, i.e., the logarithm of the relative risk, converges in distribution to a Gaussian.

Point estimates for the	Intercept	Slope
Hedges	0.203	-0.0282
DerSimonian–Laird	0.260	-0.0292
Sidik–Jonkman	0.198	-0.0281
Mandel–Paule	0.222	-0.0286
maximum likelihood	0.282	-0.0295
restricted maximum likelihood	0.251	-0.0291
generalised (via L) unadjusted	0.251	-0.0283
generalised (via R) unadjusted	0.253	-0.0289
generalised (via L) adjusted	0.239	-0.0283
generalised (via R) adjusted	0.239	-0.0287

Table 5.1 Point estimates of the intercept and slope parameters in the BCG vaccine efficacy example

Table 5.1 shows point estimates of the regression coefficients, the intercept, and the slope for different point estimators. The point estimate $\hat{B}y$ in conjunction with the DerSimonian–Laird type estimator in conjunction with the restricted maximum likelihood estimator and the generalised (unadjusted) point estimator based on \mathfrak{R}

show very similar estimates for the slope parameter. The performance study suggests to trust the Hedge type estimate.

The interval estimates for the intercept and slope parameter for the fitted random effects meta regression model are shown in Table 5.2 and Table 5.3 for different confidence levels.

Excursion 5.10 (Regression coefficients of BCG vaccine efficacy trials)

All methods which performances have been compared by the computer experiment can be accessed via the `metagen` package. The following will run all inferential methods on the BCG vaccine efficacy data set.

```
library(metagen)
bcg <- bcgVaccineData()
y <- bcg$logrisk
d <- bcg$sdiv
s <- bcg$size
x <- cbind(1, bcg$x)

set.seed(865287113)
estimation <- metagen( y=y, d=d, x=x
                      , sgnf=sgnf, s=s
                      , n=1000, adjusted=T)
```

Here, $n = 1000$ defines the number of draws taken from the generalised pivotal distributions respectively. The list of point estimates can be accessed via

```
# point estimate of the regression coefficients
estimation$pointr
```

The form of this data frame is suitable for further automated processing by R. If you are interested in a more human readable table, you can reshape the data frame into wide format by the following code.

```
# point estimates of the regression coefficients
library(reshape2)
pointr <- reshape( data = estimation$pointr
                  , v.names = "value"
                  , idvar = "type"
                  , timevar = "parameter"
                  , direction = "wide")

library(plyr)
pointr <- rename(pointr, c("value.1"="intercept", "value.2"="slope"))
pointr
```

type	method	confidence	lower	upper
unadjusted	generalised (via R)	0.990	-0.826	1.120
unadjusted	generalised (via R)	0.975	-0.705	1.019
unadjusted	generalised (via R)	0.950	-0.545	0.896
unadjusted	generalised (via R)	0.900	-0.370	0.763
Sidik-Jonkman	Knapp-Hartung adjustment	0.990	-0.866	1.263
Sidik-Jonkman	Knapp-Hartung adjustment	0.975	-0.691	1.087
Sidik-Jonkman	Knapp-Hartung adjustment	0.950	-0.556	0.953
Sidik-Jonkman	Knapp-Hartung adjustment	0.900	-0.417	0.814
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.990	-0.630	1.133
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.975	-0.485	0.988
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.950	-0.373	0.876
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.900	-0.258	0.761

Table 5.2 Interval estimates of the intercept parameter in the BCG vaccine efficacy example.

type	method	confidence	lower	upper
unadjusted	generalised (via R)	0.990	-0.0545	0.00290
unadjusted	generalised (via R)	0.975	-0.0513	-0.00385
unadjusted	generalised (via R)	0.950	-0.0475	-0.00657
unadjusted	generalised (via R)	0.900	-0.0447	-0.01108
Sidik-Jonkman	Knapp-Hartung adjustment	0.990	-0.0582	0.00210
Sidik-Jonkman	Knapp-Hartung adjustment	0.975	-0.0533	-0.00288
Sidik-Jonkman	Knapp-Hartung adjustment	0.950	-0.0495	-0.00669
Sidik-Jonkman	Knapp-Hartung adjustment	0.900	-0.0455	-0.01063
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.990	-0.0546	-0.00363
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.975	-0.0504	-0.00783
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.950	-0.0472	-0.01105
restricted maximum-likelihood	Knapp-Hartung ad hoc improvement	0.900	-0.0438	-0.01437

Table 5.3 Interval estimates of the slope parameter in the BCG vaccine efficacy example.

As a small gimmick, consider the following. Recall that the Knapp–Hartung adjustment and the Knapp–Hartung ad hoc improvement will yield the same result whenever $q_y(\hat{\tau}) \geq 11 = k - p$. The following code will print the names of the point estimates for τ for which these two interval estimates will yield the same results.

```
qfunction <- qfunc(y, d, x)
with(estimation$pointh, type[qfunction(h) >= 11])
# [1] DerSimonian-Laird Mandel-Paul maximum-likelihood restricted maximum-likelihood
```

Graphical representations of the interval estimates can be produced by

```
plotCoefficientInterval(subset(confr, parameter == 1))
plotCoefficientInterval(subset(confr, parameter == 2))
```

Section 1.2 argued that the absolute latitude may be linked to a level of natural immunity against tuberculosis within the population of a clinical trial. Formally, this translates into the testing problem

$$H_0 : \beta_2 \geq 0 \quad \text{vs.} \quad H_1 : \beta_2 < 0. \quad (5.69)$$

At an aspired confidence level of 0.01, this hypothesis is failed to be rejected when using the generalised method based on \mathfrak{R}_2 and when using the Sidik–Jonkman–Knapp–Hartung adjusted method for the interval estimation but not when using the REML–Knapp–Hartung ad hoc improvement.

6 Discussion

The text has developed new inferential methods for the random effects meta regression model and the random effects meta analysis model. In particular, new point and interval estimates were proposed for the heterogeneity parameter and the regression coefficients of the respected models based on the generalised inference principle by Tsui and Weerahandi (1989). These methods are genuine. Preliminary results, though, for the regression coefficients have been published prior to this dissertation in (Friedrich and Knapp, 2013).

Current point and interval estimates commonly found in the literature were presented in the text in a unified framework. For multiple point estimates of the heterogeneity parameter which are based on the method of moments approach, this unifying framework allowed the formulation of new iterative procedures. The convergence of these methods has not been established yet, and the presented approach should yield a fruitful connection point for further theoretical studies. It would be very interesting to see, if the resulting sequences of estimates always converge, and if the resulting limits equal already otherwise known estimates.

A concise introduction to the generalised inference principle was given. Additionally, an extension of the principle to higher dimensions was suggested which is original and has not been published yet.

All discussed methods, point and interval estimates, new and known methods, have been evaluated by two extensive simulation studies. The specific set-up of these simulation studies is genuine in this context. The set-up and the results of these studies have been presented and discussed thoroughly.

A software package called **metagen** was written for the statistical software environment R, (R Development Core Team, 2010). The package implements all discussed methods and can, thus, be used as a statistical software tool for analysing data in the meta analysis or meta regression framework. The package also allows to reproduce the simulation studies and all results presented in this text. It also allows to set up custom simulation studies including new methods which have not been coded into the base package of **metagen**. It, thus, establishes a tool box for the field in which a variety of different method can be compared in a standardised way.

Hannig (2009) argues that the generalised inference principle and the fiducial inference recipe by Fisher (1935) are not too far from each other. He underwent considerable work in producing first sketches of a unifying theory. This may also open the possibility to compare method of generalised inference with Bayesian methods on a theoretical level. Hannig (2009) points out that in some special cases Bayesian methods in conjunction with Jeffreys prior result in statistics that would otherwise also be found when applying the generalised inference principle. This conjecture

has not been applied to the random effects meta regression model yet and should be worth to be investigated.

So far, the generalised inference principle has only been defined for continuous random variables, and, indeed, is it unclear if the principle can be extended to discrete random variables.

A very different approach would be to exploit ideas by Fraser (1982) and his notion of a *structural model* and an implied *invariant group structure* that can be pulled back to the parameter space. His approach to statistical modelling resulted in quite some discussion, see (Dawid and Stone, 1982a), followed by (Barnard, 1982), and again (Dawid and Stone, 1982b). Fraser's idea of using group structures on the sample space as a guideline on how to construct pivotal quantities was pick up by Weerahandi (1995) who applied it to the generalised inference principle. These results, however, did not find their way into the unification approach by Hannig (2009). The interconnections between these theoretical approaches should be interesting to study.

The text has also defined point estimates on the basis of the distribution of generalised pivotal quantities. The first paper that can be found on this topic is (Weerahandi, 2012). As mentioned, different measures of the generalised pivotal distribution may be used for the construction such as the mean or the mode of these distributions. How these measures perform in practical examples would be a typical application for which the **metagen** package could be used.

