Calidad de Revistas Científicas Españolas
FECYT | FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA | 2011

ISSN: 1696-2281
eISSN: 2013–8830
SORT 35 (2) July-December (2011)

SORT

Statistics and Operations Research Transactions

Coediting institutions
Universitat Politècnica de Catalunya
Universitat de Barcelona
Universitat de Girona
Universitat Autònoma de Barcelona
Institut d'Estadística de Catalunya

Supporting institution
Spanish Region of the International Biometric Society

Generalitat de Catalunya
Institut d'Estadística de Catalunya


MIXT
Procedent de
fonts responsables
FSC® C081765
FSC
www.fsc.org

SORT

Volume 35
Number 2
July-December 2011
ISSN: 1696-2281
eISSN: 2013-8830

Articles

Information for authors and subscribers

SORT 35 (2) July-December 2011, 83-102

A simulation study on some confidence intervals for the population standard deviation

Moustafa Omar Ahmed Abu-Shawiesh[1], Shipra Banik[2]
and B. M. Golam Kibria[3]

Abstract

In this paper a robust estimator against outliers along with some other existing interval estimators are considered for estimating the population standard deviation. An extensive simulation study has been conducted to compare and evaluate the performance of the interval estimators. The exact and the proposed robust method are easy to calculate and are not overly computer-intensive. It appears that the proposed robust method is performing better than other confidence intervals for estimating the population standard deviation, specifically in the presence of outliers and/or data are from a skewed distribution. Some real-life examples are considered to illustrate the application of the proposed confidence intervals, which also supported the simulation study to some extent.

MSC: 62F10, 62F40

Keywords: Breakdown point, bootstrapping, confidence interval, coverage probability, standard deviation, *Qn* estimator, robust estimator, outlier, scale estimator.

1. Introduction

Point estimates are of limited value, since we cannot attach to them statements regarding the amount of confidence that they have estimated the unknown parameter. Of great value is an interval estimate, an estimate about which we can make statements of confidence (Daniel, 1990). The confidence interval is defined as an estimated range of values that is likely to include an unknown population parameter. If independent samples

[1] Department of Mathematics, Faculty of Science, Hashemite University, Zarqa 13115, Jordan, mabushawiesh@hu.edu.jo

[2] School of Engineering and Computer Science, Independent University, Bangladesh, Dhaka 1229, Bangladesh, banik@secs.iub.edu.bd

[3] Department of Mathematics and Statistics, Florida International University, Modesto A. Maidique Campus, Miami, FL 33199, USA, kibriag@fiu.edu

Received: February 2011

Accepted: May 2011

are take repeatedly from the same population, and the confidence interval is calculated for each sample, then a certain percentage, called the confidence level of the interval, will include the unknown population parameter.

Scale estimators are very important in many statistical applications. The sample standard deviation is the most common scale estimator that provides a logical point estimate of the population standard deviation, σ. Unfortunately, the sample standard deviation, S, is very sensitive to the presence of outliers in the data. Furthermore, S is not necessarily the most efficient or meaningful estimator of scale in skewed and leptokurtic distributions and it is notable that it is not robust to slight deviations from normality (Tukey, 1960). S has a good efficiency in platykurtic and moderately leptokurtic distributions but the classic inferential methods for it may perform poorly in realistically non-normal distributions (Bonett, 2006). Also, according to Gorard (2004), S has no obvious intuitive meaning because squaring before summing and then taking the square root makes the resulting figure difficult to understand, which restricts any subsequent intuitive interpretation. Nevertheless, S is the most efficient scale estimator for the normal distribution often used to construct the $100(1-\alpha)\%$ confidence interval for σ. The standard error of S is a scale multiple of the actual parameter being estimated. In this paper, we are looking for a scale estimator which is robust, has a closed form, and easy to compute as an alternative to S. The Rousseeuw-Croux estimator, Qn might be a more meaningful measure of variation and may be preferred to S. It is the most efficient scale estimator for the normal distribution often used to construct the $100(1-\alpha)\%$ confidence interval for σ.

The exact $100(1-\alpha)\%$ confidence interval for the population standard deviation, σ, is based on the assumption that the underlying distribution of the data is normal with no outliers, but what would happen if the data are not from a normal distribution instead of heavier tails or from a skewed distribution. The statistical literature shows that robust methods might give more meaningful measures of scale and are indeed more resistant to departures from normality and presence of outliers than S. Therefore, the need for alternatives to the exact $100(1-\alpha)\%$ confidence interval for σ comes to play. The statistical literature is full of robust confidence intervals for the mean, for example Tukey and McLaughlin (1963); Huber (1964); Dixon and Tukey (1968); De Wet and van Wyk (1979); Gross (1973, 1976); Bickel and Doksum (1977, p. 375); Kim (1992) and Clark (1994). For small-sample inference about variance and its transformations we refer to Longford (2010) among others. The problem of constructing robust confidence interval for the population standard deviation, σ, has received much less attention. Here, by a robust confidence interval, we mean that its actual coverage probability is close to the specified confidence level $(1-\alpha)$ with a short length of the confidence interval.

In this paper, an approximate confidence interval for the population standard deviation, σ, for one sample problems that is much less sensitive to the presence of outliers and/or to departure from normality is proposed. The proposed method provides an alternative to the exact $100(1-\alpha)\%$ confidence interval for σ based on Qn. The performance of the proposed method is investigated through a Monte Carlo simulation study based

on various evaluation criteria such as coverage probability, average width and standard deviation of the width. The coverage probability naturally varies from distribution to distribution for a given procedure, but a good procedure should keep this variation small. Furthermore, we want a confidence interval whose endpoints are generally close together, thus a small average width is good (Gross, 1976). A set of real data is employed to illustrate the results given in the paper.

The organization of the paper is as follows. In Section 2, we presented Rousseeuw-Croux estimator, *Qn*, for estimating σ, and discussed outliers. The proposed confidence intervals for σ are presented in Section 3. A Monte Carlo simulation study is conducted in Section 4. Some real life data are analyzed in Section 5. Finally, some concluding remarks based on simulation and numerical examples are given in Section 6.

2. Rousseeuw-Croux estimator, *Qn* and outliers

2.1. The Rousseeuw-Croux estimator

Rousseeuw and Croux (1993) proposed two robust estimators for scale, the *Sn* and *Qn* estimators. They can be used as initial or ancillary scale estimators in the same way as the median absolute deviation (MAD) but they are more efficient and not biased towards symmetric distributions. The breakdown point of the *Sn* estimator is 50% and its efficiency is 58%, while *Qn* has the same breakdown point but its efficiency at normal distributions is very high, about 82%. Due to its high efficiency and other good properties, the *Qn* estimator is considered in this paper. Mosteller and Tukey (1977) define two types of robustness as follows:

1. Resistance: This means that changing a small part even by a large amount of the data does not cause a large change in the estimate.

2. Robustness of efficiency: This means that the statistic has high efficiency in a variety of situations rather than in any one situation. Efficiency means that the estimate is close to the optimal estimate given that the distribution of the data is known.

Many statistics have one of these properties. However, it can be difficult to find statistics that have both resistance and robustness of efficiency. The most common estimate of scale, *S* is the most efficient estimate of scale if the data come from a normal distribution. However, *S* is not robust in the sense that changing even one value can dramatically change the computed value of *S*; that is, it has poor resistance. In addition, it does not have robustness or efficiency for non-normal data. MAD and the inter-quartile range (IQR) are the two most commonly used robust alternatives to *S*. MAD in particular is a very robust scale estimator. However, MAD does not have

particularly high efficiency for data (37% for normal data) and also MAD has an implicit assumption of symmetry, that is it measures the distance from a measure of central location (the median). Rousseeuw and Croux (1993) proposed the *Qn* estimate of scale as an alternative to MAD. It shares desirable robustness properties with MAD (50% breakdown point, bounded influence function). In addition, it has significantly better normal efficiency (82%) and it does not depend on symmetry.

2.1.1. Definition of Qn

The estimator *Qn* for a random sample $X_1, X_2, \ldots, X_n$ with model distribution F is defined as:

$$Qn = 2.2219 \left\{ |X_i - X_j| ;\ i < j \ ;\ i = 1, 2, 3, \ldots, n \ ;\ j = 1, 2, 3, \ldots, n \right\}_{\{g\}} \qquad (1)$$

where $g = \binom{h}{2}\binom{n}{2}/4$ and $h = \left[\frac{n}{2}\right] + 1$ (i.e., roughly half the number of observations). Here the symbol $(.)$ represents the combination and the symbol $[.]$ is used to take only the integer part of a fraction. The *Qn* estimator is the *g-th* order statistic of the $\binom{n}{2}$ inter-point distances. The value 2.2219 is chosen to make *Qn* a consistent estimator of scale for normal data. Rousseeuw and Croux (1993) have derived the unbiasing factor *dn* so that $dn \times Qn$ becomes an unbiased estimator of σ for the case of normal distribution. These values of *dn* are provided here in Table 2.1 as a function of *n*. The scatterplot between *n* and *dn* is presented in Figure 2.1. We can observe from both Table 2.1 and Figure 2.1 that *dn* is sensitive to the sample sizes.

Table 2.1: The values of the unbiasing factor dn.

n	*dn*	*n*	*dn*
2	0.399	14	0.787
3	0.994	15	0.915
4	0.512	16	0.808
5	0.844	17	0.924
6	0.611	18	0.826
7	0.857	19	0.931
8	0.669	20	0.840
9	0.872	21	0.938
10	0.725	22	0.853
11	0.887	23	0.943
12	0.759	24	0.863
13	0.903	25	0.647

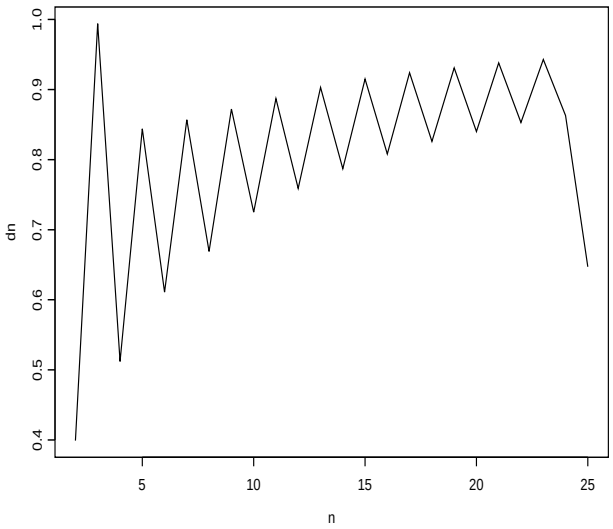


Figure 2.1: Scatterplot between n and dn.

An approximation result of dn for larger values of n is given by Croux and Rousseeuw (1992) as follows:

$$d_n = \begin{cases} \dfrac{n}{n+1.4} & \text{for odd values of } n \\ \dfrac{n}{n+3.8} & \text{for even values of } n \end{cases} \tag{2}$$

2.1.2. Properties of Qn

The Qn estimator has a simple and explicit formula, which is equally suitable for asymmetric distributions. The main properties of the Qn estimator investigated by Rousseeuw and Croux (1993) are given below:

1. For any sample $X = \{X_1, X_2, \ldots, X_n\}$ in which no two points coincide, the breakdown point of the scale estimator Qn is given by $\varepsilon*(Qn, X) = \frac{\left\lfloor \frac{n}{2} \right\rfloor}{n}$.
2. For $F = \Phi$, where Φ(x) is the standard normal distribution function, the value of d is given by $d = \frac{1}{\sqrt{2}\Phi^{-1}\left(\frac{5}{8}\right)} = 2.2219$. With this constant d, Qn has bias in small samples.
3. The influence function of Qn estimator is smooth and unbalanced (see Rousseeuw and Croux, 1993). For a model distribution F which has a density f, the influence function of Qn is given by

$$IF(x; Q, F) = d \frac{\frac{1}{4} - F(x + d^{-1}) + F(x - d^{-1})}{\int f(y + d^{-1}) f(y) dy}$$

4. The gross-error sensitivity of the Qn estimator is larger than those of MAD and Sn estimators and its value is $\gamma*(Q, \Phi) = \sup_x |IF(x; Q, \Phi)| = 2.069$.
5. The asymptotic variance of Qn in the case of normal distribution is given by $V(Q, \Phi) = 0.6077$ and this yields an efficiency of 82.27%. This is very high relative to the MAD estimator whose efficiency at normal distribution is only 36.74% and Sn whose efficiency is 58.23%. Using a simulation study, Rousseeuw (1991) concluded that the estimator Qn is more efficient than MAD and Sn estimators. However, Qn loses some of its efficiency for small sample sizes.
6. The square of the Qn estimator, that is, $(Qn)^2$, can be used as an estimate of σ^2. Even though both Qn and $(Qn)^2$ are biased estimators of σ and σ^2 respectively, they are efficient estimators of their respective targets (Rousseeuw, 1991).

2.2. Estimators and outliers

The presence of outliers in the data set is one of the most important topics in statistical inference. An outlier can be defined as observations which appear to be inconsistent with the remaining set of data. Outliers can be contaminants, i.e. arising from other distributions or can be typical observations generated from the assumed model (Barnett, 1988). Therefore, outliers need very special attention because a small departure from the assumed model can have strong negative effects on the efficiency of classical estimators for location and scale (Tukey, 1960). In this section, a simple numerical example taken from Rousseeuw (1991) is given to show the effect of outliers on S and Qn estimators. Suppose we have five measurements of a concentration without outliers given as follows:

$$5.59, 5.66, 5.63, 5.57, 5.60$$

Let us now suppose that one of these concentrations has been wrongly recorded so that the data have an outlier value and become as follows:

$$5.59, 5.66, 5.63, 55.7, 5.60$$

Based on these two data sets, the values of the two estimators are calculated and given in Table 2.2.

Table 2.2: *Values of the estimators for the example.*

	Scale Estimator	Data Set
	Without Outlier	With Outlier
S	0.0354	22.4
Qn	0.066657	0.066657

From Table 2.2, we notice that the value of the single outlier has changed the value of S, which becomes very large. The robustness of the Qn estimator is clear where the value of it is the same for the two data sets.

3. Proposed robust confidence interval for σ

3.1. Exact confidence interval for σ

Let $X_1, X_2, \ldots, X_n$be a random sample of size n from the normal distribution, i.e., $X_i \sim N(\mu, \sigma^2)$ for all i, then $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$ where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance. The exact $100(1-\alpha)\%$confidence interval for a population variance σ^2 is given as follows:

$$P\left(\frac{(n-1)S^2}{\chi^2_{(\frac{\alpha}{2},n-1)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}\right) = 1-\alpha \tag{3}$$

where $\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{1-\frac{\alpha}{2}}$ are the $(\frac{\alpha}{2})^{th}$ and $(1-\frac{\alpha}{2})^{th}$percentile points of the χ^2distribution with $(n-1)$ degrees of freedom. Taking the square root of the endpoints of equation (3) gives a $100(1-\alpha)\%$ confidence interval for σ as follows:

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{(\frac{\alpha}{2},n-1)}}} \quad , \quad \sqrt{\frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}}\right) \tag{4}$$

The exact confidence interval for σ^2 in (3) is hypersensitive to minor violations of the normality assumption. Scheffe (1959, p. 336) show that (3) has an asymptotic coverage probability of about 76, 63, 60 and 51 for the Logistic, the Student t(7), the Laplace and Student t(5) distributions respectively. The result is disturbing because these symmetric distributions are not easily distinguished from a normal distribution unless the sample size is large. Also, the exact confidence interval for σ^2 in (3) as demonstrated by Lehman (1986, p. 206) is highly sensitive to the presence of outliers and / or to departure from normality. However, as pointed out by Lehman, the sample size n may have to be rather large for the asymptotic result to give a good approximation.

3.2. Robust confidence intervals

In this section, we will propose the new robust confidence interval for estimating the population standard deviation σ. Instead of assuming $X_i \sim N(\mu, \sigma^2)$, let $X_1, X_2, \ldots, X_n$ are the random samples of size n from a continuous, independent and identically distributed random variable. The random variable T is defined as the ratio,

$$T = \frac{d_n Qn}{\sigma} \tag{5}$$

where the expression $dnQn$ acts as an unbiased estimator of σ so that E(T)=1 for normal distribution. Based on Rousseeuw and Croux(1993), for larger values of n, the following asymptotic result can be used:

$$T = \frac{d_n Qn}{\sigma} \sim N\left(1, \frac{1}{1.65n}\right) \tag{6}$$

The following approximation result can be obtained:

$$d_n Qn \sim N\left(\sigma, \frac{1}{1.65n}\sigma^2\right) \tag{7}$$

Therefore from (7), we can get the following pivotal quantity:

$$\frac{d_nQn-\sigma}{\frac{1}{1.28\sqrt{n}}\sigma} \sim N(0,\ 1) \tag{8}$$

Now, using the above pivotal quantity, we can derive the $100(1-\alpha)\%$ robust confidence interval for σ as follows:

$$P\left(q_1 < \frac{d_nQn-\sigma}{\frac{1}{1.28\sqrt{n}}\sigma} < q_2\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{q_1}{1.28\sqrt{n}}+1 < \frac{d_nQn}{\sigma} < \frac{q_2}{1.28\sqrt{n}}+1\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{1.28\sqrt{n}*d_nQn}{q_1+1.28\sqrt{n}} < \sigma < \frac{1.28\sqrt{n}*d_nQn}{q_2+1.28\sqrt{n}}\right) = 1-\alpha$$

where $q_1 = Z_{\frac{\alpha}{2}}$ and $q_2 = Z_{1-\frac{\alpha}{2}}$ are the $(\frac{\alpha}{2})^{th}$ and $(1-\frac{\alpha}{2})^{th}$percentile points of the standard normal distribution so that the length is minimum. Therefore, the $100(1-\alpha)\%$ robust confidence interval for σ, is as follows:

$$\left(\frac{1.28\sqrt{n}*d_nQn}{Z_{\frac{\alpha}{2}}+1.28\sqrt{n}}\ ,\ \frac{1.28\sqrt{n}*d_nQn}{Z_{1-\frac{\alpha}{2}}+1.28\sqrt{n}}\right) = \left(\frac{DQn}{Z_{\frac{\alpha}{2}}+D_1}\ ,\ \frac{DQn}{Z_{1-\frac{\alpha}{2}}+D_1}\right) \tag{9}$$

where the values of the factors $D = 1.28\sqrt{n}*d_n$ and $D_1 = 1.28\sqrt{n}$.

An approximation result of D for larger values of n can be calculated as follows:

$$D = \begin{cases} (1.28\sqrt{n})\left(\dfrac{n}{n+1.4}\right), & \text{for odd values of } n \\ (1.28\sqrt{n})\left(\dfrac{n}{n+3.8}\right), & \text{for even values of } n \end{cases} \tag{10}$$

The squaring of the endpoints of equation (9) gives a $100(1-\alpha)\%$confidence interval for σ^2.

3.3. Bonett confidence interval

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the normal distribution, that is, $X_i \sim N(\mu, \sigma^2)$ for all i. Scheffe (1959) found in his simulation study, the exact CI for σ

does not have an asymptotic coverage probability for non-normal distributions. Bonett (2006) proposed the following (1-α)100% confidence interval (CI) for σ as

$$LCL = \exp\left\{\ln(c\hat{\sigma}^2) - Z_{\alpha/2}se\right\} \quad \text{and} \quad UCL = \exp\left\{\ln(c\hat{\sigma}^2) + Z_{\alpha/2}se\right\}$$

where $Z_{\alpha/2}$ is two-sided critical z-value, $se = c[\{\hat{\gamma}_4(n-3)/n\}/(n-1)]^{1/2}$, $c = n/(n - Z_{\alpha/2})$ and $\hat{\gamma}_4 = n\sum_i(Y_i - \hat{\mu})^4/(\sum_i(Y_i - \hat{\mu})^2)^2$.