The correlation structure of $\mathbb{V}(\hat{\beta})$ for an estimator $\hat{\beta}$ of the regression coefficients vector of a random effects meta regression model may influence on the convergence rate of point and interval estimates. This has not been studied yet. Though the **metagen** package could be used to produce simulation results, it would be more interesting to see if results concerning the convergence rate could also be established on a theoretical level.

Another correlation structure that is worth investigating is whether an existing correlation dependence between the study responses and their respected variance estimates has an influence on the performance of the presented inferential methods. This is, for example, the case for the discussed binomial-Gaussian hierarchical model. In (Friedrich and Knapp, 2013), this had a positive effect on the performance of the generalised inference method in comparison to the Knapp–Hartung adjusted method. Whether this is generally the case, however, is not clear.

Appendix: Software documentation

Package ‘metagen’

May 9, 2014

Version 1.0

Date 2014-05

Title Inference in Meta Analysis and Meta Regression

Description Provides methods for making inference in the random effects meta regression model such as point estimates and confidence intervals for the heterogeneity parameter and the regression coefficients vector. Inference methods are based on different approaches to statistical inference. Methods from three different schools are included: methods based on the method of moments approach, methods based on likelihood, and methods based on generalised inference. The package also includes tools to run extensive simulation studies in parallel on high performance clusters in a modular way. This allows extensive testing of custom inferential methods with all implemented state-of-the-art methods in a standardised way. Tools for evaluating the performance of both point and interval estimates are provided. Also, a large collection of different pre-defined plotting functions is implemented in a ready-to-use fashion.

Author Thomas W. D. Möbius <kontakt@thomasmoebius.de>

Maintainer Thomas W. D. Möbius <kontakt@thomasmoebius.de>

URL <http://00tau.github.io/metagen/>

License GPL-3

Imports MASS, lhs, plyr, BBmisc, ParamHelpers, BatchJobs, BatchExperiments, ggplot2, metafor

LazyData yes

ByteCompile yes

Encoding UTF-8

Collate 'example-bcg.R' 'methods-for-batch-distribution.R' 'methods-for-data-generation.R' 'methods-for-experiments.R' 'methods-for-inference.R' 'plot-functions.R' 'plot-interval-estimates.R' 'plot-pivotal-distributions.R' 'plot-point-estimates.R' 'zzz.R'

2

R topics documented:

R topics documented:

bcgVaccineData	3
boxBias	4
boxByConfidence	4
boxByMethod	5
boxByType	5
boxMSE	6
boxSD	6
cbbPalette	7
cbgPalette	7
collectAllExperiments	8
collectExperiments	8
designB	9
designD	10
designY	11
dvec	12
experimentD	12
experimentY	13
formulaL	14
formulaR	15
hConfidence	17
hEstimates	17
intervalEstimates	18
joinPivotalCoefficients	19
joinPivotalHeterogeneity	20
lenBoxByMethod	21
lenBoxByType	21
lenDenByMethod	22
lenDenByType	22
makeConfInt	23
makeConfInts	23
metagen	24
metagenEmpty	25
metagenGeneralised	25
metareg	27
performance	28
performanceConfH	29
performanceConfR	29
performancePointH	30
performancePointR	30
pfunc	31
pivotalStream	32
plotCoefficientInterval	33
plotDensityH	33
plotDensityH2	34
plotDensityIntercept	34
plotDensityIntercept2	35

<i>bcgVaccineData</i>	3
plotDensitySlope	36
plotDensitySlope2	36
plotHeterogeneityInterval	37
plotIntervalEstimates	37
plotStudyForest	38
plotStudyQfuncPfunc	39
plotStudySizes	39
plotStudyUnbalance	40
qfunc	40
rB	41
rBinomGauss	42
rD	43
regressionEstimates	44
render	44
renderSVG	45
rY	46
sctBias	46
sctMSE	47
sctSD	47
sctVersusC	48
sctVersusH	48
sdmByMethod	49
sdmByType	49
sdsByMethod	50
sdsByType	50
setupExperiment	51
yvec	52
Index	53

<i>bcgVaccineData</i>	<i>Example: Setting up the BCG-data set</i>
-----------------------	---

Description

Exemplary data set of 14 clinical trials evaluating BCG vaccine efficacy.

Usage

```
bcgVaccineData(sgnf = 0.05)
```

Arguments

`sgnf` significance level of the confidence intervals for the relative risks.

4

*boxByConfidence***Details**

Reads in the BCG vaccine efficacy data from the metafor package and adds some statistics to the data such as the log-relative risk, study size, measurements of balance, confidence intervals of the responses, and the like.

Value

Returns a data set of 13 clinical trials which evaluated the efficacy of the BCG vaccine. The data set is an exact copy of the data set found in the `dat.bcg` data frame provided by the metafor package.

Examples

```
bcgVaccineData()
```

boxBias
Plotting performance: Box plots for bias

Description

Box plots for the bias.

Usage

```
boxBias(res)
```

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

boxByConfidence
Plotting performance: Box plots for target value confidence-coverage

Description

Plotting performance: Box plots for target value confidence-coverage

Usage

```
boxByConfidence(res)
```

boxByMethod

5

Arguments

res The collected results from a computer experiment.

Value

A plot object.

boxByMethod *Plotting performance: Box plots for target value confidence-coverage*

Description

Plotting performance: Box plots for target value confidence-coverage

Usage

`boxByMethod(res)`

Arguments

res The collected results from a computer experiment.

Value

A plot object.

boxByType *Plotting performance: Box plots for target value confidence-coverage*

Description

Plotting performance: Box plots for target value confidence-coverage

Usage

`boxByType(res)`

Arguments

res The collected results from a computer experiment.

Value

A plot object.

6

boxSD

boxMSE	<i>Plotting performance: Box plots for mean squared error</i>
--------	---

Description

Box plots for mean squared error.

Usage

boxMSE(res)

Arguments

res The collected results from a computer experiment.

Value

A plot object.

boxSD	<i>Plotting performance: Box plots for standard deviation</i>
-------	---

Description

Box plots for standard deviation.

Usage

boxSD(res)

Arguments

res The collected results from a computer experiment.

Value

A plot object.

`cbbPalette`

7

<code>cbbPalette</code>	<i>Colour palettes for colour blind people</i>
-------------------------	--

Description

The palette with black.

Usage

```
cbbPalette
```

Format

```
chr [1:8] "#000000" "#E69F00" "#56B4E9" "#009E73" "#F0E442" ...
```

Details

This palette is directly taken from
[http://www.cookbook-r.com/Graphs/Colors_\(ggplot2\)/](http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/)
 Hence, I don't take any credit for this.

Examples

```
scale_fill_discrete <- function(...) scale_fill_manual(...,
  values=cbbPalette)
scale_colour_discrete <- function(...) scale_fill_manual(...,
  values=cbbPalette)
```

<code>cbgPalette</code>	<i>Colour palettes for colour blind people</i>
-------------------------	--

Description

The palette with grey.

Usage

```
cbgPalette
```

Format

```
chr [1:8] "#999999" "#E69F00" "#56B4E9" "#009E73" "#F0E442" ...
```

8

*collectExperiments***Details**

This palette is directly taken from
[http://www.cookbook-r.com/Graphs/Colors_\(ggplot2\)/](http://www.cookbook-r.com/Graphs/Colors_(ggplot2)/)
 Hence, I don't take any credit for this.

Examples

```
scale_fill_discrete <- function(...) scale_fill_manual(...,
  values=cbgPalette)
scale_colour_discrete <- function(...) scale_fill_manual(...,
  values=cbgPalette)
```

collectAllExperiments *Running a computer experiment – Collect all the results*

Description

Collects all the results of all finished experiments in the given registry for all predefined algorithms.

Usage

```
collectAllExperiments(reg)
```

Arguments

`reg` A valid registry generated by 'makeExperimentRegistry'.

Value

List of data frames containing the performance measures of all point and interval estimates for the heterogeneity and the regression coefficients.

collectExperiments *Running a computer experiment – Collect specific results*

Description

Collects specific results of all finished experiments in the given registry for a given pattern.

Usage

```
collectExperiments(reg, pattern)
```

designB

9

Arguments

<code>reg</code>	A valid registry generated by 'makeExperimentRegistry'.
<code>pattern</code>	string containing the algorithm pattern for which the collection shall be performed.

Value

List of data frames containing the performance measures of all point and interval estimates for the heterogeneity and the regression coefficients.

<code>designB</code>	<i>Design: Binomial responses</i>
----------------------	-----------------------------------

Description

Method for generating a sampling design for data generation following a binomial-Gaussian model.

Usage

```
designB(n, h_bounds, a_bounds, s_bounds, r, x)
```

Arguments

<code>n</code>	resolution of the heterogeneity. <code>n</code> is the number of of different heterogeneity parameters in the design.
<code>h_bounds</code>	bounds of the heterogeneity.
<code>a_bounds</code>	bounds of the balancing factor of group assignments.
<code>s_bounds</code>	bounds of the study sizes.
<code>r</code>	fixed risk in the control.
<code>x</code>	design matrix.

Details

Generates a sampling design for the heterogeneity 'h', balancing factors 'a1', ..., 'ak' of group assignments, and study sizes 's1', ..., 'sk'. This design can be used for testing methods for inference for the random effects meta regression model since the logarithm of relative risks of each study is approximately Gaussian distributed. One may use methods that adjust for uncertainty in the heteroscedasticity estimates by additionally considering the size of the respected studies.

Points in the design are selected via a maxi-min hypercube sampling using the 'lhs' package in a predefined parameter cube.

Value

Function returns a data frame. Each line of this data frame can be an input to the function 'rB' which is used to sample data from such a design.

10

*designD***Examples**

```
dB <- designB(n=15L, h_bounds=c(0,1), a_bounds=c(-.3,3),
  s_bounds=c(200L,2000L), r=0.03, x=cbind(1,1:5))

if(!all(dim(dB) == c(15,2*dim(cbind(1,1:5))[1]+2))) {
  stop("Wrong dimension")
}
```

designD *Design: Gaussian responses (unknown heteroscedasticity)*

Description

Method for generating a sampling design for data generation following a random effects meta regression model with unknown heteroscedasticity.

Usage

```
designD(n, h_bounds, d_bounds, s_bounds, x)
```

Arguments

<code>n</code>	resolution of the heterogeneity and heteroscedasticity parameters, i.e., the number of of different (heterogeneity, heteroscedasticity, sizes) tuple in the design.
<code>h_bounds</code>	bounds of the heterogeneity.
<code>d_bounds</code>	bounds of the heteroscedasticity.
<code>s_bounds</code>	bounds of the study sizes.
<code>x</code>	design matrix.

Details

Generates a sampling design for the heterogeneity 'h', heteroscedasticity 'd1', ..., 'dk', and study sizes 's1', ..., 'sk'. This design can be used for testing methods that adjust for uncertainty in the heteroscedasticity estimates by additionally considering the size of the respected studies.

Points in the design are selected via a maxi-min hypercube sampling using the 'lhs' package in a predefined parameter cube.

Value

Function returns a data frame. Each line of this data frame can be an input to the function 'rD' which is used to sample data from such a design.

designY

11

Examples

```
dD <- designD(n=15L, h_bounds=c(0,1), d_bounds=c(0.01,2),
  s_bounds=c(200L,2000L), x=cbind(1,1:7))

if(!all(dim(dD) == c(15,2*dim(cbind(1,1:7))[1]+1))) {
  stop("Wrong dimension")
}
```

*designY**Design: Gaussian responses (known heteroscedasticity)***Description**

Method for generating a sampling design for data generation following a random effects meta regression model with known heteroscedasticity.

Usage

```
designY(n, h_bounds, d_bounds, x)
```

Arguments

<i>n</i>	resolution of the heterogeneity and heteroscedasticity parameters, i.e. the number of different (heterogeneity, heteroscedasticity) pairs in the design.
<i>h_bounds</i>	bounds of the heterogeneity.
<i>d_bounds</i>	bounds of the heteroscedasticity.
<i>x</i>	design matrix.

Details

Generates a sampling design for the heterogeneity 'h' and a heteroscedasticity 'd1', ..., 'dk'.

Points in the design are selected via a maxi-min hypercube sampling using the 'lhs' package in a predefined parameter cube.

Value

Function returns a data frame. Each line of this data frame can be an input to the function 'rY' which is used to sample data from such a design.

Examples

```
dY <- designY(n=15L, h_bounds=c(0,1), d_bounds=c(0.01,2),
  x=cbind(1,1:7))

if(!all(dim(dY) == c(15,dim(cbind(1,1:7))[1]+1))) {
  stop("Wrong dimension")
}
```

12

experimentD

dvec *Data generation: Sampling data of clinical trials*

Description

Calculates the variance estimate of log risk ratios from a study in the right format. See the example below for details.

Usage

```
dvec(study)
```

Arguments

study Study data of a clinical trial with binomial outcomes.

Examples

```
h_test <- .03
x_test <- cbind(1,1:13)
b_test <- c(0.02, 0.03)
s_test <- rep(2000, 13)
a_test <- rep(.3, 13)
rBinomGauss( h=h_test, s=s_test, a=a_test, r=0.03
             , x=x_test, b=b_test)$study -> test
yvec(test)
dvec(test)
```

experimentD *Running a computer experiment*

Description

Runs a computer experiment that evaluates the performance of different inference methods for the random effects meta regression model with respect to heterogeneity and regression coefficients.

Usage

```
experimentD(n, h, d, s, x, b, sgnf, piv_draws)
```

experimentY

13

Arguments

n number of draws.
h heterogeneity.
d heteroscedasticity.
s vector study sizes.
x design matrix.
b regression coefficients.
sgnf significance levels.
piv_draws pivotal draws.

Details

This also includes methods adjusting for uncertainty in the heteroscedasticity vector. In particular, the study sizes need to be known, here.

Value

Data frame of accumulated performance measures.

Examples

```
h_test <- 0.03
x_test <- cbind(1,1:7)
b_test <- c(.5, .25)
sgnf_test <- c(0.025, 0.01)

set.seed(5133568) # for reproducibility
d_test <- rchisq(7, df=0.02)
s_test <- runif(7, min=200, max=2000)

# In an actual computer experiment, use 'piv_draws=1000' instead!!
experimentD(n=5, h=h_test, d=d_test, s=s_test, x=x_test, b=b_test,
  sgnf=sgnf_test, piv_draws=50)
```

experimentY *Running a computer experiment*

Description

Runs a computer experiment that evaluates the performance of different inference methods for the random effects meta regression model with respect to heterogeneity and regression coefficients.

Usage

```
experimentY(n, h, d, x, b, sgnf, piv_draws)
```

14

*formulaL***Arguments**

n	number of draws.
h	heterogeneity.
d	heteroscedasticity.
x	design matrix.
b	regression coefficients.
sgnf	significance levels.
piv_draws	pivotal draws.

Value

Data frame of accumulated performance results.

Examples

```
h_test <- 0.03
x_test <- cbind(1,1:7)
b_test <- c(.5, .25)
sgnf_test <- c(0.025, 0.01)

set.seed(5133568) # for reproducibility
d_test <- rchisq(7, df=0.02)

# In an actual computer experiment, use 'piv_draws=1000' instead!!
experimentY(n=5, h=h_test, d=d_test, x=x_test, b=b_test,
  sgnf=sgnf_test, piv_draws=50)
```

formulaL

Regression coefficients: formulaL

Description

Calculate pivotal quantities for the regression coefficients using the method: *formulaL* from the dissertation.

Usage

```
formulaL(y, d, h, g, x)
```

Arguments

y	k-vector of responses.
d	k-vector of heteroscedasticity.
h	scalar of heterogeneity.
g	p-vector of some p-variate Gaussian draw.
x	design k-p-matrix.

formulaR

15

Details

Algorithm for calculating a single generalised pivotal quantity for the regression coefficients for given generalised pivotal quantities for the heterogeneity using the univariate version of the pivotal formula.

Value

A p-vector.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1,bcg$x)

# When for example using the Mandel-Paule estimate:
bcg_h <- pfunc(y=bcg_y, d=bcg_d, x=bcg_x)(dim(bcg_x)[1] -
  dim(bcg_x)[2])

set.seed(51351) # for reproducibility
random_g <- rnorm(dim(bcg_x)[2])
formulal(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=bcg_x)

# The function can also be used when planing to perform
# a meta regression with no intercept, and only a singel
# covariate (i.e. dim(x) = 1). In this case,
# the design matrix can simply be provided by a vector.
set.seed(51351) # for reproducibility
random_g <- rnorm(1)
formulal(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=bcg$x)

# When performing a meta analysis, provide the function
# with a vector of 1s.
formulal(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=rep(1,
  length(bcg_y)))
```

*formulaR**Regression coefficients: formulaR*

Description

Calculate pivotal quantities for the regression coefficients using the method: *formulaR* form the dissertation.

Usage

```
formulaR(y, d, h, g, x)
```

16

*formulaR***Arguments**

y k-vector of responses.
 d k-vector of heteroscedasticity.
 h scalar of heterogeneity.
 g p-vector of some p-variate Gaussian draw.
 x design k-p-matrix.

Details

Algorithm for calculating a single generalised pivotal quantity for the regression coefficients for given generalised pivotal quantities for the heterogeneity using the multivariate version of the pivotal formula.

Value

A p-vector.

Examples

```

bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)