3.4. Cojbasic and Tomovic (CT) CI

Based on t-statistic, Cojbasic and Tomovic (2007) proposed the following nonparametric bootstrap t CI:

$$I_{boot} = S^2 - \hat{t}^{(\alpha)}\sqrt{\widehat{var}(S^2)}$$

where $S^2 = \frac{1}{n-1}\sum_i(X_i - \bar{X})^2$ is the sample variance, $\hat{t}^{(\alpha)}$ is a α percentile of $T*$ defined as $T* = \frac{S^{2*} - S^2}{\sqrt{\widehat{var}(S^{2*})}}$, S^{2*} is a bootstrap replication of statistic S^2 and $\widehat{var}(S^2)$ is a consistent estimator of the variance, defined by $2\sigma^4/(n-1)$.

3.5. Some bootstrap CIs

Let X(*) =$X_1^{(*)}, X_2^{(*)}, \ldots, X_n^{(*)}$, where the i-th sample is denoted by $X(i)$ for $i = 1, 2, \ldots, B$ and B is the number of bootstrap samples. We proposed the following bootstrap CIs for the sample σ:

Non-parametric bootstrap CI

Compute σ for all bootstrap samples and then order the sample SDs of each bootstrap samples as follows:

$$S^*_{(1)} \leq S^*_{(2)} \leq S^*_{(3)} \cdots \leq S^*_{(B)}$$

CI for population σ:

$$LCL = S^*_{[(\alpha/2)B]} \quad \text{and} \quad UCL = S^*_{[(1-\alpha/2)B]}$$

CI for population σ:

$$LCL = S\sqrt{(n-1)/\chi^{*2}_{\alpha/2,(n-1)}} \quad \text{and} \quad UCL = S\sqrt{(n-1)/\chi^{*2}_{1-\alpha/2,(n-1)}}$$

where $\chi^{*2}_{\alpha/2}$and $\chi^{*2}_{1-\alpha/2}$ are the $(\alpha/2)$-th and $(1-\alpha/2)$-th sample quantiles of $\chi^2=\frac{(n-1)S^2}{\hat{\sigma}^2_B}$ and $\hat{\sigma}_B=\sqrt{\frac{1}{B-1}\sum_{i=1}^{B}(\bar{x}^*_i-\bar{\bar{x}})^2}$ and $\bar{x}^*_i$ is the i-th bootstrap sample mean, $\bar{\bar{x}}$ is the bootstrap mean and $\hat{\sigma}_B$ is the bootstrap standard deviation.

Bootstrap robust CI

CI for population σ:

$$LCL=\frac{DQn}{Z^*_{\alpha/2}+D_1} \quad \text{and} \quad UCL=\frac{DQn}{Z^*_{1-\alpha/2}+D_1}$$

where $Z^*_{\alpha/2}$and $Z^*_{1-\alpha/2}$ are the $(\alpha/2)$th and $(1-\alpha/2)$th sample quantiles of the bootstrap test statistic, $Z^*_i=\frac{(\bar{x}^*_i-\bar{\bar{x}})}{\hat{\sigma}_B}$.

We note that all proposed confidence intervals except exact and Bonett do not require any distributional assumptions. However, bootstrap methods are computer intensive, where as others are very easy to compute. The exact method works better for any sample size when the data are from the normal distribution.

4. Simulation study

Our basic objective is to investigate some efficient estimators of σ by a simulation study. Since a theoretical comparison among the intervals is not possible, a simulation study has been made to compare the performance of the estimators.

4.1. Simulation technique

The flowchart of our simulation is as follows:

1. We use sample sizes n = 5, 10, 20, 30, 50, 70 and 100.
2. Random samples are generated from symmetric and skewed distributions:

   (a) Normal distribution with mean 3 and SD 1.
   (b) Chi-square distribution with df 1.
   (c) Lognormal distribution with mean 1 and SD 0.80.

We used 5000 simulation replications and 1500 bootstrap samples for each n. The most common 95% confidence interval ($\alpha=0.05$) for the confidence coefficient is used. It is well known that if the data are from a symmetric distribution (or n is large), the coverage probability will be exact or close to $(1-\alpha)$. So the coverage probability is a useful criterion for evaluating the confidence interval. Another criterion is the width of

the confidence interval. A shorter length width gives a better confidence interval. It is obvious that when coverage probability is the same, a smaller width indicates that the method is appropriate for the specific sample. In order to compare the performance of the various intervals, the following criteria are considered: coverage probabilities (below, cover and above), mean and SD of the widths of the resulting confidence intervals. The

Table 4.1: *Coverage properties for N(3,1) distribution with skewness 0.*

		Sample Sizes						
Approaches	Measuring Criteria	5	10	20	30	50	70	100
Exact	Below rate	0.0230	0.0242	0.0230	0.0214	0.0250	0.0234	0.0276
	Cover rate	0.9494	0.9492	0.9524	0.9554	0.9512	0.9544	0.9484
	Over rate	0.0276	0.0266	0.0246	0.0232	0.0238	0.0222	0.0240
	Mean width	2.1400	1.1113	0.6934	0.5427	0.4075	0.3413	0.2826
	SD width	0.7807	0.2631	0.1129	0.0711	0.0412	0.0289	0.0202
Robust	Below rate	0.1142	0.0634	0.0498	0.0376	0.0318	0.0306	0.0320
	Cover rate	0.8098	0.9041	0.9160	0.9276	0.9388	0.9438	0.9402
	Over rate	0.0760	0.0352	0.0342	0.0348	0.0294	0.0256	0.0278
	Mean width	2.6107	1.2849	0.7810	0.6058	0.4528	0.3786	0.3134
	SD width	1.4262	0.3828	0.1545	0.0940	0.0528	0.0367	0.0252
Bonett	Below rate	0.4352	0.1918	0.0890	0.0594	0.0372	0.0290	0.0274
	Cover rate	0.5292	0.7972	0.9060	0.9362	0.9582	0.9674	0.9678
	Over rate	0.0356	0.0110	0.0050	0.0044	0.0046	0.0036	0.0048
	Mean width	0.5073	0.6494	0.5901	0.5238	0.4327	0.3776	0.3222
	SD width	0.1992	0.1835	0.1264	0.0954	0.0621	0.0464	0.0339
Non-para Bootstrap	Below rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.7728	0.7930	0.9998	1.0000	1.0000	1.0000	1.0000
	Over rate	0.2272	0.2070	0.0002	0.0000	0.0000	0.0000	0.0000
	Mean width	0.9662	0.7504	0.5589	0.4699	0.3727	0.3197	0.2695
	SD width	0.3723	0.2312	0.1399	0.1035	0.0666	0.0497	0.0365
Parametric Bootstrap	Below rate	0.0266	0.0000	0.0006	0.0024	0.0124	0.0036	0.0092
	Cover rate	0.8472	0.9624	0.5398	0.8606	0.9730	0.5292	0.9758
	Over rate	0.1262	0.0376	0.4596	0.1370	0.0146	0.4672	0.0150
	Mean width	1.9244	3.4834	0.8099	0.6397	0.4608	0.2805	0.3289
	SD width	0.7021	0.8246	0.1318	0.0838	0.0466	0.0238	0.0235
Robust Bootstrap	Below rate	0.1104	0.0614	0.0470	0.0324	0.0290	0.0338	0.0328
	Cover rate	0.7560	0.9002	0.9154	0.9290	0.9490	0.9372	0.9394
	Over rate	0.1336	0.0384	0.0376	0.0386	0.0220	0.0290	0.0278
	Mean width	2.7531	1.3338	0.7980	0.6210	0.4679	0.3667	0.3109
	SD width	1.5039	0.3974	0.1578	0.0964	0.0546	0.0355	0.0250
CT Bootstrap	Below rate	0.0654	0.0368	0.0010	0.0012	0.0002	0.0000	0.0000
	Cover rate	0.9346	0.9632	0.9990	0.9988	0.9998	1.0000	1.0000
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	0.9073	0.5657	0.3581	0.2655	0.1679	0.1343	0.1032
	SD width	0.3310	0.1339	0.0583	0.0348	0.0170	0.0114	0.0074

below and above rates of a confidence interval is the fraction out of 5000 samples that resulted in an interval that lies entirely above and below the true value of the population mean. The coverage probability is found as the sum of the lower rate and upper rate and then subtracted from total probability 1. Simulation results are tabulated in Tables 4.1, 4.2 and 4.3 for normal, chi-square and log-normal distributions respectively. For

Table 4.2: *Coverage properties for χ^2_1 distribution with skewness 2.83.*

		Sample Sizes						
Approaches	Measuring Criteria	5	10	20	30	50	70	100
Exact	Below rate	0.2062	0.2494	0.2694	0.2762	0.2550	0.2562	0.2532
	Cover rate	0.7084	0.6374	0.5922	0.5722	0.5714	0.5766	0.5640
	Over rate	0.0854	0.1132	0.1384	0.1516	0.1736	0.1672	0.1828
	Mean width	2.6007	1.4303	0.9242	0.7355	0.5671	0.4733	0.3952
	SD width	1.9051	0.7663	0.3623	0.2407	0.1451	0.1028	0.0719
Robust	Below rate	0.0260	0.0074	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.5410	0.6018	0.8746	0.9672	0.9972	0.9994	1.0000
	Over rate	0.4330	0.3908	0.1252	0.0328	0.0028	0.0006	0.0000
	Mean width	1.8389	0.8986	0.4863	0.3592	0.2632	0.2132	0.1747
	SD width	1.6597	0.5313	0.2027	0.1243	0.0693	0.0484	0.0335
Bonett	Below rate	0.2888	0.4280	0.2906	0.2358	0.1634	0.1406	0.1160
	Cover rate	0.6266	0.5186	0.6796	0.7388	0.8158	0.8436	0.8680
	Over rate	0.0846	0.0534	0.0298	0.0254	0.0208	0.0158	0.0160
	Mean width	0.6715	1.0754	1.1966	1.1636	1.0610	0.9628	0.8628
	SD width	0.5506	0.7581	0.7435	0.6528	0.5057	0.4333	0.3485
Non-para Bootstrap	Below rate	0.0110	0.0052	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.9876	0.9948	0.9972	0.9966	0.9924	0.9454	0.6380
	Over rate	0.0014	0.0000	0.0026	0.0034	0.0076	0.0546	0.3620
	Mean width	1.2724	1.3299	1.2379	1.1433	1.0177	0.9143	0.8157
	SD width	1.0062	0.9391	0.7437	0.6133	0.4682	0.3920	0.3163
Parametric Bootstrap	Below rate	0.0472	0.0600	0.0030	0.0052	0.0002	0.0002	0.0400
	Cover rate	0.6048	0.6202	0.6954	0.6602	0.9482	0.5050	0.7466
	Over rate	0.3480	0.3198	0.3016	0.3346	0.0516	0.4948	0.2134
	Mean width	5.9921	2.0968	3.1846	1.7225	1.5354	1.9256	0.7131
	SD width	4.3895	1.1233	1.2484	0.5638	0.3927	0.4184	0.1298
Robust Bootstrap	Below rate	0.0280	0.0082	0.0006	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.5300	0.5798	0.8400	0.9666	0.9964	0.9992	1.0000
	Over rate	0.4420	0.4120	0.1594	0.0334	0.0036	0.0008	0.0000
	Mean width	1.5955	0.8422	0.4538	0.3538	0.2615	0.2043	0.1739
	SD width	1.4400	0.4979	0.1892	0.1224	0.0689	0.0464	0.0333
CT Bootstrap	Below rate	0.0968	0.1190	0.1186	0.1674	0.1218	0.2750	0.1414
	Cover rate	0.9032	0.8810	0.8814	0.8326	0.8782	0.7250	0.8586
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	1.3284	0.7912	0.4827	0.3365	0.2444	0.1814	0.1402
	SD width	0.9731	0.4239	0.1892	0.1102	0.0625	0.0394	0.0255

all simulated distributions, we also provided the coverage probabilites in Table 4.4. For more on the simulation techniques, we refer Baklizi and Kibria (2009) and Banik and Kibria (2010a,b) and references therein.

Table 4.3: *Coverage properties for the Lognormal (1.0,0.80) distribution with skewness 3.69.*

		Sample Sizes						
Approaches	Measuring Criteria	5	10	20	30	50	70	100
Exact	Below rate	0.1710	0.2600	0.3038	0.3280	0.3266	0.3494	0.3498
	Cover rate	0.7568	0.6460	0.5640	0.5266	0.5124	0.4784	0.4720
	Over rate	0.0722	0.0940	0.1322	0.1454	0.1610	0.1722	0.1782
	Mean width	6.4430	3.4512	2.2911	1.8185	1.3942	1.1700	0.9744
	SD width	4.8451	1.9012	1.0438	0.6992	0.4350	0.3197	0.2311
Robust	Below rate	0.0268	0.0040	0.0002	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.7252	0.5714	0.7442	0.9086	0.9900	0.9986	1.0000
	Over rate	0.2480	0.4246	0.2556	0.0914	0.0100	0.0014	0.0000
	Mean width	5.6323	2.6675	1.5611	1.1971	0.8873	0.7353	0.6064
	SD width	3.9645	1.0909	0.4237	0.2594	0.1436	0.1011	0.0684
Bonett	Below rate	0.2620	0.4720	0.3510	0.3030	0.2406	0.2166	0.1816
	Cover rate	0.6678	0.4852	0.6118	0.6688	0.7382	0.7660	0.8046
	Over rate	0.0702	0.0428	0.0372	0.0282	0.0212	0.0174	0.0138
	Mean width	1.6464	2.5439	2.9784	2.9380	2.7665	2.5728	2.3409
	SD width	1.4319	1.9865	2.3489	2.1990	1.9159	1.6832	1.5210
Non-para Bootstrap	Below rate	0.0312	0.1258	0.0200	0.0086	0.0078	0.0000	0.0002
	Cover rate	0.9688	0.8742	0.9800	0.9914	0.9922	1.0000	0.9998
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	3.0850	3.0532	3.0013	2.8159	2.5739	2.3781	2.1623
	SD width	2.5675	2.3907	2.2831	1.9829	1.6510	1.4309	1.2690
Parametric Bootstrap	Below rate	0.0082	0.0052	0.0038	0.2056	0.0014	0.0674	0.0056
	Cover rate	0.8972	0.7640	0.7832	0.6352	0.6152	0.8756	0.9782
	Over rate	0.0946	0.2308	0.2130	0.1592	0.3834	0.0570	0.0162
	Mean width	20.3072	10.9107	6.9319	2.2693	4.2257	2.7495	3.6072
	SD width	15.2710	6.0106	3.1581	0.8726	1.3183	0.7512	0.8555
Robust Bootstrap	Below rate	0.0428	0.0062	0.0004	0.0000	0.0000	0.0000	0.0000
	Cover rate	0.8148	0.6230	0.7108	0.8904	0.9870	0.9984	1.0000
	Over rate	0.1424	0.3708	0.2888	0.1096	0.0130	0.0016	0.0000
	Mean width	3.5183	2.2211	1.4099	1.1484	0.8495	0.7064	0.5978
	SD width	2.4765	0.9083	0.3827	0.2488	0.1375	0.0971	0.0674
CT Bootstrap	Below rate	0.3916	0.4424	0.2970	0.3686	0.1390	0.1400	0.0182
	Cover rate	0.6084	0.5576	0.7030	0.6314	0.8610	0.8600	0.9818
	Over rate	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	Mean width	2.4976	1.6585	1.0762	0.8497	0.5779	0.4639	0.3291
	SD width	1.8782	0.9137	0.4903	0.3267	0.1803	0.1267	0.0781

Table 4.4: *Coverage Probabilities for All Approaches and Distributions.*

Distribution	Approaches	Sample Size (n)						
		5	10	20	30	50	70	100
Normal	Exact	0.9494	0.9492	0.9525	0.9554	0.9512	0.9544	0.9484
	Robust	0.8098	0.9041	0.9160	0.9276	0.9388	0.9438	0.9402
	Bonett	0.5292	0.7972	0.9060	0.9362	0.9582	0.9674	0.9678
	NP-Boot	0.7728	0.7930	0.9998	1.0000	1.0000	1.0000	1.0000
	Par-Boot	0.8472	0.9624	0.5398	0.8606	0.9730	0.5292	0.9758
	Robust Boot	0.7560	0.9002	0.9154	0.9290	0.9490	0.9372	0.9394
	CT Boot	0.9346	0.9632	0.9990	0.9988	0.9998	1.0000	1.0000
χ^2_1	Exact	0.7084	0.6374	0.5922	0.5722	0.5714	0.5766	0.5640
	Robust	0.5410	0.6018	0.8746	0.9672	0.9972	0.9994	1.0000
	Bonett	0.6266	0.5186	0.6796	0.7388	0.8158	0.8436	0.8680
	NP-Boot	0.9876	0.9948	0.9972	0.9966	0.9924	0.9454	0.6380
	Par-Boot	0.6048	0.6202	0.6954	0.6602	0.9482	0.5050	0.7466
	Robust Boot	0.5300	0.5798	0.8400	0.9666	0.9964	0.9992	1.0000
	CT Boot	0.9032	0.8810	0.8814	0.8326	0.8782	0.7250	0.8586
Log-normal	Exact	0.7568	0.6460	0.5640	0.5266	0.5124	0.4784	0.4720
	Robust	0.7252	0.5714	0.7442	0.9086	0.9900	0.9986	1.0000
	Bonett	0.6678	0.4852	0.6118	0.6688	0.7382	0.7660	0.8046
	NP-Boot	0.9688	0.8742	0.9800	0.9914	0.9922	1.0000	0.9998
	Par-Boot	0.8972	0.7640	0.7832	0.6352	0.6152	0.8756	0.9782
	Robust Boot	0.8148	0.6230	0.7108	0.8904	0.9870	0.9984	1.0000
	CT Boot	0.6084	0.5576	0.7030	0.6314	0.8610	0.8600	0.9818

4.2. Results discussion

The MATLAB programming language was used to run the simulation and to make the necessary tables. The performance of the selected techniques in Section 3 for normal distribution is examined first and the simulated results are tabulated in Table 4.1.

The results in Table 4.1 suggested that when sampling from a normal distribution, the performance of the estimators do not differ greatly. However, for small sample sizes, the exact method has coverage probability close to 0.95, followed by CT Bootstrap, the proposed robust method, the robust bootstrap method and Bonnet performed the worse. The parametric bootstrap method performed better than the non-parametric bootstrap method for all sample sizes. When measuring criterion is average width, it is observed that the CT bootstrap interval performed well as compare to others, followed by Bonett and the non-parametric interval. The average width of the exact method is observed closed to the proposed robust method.

The next simulation compares the performance of the proposed intervals for a variety of non-normal distributions. Results are depicted for chi-square and log-normal in Tables 4.2 and 4.3 respectively. The results in Table 4.2 and Table 4.3 suggested that when sampling from a skewed distribution, the proposed robust method, Non-

parametric bootstrap, robust bootstrap performed better compared to others in the sense of coverage probability and average width. It is clear that the proposed robust method is superior to the exact method when the data are from a non-normal population.

When the data are from normal distributions or sample sizes are large, the exact method would be considered as it is easy to compute and has coverage probability close to the nominal size compared to the rest. Since in real life the distributions of the data are unknown or do not follow the normality assumption for most of the cases, our proposed robust confidence interval would be recommended, as it does not required any distributional assumption and is easy to compute compared to the bootstrap methods. Even though some of the bootstrap methods are as good as our proposed robust method, it is not advisable to use them as they are very computer intensive. However, for a computer expert researcher, the non-parametric bootstrap method can be recommended.