# When, for example, using the Mandel-Paule estimate:
bcg_h <- pfunc(y=bcg_y, d=bcg_d, x=bcg_x)(dim(bcg_x)[1] -
  dim(bcg_x)[2])

set.seed(51351) # for reproducibility
random_g <- rnorm(dim(bcg_x)[2])
formulaR(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=bcg_x)

# The function can also be used when planing to perform
# a meta regression with no intercept, and only a single
# covariate (i.e. dim(x) = 1). In this case,
# the design matrix can simply be provided by a vector.
set.seed(51351) # for reproducibility
random_g <- rnorm(1)
formulaR(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=bcg$x)

# When performing a meta analysis, provide the function
# with a vector of 1s.
formulaR(y=bcg_y, d=bcg_d, h=bcg_h, g=random_g, x=rep(1,
  length(bcg_y)))

```

hConfidence

17

<i>hConfidence</i>	<i>Inference: Based on methods of moments and maximum likelihood.</i>
--------------------	---

Description

Calculates the so called Q-profiling confidence interval for the heterogeneity for data following a random effects meta regression model.

Usage

```
hConfidence(y, d, x, sgnf)
```

Arguments

y	k-vector of study responses.
d	k-vector of heteroscedasticity.
x	design k-p-matrix.
sgnf	significance levels.

Value

A data frame containing the bounds of the interval estimate.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)
sgnf_lev <- c(0.01, 0.025, 0.05, 0.01)

set.seed(865287113) # for reproducibility

hConfidence(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025)
hConfidence(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev)
```

<i>hEstimates</i>	<i>Point estimates: For the heterogeneity parameter</i>
-------------------	---

Description

Returns a list of tau estimates based on different approximative methods. Different point estimates for the heterogeneity parameter are calculated: HD (Hedges), SL (DerSimonian-Laird), SJ (Sidik-Jonkman), MP (Mandel-Paule), ML (maximum likelihood), REML (restricted maximum-likelihood). Since any of these methods may fail to converge, there result may be 'NA' in this case.

18

*intervalEstimates***Usage**

```
hEstimates(y, d, x)
```

Arguments

```

y          study responses
d          heteroscedasticity
x          design matrix

```

Value

A data frame containing point estimates. Variables are 'type' and 'h'.

Examples

```

bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)
hEstimates(y=bcg_y, d=bcg_d, x=bcg_x)

# The implementation can also handle the case in which
# a meta regression is planed with no intercept and only a
# single covariate (i.e. dim(x) = 1). In this case,
# the design matrix can simply be provided by a vector.
# (This makes no sense in this example and shall only prove
# feasibility)
hEstimates(y=bcg_y, d=bcg_d, x=bcg$x)

# When performing a meta analysis, provide the function
# with a vector of 1s.
hEstimates(y=bcg_y, d=bcg_d, x=rep(1, length(bcg_y)))

```

```
intervalEstimates      Interval estimates: For the regression coefficients
```

Description

Interval estimates: For the regression coefficients

Usage

```
intervalEstimates(y, d, h_dat, x, sgnf)
```

joinPivotalCoefficients

19

Arguments

`y` study responses.
`d` heteroscedasticity.
`h_dat` data frame of tau estimates.
`x` design matrix.
`sgnf` significance levels.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)
bcg_h <- hEstimates(y=bcg_y, d=bcg_d, x=bcg_x)
sgnf_lev <- c(0.01, 0.025, 0.05, 0.01)

intervalEstimates(y=bcg_y, d=bcg_d, h_dat=bcg_h, x=bcg_x, sgnf=0.025)
intervalEstimates(y=bcg_y, d=bcg_d, h_dat=bcg_h, x=bcg_x,
  sgnf=sgnf_lev)
```

`joinPivotalCoefficients`

Pivotal distributions: Extract pivots for regression coefficients

Description

Pivotal distributions: Extract pivots for regression coefficients

Usage

```
joinPivotalCoefficients(p0, p1)
```

Arguments

`p0` pivotal stream without adjustment.
`p1` pivotal stream with adjustment.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
```

20

joinPivotalHeterogeneity

```
adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
adjusted=TRUE)

pivr <- joinPivotalCoefficients(pivUn, pivAd)
```

joinPivotalHeterogeneity*Pivotal distributions: Extract pivots for heterogeneity*

Description

Pivotal distributions: Extract pivots for heterogeneity

Usage

```
joinPivotalHeterogeneity(p0 = NULL, p1 = NULL)
```

Arguments

p0	pivotal stream without adjustment.
p1	pivotal stream with adjustment.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$ssdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
adjusted=TRUE)

pivh <- joinPivotalHeterogeneity(pivUn, pivAd)
```

lenBoxByMethod

21

lenBoxByMethod *Plotting performance: Box plot of mean width*

Description

Plotting performance: Box plot of mean width

Usage

`lenBoxByMethod(res)`

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

lenBoxByType *Plotting performance: Box plot of mean width*

Description

Plotting performance: Box plot of mean width

Usage

`lenBoxByType(res)`

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

22

lenDenByType

<code>lenDenByMethod</code>	<i>Plotting performance: Density estimate of mean width</i>
-----------------------------	---

Description

By method.

Usage`lenDenByMethod(res)`**Arguments**`res` The collected results from a computer experiment.**Value**

A plot object.

<code>lenDenByType</code>	<i>Plotting performance: Density estimate of mean width</i>
---------------------------	---

Description

By type.

Usage`lenDenByType(res)`**Arguments**`res` The collected results from a computer experiment.**Value**

A plot object.

makeConfInt

23

<i>makeConfInt</i>	<i>Interval estimates: Generic function</i>
--------------------	---

Description

Generic function to produce interval estimates of univariate parameters based on first order limit theory.

Usage

```
makeConfInt(sgn, pst, fct, crt, name)
```

Arguments

sgn	one significance level.
pst	point estimate.
fct	standard error.
crt	function for critical value computation.
name	string: name of the method.

Details

Function for symmetric confidence intervals based on standard deviations, point estimates, and quantile functions.

Can only handle a single significance level! See 'makeConfInts' for a more flexible solution.

<i>makeConfInts</i>	<i>Interval estimates: Generic function</i>
---------------------	---

Description

Generic function to produce interval estimates of univariate parameters based on first order limit theory.

Usage

```
makeConfInts(sgn, pst, fct, crt, name)
```

Arguments

sgn	one significance level.
pst	point estimate.
fct	standard error.
crt	function for critical value computation.
name	string: name of the method.

24

*metagen***Details**

Function for symmetric confidence intervals based on standard deviations, point estimates, and quantile functions.

<code>metagen</code>	<i>Inference: Analysis of the data set</i>
----------------------	--

Description

Runs all implemented methods and combines them in a neat summary.

Usage

```
metagen(y, d, x, sgnf, s = NULL, n,
        method = list("univariate", "multivariate"),
        adjusted = FALSE)
```

Arguments

<code>y</code>	k-vector of responses.
<code>d</code>	k-vector of heteroscedasticities.
<code>x</code>	design k-p-matrix.
<code>sgnf</code>	vector of significance levels.
<code>s</code>	k-vector of study responses. Default is NULL. If 'adjusted=TRUE', this value needs to be given.
<code>n</code>	draws from the pivotal distribution.
<code>method</code>	Default is 'list("univariate", "multivariate")'.
<code>adjusted</code>	: TRUE or FALSE

Value

The same return type as the skeleton 'metagenEmpty()'.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)
sgnf_lev <- c(0.01, 0.025, 0.05, 0.01)

set.seed(865287113) # for reproducibility

# Runs a standard analysis, use n=1000 in an actual
# analysis instead!
m1 <- metagen(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025, n=50)
```

metagenEmpty

25

```

m2 <- metagen(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev, n=50)

# Runs the methods based on generalised principles via an
# adjustment for the unknown heteroscedasticity. Use
# n=1000 in an actual analysis instead!!
bcg_s <- bcg$size
m3 <- metagen(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025, s=bcg_s, n=50,
  adj=TRUE)
m4 <- metagen(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev, s=bcg_s,
  n=50, adj=TRUE)

if (!all(names(m1) == names(metagenEmpty())) stop("Name clash")
if (!all(names(m2) == names(metagenEmpty())) stop("Name clash")
if (!all(names(m3) == names(metagenEmpty())) stop("Name clash")
if (!all(names(m4) == names(metagenEmpty())) stop("Name clash")

```

*metagenEmpty**Inference: Empty skeleton*

Description

Returns an empty skeleton that has the same return type as any other 'metagenSOMETHING' function.

Usage

```
metagenEmpty()
```

Examples

```
metagenEmpty()
```

*metagenGeneralised**Inference: Based on generalised inference principles.*

Description

Inference: Based on generalised inference principles.

Usage

```
metagenGeneralised(y, d, x, sgnf, s = NULL, n,
  method = list("univariate", "multivariate"),
  adjusted = FALSE)
```

26

*metagenGeneralised***Arguments**

y k-vector of responses.
d k-vector of heteroscedasticities.
x design k-p-matrix.
sgnf vector of significance levels
s k-vector of study responses. No need to provide this, when 'adjusted==FALSE'.
 Default is NULL.
n draws from the pivotal distribution.
method Default is 'list("univariate", "multivariate")'.
adjusted TRUE or FALSE. Default is FALSE.

Examples

```

bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)
sgnf_lev <- c(0.01, 0.025, 0.05, 0.01)

set.seed(865287113) # for reproducibility

# Runs a standard analysis, use n=1000 in an actual
# analysis instead!!
g1 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025, n=50)
g2 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev,
  n=50)

# Runs the methods based on generalised principles via an
# adjustment for the unknown heteroscedasticity. Use n=1000 in an
# actual analysis instead!!
bcg_s <- bcg$size
g3 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025,
  s=bcg_s, n=50, adj=TRUE)
g4 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev,
  s=bcg_s, n=50, adj=TRUE)

# The implementation can also handle the case in which
# a meta regression is planned with no intercept and only a
# single covariate (i.e. dim(x) = 1). In this case,
# the design matrix can simply be provided by a vector.
# (This makes no sense in this example and shall only prove
# feasibility)
g5 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=bcg$x, sgnf=0.025, n=50)

# When performing a meta analysis, provide the function
# with a vector of 1s.
g6 <- metagenGeneralised(y=bcg_y, d=bcg_d, x=rep(1, length(bcg_y)),
  sgnf=0.025, n=50)

```

metareg

27

```

if (!all(names(g1) == names(metagenEmpty())) stop("Name clash")
if (!all(names(g2) == names(metagenEmpty())) stop("Name clash")
if (!all(names(g3) == names(metagenEmpty())) stop("Name clash")
if (!all(names(g4) == names(metagenEmpty())) stop("Name clash")
if (!all(names(g5) == names(metagenEmpty())) stop("Name clash")
if (!all(names(g6) == names(metagenEmpty())) stop("Name clash")

```

*metareg**Inference: Based on methods of moments and maximum likelihood.***Description**

Calculates common statistics for point and confidence interval estimates for the heterogeneity and the regression coefficients of the random effects meta regression model based on the given data.

Usage

```
metareg(y, d, x, sgnf)
```

Arguments

<i>y</i>	k-vector of study responses.
<i>d</i>	k-vector of heteroscedasticity.
<i>x</i>	design k-p-matrix.
<i>sgnf</i>	significance levels.

Value

The same return type as the skeleton 'metagenEmpty'.

Examples

```

bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)
sgnf_lev <- c(0.01, 0.025, 0.05, 0.01)

set.seed(865287113) # for reproducibility

c1 <- metareg(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=0.025)
c2 <- metareg(y=bcg_y, d=bcg_d, x=bcg_x, sgnf=sgnf_lev)

# When performing a meta analysis, provide the function
# with a vector of 1s.
if (!all(names(c1) == names(metagenEmpty())) stop("Name clash")
if (!all(names(c2) == names(metagenEmpty())) stop("Name clash")

```

28

performance

 performance *Running a computer experiment*

Description

Adding performance measures to the results

Usage

```
performance(results, b, h)
```

Arguments

results	Needs to be of the same type as, for example, the return value of the computer experiments 'experimentY', 'experimentD'.
b	true regression coefficients.
h	true heterogeneity.

Details

Calculating performance measurements from a computer experiment.

Value

Data frame containing performance measurements of inference methods based on the results of the computer experiment given by 'results'.

Examples

```
h_test <- 0.03
x_test <- cbind(1,1:7)
b_test <- c(.5, .25)
sgnf_test <- c(0.025, 0.01)

set.seed(5133568) # for reproducibility
d_test <- rchisq(7, df=0.02)

# In an actual computer experiment, use 'piv_draws=1000' instead!!
eY <- experimentY(n=5, h=h_test, d=d_test, x=x_test, b=b_test,
  sgnf=sgnf_test, piv_draws=50)

performance(results=eY, b=b_test, h=h_test)
```

performanceConfH

29

performanceConfH *Running a computer experiment: Adding performance measures*

Description

Adding performance measurements to accumulated results of a computer experiment running multiple analysis of different simulated data following a random effects meta regression model.

Usage

```
performanceConfH(accum_int, true)
```

Arguments

`accum_int` accumulated interval estimates. At least the following columns need to be present: lower and upper.
`true` true parameter.

Details

Adds performance measurements to interval estimates of the heterogeneity.

Examples

```
# For an example, see the 'performance' function.
```

performanceConfR *Running a computer experiment: Adding performance measures*

Description

Adding performance measurements to accumulated results of a computer experiment running multiple analysis of different simulated data following a random effects meta regression model.

Usage

```
performanceConfR(accum_int, true)
```

Arguments

`accum_int` accumulated interval estimates. At least the following columns need to be present: lower and upper and parameter.
`true` true parameter.

30

*performancePointR***Details**

Adds performance measurements to interval estimates of the regression coefficients.

Examples

For an example, see the 'performance' function.

`performancePointH` *Running a computer experiment: Adding performance measures*

Description

Adding performance measurements to accumulated results of a computer experiment running multiple analysis of different simulated data following a random effects meta regression model.

Usage

```
performancePointH(point, h)
```

Arguments

`point` accumulated point estimates.
`h` true parameter.

Details

Adds performance measurements to point estimates of the heterogeneity.

Examples

For an example, see the 'performance' function.

`performancePointR` *Running a computer experiment: Adding performance measures*

Description

Adding performance measurements to accumulated results of a computer experiment running multiple analysis of different simulated data following a random effects meta regression model.

Usage

```
performancePointR(point, b)
```


32

pivotalStream

pivotalStream *Streams of pivotal quantities of the regression coefficient*

Description

Algorithm for generating a stream of generalised pivotal quantities for the regression coefficients. If `adjusted=FALSE`, then no adjustments are made for the uncertainty in the heteroscedasticity estimates `d`. If `adjusted=TRUE`, then adjustments are performed. In this case, 's' needs to be provided.

Usage

```
pivotalStream(n, y, d, x, s = NULL,
              method = list("univariate", "multivariate"), adjusted)
```

Arguments

<code>n</code>	length of stream.
<code>y</code>	k-vector of responses.
<code>d</code>	k-vector of heteroscedasticity.
<code>x</code>	design (k,p)-matrix.
<code>s</code>	k-vector of study responses. No need to provide this, when <code>adjusted=FALSE</code> . Default is <code>NULL</code> .
<code>method</code>	A list. Used to choose the methods for calculating the pivotal quantities of the regression coefficients. Default is <code>'method=list("univariate", "multivariate")'</code> .
<code>adjusted</code>	TRUE or FALSE. Default is <code>FALSE</code> .

Value

If `method=="univariate"` or `method=="multivariate"`, then the return is a (p+1)-n-matrix. The first row contains pivotal quantities of the heterogeneity, the rest of the rows pivotal quantities of the regression coefficients. Each column is an independent draw.

If `'method==list("univariate", "multivariate")'`, then the return is a (2p+1)-n-matrix. Of each column, the first element is a pivotal for the heterogeneity, the next 'p' elements is a pivotal vector for the regression coefficients based on "univariate", the last 'p' elements are a pivotal vector for the regression coefficients based on "multivariate"

plotCoefficientInterval

33

`plotCoefficientInterval`

Plot pivots: Interval estimates of the heterogeneity

Description

Plot pivots: Interval estimates of the heterogeneity

Usage

```
plotCoefficientInterval(cnfr)
```

Arguments

`cnfr` interval estimates of the heterogeneity.