5. Applications to real data

In this section, we will present some real life examples to illustrate the application and the performance of the selected intervals.

5.1. Example 1

This example is taken from Hogg and Tanis (2001, page 359). The data set represents the amount of butterfat in pounds produced by a typical cow during a 305-day milk production period between her first and second calves. The butterfat production for a random sample of size $n = 20$ cows measured by a farmer yielding the following observations:

$$481, 537, 513, 583, 453, 510, 570, 500, 457, 555$$

$$618, 327, 350, 643, 499, 421, 505, 637, 599, 392$$

The sample mean, standard deviation and skewness of data are 507.5, 89.75 and -0.3804 and respectively. Shapiro-Wilk Normality Test (W = 0.9667, p-value = 0.6834) suggested that the data follow a normal distribution. The resulting 95% confidence intervals for different methods and the corresponding confidence widths are given in Table 5.1.

From Table 5.1, we observed that when the data under consideration has a normal distribution, the confidence intervals widths for both exact and robust methods are approximately the same, but as expected, the exact method provided the shortest intervals widths among the two methods. From the above Table, we observed that the non-parametric bootstrap interval has the narrowest width followed by the Bonett interval. It is noted that the CT bootstrap has the widest width.

Table 5.1: *The 95% Confidence Intervals for the Butterfat Data.*

Method	95% Confidence Interval	Width
Exact	(68.255,131.087)	62.832
Robust	(63.261,129.137)	65.876
Bonett	(63.910,114.789)	50.879
Non-para Bootstrap	(62.515,106.951)	44.435
Parametric Bootstrap	(74.656,130.790)	56.133
Robust Bootstrap	(74.970,155.960)	80.989
CT Bootstrap	(76.851,169.375)	92.523

5.2. Example 2

This example is taken from Weiss (2002, page 291). The data set represents the last year's chicken consumption in pounds for people on USA published by the USA Department of Agriculture in Food Consumption, Prices, and Expenditures. The last year's chicken consumption, in pounds, for a random sample of size $n = 17$ people yielded the following observations:

$$47, 39, 62, 49, 50, 70, 59, 53, 55, 0, 65, 63, 53, 51, 50, 72, 45$$

The sample mean, standard deviation and skewness of these data are 51.94, 16.08 and -2.11 respectively. Shapiro-Wilk Normality Test ($W = 0.8013$, p-value = 0.0021) suggested that the data do not follow a normal distribution. The zero value may be a recording error or due to a person in the random sample who does not eat chicken for some reason (e.g., a vegetarian) and may be considered as an outlier. Now, if we remove the outlier 0 pound from the sample data, the sample mean, standard deviation and skewness of data are 55.19, 9.21 and 0.33 respectively. Shapiro-Wilk Normality Test ($W = 0.9651$, p-value = 0.7539) suggested that the data follow a normal distribution. The resulting 95% confidence intervals for different methods and the corresponding confidence widths based on the two types of data are given in Table 5.2.

From Table 5.2, we observed that the non-parametric bootstrap interval has the narrowest width followed by the Bonett interval. It is also noted that the CT bootstrap has the widest width than others. From Table 5.2 it is also observed that when the outlier is removed from the data, then the confidence interval for exact and robust methods are very similar and approximately have the same interval width, although the robust interval width is slightly shorter. The value of the outlier does not affect so much the proposed robust confidence interval and therefore the exact confidence interval for the population standard deviation, σ, should be avoided in the presence of outliers. In general, an outlier should not be removed without careful consideration. Simply removing an outlier because it is an outlier is unacceptable statistical practice.

Table 5.2

Method	95% Confidence Interval		Width	
	Original Data	Abridged Data	Original Data	Abridged Data
Exact	**(**11.978 , 24.478**)**	**(**7.485 , 16.330**)**	12.50	8.845
Robust	(6.804, 14.255)	(5.193, 11.635)	7.451	6.442
Bonett	(8.798,26.362)	(6.396,11.813)	17.564	5.416
Non-para Bootstrap	(6.789,23.977)	(6.158,11.470)	17.188	5.311
Parametric Bootstrap	(10.676,37.332)	(7.506,13.971)	26.655	6.464
Robust Bootstrap	(7.800,17.840)	(6.398,13.991)	10.040	7.593
CT Bootstrap	(16.374,40.284)	(9.737, 18.698)	23.910	8.961

Also, our result if we had blindly finding a confidence interval without first examining the data would have been invalid and misleading. In this case we can use the proposed robust confidence interval which is resistant to outliers.

6. Concluding remarks

This paper proposes an approximate confidence interval for estimating the population standard deviation, σ, based on a robust estimator and compares its performance with other proposed intervals. A simulation study has been conducted to compare the performance of the estimators, and shows that the proposed robust confidence interval for all distributions considered performs well and had a good coverage probability compared to the exact method especially for non-normal distributions. It appears that the sample size (n) has significant effect on the proposed confidence interval. We observed that if the population is really normal, the exact confidence interval for the population standard deviation, σ, performs slightly better than the proposed robust method. If the distribution is highly skewed, the coverage probability of the proposed robust method becomes close to $1-\alpha$ and improves as the sample size increases. Actually, if the population is really non-normal, the exact confidence interval for the population standard deviation, σ, can be arbitrarily bad. A single outlier makes it worse than useless. To illustrate the findings of the paper we considered some real life examples which also supported the simulation study to some extent. Finally, among all proposed intervals, the robust, exact, non-parametric boot strap and CT bootstrap intervals are promising and can be recommended for the practitioners. However, both exact and proposed robust intervals are easy to compute and are not computer intensive.

Acknowledgements

The authors are thankful to the referees and editor of the journal for their valuable and constructive comments which certainly improved the quality and presentation of the paper. This paper was completed while Dr. B. M. Golam Kibia was on sabbatical leave (2010-2011). He is grateful to Florida International University for awarding him the sabbatical leave, which gave him excellent research facilities.

References

Baklizi, A. and Kibria, B. M. G. (2009). One and two sample confidence intervals for estimating the mean of skewed populations: an empirical comparative study. *Journal of Applied Statistics*, 36, 601–609.

Banik, S. and Kibria, B. M. G. (2010a). Comparison of some parametric and nonparametric type one sample confidence intervals for estimating the mean of a positively skewed distribution. *Communications in Statistics-Simulation and Computation*, 39, 361–389.

Banik, S. and Kibria, B. M. G. (2010b). Comparison of some test statistics for testing the mean of a right skewed distribution. *Journal of Statistical Theory and Applications*, 9, 77–90.

Barnett, V. (1988). Outliers and order statistics. *Communication in Statistics: Theory and Methods*, 17, 2109–2118.

Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden Day, San Francisco, USA.

Bonett, D. G. (2006). Approximate confidence interval for standard deviation of nonnormal distributions. *Computational Statistics and Data Analysis*, 50, 775–782.

Clark, B. R. (1994). Empirical evidence for adaptive confidence interval and identification of outliers using methods of trimming, *Australian Journal of Statistics*, 36(1): 45–58.

Cojbasic, V. and Tomovic, A. (2007). Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples, *Computational Statistics & Data Analysis*, 51, 5562–5578.

Croux, C. and P. J. Rousseeuw (1992). Time efficient algorithms for two highly robust estimators of scale. Computational Statistics, Volume 1, eds. Y. Dodge and J. Whittaker, Heidelberg: Physika-Verlag, 411–428.

Daniel, W. W. (1990). *Applied Nonparametric Statistics*, second edition, Duxbury, USA.

De Wet, T. and van Wyk, W.J. (1979). Efficiency and robustness of Hogg's adaptive trimmed means. *Communications in Statistics, Theory and Methods*, 8, 117–128.

Dixon, W. J. and Tukey, J. W. (1968). Approximate behavior of the distribution of winsorized t (trimming/winsorizing 2), *Technometrics*, 10, 83–93.

Gorard, S. (2004). Revisiting a 90-year-old debate: the advantages of the mean deviation. *Paper presented at the British Educational Research Association Annual Conference*. University of Manchester, UK.

Gross, A. M. (1973). A Monte Carlo swindle for estimators of locations. *Applied Statistics*, 22, 347–353.

Gross, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions, *American Statistical Associations*, 71, 409–416.

Hogg, R. V. and E. A. Tanis (2001). *Probability and Statistical Inference*. Sixth edition, Prentice-Hall Inc., Upper Saddle River, New Jersey, USA.

Huber, P. J. (1964). Estimation of a location parameter, *The Annals of Mathematical Statistics*, 35, 73–101.

Kim, Seong-Ju (1992). The metrically trimmed mean as a robust estimator of location, *The Annals of Statistics*, 20, 1534–1547.

Lehman, E. L. (1986). *Testing Statistical Hypothesis*, John Wiley, New York.

Longford, N. T. (2010). Small-sample inference about variance and its transformations, *SORT*, 34, 1, 3–20.

Mosteller and Tukey (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, USA.

Rousseuw, P. J. (1991). Tutorial to robust statistics. *Journal of Chemometrics*, 5, 1–20.

Rousseuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88, 1273–1283.

Scheffe, H. (1959). *The Analysis of Variance*. Wiley, New York.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In Olkin, I., *et al.* (Eds.) *Contributions to Probability and Statistics, Essays in Honor of Harold Hotelling*. Stanford: Stanford University Press. 448–485.

Tukey, J. W. and McLaughlin, D. H. (1963). Less confidence and significance procedures for location based on a single sample: trimming / winsorizing 1, *Sankhya A*, 25, 331–352.

Weiss, A. N. (2002). *Introductory Statistics*, Sixth edition, Addison-Wesley, USA.

SORT 35 (2) July-December 2011, 103-124

Stress-strength reliability of Weibull distribution based on progressively censored samples

Akbar Asgharzadeh[1,*], Reza Valiollahi[2] and Mohammad Z. Raqab[3]

Abstract

Based on progressively Type-II censored samples, this paper deals with inference for the stress-strength reliability $\boldsymbol{R} = \boldsymbol{P}(\boldsymbol{Y} < \boldsymbol{X})$ when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two independent Weibull distributions with different scale parameters, but having the same shape parameter. The maximum likelihood estimator, and the approximate maximum likelihood estimator of $\boldsymbol{R}$ are obtained. Different confidence intervals are presented. The Bayes estimator of $\boldsymbol{R}$ and the corresponding credible interval using the Gibbs sampling technique are also proposed. Further, we consider the estimation of $\boldsymbol{R}$ when the same shape parameter is known. The results for exponential and Rayleigh distributions can be obtained as special cases with different scale parameters. Analysis of a real data set as well a Monte Carlo simulation have been presented for illustrative purposes.

MSC: 62F10, 62F15, 62N01

Keywords: Maximum likelihood estimator, Approximate maximum likelihood estimator, Bootstrap confidence interval, Bayesian estimation, Metropolis-Hasting method, Progressive Type-II censoring.

1. Introduction

The Weibull distribution is one of the most widely used distributions in the reliability and survival studies. The two-parameter Weibull distribution denoted by $W(\alpha, \theta)$ has the probability density function (pdf)

$$f(x, \alpha, \theta) = \frac{\alpha}{\theta} x^{\alpha-1} e^{-\frac{x^\alpha}{\theta}}, \quad x > 0,\ \alpha,\ \theta > 0, \tag{1}$$

*Address for correspondence: a.asgharzadeh@umz.ac.ir (A. Asgharzadeh), rvaliollahi@semnan.ac.ir (R. Valiollahi), mraqab@ju.edu.jo (M. Z. Raqab).

[1] Department of Statistics, University of Mazandaran, Babolsar, Iran.

[2] Department of Statistics, University of Semnan, Semnan, Iran.

[3] Department of Mathematics, University of Jordan, Amman 11942, JORDAN.

Received: February 2011
Accepted: July 2011

and the cumulative distribution function (cdf)

$$F(x,\alpha,\theta) = 1 - e^{-\frac{x^{\alpha}}{\theta}}, \qquad x > 0,\ \alpha,\ \theta > 0. \tag{2}$$

Here α is the shape parameter and θ is the scale parameter.

It has been used very effectively for analyzing lifetime data, particularly when the data are censored. Among various censoring schemes, the Type II progressive censoring scheme has become very popular one in the last decade. It can be described as follows: Consider N units are placed under a study and only $n(< N)$ units are completely observed until failure. At the time of the first failure (the first stage), r_1 of the $N-1$ surviving units are randomly withdrawn (censored intentionally) from the experiment. At the time of the second failure (the second stage), r_2 of the $N-2-r_1$ surviving units are withdrawn and so on. Finally, at the time of the nth failure (the nth stage), all the remaining $r_n = N-n-r_1-\cdots-r_{n-1}$ surviving units are withdrawn. We will refer to this as progressive Type-II right censoring with scheme $(r_1, r_2, \ldots, r_n)$. It is clear that this scheme includes the conventional Type-II right censoring scheme (when $r_1 = r_2 = \cdots = r_{n-1} = 0$ and $r_n = N-n$) and complete sampling scheme (when $N = n$ and $r_1 = r_2 = \cdots = r_n = 0$). For further details on progressively censoring and relevant references, the reader may refer to the book by Balakrishnan and Aggarwala (2000).

In the stress-strength modelling, $R = P(Y < X)$ is a measure of component reliability when it is subjected to random stress Y and has strength X. For a particular situation, consider Y as the pressure of a chamber generated by ignition of a solid propellant and X as the strength of the chamber. Then R represents the probability of successful firing of the engine. In this context, R can be considered as a measure of system performance and it is naturally arise in electrical and electronic systems. It may be mentioned that R is of greater interest than just reliability since it provides a general measure of the difference between two populations and has applications in many area. For example, if X is the response for a control group, and Y refers to a treatment group, R is a measure of the effect of the treatment. Also, it may be mentioned that R equals the area under the receiver operating characteristic (ROC) curve for diagnostic test or bio-markers with continuous outcome (see Bamber (1975)). The ROC curve is widely used, in biological, medical and health service research, to evaluate the ability of diagnostic tests or bio-markers to distinguish between two groups of subjects, usually non-diseased and diseased subjects. For more application of R, see Kotz et al. (2003). Many authors have studied the stress-strength parameter R. Among them, Ahmad et al. (1997), Awad et al. (1981), Kundu and Gupta (2005, 2006), Adimari and Chiogna (2006), Baklizi (2008), Raqab et al. (2008) and Rezaei et al. (2010).

Based on complete X-sample and Y-sample, Kundu and Gupta (2006) considered the estimation of $R = P(Y < X)$ when $X \sim W(\alpha, \theta_1)$ and $Y \sim W(\alpha, \theta_2)$ are two independent Weibull distributions with different scale parameters, but having the same shape parameter. In this paper, we extend their results for the case when the samples are progressively Type-II censored. The layout of this paper is as follows: In Section 2,

we derive the maximum likelihood estimator (MLE) of R. It is observed that the MLE can be obtained using an iterative procedure. We further propose an approximate MLE (AMLE) of R, which can be obtained explicitly. Different confidence intervals (C.I.'s) are presented in Section 3. A Bayes estimator of R and the corresponding credible interval using the Gibbs sampling technique have been proposed in Section 4. Analysis of a real data set as well a Monte Carlo simulation based comparison of the proposed methods are performed in Section 5. Finally, we conclude the paper in Section 6.

2. Maximum likelihood estimator of R

Let $X \sim W(\alpha,\theta_1)$ and $Y \sim W(\alpha,\theta_2)$ be independent random variables. Then it can be easily seen that

$$R = P(Y < X) = \frac{\theta_1}{\theta_1+\theta_2}. \tag{3}$$

Our interest is in estimating R based on progressive Type-II censored data on both variables. To derive the MLE of R, first we obtain the MLE's of α, θ_1 and θ_2. Suppose $\mathbf{X} = (X_{1:N}, X_{2:N}, \ldots, X_{n:N})$ is a progressively Type-II censored sample from $W(\alpha,\theta_1)$ with censored scheme $\mathbf{r} = (r_1, r_2, \ldots, r_n)$ and $\mathbf{Y} = (Y_{1:M}, Y_{2:M}, \ldots, Y_{m:M})$ is a progressively Type-II censored sample from $W(\alpha,\theta_2)$ with censored scheme $\mathbf{r}' = (r'_1, r'_2, \ldots, r'_m)$. For notation simplicity, we will write $(X_1, X_2, \ldots, X_n)$ for $(X_{1:N}, X_{2:N}, \ldots, X_{n:N})$ and $(Y_1, Y_2, \ldots, Y_m)$ for $(Y_{1:M}, Y_{2:M}, \ldots, Y_{m:M})$. Therefore, the likelihood function of α, θ_1 and θ_2 is given (see Balakrishnan and Aggarwala (2000)) by

$$L(\alpha,\theta_1,\theta_2) = \left[c_1 \prod_{i=1}^{n} f(x_i)[1-F(x_i)]^{r_i}\right] \times \left[c_2 \prod_{j=1}^{m} f(y_j)[1-F(y_j)]^{r'_j}\right], \tag{4}$$

where

$$\begin{aligned} c_1 &= N(N-1-r_1)(N-2-r_1-r_2)\cdots(N-n+1-r_1\cdots-r_{n-1}),\\ c_2 &= M(M-1-r'_1)(M-2-r'_1-r'_2)\cdots(M-m+1-r'_1\cdots-r'_{m-1}). \end{aligned}$$

Upon using (1) and (2), we immediately have the likelihood function of the observed data as follows:

$$\begin{aligned} L(data|\alpha,\theta_1,\theta_2) = c_1c_2\alpha^{n+m}\theta_1^{-n}\theta_2^{-m}\prod_{i=1}^{n}x_i^{\alpha-1}\prod_{j=1}^{m}y_j^{\alpha-1}\\ \times \exp\left\{-\frac{1}{\theta_1}\sum_{i=1}^{n}(r_i+1)x_i^{\alpha} - \frac{1}{\theta_2}\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}\right\}. \end{aligned} \tag{5}$$

From (5), the log-likelihood function is

$$l(\alpha,\theta_1,\theta_2) \propto (n+m)\ln\alpha - n\ln(\theta_1) - m\ln(\theta_2) + (\alpha-1) \times \left[\sum_{i=1}^{n}\ln(x_i) + \sum_{j=1}^{m}\ln(y_j)\right] - \frac{1}{\theta_1}\sum_{i=1}^{n}(r_i+1)x_i^{\alpha} - \frac{1}{\theta_2}\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}.$$

The MLEs of α, θ_1 and θ_2, say $\widehat{\alpha}$, $\widehat{\theta}_1$ and $\widehat{\theta}_2$ respectively, can be obtained as the solution of

$$\frac{\partial l}{\partial\alpha} = \frac{n+m}{\alpha} + \left[\sum_{i=1}^{n}\ln(x_i) + \sum_{j=1}^{m}\ln(y_j)\right] - \frac{1}{\theta_1}\sum_{i=1}^{n}(r_i+1)x_i^{\alpha}\ln(x_i) - \frac{1}{\theta_2}\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}\ln(y_j) = 0, \tag{6}$$

$$\frac{\partial l}{\partial\theta_1} = -\frac{n}{\theta_1} + \frac{1}{\theta_1^2}\sum_{i=1}^{n}(r_i+1)x_i^{\alpha} = 0, \tag{7}$$

$$\frac{\partial l}{\partial\theta_2} = -\frac{m}{\theta_2} + \frac{1}{\theta_2^2}\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha} = 0. \tag{8}$$