`plotDensityH`

Pivotal distributions: Plot pivotal distribution of heterogeneity

Description

Pivotal distributions: Plot pivotal distribution of heterogeneity

Usage

```
plotDensityH(pivh)
```

Arguments

`pivh` pivotal stream with or without adjustment of independent draws of a pivotal quantity of the heterogeneity.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1,bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
  adjusted=TRUE)

pivh <- joinPivotalHeterogeneity(pivUn, pivAd)
plotDensityH(pivh)
```

34

plotDensityIntercept

`plotDensityH2` *Pivotal distributions: Plot pivot density of the heterogeneity*

Description

Pivotal distributions: Plot pivot density of the heterogeneity

Usage

```
plotDensityH2(pivh)
```

Arguments

`pivh` pivotal stream with or without adjustment of independent draws of a pivotal quantity of the heterogeneity.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
pivh <- joinPivotalHeterogeneity(pivUn)
plotDensityH2(pivh)
```

`plotDensityIntercept` *Pivotal distributions: Plot pivotal distribution of regression coefficients*

Description

Pivotal distributions: Plot pivotal distribution of regression coefficients

Usage

```
plotDensityIntercept(pivr)
```

Arguments

`pivr` data frame of independent draws from of pivots.

plotDensityIntercept2

35

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1,bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
  adjusted=TRUE)

pivr <- joinPivotalCoefficients(pivUn, pivAd)
plotDensityIntercept2(pivr)
```

plotDensityIntercept2 *Pivotal distributions: Plot pivotal distribution of regression coefficients*

Description

Pivotal distributions: Plot pivotal distribution of regression coefficients

Usage

```
plotDensityIntercept2(pivr)
```

Arguments

pivr data frame of independent draws from of pivots.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1,bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
  adjusted=TRUE)

pivr <- joinPivotalCoefficients(pivUn, pivAd)
plotDensityIntercept2(pivr)
```

36

plotDensitySlope2

`plotDensitySlope` *Pivotal distributions: Plot pivotal distribution of regression coefficients*

Description

Pivotal distributions: Plot pivotal distribution of regression coefficients

Usage

```
plotDensitySlope(pivr)
```

Arguments

`pivr` data frame of independent draws from of pivots.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$ssize
bcg_x <- cbind(1,bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
  adjusted=TRUE)

pivr <- joinPivotalCoefficients(pivUn, pivAd)
plotDensitySlope(pivr)
```

`plotDensitySlope2` *Pivotal distributions: Plot pivotal distribution of regression coefficients*

Description

Pivotal distributions: Plot pivotal distribution of regression coefficients

Usage

```
plotDensitySlope2(pivr)
```

plotHeterogeneityInterval

37

Arguments

`pivr` data frame of independent draws from of pivots.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)

set.seed(865287113)
pivUn <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x,
  adjusted=FALSE)
set.seed(865287113)
pivAd <- pivotalStream(50, y=bcg_y, d=bcg_d, x=bcg_x, s=bcg_s,
  adjusted=TRUE)

pivr <- joinPivotalCoefficients(pivUn, pivAd)
plotDensitySlope2(pivr)
```

`plotHeterogeneityInterval`

Plot pivots: Interval estimates of the heterogeneity

Description

Plot pivots: Interval estimates of the heterogeneity

Usage

```
plotHeterogeneityInterval(cnfh)
```

Arguments

`cnfh` interval estimates of the heterogeneity.

`plotIntervalEstimates` *Example: Plotting interval estimates*

Description

Plots a graphical representation of interval estimates in the data frame 'cnf' by type of method used for the estimation.

38

*plotStudyForest***Usage**`plotIntervalEstimates(cnf)`**Arguments**`cnf` data frame of interval estimates**Value**An object created by `ggplot2`.

`plotStudyForest` *Example: Plotting a forest plot of a data frame*

Description

Example: Plotting a forest plot of a data frame

Usage`plotStudyForest(dat)`**Arguments**`dat` data frame of study responses of binomial type.**Value**An object created by `ggplot2`.**Examples**

```
bcg <- bcgVaccineData()
plotStudyForest(bcg)
```


`plotStudyQfuncPfunc`

39

`plotStudyQfuncPfunc` *Example: Plotting the q- and p-function from the dissertation*

Description

Example: Plotting the q- and p-function from the dissertation

Usage

```
plotStudyQfuncPfunc(y, d, x, n)
```

Arguments

<code>y</code>	a vector of responses.
<code>d</code>	a vector of heteroscedasticity.
<code>x</code>	a design matrix.
<code>n</code>	number of points to interpolate along.

Value

A list of objects created by ggplot2.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_s <- bcg$size
bcg_x <- cbind(1, bcg$x)
p <- plotStudyQfuncPfunc(y=bcg_y, d=bcg_d, x=bcg_x, n=500)
p[1] # plot of the q-function
p[2] # plot of the p-funciton
```

`plotStudySizes` *Example: Plotting study sizes*

Description

Example: Plotting study sizes

Usage

```
plotStudySizes(dat)
```

40

*qfunc***Arguments**

`dat` data frame of study responses of binomial type.

Value

An object created by `ggplot2`.

Examples

```
bcg <- bcgVaccineData()
plotStudySizes(bcg)
```

`plotStudyUnbalance` *Example: Plotting study unbalances in group assignments*

Description

Example: Plotting study unbalances in group assignments

Usage

```
plotStudyUnbalance(dat)
```

Arguments

`dat` data frame of study responses of binomial type.

Value

An object created by `ggplot2`.

`qfunc` *The $q_{\Delta(\tau)}$ function.*

Description

Returns the q -function.

Usage

```
qfunc(y, d, x)
```

Arguments

`y` study responses.
`d` heteroscedasticity.
`x` design matrix.

rB

41

Value

A vector valued function.

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$sdiv
bcg_x <- cbind(1, bcg$x)
qfunc(y=bcg_y, d=bcg_d, x=bcg_x)
```

*rB**Data generation: Log-risk-ratio of a binomial-Gaussian model***Description**

Random draws of log risk ratios from a hierarchical binomial Gaussian model.

Usage

```
rB(n, h, s, a, r, x, b)
```

Arguments

n	number of draws.
h	heterogeneity.
s	study sizes.
a	balance of group assignments.
r	fixed risk in the treatment group.
x	design matrix.
b	regression coefficients.

Details

It is always assumed that at least one response in a study has happend, i.e., a response of 0 in a treatment or control group is rounded up to 1. Note that this may lead to an overestimation of small risks. If possible, make sure your sample sizes are large enough to compensate for this effect.

Value

A (2k,n) matrix. Each column is an independent draw.

42

*rBinomGauss***Examples**

```
h_test <- .03
x_test <- cbind(1,1:13)
b_test <- c(0.02, 0.03)
s_test <- rep(2000, 13)
a_test <- rep(.3, 13)
rB(n=10, h=h_test, s=s_test, a=a_test, r=.3, x=x_test, b=b_test)
```

*rBinomGauss**Data generation: Sampling data of clinical trials*

Description

A random draw of a hierarchical binomial Gaussian model.

Usage

```
rBinomGauss(h, s, a, r, x, b)
```

Arguments

<code>h</code>	heterogeneity.
<code>s</code>	study sizes.
<code>a</code>	balance of group assignments.
<code>r</code>	fixed risk in the control group.
<code>x</code>	design matrix.
<code>b</code>	regression coefficients.

Details

It is always assumed that at least one response in a study has happened, i.e., a response of 0 in a treatment or control group is rounded up to 1. Note that this may lead to an overestimation of small risks. If possible, make sure your sample sizes are large enough to compensate for this effect.

You may work around this by increasing study sizes.

Value

A list containing the risk and a data frame with the studies.

rD

43

Examples

```

h_test <- .03
s_test <- rep(2000, 13)
a_test <- rep(.3, 13)
x_test <- cbind(1,1:13)
b_test <- c(0.02, 0.03)
dat <- rBinomGauss(h=h_test, s=s_test, a=a_test, r=0.03, x=x_test,
b=b_test)$study

if(!all(dim(dat) == c(dim(x_test)[1], 4))) stop("Wrong dimension")

```

*rD**Data generation: Gaussian-Gaussian model***Description**

Random draws of heteroscedasticity responses of studies, where each study in a random effects meta regression model follows a Gaussian response. Thus $D = (d * X) / (s-1)$ where X is chi-squared distributed.

Usage

```
rD(n, d, s)
```

Arguments

n	number of draws.
d	heteroscedasticity.
s	study sizes.

Value

A (k,n)-matrix. Each column is an independent draw.

Examples

```

d_test = rchisq(13, df=0.02)
s_test = rep(100, 13)
rD(n=10, d=d_test, s=s_test)

```

44

render

regressionEstimates *Point estimates: For the regression coefficients*

Description

Calculates point estimates for the regression coefficient for given point estimates of the variance components 'd' and a data frame of different estimates of the heterogeneity 'h'.

Usage

```
regressionEstimates(y, d, h_dat, x)
```

Arguments

y	study responses, k-vector of responses.
d	heteroscedasticity, k-vector of heteroscedasticities.
h_dat	Here, 'h_dat' should be a data frame with variables 'type' and 'h'. Thus, one may use <code>h_dat = hEstimates(y, d, x)</code> .
x	design matrix, k-p-matrix.

Value

A list of estimates for the regression coefficients.

Here, 'h_dat' should be a data frame with variables 'type' and 'h', thus, we may use `h_dat = hEstimates(y, d, x)`

Examples

```
bcg <- bcgVaccineData()
bcg_y <- bcg$logrisk
bcg_d <- bcg$ssdiv
bcg_x <- cbind(1, bcg$x)
bcg_h <- hEstimates(y=bcg_y, d=bcg_d, x=bcg_x)
regressionEstimates(y=bcg_y, d=bcg_d, h_dat=bcg_h, x=bcg_x)
```

render

Render plot: To PDF

Description

Renders obj into a pdf-file of name: `path++name`. Neat feature is that the default size in A4. Simply use the 'scale' parameter to adjust the size of the plot to a fraction of a page.

renderSVG

45

Usage

```
render(name, plotObj, path, scale = 1, height = 11.6,
       width = 8.2)
```

Arguments

name	Should be self explanatory.
plotObj	Should be self explanatory.
path	Should be self explanatory.
scale	Should be self explanatory.
height	Should be self explanatory.
width	Should be self explanatory.