From (7) and (8), we obtain

$$\widehat{\theta}_1(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(r_i+1)x_i^{\alpha}, \text{ and } \widehat{\theta}_2(\alpha) = \frac{1}{m}\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}. \tag{9}$$

Substituting the expressions of $\widehat{\theta}_1(\alpha)$ and $\widehat{\theta}_2(\alpha)$ into (6), $\widehat{\alpha}$ can be obtained as a fixed point solution of the following equation:

$$k(\alpha) = \alpha, \tag{10}$$

where

$$k(\alpha) = \frac{n+m}{\frac{n\sum_{i=1}^{n}(r_i+1)x_i^{\alpha}\ln(x_i)}{\sum_{i=1}^{n}(r_i+1)x_i^{\alpha}} + \frac{m\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}\ln(y_j)}{\sum_{j=1}^{m}(r'_j+1)y_j^{\alpha}} - \left[\sum_{i=1}^{n}\ln(x_i) + \sum_{j=1}^{m}\ln(y_j)\right]}.$$

A simple iterative procedure $k(\alpha^{(j)}) = \alpha^{(j+1)}$ where $\alpha^{(j)}$ is the j-th iterate, can be used to find the solution of (10). Once we obtain $\widehat{\alpha}_{ML}$, the MLE of θ_1 and θ_2, can be deduced from (9) as $\widehat{\theta}_{1ML} = \widehat{\theta}_1(\widehat{\alpha}_{ML})$ and $\widehat{\theta}_{2ML} = \widehat{\theta}_2(\widehat{\alpha}_{ML})$. Therefore, we compute the MLE of R as

$$\widehat{R}_{ML} = \frac{\frac{1}{n}\sum_{i=1}^{n}(r_i+1)x_i^{\widehat{\alpha}_{ML}}}{\frac{1}{n}\sum_{i=1}^{n}(r_i+1)x_i^{\widehat{\alpha}_{ML}} + \frac{1}{m}\sum_{j=1}^{m}(r'_j+1)y_j^{\widehat{\alpha}_{ML}}}. \tag{11}$$

Here the maximum likelihood approach does not give an explicit estimator for α and hence for R, based on a progressively Type-II censored sample. Now, we approximate the likelihood equation analogously to Kundu and Gupta (2006). It is based on the fact that if the random variable X has $W(\alpha,\theta)$, then $V = \ln X$, has the extreme value distribution with pdf as

$$f(v;\mu,\sigma) = \frac{1}{\sigma} e^{\frac{v-\mu}{\sigma} - e^{\frac{v-\mu}{\sigma}}}, \quad -\infty < v < \infty, \tag{12}$$

where $\mu = \frac{1}{\alpha}\ln\theta$ and $\sigma = 1/\alpha$. The density function (12) is known as the density function of an extreme value distribution, with location, and scale parameters as μ and σ respectively. The standard extreme value distribution has the pdf and cdf as

$$g(v) = e^{v-e^v}, \qquad\qquad G(v) = 1 - e^{-e^v}.$$

Suppose $X_1 < X_2 < \cdots < X_n$ is a progressively Type-II censored sample from $W(\alpha,\theta_1)$ with censored scheme $(r_1, r_2, \ldots, r_n)$ and $Y_1 < Y_2 < \cdots < Y_m$ is a progressively Type-II censored sample from $W(\alpha,\theta_2)$ with censored scheme $(r'_1, r'_2, \ldots, r'_m)$. Furthermore, we use the following notation for this subsection. $T_i = \ln X_i$, $Z_i = \frac{T_i-\mu_1}{\sigma}$, $i = 1,\ldots,n$ and $S_j = \ln Y_j$, $W_j = \frac{S_j-\mu_2}{\sigma}$, $j = 1,\ldots,m$, where $\mu_1 = \frac{1}{\alpha}\ln\theta_1$, $\mu_2 = \frac{1}{\alpha}\ln\theta_2$ and $\sigma = \frac{1}{\alpha}$.
The log-likelihood function of the observed data $T_1,\ldots,T_n$ and $S_1,\ldots,S_m$ is

$$\begin{aligned} l^*(\mu_1,\mu_2,\sigma) \propto & -(n+m)\ln\sigma + \sum_{i=1}^{n}\ln(g(z_i)) + \sum_{i=1}^{n} r_i \ln(1-G(z_i)) \\ & + \sum_{j=1}^{m}\ln(g(w_j)) + \sum_{j=1}^{m} r'_j \ln(1-G(w_j)). \end{aligned} \tag{13}$$

Differentiating (13) with respect to μ_1, μ_2 and σ, we obtain the likelihood equations as

$$\frac{\partial l^*}{\partial \mu_1} = -\frac{1}{\sigma}\sum_{i=1}^{n}\frac{g'(z_i)}{g(z_i)} + \frac{1}{\sigma}\sum_{i=1}^{n} r_i \frac{g(z_i)}{1-G(z_i)} = 0, \tag{14}$$

$$\frac{\partial l^*}{\partial \mu_2} = -\frac{1}{\sigma}\sum_{j=1}^{m}\frac{g'(w_j)}{g(w_j)} + \frac{1}{\sigma}\sum_{j=1}^{m} r'_j \frac{g(w_j)}{1-G(w_j)} = 0, \tag{15}$$

$$\begin{aligned} \frac{\partial l^*}{\partial \sigma} = & -\frac{n+m}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n} z_i \frac{g'(z_i)}{g(z_i)} + \frac{1}{\sigma}\sum_{i=1}^{n} r_i z_i \frac{g(z_i)}{1-G(z_i)} \\ & - \frac{1}{\sigma}\sum_{j=1}^{m} w_j \frac{g'(w_j)}{g(w_j)} + \frac{1}{\sigma}\sum_{j=1}^{m} r'_j w_j \frac{g(w_j)}{1-G(w_j)} = 0. \end{aligned} \tag{16}$$

We approximate the terms $p(z_i) = \frac{g'(z_i)}{g(z_i)}$ and $q(z_i) = \frac{g(z_i)}{1-G(z_i)}$ by expanding them in a Taylor series around $\nu_i = E(Z_i)$. Further, we also approximate the terms $\bar{p}(w_j) = \frac{g'(w_j)}{g(w_j)}$, and $\bar{q}(w_j) = \frac{g(w_j)}{1-G(w_j)}$ by expanding them in a Taylor series around $\bar{\nu}_j = E(W_j)$. It is known that $Z_i \overset{d}{=} G^{-1}(U_i)$, where U_i is the i-th progressively Type-II censored order statistic from the uniform $U(0,1)$ distribution. Therefore,

$$\nu_i = E(Z_i) \approx G^{-1}(\eta_i),$$

where $\eta_i = E(U_i)$. From Balakrishnan and Aggarwala (2000),

$$\eta_i = 1 - \prod_{k=n-i+1}^{n} \frac{k + R_{n-k+1} + \cdots + R_n}{k+1+R_{n-k+1}+\cdots+R_n}, \quad i = 1,\ldots,n.$$

Since, $G^{-1}(u) = \ln(-\ln(1-u))$, we can approximate ν_i by $\ln(-\ln(1-\eta_i))$. Similarly, we approximate $\bar{\nu}_j$ by $\ln(-\ln(1-\eta_j))$. Now, upon expanding the function $p(z_i)$, $\bar{p}(w_j)$, $q(z_i)$ and $\bar{q}(w_j)$ keeping only the first two terms, we get

$$\begin{aligned} p(z_i) &\approx \alpha_i + \beta_i z_i, & \bar{p}(w_j) &\approx \bar{\alpha}_j + \bar{\beta}_j w_j, \\ q(z_i) &\approx \gamma_i + \delta_i z_i, & \bar{q}(w_j) &\approx \bar{\gamma}_j + \bar{\delta}_j w_j, \end{aligned}$$

where

$$\begin{aligned} \alpha_i &= p(\nu_i) - \nu_i p'(\nu_i) = 1 + (\nu_i - 1)e^{\nu_i}, & \beta_i &= p'(\nu_i) = -e^{\nu_i}, \\ \bar{\alpha}_j &= \bar{p}(\bar{\nu}_j) - \bar{\nu}_j \bar{p}'(\bar{\nu}_j) = 1 + (\bar{\nu}_j - 1)e^{\bar{\nu}_j}, & \bar{\beta}_j &= \bar{p}'(\bar{\nu}_j) = -e^{\bar{\nu}_j}, \\ \gamma_i &= q(\nu_i) - \nu_i q'(\nu_i) = (1-\nu_i)e^{\nu_i}, & \delta_i &= q'(\nu_i) = e^{\nu_i}, \\ \bar{\gamma}_j &= \bar{q}(\bar{\nu}_j) - \bar{\nu}_j \bar{q}'(\bar{\nu}_j) = (1-\bar{\nu}_j)e^{\bar{\nu}_j}, & \bar{\delta}_j &= \bar{q}'(\bar{\nu}_j) = e^{\bar{\nu}_j}. \end{aligned}$$

Therefore, (14), (15), and (16) can be approximated respectively as

$$\frac{\partial l^*}{\partial \mu_1} = -\frac{1}{\sigma}\sum_{i=1}^{n}[(\alpha_i - r_i\gamma_i) + z_i(\beta_i - r_i\delta_i)] = 0, \tag{17}$$

$$\frac{\partial l^*}{\partial \mu_2} = -\frac{1}{\sigma}\sum_{j=1}^{m}\left[(\bar{\alpha}_j - r'_j\bar{\gamma}_j) + w_j(\bar{\beta}_j - r'_j\bar{\delta}_j)\right] = 0, \tag{18}$$

$$\begin{aligned} \frac{\partial l^*}{\partial \sigma} = &-\frac{n+m}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n}\left[z_i(\alpha_i - r_i\gamma_i) + z_i^2(\beta_i - r_i\delta_i)\right] \\ &- \frac{1}{\sigma}\sum_{j=1}^{m}\left[w_j(\bar{\alpha}_j - r'_j\bar{\gamma}_j) + w_j^2(\bar{\beta}_j - r'_j\bar{\delta}_j)\right] = 0. \end{aligned} \tag{19}$$

If we denote $\tilde{\mu}_1$, $\tilde{\mu}_2$ and $\tilde{\sigma}_1$ as the solutions of (17), (18) and (19) respectively, then observe that

$$\tilde{\mu}_1 = A_1 + B_1\tilde{\sigma},\ \tilde{\mu}_2 = A_2 + B_2\tilde{\sigma},\ \text{and}\ \tilde{\sigma} = \frac{-D + \sqrt{D^2 - 4(n+m)E}}{2(n+m)},$$

where

$$A_1 = \frac{\sum_{i=1}^{n} t_i(\beta_i - r_i\delta_i)}{\sum_{i=1}^{n}(\beta_i - r_i\delta_i)}, \qquad B_1 = \frac{\sum_{i=1}^{n}(\alpha_i - r_i\gamma_i)}{\sum_{i=1}^{n}(\beta_i - r_i\delta_i)},$$

$$A_2 = \frac{\sum_{j=1}^{m} s_j(\bar{\beta}_j - r'_j\bar{\delta}_j)}{\sum_{j=1}^{m}(\bar{\beta}_j - r'_j\bar{\delta}_j)}, \qquad B_2 = \frac{\sum_{j=1}^{m}(\bar{\alpha}_j - r'_j\bar{\gamma}_j)}{\sum_{j=1}^{m}(\bar{\beta}_j - r'_j\bar{\delta}_i)},$$

$$\begin{aligned} D &= \sum_{i=1}^{n} t_i(\alpha_i - r_i\gamma_i) - 3A_1\sum_{i=1}^{n}(\alpha_i - r_i\gamma_i) + 2A_1B_1\sum_{i=1}^{n}(\beta_i - r_i\delta_i) \\ &\quad + \sum_{j=1}^{m} s_j(\bar{\alpha}_j - r'_j\bar{\gamma}_j) - 3A_2\sum_{j=1}^{m}(\bar{\alpha}_j - r'_j\bar{\gamma}_j) + 2A_2B_2\sum_{j=1}^{m}(\bar{\beta}_j - r'_j\bar{\delta}_j), \\ E &= \sum_{i=1}^{n} t_i^2(\beta_i - r_i\delta_i) - A_1\sum_{i=1}^{n} t_i(\beta_i - r_i\delta_i) \\ &\quad + \sum_{j=1}^{m} s_j^2(\bar{\beta}_j - r'_j\bar{\delta}_j) - A_2\sum_{j=1}^{m} s_j(\bar{\beta}_j - r'_j\bar{\delta}_j). \end{aligned}$$

Once $\tilde{\sigma}$ is obtained, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ can be obtained immediately. Therefore, the approximate MLE of R is given by

$$\tilde{R} = \frac{\tilde{\theta}_1}{\tilde{\theta}_1 + \tilde{\theta}_2},$$

where

$$\tilde{\theta}_1 = \exp\left(\frac{1}{\tilde{\sigma}}(A_1 + B_1\tilde{\sigma})\right),\ \text{and}\ \tilde{\theta}_2 = \exp\left(\frac{1}{\tilde{\sigma}}(A_2 + B_2\tilde{\sigma})\right).$$

3. C.I.'s of R

Based on asymptotic behavior of R, we present an asymptotic C.I. of R. We further, propose two C.I.'s based on the non-parametric bootstrap method.

3.1. Asymptotic C.I. of R

Let us first start with obtaining the Fisher information matrix of $\theta = (\alpha, \theta_1, \theta_2)$. If $X_{1:N} < X_{2:N} < \cdots < X_{n:N}$ is a progressively Type-II censored sample from the $W(\alpha, \theta)$ distribution with censored scheme $\mathbf{r} = (r_1, r_2, \ldots, r_n)$. Then $Z_{1:N} < Z_{2:N} < \cdots < Z_{n:N}$, where $Z_{i:N} = \frac{X_{i:N}^{\alpha}}{\theta}$ $(i = 1, \ldots, n)$ is a progressively Type-II censored sample from the standard Exponential distribution with censored scheme $\mathbf{r} = (r_1, r_2, \ldots, r_n)$. Hence (see Balakrishnan and Aggarwala (2000), p. 19.)

$$E[(X_{i:N})^{\alpha}] = \theta E(Z_{i:N}) = \theta \mu_i,$$

where

$$\mu_i = \sum_{k=1}^{i} \frac{1}{N - \sum_{s=0}^{k-1} r_s - k + 1}, \quad i = 1, \ldots, n.$$

The pdf of $X_{i:N}$ (see, for example, Kamps and Cramer (2001)) is

$$f_{X_{i:N}}(x) = c_{i-1} \sum_{k=1}^{i} a_{k,i} \frac{\alpha}{\theta} x^{\alpha-1} e^{-\gamma_k \frac{x^{\alpha}}{\theta}}, \; x > 0,$$

where

$$\gamma_k = N - k + 1 + \sum_{s=k}^{n} r_s, \;\; c_{i-1} = \prod_{s=1}^{i} \gamma_s, \;\; \text{and} \;\; a_{k,i} = \prod_{\substack{s=1 \\ s \neq k}}^{i} \frac{1}{\gamma_s - \gamma_i}.$$

The Fisher information matrix of $\theta = (\alpha, \theta_1, \theta_2)$ (cf. Kundu and Gupta (2005, 2006)) is obtained to be

$$I(\theta) = -\begin{pmatrix} E\left(\frac{\partial^2 l}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 l}{\partial \alpha \partial \theta_1}\right) & E\left(\frac{\partial^2 l}{\partial \alpha \partial \theta_2}\right) \\ E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \alpha}\right) & E\left(\frac{\partial^2 l}{\partial \theta_1^2}\right) & E\left(\frac{\partial^2 l}{\partial \theta_1 \partial \theta_2}\right) \\ E\left(\frac{\partial^2 l}{\partial \theta_2 \partial \alpha}\right) & E\left(\frac{\partial^2 l}{\partial \theta_2 \partial \theta_1}\right) & E\left(\frac{\partial^2 l}{\partial \theta_2^2}\right) \end{pmatrix} = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}$$

where

$$\begin{aligned} -I_{11} &= -\frac{n+m}{\alpha^2} - \frac{1}{\theta_1} \sum_{i=1}^{n} \left[(r_i + 1) E[X_i^{\alpha} (\ln(X_i))^2]\right] - \frac{1}{\theta_2} \sum_{j=1}^{m} \left[(r'_j + 1) E[Y_j^{\alpha} (\ln(Y_j))^2]\right] \\ &= -\frac{1}{\alpha^2} \sum_{i=1}^{n} \left[(r_i + 1) c_{i-1} \sum_{k=1}^{i} \frac{a_{k,i}}{\gamma_k^2} \left[\Gamma''(2) + 2 \ln(\frac{\theta_1}{\gamma_k}) \Gamma'(2) + (\ln(\frac{\theta_1}{\gamma_k}))^2 \Gamma(2)\right]\right] \\ &\quad - \frac{1}{\alpha^2} \sum_{j=1}^{m} \left[(r'_j + 1) c'_{j-1} \sum_{k=1}^{j} \frac{a'_{k,j}}{\gamma_k'^2} \left[\Gamma''(2) + 2 \ln(\frac{\theta_2}{\gamma'_k}) \Gamma'(2) + (\ln(\frac{\theta_2}{\gamma'_k}))^2 \Gamma(2)\right]\right] - \frac{n+m}{\alpha^2}, \end{aligned}$$

$$-I_{22} = \frac{n}{\theta_1^2} - \frac{2}{\theta_1^3}\sum_{i=1}^{n}(r_i+1)E[X_i^{\alpha}] = \frac{1}{\theta_1^2}\left[n - 2\sum_{i=1}^{n}(r_i+1)\mu_i\right],$$

$$-I_{33} = \frac{m}{\theta_2^2} - \frac{2}{\theta_2^3}\sum_{j=1}^{m}(r'_j+1)E[Y_j^{\alpha}] = \frac{1}{\theta_2^2}\left[m - 2\sum_{j=1}^{m}(r'_j+1)\mu'_j\right],$$

$$-I_{12} = \frac{1}{\theta_1^2}\sum_{i=1}^{n}(r_i+1)E[X_i^{\alpha}\ln(X_i)]$$

$$= \frac{1}{\alpha\theta_1}\sum_{i=1}^{n}(r_i+1)c_{i-1}\sum_{k=1}^{i}\frac{a_{k,i}}{\gamma_k^2}\left[\Gamma'(2) + \ln(\frac{\theta_1}{\gamma_k})\Gamma(2)\right] = -I_{21}.$$

$$-I_{13} = \frac{1}{\theta_2^2}\sum_{j=1}^{m}(r'_j+1)E[Y_j^{\alpha}\ln(Y_j)]$$

$$= \frac{1}{\alpha\theta_2}\sum_{j=1}^{m}(r'_j+1)c'_{j-1}\sum_{k=1}^{j}\frac{a'_{k,j}}{\gamma'^2_k}\left[\Gamma'(2) + \ln(\frac{\theta_2}{\gamma'_k})\Gamma(2)\right] = -I_{31},$$

$$I_{23} = I_{32} = 0,$$

where

$$\mu'_j = \sum_{k=1}^{j}\frac{1}{M - \sum_{s=0}^{k-1} r'_s - k + 1} \qquad \gamma'_k = M - k + 1 + \sum_{s=k}^{m} r'_s$$

$$c'_{j-1} = \prod_{s=1}^{j}\gamma'_s \qquad a'_{k,j} = \prod_{\substack{s=1\\ s\neq k}}^{j}\frac{1}{\gamma'_s - \gamma'_j}.$$

From the asymptotic properties of the MLE's and the fact that the two-parameter Weibull distribution satisfies all the regularity conditions (cf. Bain (1978)), we state the following theorem.

Theorem 1. *For $n \to \infty$ and $m \to \infty$,*

$$\left(\sqrt{m}(\widehat{\alpha} - \alpha),\ \sqrt{n}(\widehat{\theta}_1 - \theta_1),\ \sqrt{n}(\widehat{\theta}_2 - \theta_2)\right) \xrightarrow{d} N_3(0, A^{-1}(\alpha, \theta_1, \theta_2))$$

where

$$A(\alpha, \theta_1, \theta_2) = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix},$$

with

$$a_{11} = \frac{I_{11}}{m},\ a_{22} = \frac{I_{22}}{n},\ a_{33} = \frac{I_{33}}{n},\ a_{12} = a_{21} = \frac{I_{12}}{\sqrt{nm}},\ a_{13} = a_{31} = \frac{I_{13}}{\sqrt{nm}}.$$

Theorem 2. *For $n \to \infty$, $m \to \infty$, $\sqrt{m}(\widehat{R}-R) \to N(0,B)$, where*

$$B = \frac{1}{u(\theta_1+\theta_2)^4}\left[(a_{11}a_{22}-a_{12}^2)\theta_1^2 - 2a_{12}a_{13}\theta_1\theta_2 + (a_{11}a_{33}-a_{13}^2)\theta_2^2\right],$$

and $u = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{31}a_{22}$.

Proof. See the Appendix.

From Theorem 2, we construct the asymptotic C.I. of R. Using the MLE's of α, θ_1 and θ_2, B can be estimated. As a consequence of that, a $100(1-\gamma)\%$ asymptotic C.I. for R can be presented of the form,

$$(\widehat{R} - z_{1-\frac{\gamma}{2}}\frac{\sqrt{\widehat{B}}}{\sqrt{m}}, \widehat{R} + z_{1-\frac{\gamma}{2}}\frac{\sqrt{\widehat{B}}}{\sqrt{m}}),$$

where z_γ is 100γth percentile of $N(0,1)$. The C.I. of R by using the asymptotic distribution of the AMLE of R can be obtained similarly by replacing α, θ_1 and θ_2 in B by their respective AMLE's.
It is of interest to observe that when the shape parameter α is known, the MLE of R can be obtained explicitly as

$$\hat{R}_{ML} = \frac{S_1(\mathbf{x})}{S_1(\mathbf{x}) + \frac{n}{m}S_2(\mathbf{y})}, \tag{20}$$

where $S_1(\mathbf{x}) = \sum_{i=1}^{n}(r_i+1)x_i^\alpha$ and $S_2(\mathbf{y}) = \sum_{j=1}^{m}(r'_j+1)y_j^\alpha$. It is easily checked that $(2/\theta_1)S_1(\mathbf{X})$ and $(2/\theta_2)S_2(\mathbf{Y})$ have chi-square distribution with $2n$ and $2m$ degrees of freedom, respectively. Alternatively, we have

$$\widehat{R}_{ML} \overset{d}{=} \frac{1}{1+\frac{\theta_2}{\theta_1}W},$$

or

$$W \overset{d}{=} \frac{1-\widehat{R}_{ML}}{\widehat{R}_{ML}}.\frac{R}{1-R},$$

where W has an F distribution with $2m$ and $2n$ degrees of freedom. Then, a $100(1-\gamma)\%$ C.I. for R can be presented as

$$\left(\frac{1}{1+\left[\frac{1-\widehat{R}_{ML}}{\widehat{R}_{ML}}\right]F_{1-\frac{\gamma}{2},2n,2m}}, \frac{1}{1+\left[\frac{1-\widehat{R}_{ML}}{\widehat{R}_{ML}}\right]F_{\frac{\gamma}{2},2n,2m}}\right),$$

where $F_{\frac{\gamma}{2},2n,2m}$ and $F_{1-\frac{\gamma}{2},2n,2m}$ are the lower and upper $\frac{\gamma}{2}th$ percentile points of an F distribution with $2n$ and $2m$ degrees of freedom.

3.2. Bootstrap C.I.'s

It is evident that the C.I.'s based on the asymptotic results do not perform very well for small sample size. For this, we propose two C. I.'s based on the non-parametric bootstrap methods: (i) percentile bootstrap method (we call it Boot-p) based on the idea of Efron (1982), and (ii) bootstrap-t method (we refer to it as Boot-t) based on the idea of Hall (1988). We illustrate briefly how to estimate C.I.'s of R using both methods.

(i) Boot-p method

1. Generate a bootstrap sample of size n, $\{x_1^*,\ldots,x_n^*\}$ from $\{x_1,\ldots,x_n\}$, and generate a bootstrap sample of size m, $\{y_1^*,\ldots,y_m^*\}$ from $\{y_1,\ldots,y_m\}$. Based on $\{x_1^*,\ldots,x_n^*\}$ and $\{y_1^*,\ldots,y_m^*\}$, compute the bootstrap estimate of R say $\widehat{R}^*$ using (11).
2. Repeat 1 NBOOT times.
3. Let $H_1(x) = P(\widehat{R}^* \le x)$ be the cumulative distribution function of $\widehat{R}^*$. Define $\widehat{R}_{Bp}(x) = H_1^{-1}(x)$ for a given x. The approximate $100(1-\gamma)\%$ confidence interval of R is given by

$$\left(\widehat{R}_{Bp}(\frac{\gamma}{2})\ ,\ \widehat{R}_{Bp}(1-\frac{\gamma}{2})\right)$$

(ii) Boot-t method

1. From the sample $\{x_1,\ldots,x_n\}$ and $\{y_1,\ldots,y_m\}$, compute $\widehat{R}$.
2. Same as in Boot-p method, first generate bootstrap sample $\{x_1^*,\ldots,x_n^*\}$, $\{y_1^*,\ldots,y_m^*\}$ and then compute $\widehat{R}^*$, the bootstrap estimate of R. Also, compute the statistic

$$T^* = \frac{\sqrt{m}(\widehat{R}^* - \widehat{R})}{\sqrt{Var(\widehat{R}^*)}}$$

   Compute $Var(\widehat{R}^*)$ using Theorem 2.
3. Repeat 1 and 2 NBOOT times.
4. Let $H_2(x) = P(T^* \le x)$ be the cumulative distribution function of T^*. For a given x define $\widehat{R}_{Bt}(x) = \widehat{R} + H_2^{-1}(x)\sqrt{\frac{Var(\widehat{R})}{m}}$. The approximate $100(1-\gamma)\%$ C.I. of R is given by

$$\left(\widehat{R}_{Bt}(\frac{\gamma}{2})\ ,\ \widehat{R}_{Bt}(1-\frac{\gamma}{2})\right).$$

4. Bayes estimation of R

In this section, we obtain the Bayes estimation of R under assumption that the shape parameter α and scale parameters θ_1 and θ_2 are random variables. Following the approach of Berger and Sun (1993), it is assumed that the prior density of θ_j is the inverse gamma $IG(a_j, b_j)$, $j = 1, 2$ with density function

$$\pi_j(\theta_j) = \pi(\theta_j|a_j, b_j) = e^{-\frac{b_j}{\theta_j}} \frac{\theta_j^{-a_j-1} b_j^{a_j}}{\Gamma(a_j)},$$

and α has the gamma $G(a_3, b_3)$ with density function

$$\pi_3(\alpha) = \pi(\alpha|a_3, b_3) = e^{-b_3\alpha} \frac{\alpha^{a_3-1} b_3^{a_3}}{\Gamma(a_3)}.$$

Moreover, it is assumed that θ_1, θ_2 and α are independent. Therefore the joint posterior density of α, θ_1 and θ_2 given the data is

$$L(\alpha, \theta_1, \theta_2|\text{data}) = \frac{L(\text{data}|\alpha, \theta_1, \theta_2)\ \pi(\alpha)\ \pi_1(\theta_1)\ \pi_2(\theta_2)}{\int_0^\infty \int_0^\infty \int_0^\infty L(\text{data}|\alpha, \theta_1, \theta_2)\pi_1(\theta_1)\ \pi_2(\theta_2)\ \pi(\alpha)\ d\alpha\ d\theta_1 d\theta_2}. \quad (21)$$

From (21), it is obvious that the form of the posterior density function will not be tractable and the computation of its respective Bayes estimate will not be analytically obtained. Consequently, we opt for stochastic simulation procedures, namely, the Gibbs and Metropolis samplers (Gilks et al., 1995) to generate samples from the posterior distributions and then compute the Bayes estimate of R and the corresponding credible interval of R. The posterior pdfs of θ_1 and θ_2 are as follows:

$$\theta_1|\alpha, \theta_2, \text{data} \sim IG\left(n + a_1, b_1 + \sum_{i=1}^{n} (r_i + 1) x_i^\alpha\right),$$

$$\theta_2|\alpha, \theta_1, \text{data} \sim IG\left(m + a_2, b_2 + \sum_{j=1}^{m} (r'_j + 1) y_j^\alpha\right),$$

and

$$f(\alpha|\theta_1, \theta_2, \text{data}) \propto \alpha^{n+m+a_3-1} \prod_{i=1}^{n} x_i^{\alpha-1} \prod_{j=1}^{m} y_j^{\alpha-1} \times \exp\left\{-b_3\alpha - \frac{1}{\theta_1} \sum_{i=1}^{n} (r_i + 1) x_i^\alpha - \frac{1}{\theta_2} \sum_{j=1}^{m} (r'_j + 1) y_j^\alpha\right\}.$$

The posterior pdf of α is not known, but the plot of its show that it is similar to normal distribution. So to generate random numbers from this distributions, we use the Metropolis-Hastings method with normal proposal distribution. Therefore the algorithm of Gibbs sampling is as follows:

1. Start with an initial guess $(\alpha^{(0)}, \theta_1^{(0)}, \theta_2^{(0)})$.
2. Set t = 1.
3. Using Metropolis-Hastings, generate $\alpha^{(t)}$ from $f(\alpha|\theta_1^{(t-1)}, \theta_2^{(t-1)}, \text{data})$ with the $N(\alpha^{(t-1)}, 1)$ proposal distribution.
4. Generate $\theta_1^{(t)}$ from $IG(n+a_1, b_1 + \sum_{i=1}^{n}(r_i+1)x_i^{\alpha^{(t-1)}})$.
5. Generate $\theta_2^{(t)}$ from $IG(m+a_2, b_2 + \sum_{j=1}^{m}(r'_j+1)y_j^{\alpha^{(t-1)}})$.
6. Compute $R^{(t)}$ from (3).
7. Set $t = t+1$.
8. Repeat steps 3-7, T times.

Now the approximate posterior mean, and posterior variance of R become

$$\widehat{E}(R|\text{data}) = \frac{1}{T}\sum_{t=1}^{T} R^{(t)},$$

$$\widehat{Var}(R|\text{data}) = \frac{1}{T}\sum_{t=1}^{T}\left(R^{(t)} - \widehat{E}(R|\text{data})\right)^2.$$

Based on T and R values, using the method proposed by Chen and Shao (1999), a $100(1-\gamma)\%$ HPD credible interval can be constructed as $\left(R_{[\frac{\gamma}{2}T]}, R_{[(1-\frac{\gamma}{2})T]}\right)$, where $R_{[\frac{\gamma}{2}T]}$ and $R_{[(1-\frac{\gamma}{2})T]}$ are the $[\frac{\gamma}{2}T]$-th smallest integer and the $[(1-\frac{\gamma}{2})T]$-th smallest integer of $\{R_t,\ t = 1,2,\ldots,T\}$, respectively.

Here we obtain the Bayes estimation of R under the assumptions that the scale parameters θ_1 and θ_2 are random variables and the shape parameter α is known. It is assumed that θ_1 and θ_2 have independent inverted Gamma priors with parameters (a_1, b_1) and (a_2, b_2), respectively. The posterior pdf's of θ_1 and θ_2 can be shown to be IGamma$(n+a_1, b_1+S_1(\mathbf{x}))$ and IGamma$(m+a_2, b_2+S_2(y))$ respectively. Since the priors θ_1 and θ_2 are independent, the posterior pdf of R becomes

$$f_R(z) = A\frac{z^{m+a_2-1}(1-z)^{n+a_1-1}}{[(b_1+S_1(\mathbf{x}))(1-z)+(b_2+S_2(\mathbf{y}))z]^{n+m+a_1+a_2}}, \qquad 0<z<1,$$

where $A = \frac{\Gamma(n+m+a_1+a_2)}{\Gamma(a_1+n)\Gamma(a_2+m)}[b_1+S_1(\mathbf{x})]^{n+a_1}[b_2+S_2(\mathbf{y})]^{m+a_2}$.

The Bayes estimate of R under the squares error loss function is the posterior mean

$$\widehat{R}_{BS} = \int_0^1 z f_R(z)dz. \tag{22}$$

Since the Bayes estimate of R under the squared erros loss can not be obtained analytically, we can obtain the approximate Bayes estimate of R by using the method of Lindley (1980). It can be shown that the approximate Bayes estimate of R, say $\widehat{R}_{ABS}$, under squared error loss function is

$$\widehat{R}_{ABS} = \tilde{R}\left[1 + \frac{\tilde{\theta}_2\tilde{R}^2}{\tilde{\theta}_1^2} \times \frac{(n+a_1-1)\tilde{\theta}_2 - (m+a_2-2)\tilde{\theta}_1}{(n+a_1-1)(m+b_2-1)}\right], \tag{23}$$

where $\tilde{\theta}_1 = \frac{n+a_1-1}{b_1+S_1(\mathbf{x})}$, $\tilde{\theta}_2 = \frac{m+a_2-1}{b_2+S_2(\mathbf{y})}$ and $\tilde{R} = \frac{\tilde{\theta}_1}{\tilde{\theta}_1+\tilde{\theta}_2}$.

The $100(1-\gamma)\%$ Bayesian interval for R is given by (L,U), where L and U are the lower and upper bounds, respectively, satisfying

$$P[R \le L|\text{data}] = \frac{\gamma}{2}, \quad \text{and} \quad P[R \le U|\text{data}] = 1 - \frac{\gamma}{2}.$$

By using some numerical integration methods, we can obtain Bayesian interval estimation of R.

5. Data analysis and comparison study

In this section, a Monte Carlo simulation study and a real data set are presented to illustrate all the estimation methods described in the preceding sections. All the computations are performed using Visual Maple (V12) package. The codes themselves are available from the authors.

5.1. Numerical comparison study

In this subsection, we present some results based on Monte Carlo simulations to compare the performance of the different methods for different censoring schemes, and for different parameter values. We compare the performances of the MLE, AMLE, and the Bayes estimates (with respect to the squared error loss function) in terms of biases, and mean squares errors (MSE). We also compare different C.I.'s, namely the C.I.'s obtained by using asymptotic distributions of the MLE and AMLE, bootstrap C.I.'s and the HPD credible intervals in terms of the average confidence lengths, and coverage percentages. We use different parameter values, different hyper parameters and different sampling schemes. We used three sets of parameter values $(\alpha = 0.5, \theta_1 = 1, \theta_2 = 1)$, $(\alpha = 1.5, \theta_1 = 1, \theta_2 = 1)$ and $(\alpha = 2.5, \theta_1 = 1, \theta_2 = 1)$ mainly to compare the MLEs and different Bayes estimators. For computing the Bayes estimators and HPD credible intervals, we assume 3 priors as follows:

$$\begin{array}{lllll} \text{Prior 1:} & a_j = 0, & b_j = 0, & j = 1,2,3, \\ \text{Prior 2:} & a_j = 1, & b_j = 2, & j = 1,2,3, \\ \text{Prior 3:} & a_j = 2, & b_j = 3, & j = 1,2,3. \end{array}$$

Prior 1 is the non-informative gamma prior for both the shape and scale parameters. Priors 2 and 3 are informative gamma priors. We also use three censoring schemes as given in Table 1.

Table 1: *Censoring schemes.*

	(n,N)	C. S.
r_1	(10, 30)	(0,0,0,0,0,0,0,0,0,20)
r_2	(10, 30)	(20,0,0,0,0,0,0,0,0,0)
r_3	(10, 30)	(2,2,2,2,2,2,2,2,2,2)

For different parameter values, different censoring schemes and different priors, we report the average biases, and MSE of the MLE, AMLE, and Bayes estimates of R over 1000 replications. The results are reported in Table 2. In our simulation experiments for both the bootstrap methods, we have computed the confidence intervals based on 250 re-sampling. The Bayes estimates and the corresponding credible intervals are based on 1000 sampling, namely $T = 1000$.

From Table 2, we observe that the MLE and AMLE compare very well with the Bayes estimator in terms of biases and MSEs. We also observe that the MSE, and biases of the MLE, and AMLE are very close. Comparing the two Bayes estimators based on two informative gamma priors clearly shows that the Bayes estimators based on prior 3 perform better than the Bayes estimators based on prior 2, in terms of both biases and MSEs. The Bayes estimators based on both priors perform better than the ones obtained using the noninformative prior 1.

We also computed the 95% C.I.'s for R based on the asymptotic distributions of the MLE and AMLE. We further compute Boot-p, and Boot-t C.I.'s, and the HPD credible intervals. In Table 3, we presented the average confidence credible lengths, and the corresponding coverage percentages. The nominal level for the C.I.'s or the credible intervals is 0.95 in each case. From Table 3, we observe that the bootstrap C.I.'s are wider than the other C.I.'s. We also observe that the HPD intervals provide the smallest average confidence credible lengths for different censoring schemes, and for different parameter values. The asymptotic C.I.'s MLE and AMLE are the second best CIs. It is also observed that Boot-p C.I.'s perform better than the Boot-t C.I.'s. From Table 3, it is evident that the the Boot-t credible intervals provide the most coverage probabilities in most cases considered.

Table 2: Biases and MSE of the MLE, AMLE, and Bayes estimates of R.

$(\alpha, \theta_1, \theta_2)$	C.S.		MLE	AMLE	BS		
					prior 1	prior 2	prior 3
(0.5 , 1 , 1)	(r_1 , r_1)	Bias	−0.051	−0.053	−0.064	−0.60	−0.055
		MSE	0.019	0.020	0.022	0.021	0.020
	(r_1 , r_2)	Bias	−0.046	−0.050	−0.060	−0.056	−0.051
		MSE	0.012	0.014	0.017	0.016	0.015
	(r_1 , r_3)	Bias	−0.032	−0.035	−0.047	−0.043	−0.039
		MSE	0.020	0.021	0.026	0.024	0.023
	(r_2 , r_2)	Bias	−0.015	−0.022	−0.024	−0.023	−0.021
		MSE	0.011	0.013	0.016	0.016	0.015
	(r_2 , r_3)	Bias	−0.013	−0.015	−0.021	−0.019	−0.018
		MSE	0.013	0.014	0.017	0.016	0.016
	(r_3 , r_3)	Bias	−0.031	−0.036	−0.047	−0.043	−0.038
		MSE	0.012	0.014	0.018	0.017	0.015
(1.5 , 1 , 1)	(r_1 , r_1)	Bias	−0.046	−0.048	−0.060	−0.057	−0.053
		MSE	0.021	0.021	0.025	0.024	0.023
	(r_1 , r_2)	Bias	−0.034	−0.039	−0.050	−0.047	−0.045
		MSE	0.017	0.018	0.023	0.022	0.021
	(r_1 , r_3)	Bias	−0.027	−0.028	−0.034	−0.031	−0.030
		MSE	0.013	0.014	0.018	0.016	0.015
	(r_2 , r_2)	Bias	−0.017	−0.023	−0.034	−0.031	−0.029
		MSE	0.010	0.011	0.017	0.016	0.014
	(r_2 , r_3)	Bias	−0.013	−0.017	−0.031	−0.025	−0.021
		MSE	0.008	0.009	0.015	0.012	0.011
	(r_3 , r_3)	Bias	−0.035	−0.039	−0.049	−0.047	−0.043
		MSE	0.021	0.023	0.024	0.024	0.022
(2.5 , 1 , 1)	(r_1 , r_1)	Bias	−0.029	−0.032	−0.036	−0.034	−0.032
		MSE	0.020	0.022	0.026	0.025	0.023
	(r_1 , r_2)	Bias	−0.022	−0.027	−0.039	−0.037	−0.034
		MSE	0.011	0.013	0.018	0.018	0.016
	(r_1 , r_3)	Bias	−0.017	−0.019	−0.025	−0.024	−0.022
		MSE	0.007	0.008	0.010	0.010	0.009
	(r_2 , r_2)	Bias	−0.016	−0.018	−0.024	−0.023	−0.022
		MSE	0.008	0.009	0.012	0.011	0.011
	(r_2 , r_3)	Bias	−0.014	−0.017	−0.025	−0.021	−0.020
		MSE	0.012	0.015	0.020	0.018	0.017
	(r_3 , r_3)	Bias	−0.033	−0.038	−0.050	−0.047	−0.045
		MSE	0.017	0.018	0.025	0.024	0.022

Table 3: Average confidence/credible length and coverage percentage for estimators of R.