*renderSVG**Render plot: To SVG*

Description

Renders obj into a svg-file of name: path++name. Neat feature is that the default size in A4. Simply use the 'scale' parameter to adjust the size of the plot to a fraction of a page.

Usage

```
renderSVG(name, plotObj, path, scale = 1, height = 11.6,
          width = 8.2)
```

Arguments

name	Should be self explanatory.
plotObj	Should be self explanatory.
path	Should be self explanatory.
scale	Should be self explanatory.
height	Should be self explanatory.
width	Should be self explanatory.

46

sctBias

rY	<i>Data generation: Gaussian-Gaussian model</i>
----	---

Description

Random draws of response vectors y following the distribution of a random effects meta regression model. Each column is an independent draw.

Usage

```
rY(n, h, d, x, b)
```

Arguments

n	number of draws.
h	heterogeneity.
d	heteroscedasticity.
x	design matrix.
b	regression coefficients.

Value

A (k,n)-matrix. Each column is an independent draw.

Examples

```
x_test = cbind(1,1:13)
h_test = .03
d_test = rchisq(13, df=0.02)
b_test = c(0.02, 0.03)
rY(n=10, h=h_test, d=d_test, x=x_test, b=b_test)
```

sctBias	<i>Plotting performance: Scatter plots against heterogeneity</i>
---------	--

Description

Scatter plots of heterogeneity and bias.

Usage

```
sctBias(res, ...)
```


48

sctVersusH

sctVersusC *Plotting performance: Scatter plot against heterogeneity*

Description

Plotting performance: Scatter plot against heterogeneity

Usage

`sctVersusC(res)`

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

sctVersusH *Plotting performance: Scatter plot against heterogeneity*

Description

Plotting performance: Scatter plot against heterogeneity

Usage

`sctVersusH(res)`

Arguments

`res` The collected interval results from a computer experiment.

Value

A plot object.

sdmByMethod

49

sdmByMethod *Plotting performance: Scatter plot against heterogeneity*

Description

Plotting performance: Scatter plot against heterogeneity

Usage

`sdmByMethod(res)`

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

sdmByType *Plotting performance: Scatter plot against heterogeneity*

Description

Plotting performance: Scatter plot against heterogeneity

Usage

`sdmByType(res)`

Arguments

`res` The collected results from a computer experiment.

Value

A plot object.

50

sdsByType

<code>sdsByMethod</code>	<i>Plotting performance: Scatter plot against heteroscedasticity</i>
--------------------------	--

Description

Plotting performance: Scatter plot against heteroscedasticity

Usage

```
sdsByMethod(res)
```

Arguments

<code>res</code>	The collected results from a computer experiment.
------------------	---

Value

A plot object.

<code>sdsByType</code>	<i>Plotting performance: Scatter plot against heteroscedasticity</i>
------------------------	--

Description

Plotting performance: Scatter plot against heteroscedasticity

Usage

```
sdsByType(res)
```

Arguments

<code>res</code>	The collected results from a computer experiment.
------------------	---

Value

A plot object.

setupExperiment

51

setupExperiment *Running a computer experiment in batch mode*

Description

Sets up a computer experiment evaluating the performance of different inference methods in the random effects meta regression model.

Usage

```
setupExperiment(name, seed, n, resolution, bounds, x, b,
               sgnf, piv_draws, ...)
```

Arguments

name	Reference name for the experiment.
seed	Random seed for the experiment.
n	number of simulations to at each parameter configuration.
resolution	list of number of parameter configurations in each design, e.g. resolution=list(h=5L, d=3L)
bounds	list of parameter bounds used for experimental design, e.g. bounds=list(h=c(0,1), d=c(0.001, 2), s=c(200L, 2000L)) where - h : bounds of the heterogeneity. - d : bounds of the heteroscedasticity. - a : bounds of the balancing factor of group assignments. - s : bounds of the study sizes. - r : fixed risk in the control.
x	design matrix.
b	regression coefficients.
sgnf	levels of significance.
piv_draws	number of pivotal draws.
...	further arguments to makeExperimentRegistry, e.g. file.dir=tempfile().

Value

The registry.

52

yvec

`yvec`*Data generation: Sampling data of clinical trials*

Description

Calculates log risk ratios from a study in the right format.

Usage

```
yvec(study)
```

Arguments

`study` Study data of a clinical trial with binomial outcomes.

Examples

```
h_test <- .03
x_test <- cbind(1,1:13)
b_test <- c(0.02, 0.03)
s_test <- rep(2000, 13)
a_test <- rep(.3, 13)
rBinomGauss( h=h_test, s=s_test, a=a_test, r=0.03
, x=x_test, b=b_test)$study -> test
yvec(test)
dvec(test)
```

Index

*Topic **datasets**
 cbbPalette, 7
 cbgPalette, 7

bcgVaccineData, 3
 boxBias, 4
 boxByConfidence, 4
 boxByMethod, 5
 boxByType, 5
 boxMSE, 6
 boxSD, 6

cbbPalette, 7
 cbgPalette, 7
 collectAllExperiments, 8
 collectExperiments, 8

designB, 9
 designD, 10
 designY, 11
 dvec, 12

experimentD, 12
 experimentY, 13

formulaL, 14
 formulaR, 15

hConfidence, 17
 hEstimates, 17

intervalEstimates, 18

joinPivotalCoefficients, 19
 joinPivotalHeterogeneity, 20

lenBoxByMethod, 21
 lenBoxByType, 21
 lenDenByMethod, 22
 lenDenByType, 22

makeConfInt, 23

makeConfInts, 23
 metagen, 24
 metagenEmpty, 25
 metagenGeneralised, 25
 metareg, 27

performance, 28
 performanceConfH, 29
 performanceConfR, 29
 performancePointH, 30
 performancePointR, 30
 pfunc, 31
 pivotalStream, 32
 plotCoefficientInterval, 33
 plotDensityH, 33
 plotDensityH2, 34
 plotDensityIntercept, 34
 plotDensityIntercept2, 35
 plotDensitySlope, 36
 plotDensitySlope2, 36
 plotHeterogeneityInterval, 37
 plotIntervalEstimates, 37
 plotStudyForest, 38
 plotStudyQfuncPfunc, 39
 plotStudySizes, 39
 plotStudyUnbalance, 40

qfunc, 40

rB, 41
 rBinomGauss, 42
 rD, 43
 regressionEstimates, 44
 render, 44
 renderSVG, 45
 rY, 46

sctBias, 46
 sctMSE, 47
 sctSD, 47

54

INDEX

[sctVersusC](#), [48](#)
[sctVersusH](#), [48](#)
[sdmByMethod](#), [49](#)
[sdmByType](#), [49](#)
[sdsByMethod](#), [50](#)
[sdsByType](#), [50](#)
[setupExperiment](#), [51](#)

[yvec](#), [52](#)

List of figures

1.1	Differences in total study sizes of 13 clinical trials evaluating the efficacy of the BCG vaccine.	5
1.2	Differences in subject assignments of 13 clinical trials evaluating the efficacy of the BCG vaccine. The measure of balance is $\frac{v-\neg v}{v+\neg v}$ where v is the number of vaccinated and $\neg v$ the number of non-vaccinated subjects in a study.	5
1.3	Forest plot of 13 clinical trials evaluating the efficacy of the BCG vaccine. Trials are sorted top to bottom by their absolute latitude. Intervals show 95% confidence intervals for the logarithm of relative risk.	6
2.1	For the BCG vaccine efficacy data, the squared length of the residual vector $q_y = q_y(\tau)$ as a function of the unknown heterogeneity τ in the data is plotted for different values of τ . On the right, the inverse function $p_y = p_y(\eta)$ of q_y is plotted for different values of η .	22
4.1	Density histogram and kernel density estimate of the density of the generalised pivotal quantity \tilde{T} for the heterogeneity parameter τ in the BCG vaccine efficacy example. The estimates are based on $m = 1000$ independent draws from the pivotal distribution using Algorithm 4.8.	54
4.2	Two kernel density estimates of the densities of generalised pivotal quantities for the heterogeneity parameter τ , one based on Algorithm 4.8, the other additionally adjusted for uncertainty in the heteroscedasticity estimate.	55
4.3	Box plots of the estimated mean squared errors of different point estimators for the heterogeneity parameter τ .	63
4.4	Box plots of the estimated bias of different point estimators for the heterogeneity parameter τ .	64
4.5	Box plots of the estimated standard deviation of different point estimators for the heterogeneity parameter τ .	65