$(\alpha = 0.5, \quad \theta_1 = 1, \quad \theta_2 = 1)$							
C.S.	ML	AML	Boot-p	Boot-t	BS		
					prior 1	prior 2	prior 3
(r_1 , r_1)	0.376(0.925)	0.378(0.926)	0.389(0.937)	0.396(0.951)	0.351(0.950)	0.346(0.950)	0.337(0.947)
(r_1 , r_2)	0.373(0.922)	0.376(0.924)	0.386(0.942)	0.391(0.946)	0.349(0.945)	0.341(0.942)	0.335(0.942)
(r_1 , r_3)	0.361(0.924)	0.365(0.927)	0.372(0.947)	0.382(0.959)	0.342(0.945)	0.338(0.943)	0.334(0.942)
(r_2 , r_2)	0.354(0.947)	0.356(0.951)	0.368(0.948)	0.375(0.957)	0.337(0.952)	0.329(0.951)	0.325(0.948)
(r_2 , r_3)	0.347(0.932)	0.350(0.934)	0.361(0.939)	0.366(0.950)	0.331(0.949)	0.326(0.947)	0.321(0.944)
(r_3 , r_3)	0.377(0.930)	0.380(0.932)	0.389(0.932)	0.391(0.949)	0.345(0.936)	0.342(0.934)	0.338(0.934)
$(\alpha = 1.5, \quad \theta_1 = 1, \quad \theta_2 = 1)$							
C.S.	ML	AML	Boot-p	Boot-t	BS		
					prior 1	prior 2	prior 3
(r_1 , r_1)	0.333(0.941)	0.336(0.943)	0.345(0.947)	0.352(0.954)	0.321(0.948)	0.313(0.945)	0.304(0.947)
(r_1 , r_2)	0.317(0.943)	0.319(0.946)	0.327(0.946)	0.340(0.953)	0.311(0.951)	0.305(0.950)	0.301(0.948)
(r_1 , r_3)	0.308(0.946)	0.311(0.949)	0.320(0.951)	0.334(0.953)	0.301(0.957)	0.295(0.953)	0.285(0.951)
(r_2 , r_2)	0.299(0.940)	0.304(0.944)	0.312(0.944)	0.323(0.951)	0.288(0.954)	0.280(0.952)	0.277(0.950)
(r_2 , r_3)	0.290(0.939)	0.296(0.937)	0.306(0.946)	0.317(0.951)	0.282(0.953)	0.274(0.949)	0.272(0.948)
(r_3 , r_3)	0.322(0.940)	0.325(0.946)	0.332(0.949)	0.340(0.956)	0.313(0.946)	0.307(0.948)	0.303(0.942)
$(\alpha = 2.5, \quad \theta_1 = 1, \quad \theta_2 = 1)$							
C.S.	ML	AML	Boot-p	Boot-t	BS		
					prior 1	prior 2	prior 3
(r_1 , r_1)	0.255(0.935)	0.260(0.938)	0.272(0.949)	0.285(0.951)	0.247(0.947)	0.243(0.942)	0.236(0.946)
(r_1 , r_2)	0.243(0.941)	0.247(0.943)	0.254(0.944)	0.265(0.953)	0.238(0.947)	0.233(0.948)	0.226(0.946)
(r_1 , r_3)	0.221(0.946)	0.225(0.945)	0.237(0.947)	0.249(0.949)	0.215(0.945)	0.211(0.944)	0.205(0.948)
(r_2 , r_2)	0.208(0.947)	0.211(0.951)	0.218(0.948)	0.234(0.953)	0.204(0.951)	0.197(0.953)	0.189(0.950)
(r_2 , r_3)	0.190(0.941)	0.192(0.945)	0.203(0.946)	0.211(0.954)	0.183(0.943)	0.179(0.946)	0.175(0.948)
(r_3 , r_3)	0.244(0.940)	0.247(0.941)	0.255(0.942)	0.268(0.945)	0.241(0.943)	0.239(0.944)	0.232(0.946)

Table 4: Biases and MSE of the MLE and Bayes estimators of R and average confidence length and coverage percentage when α is known and $\theta_1 = \theta_2 = 1$.

	C.S		MLE	BS	ABS	Exact con.
$\alpha = 1$ (exponential case)	(r_1, r_1)	Bias	−0.0016	−0.0015	−0.0026	Mean 0.405
		MSE	0.0122	0.0112	0.0157	Cov.Prob. 0.949
	(r_1, r_2)	Bias	−0.0014	−0.0013	−0.0023	Mean 0.405
		MSE	0.0123	0.0113	0.0158	Cov.Prob. 0.949
	(r_1, r_3)	Bias	0.0007	0.0007	0.0009	Mean 0.405
		MSE	0.0111	0.0101	0.0137	Cov.Prob. 0.957
	(r_2, r_2)	Bias	0.0028	0.0027	0.0061	Mean 0.405
		MSE	0.0122	0.0112	0.0189	Cov.Prob. 0.944
	(r_2, r_3)	Bias	−0.0037	−0.0035	−0.0059	Mean 0.407
		MSE	0.0112	0.0103	0.0181	Cov.Prob. 0.958
	(r_3, r_3)	Bias	−0.0036	−0.0034	−0.0049	Mean 0.405
		MSE	0.0122	0.0112	0.0149	Cov.Prob. 0.950
$\alpha = 2$ (Rayleigh case)	(r_1, r_1)	Bias	0.0029	0.0027	0.0069	Mean 0.405
		MSE	0.0126	0.0116	0.0173	Cov.Prob. 0.947
	(r_1, r_2)	Bias	0.0016	0.0015	0.0028	Mean 0.406
		MSE	0.0113	0.0104	0.0141	Cov.Prob. 0.953
	(r_1, r_3)	Bias	−0.0003	−0.0003	−0.0017	Mean 0.404
		MSE	0.0110	0.0101	0.0128	Cov.Prob. 0.948
	(r_2, r_2)	Bias	−0.0024	−0.0023	−0.0038	Mean 0.406
		MSE	0.0114	0.0105	0.0133	Cov.Prob. 0.950
	(r_2, r_3)	Bias	0.0008	0.0007	0.0019	Mean 0.406
		MSE	0.0115	0.0106	0.0166	Cov.Prob. 0.948
	(r_3, r_3)	Bias	0.0047	0.0045	0.0086	Mean 0.405
		MSE	0.0124	0.0114	0.0167	Cov.Prob. 0.944
$\alpha = 2.5$	(r_1, r_1)	Bias	−0.0039	−0.0037	−0.0080	Mean 0.404
		MSE	0.0128	0.0117	0.0193	Cov.Prob. 0.945
	(r_1, r_2)	Bias	−0.0029	−0.0027	−0.0045	Mean 0.405
		MSE	0.0119	0.0110	0.0132	Cov.Prob. 0.948
	(r_1, r_3)	Bias	0.0002	0.0002	0.016	Mean 0.404
		MSE	0.0110	0.0101	0.0129	Cov.Prob. 0.946
	(r_2, r_2)	Bias	0.0059	0.0057	0.0092	Mean 0.407
		MSE	0.0118	0.0109	0.0153	Cov.Prob. 0.956
	(r_2, r_3)	Bias	0.0021	0.0020	0.0031	Mean 0.405
		MSE	0.0123	0.0112	0.0153	Cov.Prob. 0.944
	(r_3, r_3)	Bias	0.0024	0.0023	0.0055	Mean 0.405
		MSE	0.0122	0.0112	0.0152	Cov.Prob. 0.941

Now let us consider the case when the common shape parameter α is known. In this case, we obtain the MLE of R using (20). Since we do not have any prior information on R, we prefer to use the non-informative prior i.e $a_1 = b_1 = a_2 = b_2 = 0$ to compute Bayes estimates. Under the same prior distributions, we compute Bayes estimates and approximate Bayes estimates of R using (22) and (23), respectively. We report the average biases and MSEs based on 2000 replications. The results are reported in Table 4. From Table 4, comparing the MLE, Bayes and approximate Bayes estimators, we observe that Bayes estimators provides the smallest biases and MSE's. The MLE's are the best second estimators. Comparing different censoring schemes, we observe that the scheme (r_1, r_3) provides the smallest biases and MSEs.

5.2. Example (real data set)

Here we present a data analysis of the strength data reported by Badar and Priest (1982). This data, represent the strength measured in GPA for single carbon fibers, and impregnated 1000-carbon fiber tows. Single fibers were tested under tension at gauge lengths of 20mm (Data Set 1) and 10mm (Data Set 2). These data have been used previously by Raqab and Kundu (2005), Kundu and Gupta (2006) and Kundu and Raqab (2009). The data are presented in Tables 5 and 6.

Table 5: Data Set 1 (gauge lengths of 20 mm).

1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958
1.966	1.997	2.006	2.021	2.027	2.055	2.063	2.098	2.140	2.179
2.224	2.240	2.253	2.270	2.272	2.274	2.301	2.301	2.359	2.382
2.382	2.426	2.434	2.435	2.478	2.490	2.511	2.514	2.535	2.554
2.566	2.570	2.586	2.629	2.633	2.642	2.648	2.684	2.697	2.726
2.770	2.773	2.800	2.809	2.818	2.821	2.848	2.880	2.954	3.012
3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.585	

Table 6: Data Set 2 (gauge lengths of 10 mm).

1.901	2.132	2.203	2.228	2.257	2.350	2.361	2.396	2.397	2.445
2.454	2.474	2.518	2.522	2.525	2.532	2.575	2.614	2.616	2.618
2.624	2.659	2.675	2.738	2.740	2.856	2.917	2.928	2.937	2.937
2.977	2.996	3.030	3.125	3.139	3.145	3.220	3.223	3.235	3.243
3.264	3.272	3.294	3.332	3.346	3.377	3.408	3.435	3.493	3.501
3.537	3.554	3.562	3.628	3.852	3.871	3.886	3.971	4.024	4.027
4.225	4.395	5.020							

Kundu and Gupta (2006) analyzed these data sets using two-parameter Weibull distribution after subtracting 0.75 from both these data sets. After subtracting 0.75 from all the points of these data sets, Kundu and Gupta (2006) observed that the Weibull distributions with equal shape parameters fit to both these data sets. We have generated two different progressively censored samples using two different sampling schemes from above data sets in Tables 5 and 6. The generated data and corresponding censored

Table 7: Data and the corresponding censored schemes.

i, j	1	2	3	4	5	6	7	8	9	10
x_i	1.312	1.479	1.552	1.803	1.944	1.858	1.966	2.027	2.055	2.098
r_i	1	0	1	2	0	0	3	0	1	50
y_j	1.901	2.132	2.257	2.361	2.396	2.445	2.373	2.525	2.532	2.575
r'_j	0	2	1	0	1	1	2	0	0	44

schemes have been presented in Table 7. The ML, AML and Bayes estimations of R become 0.176, 0.179 and 0.328; and the corresponding 95% C.I.'s become (0.069, 0.283), (0.076, 0.284) and (0.097, 0.527) respectively. We also obtain the 95% Boot-p and Boot-t confidence intervals as (0.064, 0.310) and (0.063, 0.342) respectively.

6. Some concluding remarks

Based on progressively censored samples, this paper considers estimation of $R = P(Y < X)$ by different methods when X and Y are two independent Weibull distributions with different scale parameters, but having the same shape parameter. It is observed that the MLE of R can be obtained using an iterative procedure. The proposed AMLE of R can be obtained explicitly. It is observe that the MSE, and biases of the MLE, and AMLE are very close. The Bayes estimate of R, and the corresponding credible interval can be obtained using the Gibbs sampling technique. It is also observe that the MLE and AMLE compare very well with the Bayes estimator in terms of biases and MSEs. Note that the results for exponential and Rayleigh distributions can be obtained as special cases with different scale parameters.

Appendix

Proof of Theorem 2

On using Theorem 1 and applying delta method, we can describe the asymptotic distribution of $\widehat{R} = g(\widehat{\alpha}, \widehat{\theta}_1, \widehat{\theta}_2)$, where $g(\alpha, \theta_1, \theta_2) = \theta_1/(\theta_1 + \theta_2)$ as the following:

$$\sqrt{m}(\widehat{R} - R) \xrightarrow{D} N(0, B),$$

where $B = \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}$, with

$$\mathbf{b} = \begin{pmatrix} \frac{\partial g}{\partial \alpha} \\ \frac{\partial g}{\partial \theta_1} \\ \frac{\partial g}{\partial \theta_2} \end{pmatrix} = \frac{1}{(\theta_1+\theta_2)^2} \begin{pmatrix} 0 \\ \theta_2 \\ -\theta_1 \end{pmatrix},$$

$$\mathbf{A}^{-1} = \frac{1}{u} \begin{pmatrix} a_{22}a_{33} & -a_{12}a_{33} & -a_{22}a_{13} \\ -a_{21}a_{33} & a_{11}a_{33} - a_{13}a_{31} & a_{21}a_{13} \\ -a_{22}a_{31} & a_{21}a_{13} & a_{11}a_{22} - a_{12}a_{21} \end{pmatrix}.$$

and $u = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{31}a_{22}$. Therefore

$$B = \mathbf{b}'\mathbf{A}^{-1}\mathbf{b} = \frac{1}{u(\theta_1+\theta_2)^4}\left[(a_{11}a_{22}-a_{12}^2)\theta_1^2 - 2a_{12}a_{13}\theta_1\theta_2 + (a_{11}a_{33}-a_{13}^2)\theta_2^2\right].$$

The proof is thus obtained.

Acknowledgements

The authors are thankful to the referees for their helpful comments and suggestions that led to substantive improvements in the article. The third author is grateful to the University of Jordan for supporting this research work.

References

Adimari, G. and Chiogna, M. (2006). Partially parametric interval estimation of $Pr(> X)$, *Computational Statistics and Data Analysis* 51, 1875–1891.

Ahmad, K. E., Fakhry, M. E. and Jaheen, Z. F. (1997). Empirical Bayes estimation of $P(Y < X)$ and characterizations of Burr-type X model, *Journal of Statistical Planning and Inference*, 64, 297–308.

Awad, A. M., Azzam, M. M. and Hamdan, M. A. (1981). Some inference results on $P(Y < X)$ in the bivariate exponential model, *Communications in Statistics-Theory and Methods*, 10, 2515–2525.

Badar, M. G. and Priest, A. M. (1982). Statistical aspects of fiber and bundle strength in hybrid composites. In: Hayashi, T., Kawata, K., Umekawa, S. (Eds.), *Progress in Science and Engineering Composites*, ICCM-IV, Tokyo, 1129–1136.

Bain, L. J. (1978). *Statistical Analysis of Reliability and Life Testing Model*, NewYork: Marceland Dekker Inc.

Baklizi, A. (2008). Likelihood and Bayesian estimation of $Pr(X < Y)$ using lower record values from the generalized exponential distribution, *Computational Statistics and Data Analysis*, 52, 3468–3473.

Balakrishnan, N. and Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods and Applications*, Boston: Birkhäuser.

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating graph, *Journal of Mathematical Psychology*, 12, 387–415.

Berger, J. O. and Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, 88, 1412–1418.

Chen, M. H. and Shao, Q. M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals, *Journal of Computational and Graphical Statistics*, 8, 69–92.

Efron, B. (1982). The jackknife, the bootstrap and other re-sampling plans, *Philadelphia, PA: SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics*, 34.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals, *Annals of Statistics*, 16, 927–953.

Kamps, U. and Cramer, E. (2001). On distributions of generalized order statistics, *Statistics*, 35(3), 269–280.

Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. New York: World Scientific.

Kundu, D. and Gupta, R. D. (2005). Estimation of $P(Y < X)$ for the generalized exponential distribution, *Metrika*, 61, 291–308.

Kundu, D. and Gupta, R. D. (2006). Estimation of $P(Y < X)$ for Weibull distributions, *IEEE Transactions on Reliability*, 55(2), 270–280.

Kundu, D. and Raqab, M. Z. (2009). Estimation of $R = P(Y < X)$ for three-parameter Weibull distribution, *Statistics and Probability Letters*, 79, 1839–1846.

Lindley, D. V. (1980). Approximate Bayesian method. *Trabajos de Estadística y de Investigació*n Operativa 31, 223–237.

Raqab, M. Z. and Kundu, D. (2005). Comparison of different estimators of $P(Y < X)$ for a scaled Burr type X distribution, *Communications in Statistics-Simulation and Computation*, 34(2), 465–483.

Raqab, M. Z, Madi, T. and Kundu, D. (2008). Estimation of $P(Y < X)$ for the three-parameter generalized exponential distribution, *Communications in Statistics-Theory and Methods*, 37, 2854–2865.

Rezaei. S., Tahmasbi. and R. Mahmoodi. M. (2010). Estimation of $P(Y < X)$ for generalized Pareto distribution, *Journal of Statistical Planning and Inference*, 140, 480–494.

SORT 35 (2) July-December 2011, 125-144

Diagnostic measures for linear mixed measurement error models

Karim Zare and Abdolrahman Rasekh*

Abstract

In this paper, we present case deletion and mean shift outlier models for linear mixed measurement error models using the corrected likelihood of Nakamura (1990). We derive the corrected score test statistic for outliers detection based on mean shift outlier models. Furthermore, several case deletion diagnostics are constructed as a tool for influence diagnostics. It is found that they can be written in terms of studentized residuals of model, error contrast matrix and the inverse of the response variable covariance matrix. Our influence diagnostics are illustrated through a real data set.

MSC: 62J20 and 62J99

Keywords: Case deletion, corrected score method, influential points, linear mixed measurement error models, mean shift outlier model.

1. Introduction

Since all the observations in a data set do not play an equal role in determining estimators, tests and other statistics, it is important to consider influential points in data analysis. To identify anomalous observations, various approaches, including case deletion model (CDM) and mean shift outlier model (MSOM), have been proposed in the literature (Cook and Weisberg, 1982).

In linear mixed models, CDM, MSOM and related diagnostics are studied more widely by different authors including, Christensen et al. (1992), Banerjee and Frees (1997), Zhong and Wei (1999), Haslett and Dillane (2004), Zewotir and Galpin (2005) and Li et al. (2009).

*Address for correspondence: Department of Statistics, Shahid Chamran University, Ahvaz, Iran.
Rasekh_a@scu.ac.ir, Tel: +98-611-3331043, Fax: +98-611-3331043.
Received: March 2011
Accepted: August 2011

Christensen et al. (1992) proposed case deletion diagnostics for both fixed effects and variance components. Banerjee and Frees (1997) proposed case deletion diagnostics for both fixed effects and random subject effects in linear longitudinal models. Zhong and Wei (1999) presented a unified diagnostic method for linear mixed models based upon the joint likelihood given by Robinson (1991). They showed that the estimates of parameters in CDM are equivalent to those in MSOM. Haslett and Dillane (2004) proved a 'delete = replace' identity in linear models and applied it to deletion diagnostics for estimators of variance components. Zewotir and Galpin (2005) provided routine diagnostic tools for fixed effects, random effects and variance components, which are computationally inexpensive. Li et al. (2009) considered subset deletion diagnostics for fixed effects, random effects and one variance component in varying coefficient mixed models.

As pointed out by Davidian and Giltinan (1995), independent variables in the models are often measured with non-negligible errors. Hence it is of great interest to study the measurement error models. On regression diagnostics for linear measurement error models, only some works has been done by Kelly (1984), Fuller (1987), Wellman and Gunst (1991), Zhong et al. (2000). Zhong et al. (2000) obtained CDM and MSOM for linear measurement error models. Also, they derived several diagnostics via CDM.

In linear mixed measurement error models the only work is due to Fung et al. (2003). However, in this paper, the corrected score function and the other relevant relations are not derived correctly. This problem also exists in diagnostic methods such as case deletion diagnostic on fixed effects. Furthermore, some of the relations in Fung et al. (2003) are somewhat different with Zhong et al. (2002).

Since there is no outstanding work in diagnostic methods for linear mixed measurement error models, in this paper, we concentrate on diagnostic methods for these models upon the corrected score function of Nakamura (1990). In Section 2, we present the model and the corrected score method for estimation of parameters. By using the corrected score method, Section 3 deals with two diagnostic models: CDM and MSOM. Besides, since MSOM is efficient to detect outliers, we construct a corrected score test for detecting outliers. In Section 4, we develop case deletion diagnostics for detecting influential points in linear mixed measurement error models. The given diagnostics are similar to diagnostics in linear mixed models and so are easy to compute. An influence analysis of a data set on hedonic housing-prices is given to illustrate the results in Section 5. Concluding remarks are given in Section 6.

2. Model definition and estimation

Consider the following linear mixed model with measurement errors in fixed effects:

$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\beta} + \mathbf{U}\,\mathbf{b} + \boldsymbol{\varepsilon}, \\ \mathbf{X} &= \mathbf{Z} + \boldsymbol{\Delta}. \end{aligned} \tag{1}$$

In this model $\boldsymbol{\beta}$ is a $p \times 1$ vector of unobservable parameters, which are called fixed effects; $\mathbf{Z}$ and $\mathbf{U} = [\mathbf{U}_1 | \mathbf{U}_2 | \ldots | \mathbf{U}_m]$ are $n \times p$ and $n \times q$ matrices of "regressors", respectively, where $\mathbf{U}_i$ is an $n \times q_i$ known design matrix of the random effect factor i; $\mathbf{b}^{\mathsf{T}} = \left(\mathbf{b}_1^{\mathsf{T}}, \mathbf{b}_2^{\mathsf{T}}, \ldots, \mathbf{b}_m^{\mathsf{T}}\right)$, where $\mathbf{b}_i$ is a $q_i \times 1$ vector of unobservable random effects from $N\left(\mathbf{0}, \sigma_i^2 \mathbf{I}\right)$, $i = 1, \ldots, m$; $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of unobservable random errors from $N\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$. The variances σ^2 and σ_i^2, $i = 1, \ldots, m$ are called variance components. $\mathbf{X}$ is the observed value of $\mathbf{Z}$ with the measurement error $\boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ is an $n \times p$ random matrix from $N(\mathbf{0}, \mathbf{I} \otimes \boldsymbol{\Lambda})$. We assume that $\mathbf{b}_i, \boldsymbol{\varepsilon}$ and $\boldsymbol{\Delta}$ are mutually independent. One may also write $\mathbf{b} \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\Sigma}$ is a block diagonal matrix with the ith block being $\gamma_i \mathbf{I}$, for $\gamma_i = \sigma_i^2 / \sigma^2$, so that $\mathbf{y}$ has a multivariate normal distribution with $E(\mathbf{y}) = \mathbf{Z}\boldsymbol{\beta}$ and $\mathrm{Var}(\mathbf{y}) = \sigma^2 \mathbf{V}$, in which $\mathbf{V} = \mathbf{I} + \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}} = \mathbf{I} + \sum_{i=1}^{m} \gamma_i \mathbf{U}_i \mathbf{U}_i^{\mathsf{T}}$. The conditional distribution of $\mathbf{b}|\mathbf{y}$ is $\mathbf{b}|\mathbf{y} \sim N\left(\boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}), \sigma^2 \boldsymbol{\Sigma}\mathbf{T}\right)$ where $\mathbf{T} = \left(\mathbf{I} + \mathbf{U}^{\mathsf{T}}\mathbf{U}\boldsymbol{\Sigma}\right)^{-1}$. The log-likelihood of $\mathbf{y}$ is given by

$$l\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}; \mathbf{Z}, \mathbf{y}\right) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log\left(|\mathbf{V}|\right) - \frac{1}{2\sigma^2}\left[(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})^{\mathsf{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right],$$

where $\left(\sigma^2, \boldsymbol{\gamma}\right) = \left(\sigma^2, \gamma_1, \ldots, \gamma_m\right)$ belongs to $\Omega = \left\{\left(\sigma^2, \boldsymbol{\gamma}\right) : \sigma^2 > 0, \gamma_i \geq 0 \ (i = 1, \ldots, m)\right\}$. Also, the conditional log-likelihood of $\mathbf{b}|\mathbf{y}$ is given by

$$l_{\mathbf{b}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}; \mathbf{Z}, \mathbf{y}\right) = -\frac{q}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log\left(|\boldsymbol{\Sigma}\mathbf{T}|\right)$$
$$- \frac{1}{2\sigma^2}\left\{\left[\mathbf{b} - \boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right]^{\mathsf{T}}(\boldsymbol{\Sigma}\mathbf{T})^{-1}\left[\mathbf{b} - \boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})\right]\right\}.$$

Suppose that as in the model (1), the covariate $\mathbf{Z}$ is measured with error and the correlated structure arises from the random effects. If we simply replace $\mathbf{Z}$ by $\mathbf{X}$, then the estimates obtained from the score functions are not consistent in general. Various ways are proposed in dealing with measurement error models. In this paper, we use corrected score method proposed by Nakamura (1990) that is a common approach in measurement error models (see also Nakamura, 1992; Hanfelt and Liang, 1997; Gimenz and Bolfarine, 1997 and Zhong et al., 2000). In this method, we have to find a corrected score function whose expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true independent variables. For the model (1), Zhong et al. (2002) derived the corrected score estimates of fixed and random effects.

Let E^* denotes the conditional mean with respect to $\mathbf{X}$ given $\mathbf{y}$. The corrected log-likelihood $l^*\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}; \mathbf{X}, \mathbf{y}\right)$ for our model should satisfy

$$E^*\left[\partial l^*\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}; \mathbf{X}, \mathbf{y}\right) / \partial \boldsymbol{\beta}\right] = \partial l\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\gamma}; \mathbf{Z}, \mathbf{y}\right) / \partial \boldsymbol{\beta},$$
$$E^*\left[\partial l_1^*\left(\sigma^2, \boldsymbol{\gamma}; \mathbf{X}, \mathbf{y}\right) / \partial \sigma^2\right] = \partial l_1\left(\sigma^2, \boldsymbol{\gamma}; \mathbf{Z}, \mathbf{y}\right) / \partial \sigma^2$$

and

$$E^*\left[\partial l_1^*\left(\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\gamma_i\right]=\partial l_1\left(\sigma^2,\boldsymbol{\gamma};\mathbf{Z},\mathbf{y}\right)/\partial\gamma_i,\quad i=1,\ldots,m,$$

where $l_1\left(\sigma^2,\boldsymbol{\gamma};\mathbf{Z},\mathbf{y}\right)=l\left(\hat{\boldsymbol{\beta}}\left(\boldsymbol{\gamma}\right),\sigma^2,\boldsymbol{\gamma};\mathbf{Z},\mathbf{y}\right)$, in which $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}\left(\boldsymbol{\gamma}\right)$ is maximum likelihood estimate of $\boldsymbol{\beta}$ and $l_1^*\left(\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)=l^*\left(\hat{\boldsymbol{\beta}}\left(\boldsymbol{\gamma}\right),\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)$, in which $\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}\left(\boldsymbol{\gamma}\right)$ is the solution of the equation $\partial l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}=\mathbf{0}$. Also, the conditional corrected log-likelihood $l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)$ should satisfy

$$E^*\left[\partial l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\right]=\partial l_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{Z},\mathbf{y}\right)/\partial\mathbf{b}.$$

The following equation is useful to find such l^* and $l_{\mathbf{b}}^*$,

$$E^*\left(\mathbf{X}^\mathsf{T}\mathbf{A}\mathbf{X}\right)=\mathbf{Z}^\mathsf{T}\mathbf{A}\mathbf{Z}+\mathrm{tr}\left(\mathbf{A}\right)\boldsymbol{\Lambda}.$$

Given $\boldsymbol{\Lambda}$, l^* and $l_{\mathbf{b}}^*$ are obtained as

$$\begin{aligned}l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)=&-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\mathbf{V}|\right)\\&-\frac{1}{2\sigma^2}\left\{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^\mathsf{T}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\beta}^\mathsf{T}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}\end{aligned}$$

and

$$\begin{aligned}l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)=&-\frac{q}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\boldsymbol{\Sigma}\mathbf{T}|\right)\\&-\frac{1}{2\sigma^2}\left\{\left[\mathbf{b}-\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]^\mathsf{T}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\left[\mathbf{b}-\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right]\right.\\&\left.-\mathrm{tr}\left(\mathbf{I}-\mathbf{V}^{-1}\right)\boldsymbol{\beta}^\mathsf{T}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}.\end{aligned}$$

If the γ_i's (and hence $\mathbf{V}$) are known, by solving the equations $\partial l^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}=\mathbf{0}$, $\partial l_1^*\left(\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\sigma^2=0$ and $\partial l_{\mathbf{b}}^*\left(\boldsymbol{\beta},\sigma^2,\boldsymbol{\gamma};\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}=\mathbf{0}$, the corrected score estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$, respectively, are given by (See Zhong et al., 2002 and Zare et al., 2011 for more details)

$$\begin{aligned}\hat{\boldsymbol{\beta}}&=\left(\mathbf{X}^\mathsf{T}\mathbf{V}^{-1}\mathbf{X}-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{X}^\mathsf{T}\mathbf{V}^{-1}\mathbf{y},\\\hat{\sigma}^2&=\frac{1}{n}\left[\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^\mathsf{T}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\hat{\boldsymbol{\beta}}^\mathsf{T}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}\right],\\\tilde{\mathbf{b}}&=\boldsymbol{\Sigma}\mathbf{U}^\mathsf{T}\mathbf{V}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right).\end{aligned}$$

If the γ_i's are unknown, the corrected score estimates are substituted back into $\boldsymbol{\Sigma}$ to obtain $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\tilde{\mathbf{b}}$. For the estimation of γ_i's, we can use the corrected score estimates of $\sigma_1^2,\ldots,\sigma_m^2$ that are given by (Zare et al., 2011)

$$\hat{\sigma}_i^2 = \frac{1}{q_i - \operatorname{tr}(\mathbf{T}_{ii})}\left[\tilde{\mathbf{b}}_i^{\top}\tilde{\mathbf{b}}_i - \operatorname{tr}\left(\hat{\mathbf{D}}_i^{\top}\hat{\mathbf{D}}_i\right)\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}\right], \quad i = 1,\ldots,m,$$

where $\mathbf{T}_{ij}$ is ijth block of matrix $\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \cdots & \mathbf{T}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{T}_{m1} & \cdots & \mathbf{T}_{mm} \end{bmatrix}$, $\hat{\mathbf{D}}_i = \hat{\gamma}_i\mathbf{U}_i^{\top}\mathbf{V}^{-1} = \left(\hat{\sigma}_i^2/\hat{\sigma}^2\right)\mathbf{U}_i^{\top}\mathbf{V}^{-1}$ and $\tilde{\mathbf{b}}_i = \hat{\mathbf{D}}_i\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$.

The above results show that we must use an iterative numerical procedure to obtain the corrected score estimates of parameters. We use the iterative algorithm given in Zare et al. (2011). Also, Zare et al. (2011) showed the corrected score estimates of γ_i's are consistent. In continuing, we assume that the γ_i's are known.

For notational simplicity, $\mathbf{A}_{(i)}$ denotes an $n \times m$ matrix $\mathbf{A}$ with ith row removed, $\mathbf{A}_{[i]}$ denotes a matrix $\mathbf{A}$ with the ith row and column removed, $\mathbf{a}_i^{\top}$ denotes the ith row of $\mathbf{A}$ and a_{ij} denotes the ijth element of $\mathbf{A}$. Similarly, $\mathbf{a}_{(i)}$ denotes vector $\mathbf{a}$ with the ith element removed and a_i denotes the ith element of $\mathbf{a}$. Without loss of generality, we partition the matrices as if the ith deleted case is the first row; i.e. $i = 1$. Then

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_i^{\top} \\ \mathbf{X}_{(i)} \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} \mathbf{z}_i^{\top} \\ \mathbf{Z}_{(i)} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_i \\ \mathbf{y}_{(i)} \end{bmatrix} \text{ and } \mathbf{C} = \mathbf{V}^{-1} = \begin{bmatrix} c_{ii} & \mathbf{c}_{i(i)}^{\top} \\ \mathbf{c}_{i(i)} & \mathbf{V}_{[i]}^{-1} + \mathbf{c}_{i(i)}\mathbf{c}_{i(i)}^{\top}/c_{ii} \end{bmatrix}.$$

3. Mean shift outlier and case deletion model

In regression diagnostics, there are two commonly used models: CDM amd MSOM (Cook and Weisberg, 1982). Each of models has its own advantage in practice. CDM's are used to obtain case deletion diagnostics for detecting influential observations. MSOM's are used for detecting outlier observations. It is well known that in linear (mixed) models maximum likelihood estimates of parameters in CDM and MSOM are equal. In linear measurement error models, the estimates are approximately equal.

3.1. Mean shift outlier model

A commonly used diagnostic model is MSOM (Cook and Weisberg, 1982). MSOM can be represented as

$$\begin{aligned} &y_j = \mathbf{z}_j^{\top}\boldsymbol{\beta} + \mathbf{u}_j^{\top}\mathbf{b} + \varepsilon_j \quad \text{for} \quad j \neq i,\ j = 1,\ldots,n,\ y_i = \mathbf{z}_i^{\top}\boldsymbol{\beta} + \mathbf{u}_i^{\top}\mathbf{b} + \tau + \varepsilon_i, \\ &\mathbf{x}_k^{\top} = \mathbf{z}_k^{\top} + \boldsymbol{\delta}_k^{\top} \quad \text{for} \quad k = 1,\ldots,n, \end{aligned} \tag{2}$$

where τ is an extra parameter to indicate the presence of an outlier (Cook and Weisberg, 1982). Obviously, if value of τ is nonzero, then it no longer comes from the original model, and so ith case may be an outlier. An outlier test can be formulated as a test of the null hypothesis that $\tau = 0$. The corrected likelihood estimates of $\boldsymbol{\beta}$, σ^2, τ, and $\mathbf{b}$ in (2) are denoted by $\hat{\boldsymbol{\beta}}_{mi}$, $\hat{\sigma}^2_{mi}$, $\hat{\tau}_{mi}$ and $\tilde{\mathbf{b}}_{mi}$, respectively.

Theorem 1 *For model (2), we have*

$$\hat{\boldsymbol{\beta}}_{mi} = \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - tr\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

$$\hat{\tau}_{mi} = \frac{\hat{v}_i}{r_{ii}}, \quad \hat{\sigma}^2_{mi} = \frac{n - t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n}\hat{\sigma}^2 \quad \text{and} \quad \tilde{\mathbf{b}}_{mi} = \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\frac{\hat{v}_i}{r_{ii}},$$

where $\mathbf{c}_i^{\mathsf{T}}$ *and* $\mathbf{r}_i^{\mathsf{T}}$ *are ith rows of* $\mathbf{V}^{-1}$ *and* $\mathbf{R} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - tr\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}$*, respectively,* c_{ii} *and* r_{ii} *are the ith diagonal elements of* $\mathbf{V}^{-1}$ *and* $\mathbf{R}$*,* $\hat{v}_i = y_i - \mathbf{x}_i^{\mathsf{T}}\hat{\boldsymbol{\beta}} - \mathbf{u}_i^{\mathsf{T}}\tilde{\mathbf{b}}$ *is ith residual of model and* $t_i = \hat{v}_i/\left(\hat{\sigma}_v\sqrt{r_{ii}}\right)$ *is ith studentized residual of model, in which* $\hat{\sigma}_v^2 = \hat{\sigma}^2 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}$.

Theorem 2 *For MSOM, the score test statistic for the hypothesis* $H_0 : \tau = 0$ *is given by*

$$SC_i = \frac{\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}} = t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right).$$

This theorem shows that score statistic SC_i is a multiple of the square of studentized residual of model that is an adequate diagnostic statistic as often used in linear regression diagnostics.

3.2. Case deletion model

As mentioned, CDM is the fundamental basis for constructing effective diagnostics. CDM can be represented as

$$y_j = \mathbf{z}_j^{\mathsf{T}}\boldsymbol{\beta} + \mathbf{u}_j^{\mathsf{T}}\mathbf{b} + \varepsilon_j,\ \mathbf{x}_j^{\mathsf{T}} = \mathbf{z}_j^{\mathsf{T}} + \boldsymbol{\delta}_j^{\mathsf{T}} \quad \text{for} \quad j \neq i,\ j = 1, \ldots, n.$$

Let $\hat{\boldsymbol{\beta}}_{(i)}$, $\hat{\sigma}^2_{(i)}$ and $\tilde{\mathbf{b}}_{(i)}$ denote the estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$ when the ith case is deleted, respectively.

Theorem 3 *For model (3), we have*

$$\hat{\boldsymbol{\beta}}_{(i)} \approx \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

$$\hat{\sigma}^2_{(i)} \approx \frac{n}{n-1}\hat{\sigma}^2 - \frac{\hat{v}_i^2}{(n-1)\,r_{ii}} = \frac{n - t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\hat{\sigma}^2,$$

$$\tilde{\mathbf{b}}_{(i)} \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\,\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\frac{\hat{v}_i}{r_{ii}}.$$

Comparing results of the theorems 1 and 3, it is obvious that the estimates of the parameters are approximately the same. In the following section we derive different diagnostic measures based on CDM.