4.6	Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ which are based on method of moments.	66
4.7	Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ which are based on likelihood methods.	67
4.8	Scatter plots of the estimated mean squared errors of point estimators for the heterogeneity parameter τ based on generalised inference principles.	67
4.9	Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on method of moments.	68
4.10	Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on likelihood methods.	68
4.11	Scatter plots of the estimated bias of different point estimators for the heterogeneity parameter τ which are based on generalised inference principles.	69
4.12	Scatter plots of the estimated standard deviations of point estimators for the heterogeneity parameter τ which are based on method of moments.	69
4.13	Scatter plots of the estimated standard deviation of point estimators for the heterogeneity parameter τ which are based on likelihood methods.	70
4.14	Scatter plots of the estimated standard deviations of point estimators for the heterogeneity parameter τ which are based on generalised inference principles.	70
4.15	Box plots of the estimated coverage of the interval estimators for the heterogeneity parameter τ .	72
4.16	Scatter plots of the estimated coverage of the interval estimators for the heterogeneity parameter τ .	73
4.17	Box plots of the estimated mean width of the discussed interval estimators for the heterogeneity parameter τ .	73

5.1	Density estimates of the distributions \mathcal{L}_2 and \mathfrak{R}_2 for the slope in the BCG vaccine efficacy example.	94
5.2	Density estimates of the distributions \mathcal{L}_2 and \mathfrak{R}_2 for the slope in the BCG vaccine efficacy example.	95
5.3	Box plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.	99
5.4	Box plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.	99
5.5	Box plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.	100
5.6	Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.	101
5.7	Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.	101
5.8	Scatter plots of estimated mean squared error of different point estimators for the slope parameter for different model parameter configurations.	102
5.9	Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.	102
5.10	Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.	103
5.11	Scatter plots of estimated bias of different point estimators for the slope parameter for different model parameter configurations.	103
5.12	Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.	104
5.13	Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations.	104

- 5.14 Scatter plots of estimated standard deviation of different point estimators for the slope parameter for different model parameter configurations. 105
- 5.15 Boxplots of **confidence level - coverage probability** for different model parameters for $k = 7$ number of studies and different inferential methods shown for the slope parameter. The optimal line is the black vertical line at zero in each of the plots. Being to the left of this line is preferable. 107
- 5.16 Boxplots of **confidence level - coverage probability** for different model parameters for $k = 13$ number of studies and different inferential methods shown for the slope parameter. The optimal line is the black vertical line at zero in each of the plots. Being to the left of this line is preferable. 108
- 5.17 Scatter plots of estimated coverage of different interval estimators for the slope parameter for different model parameter configurations. 109
- 5.18 Density estimates of estimated coverage of different interval estimators for the slope parameter for different model parameter configurations. 110
- 5.19 Box plots of estimated average width of different interval estimators for the slope parameter for different model parameter configurations. 111

List of tables

1.1	The results of 13 clinical trials evaluating the efficacy of the BCG vaccine.	4
4.1	Point estimates of the heterogeneity parameters in the BCG vaccine efficacy example.	74
4.2	Interval estimates of the heterogeneity parameter in the BCG vaccine efficacy example.	74
5.1	Point estimates of the intercept and slope parameters in the BCG vaccine efficacy example	112
5.2	Interval estimates of the intercept parameter in the BCG vaccine efficacy example.	114
5.3	Interval estimates of the slope parameter in the BCG vaccine efficacy example.	115

References

- Barnard, G. A. (1982). Discussion: The functional-model basis of fiducial inference. *The Annals of Statistics*, 10, 1068-1069.
- Berkey, C. S., Hoaglin, D. C., Mosteller, F. and Colditz, C. A. (1995). A random-effects regression model for meta-analysis. *Statistics in Medicine*, 14, 395-411.
- Bischl, B., Lang, M., Mersmann, O., Rahnenfuehrer, J. and Weihs, C. (2012). Computing on high performance clusters with R: Packages BatchJobs and BatchExperiments. Technical Report SFB 876, TU Dortmund University.
- Brent, R. P. (2002). *Algorithms for Minimization Without Derivatives*. New York: Dover Publications.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics*. Cambridge: Cambridge University Press.
- Carnell, R. (2012). *lhs: Latin Hypercube Samples*. Accessible via <http://cran.r-project.org/package=lhs>. R package version 0.10.
- Dawid, A. P. and Stone, M. (1982a). The functional-model basis of fiducial inference. *The Annals of Statistics*, 10, 1054-1067.
- Dawid, A. P. and Stone, M. (1982b). Reply: The functional-model basis of fiducial inference. *The Annals of Statistics*, 10, 1074.
- DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3), 177-188.
- Erez, A., Bloom, M. C. and Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: implications for situational specificity and validity generalization. *Personnel Psychology*, 49, 275-306.
- Fisher, R. A. (1935). The fiducial argument in statistical inference. *Annals of Eugenics Cambridge*, 6, 391-398.
- Fraser, D. A. S. (1982). Discussion: The functional-model basis of fiducial inference. *The Annals of Statistics*, 10, 1070-1073.
- Friedrich, T. and Knapp, G. (2013). Generalised interval estimation in the random effects meta regression model. *Computational Statistics and Data Analysis*, 64, 165-179.
- Ginsberg, A. M. (1998). The tuberculosis epidemic: Scientific challenges and opportunities. *Public Health Reports*, 113, 128-136.
- Graybill, F. A. (1961). *An introduction to linear statistical models*, volume 1. New York: McGraw-Hill Book Company.
- Hannig, J. (2009). On generalized fiducial inference. *Statistica Sinica*, 19, 491-544.
- Harbord, R. M. and Higgins, J. P. T. (2008). Meta-regression in stata. *The Stata Journal*, 8, 493-519.
- Hartl, M. (2010). The Tau Manifesto. On 2010-11-23, 15:35:41 accessed via <http://tauday.com/pdf>.

- Hartung, J. and Knapp, G. (2005). On confidence intervals for the among-group variance in the one-way random effects model with unequal error variances. *Journal of Statistical Planning and Inference*, 127, 157-177.
- Hartung, J., Knapp, G. and Sinha, B. K. (2008). *Statistical Meta-Analysis with Applications*. Hoboken: John Wiley & Sons.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358), 320-338.
- Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, 93(2), 388-395.
- Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.
- Iyer, H. K., Wang, C. M. J. and Mathew, T. (2004). Models and confidence intervals for true values in interlaboratory trials. *Journal of the American Statistical Association*, 99, 1060-1071.
- Khatri, C. G. (1966). A note on a manova model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, 18, 75-86.
- Knapp, G., Biggerstaff, B. and Hartung, J. (2006). Assessing the amount of heterogeneity in random-effects meta-analysis. *Biometrical Journal*, 48, 271-285.
- Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22, 2693-2710.
- LaMotte, L. R. (1973). On non-negative quadratic unbiased estimation of variance components. *American Statistical Association*, 68, 728-730.
- Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2(1), 49-55.
- Mandel, J. and Paule, R. C. (1970). Interlaboratory evaluation of a material with unequal number of replicates. *Analytical Chemistry*, 42, 1194-1197.
- Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications (with discussion). *Journal of the American Statistical Association*, 78, 47-65.
- Möbius, T. W. D. (2014). *metagen: Inference in Meta Analysis and Meta Regression*. Accessible via <http://00tau.github.io/metagen/>. R package version 1.0.
- Palais, B. (2001). π is wrong!. *The Mathematical Intelligencer*, 23, 7-8.
- Paule, R. C. and Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, 87, 377-385.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna.: R Foundation for Statistical Computing
- Santner, T. J., Williams, B. J. and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.
- Sidik, K. and Jonkman, J. N. (2005). Simple heterogeneity variance estimation for meta-analysis. *Applied Statistical Science*, 54, 367-384.

- Stein, M. (1987). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29, 143-151.
- Thompson, S. G. and Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine*, 18, 2693-2708.
- Tian, L. (2008a). Generalized inferences on the overall treatment effect in meta-analysis with normally distributed outcomes. *Biometrical Journal*, 50, 237-247.
- Tian, L. (2008b). Inferences about the between-study variance in meta-analysis with normally distributed outcomes. *Biometrical Journal*, 50, 248-256.
- Tsui, K.-W. and Weerahandi, S. (1989). Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 84, 602-607.
- Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, 30, 261-293.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1-48.
- Weerahandi, S. (1993). Generalized confidence intervals. *Journal of the American Statistical Association*, 88, 899-905.
- Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. New York: Springer.
- Weerahandi, S. (2012). Generalized point estimation with application to small response estimation. *Communications in Statistics – Theory and Methods*, 22, 4069-4095.
- Whitehead, A. (2002). *Meta-Analysis of Controlled Clinical Trials*. Hoboken: John Wiley & Sons.

Erklärung gemäß § 9 Abs. 2 der Promotionsordnung

Ich versichere an Eides statt, dass meine Dissertation selbständig verfasst und alle in Anspruch genommenen Quellen und Hilfen in der Dissertation vermerkt wurden.