4. Influence diagnostics

It is well known that results from an analysis can be substantially influenced by one or a few observations; that is, all the observations have not equal effect in statistical models. Case deletion diagnostics are the usual methods to measure the influence of individual observations in the statistical models with dropping the observation from data set and computing a convenient norm of the change in the parameters. Let the corrected Fisher information matrix of $\mathbf{y}$ for $\boldsymbol{\beta}$ be $\mathbf{I}^*(\boldsymbol{\beta})$, then

$$\mathbf{I}^*(\boldsymbol{\beta}) = \frac{1}{\sigma^2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right].$$

Also, the corrected Fisher information matrix of $\mathbf{y}$ for $\mathbf{b}$ is

$$\mathbf{I}^*(\mathbf{b}) = \frac{1}{\sigma^2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U} + \boldsymbol{\Sigma}^{-1}\right).$$

4.1. Analogue of generalized Cook's distance

4.1.1. Analogue of generalized Cook's distance for fixed effects

The generalized Cook (1977) distance is the norm of $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$ with respect to certain weight matrix $\mathbf{M} > \mathbf{0}$, i.e.

$$CD_i(\boldsymbol{\beta}) = \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\mathbf{M}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right).$$

Choosing $\mathbf{M} = \hat{\mathbf{I}}^*(\boldsymbol{\beta}) = \hat{\sigma}^{-2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]$, where $\hat{\mathbf{I}}^*(\boldsymbol{\beta})$ is estimate of $\mathbf{I}^*(\boldsymbol{\beta})$, yields

$$CD_i(\boldsymbol{\beta}) = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}^2}.$$

Since

$$\hat{\boldsymbol{\beta}}_{(i)} \approx \hat{\boldsymbol{\beta}} - \left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

we can get, approximately,

$$CD_i\left(\boldsymbol{\beta}\right) = \frac{\left(c_{ii} - r_{ii}\right)\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}^2} = \frac{c_{ii} - r_{ii}}{r_{ii}}t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right).$$

Let $\mathbf{d}_k$ be a p-vector with 1 at the kth position and zero elsewhere, then $\hat{\beta}_k = \mathbf{d}_k^{\mathsf{T}}\hat{\boldsymbol{\beta}}$ has the standard error s_k and t-value $t_k = \hat{\beta}_k/s_k$, where $s_k^2 = \hat{\sigma}^2\mathbf{d}_k^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{d}_k$. The joint $100\left(1-\alpha\right)\%$ confidence region for parameter $\boldsymbol{\beta}$ is

$$\left\{\boldsymbol{\beta} : \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \leq p\hat{\sigma}^2 F\left(p, n-p, \alpha\right)\right\},$$

where $F\left(p, n-p, \alpha\right)$ denotes the upper α percentile of the Fisher's distribution with p and $n-p$ degrees of freedom. Suppose that $CD_i\left(\boldsymbol{\beta}\right) \simeq pF\left(p, n-p, \alpha\right)$, then the removal of the ith case moves corrected score estimate to the edge of the $100\left(1-\alpha\right)\%$ confidence region. Such a situation may be cause for concern and so more attention should be paid to that case. Usually, one would like each $\hat{\boldsymbol{\beta}}_{(i)}$ to stay well within a 90%, say, confidence region. Then case i can be considered a highly influential point if

$$CD_i\left(\boldsymbol{\beta}\right) > pF\left(p, n-p, 0.1\right).$$

Based on Cook (1977), we have $CD_i\left(\beta_k\right) = t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)G_i^2\left(\mathbf{d}_k^{\mathsf{T}}\right)$, where, for any $q' \times p$ matrix $\mathbf{A}$ of rank q', $G_i\left(\mathbf{A}\right)$ is defined as

$$G_i\left(\mathbf{A}\right) = \frac{1}{\sqrt{r_{ii}}}\left[\mathbf{A}\left(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{A}^{\mathsf{T}}\right]^{-1/2}\mathbf{A}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i.$$

As similar, case i can be considered a highly influential point if $CD_i\left(\beta_k\right) > F\left(1, n-1, 0.1\right)$, since this case, if deleted, would move the estimate of $\hat{\beta}_k$ to the edge of the 90% confidenc region. Rio (1988) argued that $G_i^2\left(\mathbf{u}^{\mathsf{T}}\right)$ can be used to measure the influence of case i on the precision of the estimation of $\mathbf{u}^{\mathsf{T}}\boldsymbol{\beta}$. Therefore, based on Rio (1988), ith case is said to have high influential on the estimate of $\hat{\beta}_k$ if $G_i^2\left(\mathbf{d}_k^{\mathsf{T}}\right)$ is sufficiently large.

Let $\mathbf{A}$ denote a $q' \times p$ rank q' matrix and let $\Psi = \mathbf{A}\boldsymbol{\beta}$ denote the combinations of interest. A generalized measure of the importance of the ith case is defined as

$$CD_i\left(\Psi\right) = \frac{\left(\hat{\Psi}_{(i)} - \hat{\Psi}\right)^{\mathsf{T}}\left[\mathbf{A}\left(\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right)^{-1}\mathbf{A}^{\mathsf{T}}\right]^{-1}\left(\hat{\Psi}_{(i)} - \hat{\Psi}\right)}{q'\hat{\sigma}^2},$$

where $\hat{\Psi}_{(i)} = \mathbf{A}\hat{\boldsymbol{\beta}}_{(i)}$ and $\hat{\Psi} = \mathbf{A}\hat{\boldsymbol{\beta}}$. Since

$$\hat{\Psi}_{(i)} - \hat{\Psi} \approx \mathbf{A}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},$$

we can get, approximately,

$$CD_i(\Psi) = \frac{t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{q'}G_i^{\mathsf{T}}(\mathbf{A})\,G_i(\mathbf{A}).$$

To obtain the levels of significance the values of this generalized measure should be compared to the probability points of the central Fisher distribution with q' and $n - q'$ degrees of freedom.

4.1.2. Analogue of generalized Cook's distance for random effects

The proposed diagnostic measure examines the squared distance from the complete data predictor of the random effects to *i*th case deleted predictor of the random effects, relative to $\mathbf{M} = \hat{\mathbf{I}}^{*}(\mathbf{b}) = \hat{\sigma}^{-2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U} + \boldsymbol{\Sigma}^{-1}\right)$. This is the generalized Cook distance and can be written as

$$CD_i(\mathbf{b}) = \left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}}\mathbf{M}\left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right) = \frac{\left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U} + \boldsymbol{\Sigma}^{-1}\right)\left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)}{\hat{\sigma}^2}.$$

Since $\tilde{\mathbf{b}}_{(i)} \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma}\,\mathbf{U}^{\mathsf{T}}\mathbf{r}_i\frac{\hat{v}_i}{r_{ii}}$, we can get, approximately,

$$\begin{aligned} CD_i(\mathbf{b}) &= \mathbf{r}_i^{\mathsf{T}}(\mathbf{V} - \mathbf{I})\mathbf{V}\mathbf{r}_i\frac{\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}^2} \\ &= \mathbf{r}_i^{\mathsf{T}}(\mathbf{V} - \mathbf{I})\mathbf{V}\mathbf{r}_i\frac{t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{r_{ii}}. \end{aligned} \tag{4}$$

Also, from (4) we have

$$CD_i(\mathbf{b}_j) = \mathbf{r}_i^{\mathsf{T}}\mathbf{U}_j\mathbf{T}_{jj}^{-1}\mathbf{U}_j^{\mathsf{T}}\mathbf{r}_i\frac{\gamma_j t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{r_{ii}}, \quad \text{for} \quad j = 1, \ldots, m.$$

4.2. Analogue of Welsch's distance

Welsch (1982) has suggested using Welsch's distance as a diagnostic tool and, for $n > 15$, using $3\sqrt{p}$ as a cutoff point for linear models. Welsch's distance gives more emphasize to high leverage points. It has similar rationale as Cook's distance Essential difference between these two methods is in the choice of scale. (Chatterjee and Hadi, 1986). For the fixed effects it is given as

$$W_i(\boldsymbol{\beta}) = \left[(n-1) \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}} \left(\mathbf{X}_{(i)}^{\mathsf{T}} \mathbf{V}_{[i]}^{-1} \mathbf{X}_{(i)} - \operatorname{tr}\left(\mathbf{V}_{[i]}^{-1}\right) \boldsymbol{\Lambda}\right) \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}^2_{(i)}} \right]^{1/2}$$

$$\approx \left[(n-1) \frac{c_{ii} - r_{ii}}{c_{ii} r_{ii} \hat{\sigma}^2_{(i)}} \right]^{1/2} |\hat{v}_i| .$$

Welsch (1982) suggested using W_i as a diagnostic tool. The analogue of Welsch's distance for random effects, is

$$W_i(\mathbf{b}) = \left[(n-1) \frac{\left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)^{\mathsf{T}} \left(\mathbf{U}_{(i)}^{\mathsf{T}} \mathbf{U}_{(i)} + \boldsymbol{\Sigma}^{-1}\right) \left(\tilde{\mathbf{b}} - \tilde{\mathbf{b}}_{(i)}\right)}{\hat{\sigma}^2_{(i)}} \right]^{1/2}$$

$$\approx \left[(n-1) \frac{\mathbf{r}_i^{\mathsf{T}} (\mathbf{V} - \mathbf{I}) \mathbf{V} \mathbf{r}_i - \left(r_{ii} - \mathbf{c}_i^{\mathsf{T}} \mathbf{r}_i\right)^2}{r_{ii}^2 \hat{\sigma}^2_{(i)}} \right]^{1/2} |\hat{v}_i| .$$

4.3. Analogue of the likelihood distance

Another popular measure to assess the influence of the ith case on corrected score estimate is the likelihood distance (Cook and Weisberg, 1982). Let $l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ and $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ be the corrected log-likelihood evaluated at $\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2\right)$ and $\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2\right)$, respectively. A measure of the influence of the ith case on $\hat{\boldsymbol{\beta}}$ can be derived based on the distance between $l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ and $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$. The likelihood distance is defined as

$$LD_i(\boldsymbol{\beta}) = 2\left[l^*\left(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right) - l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right) \right].$$

Taylor expansion of $l^*\left(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2; \mathbf{X}, \mathbf{y}\right)$ at $\hat{\boldsymbol{\beta}}$ gives

$$LD_i(\boldsymbol{\beta}) = 2 \left\{ \left[\frac{\partial l^*\left(\boldsymbol{\beta}, \sigma^2; \mathbf{X}, \mathbf{y}\right)}{\partial \boldsymbol{\beta}} \bigg|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \sigma^2 = \hat{\sigma}^2} \right]^{\mathsf{T}} \left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}\right) \right.$$

$$+\frac{1}{2}\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[-\frac{\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}}\bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}},\sigma^2=\hat{\sigma}^2}\right]\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)\Bigg\}$$

$$=\frac{\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)^{\mathsf{T}}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_{(i)}\right)}{\hat{\sigma}^2}.$$

This result is exact because the third derivative is zero. As seen, we have $LD_i(\boldsymbol{\beta}) = CD_i(\boldsymbol{\beta})$ As before, it can be shown that $LD_i(\mathbf{b}) = CD_i(\mathbf{b})$.

4.4. Analogue of the corrected Fisher information ratio

4.4.1. Analogue of corrected Fisher information ratio for fixed effects

As suggested by Belsley et al. (1980), the influence of the ith case on corrected Fisher information matrix for $\boldsymbol{\beta}$ can be measured by comparing the ratio of $\left|\hat{\mathbf{I}}^*(\boldsymbol{\beta})\right|$ to $\left|\hat{\mathbf{I}}^*_{ci}(\boldsymbol{\beta})\right|$; that is,

$$CFIR1_i(\boldsymbol{\beta}) = \frac{\left|-\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2}}{\left|-\partial^2 l^*_{ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2_{(i)}}}$$

$$=\frac{\left|\hat{\sigma}^{-2}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]\right|}{\left|\hat{\sigma}^{-2}_{(i)}\left[\mathbf{X}^{\mathsf{T}}_{(i)}\mathbf{V}^{-1}_{[i]}\mathbf{X}_{(i)}-\operatorname{tr}\left(\mathbf{V}^{-1}_{[i]}\right)\boldsymbol{\Lambda}\right]\right|}$$

$$=\left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^{p}\frac{\left|\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\operatorname{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right|}{\left|\mathbf{X}^{\mathsf{T}}_{(i)}\mathbf{V}^{-1}_{[i]}\mathbf{X}_{(i)}-\operatorname{tr}\left(\mathbf{V}^{-1}_{[i]}\right)\boldsymbol{\Lambda}\right|}.$$

We can get, approximately,

$$CFIR1_i(\boldsymbol{\beta}) = \left(\frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\right)^{p}\frac{c_{ii}}{r_{ii}}.$$

As this is close to 1 if the point is not influential, it seems sensible to use the relative measure $|CFIR1_i(\boldsymbol{\beta})-1|$ as a criterion for assessing the influence of the ith case on $\mathbf{I}^*(\boldsymbol{\beta})$. The larger the statistic $|CFIR1_i(\boldsymbol{\beta})-1|$, the higher the influence of the ith case.

If one uses the trace instead of the determinant, the corrected fisher information ratio becomes

$$CFIR2_i(\boldsymbol{\beta}) = \mathrm{tr}\Bigg\{ \left[-\partial^2 l^*\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right]\Big|_{\sigma^2=\hat{\sigma}^2} \left[-\partial^2 l^*_{ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}^{\mathsf{T}}\right]^{-1}\Big|_{\sigma^2=\hat{\sigma}^2_{(i)}} \Bigg\}$$
$$\approx \frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\left(\frac{c_{ii}}{r_{ii}}+p-1\right).$$

If removing the *i*th case does not change the trace, $CFIR2_i(\boldsymbol{\beta})$ will be close to p and so we could use the relative measure $|CFIR2_i(\boldsymbol{\beta})-p|$ as a criterion for assessing the influence of the *i*th case on the corrected Fisher information for fixed effects.

4.4.2. Analogue of corrected Fisher information ratio for random effects

As similar, the influence of the *i*th case on corrected Fisher information matrix for $\mathbf{b}$ can be measured by comparing the ratio of $\left|\hat{\mathbf{I}}^*(\mathbf{b})\right|$ to $\left|\hat{\mathbf{I}}^*_{ci}(\mathbf{b})\right|$; that is,

$$CFIR1_i(\mathbf{b}) = \frac{\left|-\partial^2 l^*_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2}}{\left|-\partial^2 l^*_{\mathbf{b}_{ci}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right|_{\sigma^2=\hat{\sigma}^2_{(i)}}}$$
$$= \frac{\left|\hat{\sigma}^{-2}\left(\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right)\right|}{\left|\hat{\sigma}^{-2}_{(i)}\left[\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right]\right|} = \left(\frac{\hat{\sigma}^2_{(i)}}{\hat{\sigma}^2}\right)^q \frac{\left|\mathbf{U}^{\mathsf{T}}\mathbf{U}+\boldsymbol{\Sigma}^{-1}\right|}{\left|\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right|}.$$

We can get, approximately,

$$CFIR1_i(\mathbf{b}) = \left(\frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\right)^q \frac{1}{c_{ii}}.$$

Also, if one uses the trace instead of the determinant, the corrected fisher information ratio becomes

$$CFIR2_i(\mathbf{b}) = \mathrm{tr}\Bigg\{ \left[-\partial^2 l^*_{\mathbf{b}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right]\Big|_{\sigma^2=\hat{\sigma}^2} \left[-\partial^2 l^*_{\mathbf{b}_{ci}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)/\partial\mathbf{b}\partial\mathbf{b}^{\mathsf{T}}\right]^{-1}\Big|_{\sigma^2=\hat{\sigma}^2_{(i)}} \Bigg\}$$
$$\approx \frac{n-t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right)}{n-1}\left(c_{ii}^{-1}+q-1\right).$$

If removing the ith case does not change the trace, $CFIR2_i(\mathbf{b})$ will be close to q. Hence, we could use the relative measure $|CFIR2_i(\mathbf{b}) - q|$ as a criterion for assessing the influence of the ith case on the corrected Fisher information for random effects The ith observation is influential observation if $|CFIR2_i(\mathbf{b}) - q|$ is sufficiently large.

5. Example

Diagnostic measures developed in the previous sections are applied to analyse a set of real data which is known as the Boston Housing data set. This data set was the basis for a 1978 paper by Harrison and Rubinfeld, which discussed approaches for using housing market data to estimate the willingness to pay for clean air. The authors employed a hedonic price model, based on the premise that the price of the property is determined by structural attributes (such as size, age, condition) as well as neighborhood attributes (such as crime rate, accessibility, environmental factors). This type of approach is often used to quantify the effects of environmental factors that affect the price of a property. A description of this data set can be found in Harrison and Rubinfeld (1978) and Belsley et al. (1980).

Zhong et al. (2002) considered this data set and used the data of $n = 132$ census tracts within the 15 districts of the Boston city (as a part of 506 observations on census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970). They followed the regression model of Harrison and Rubinfeld (1978). However, the census tracts within districts are taken as repeated measurements. All independent variables can be measured precisely except the pollution variable NOXSQ which is taken to have measurement errors. Therefore, a linear mixed measurement error model was employed.

Now, we consider the same data set and derive different diagnostic measures for linear mixed measurement error model given in previous section. Figures 1-3 give the index plot of the diagnostic measures for fixed effects and Figures 4-6 give the index plot of the diagnostic measures for random effects, respectively. Based on generalized Cook's distance, a glance at Figures 1 and 4 shows that cases 9 and 15 have more influence on fixed effect and cases 9, 16 and 36 have more influence on random effects. The plots for W_i in Figures 2 and 5, respectively, for fixed and random effects have almost identical behavior as CD_i

Table 1 gives the parameter estimates from corrected score method with the full data and with only case 9 deleted. As seen, after deleting case 9, the NOXSQ variable does not have any significant effect. The RM, AGE and CHAS variables, in each case, do not have any significant effects but after deleting case 9 their signs have been changed. Figure 3 show that case 36 is the most influential point on corrected Fisher information for fixed effects, while Figure 6 indicate that this case is the most influential point on corrected Fisher information for random effects.

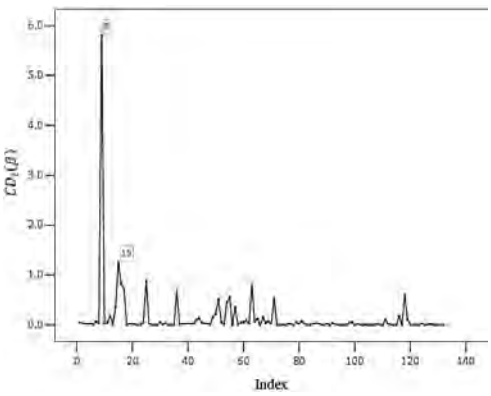


Figure 1: *Index plot of* $CD_i\,(\boldsymbol{\beta})$.

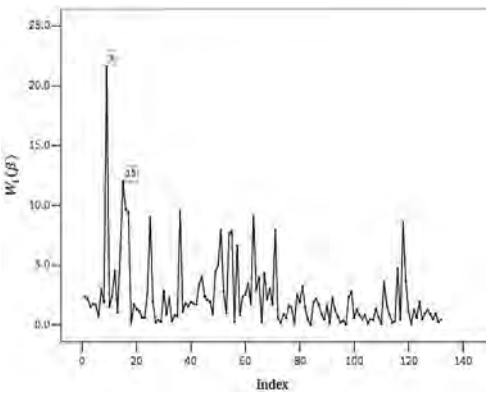


Figure 2: *Index plot of* $W_i\,(\boldsymbol{\beta})$.

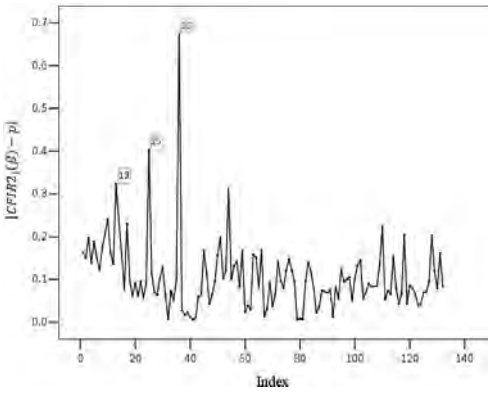


Figure 3: *Index plot of* $|CFIR2_i\,(\boldsymbol{\beta}) - p|$.

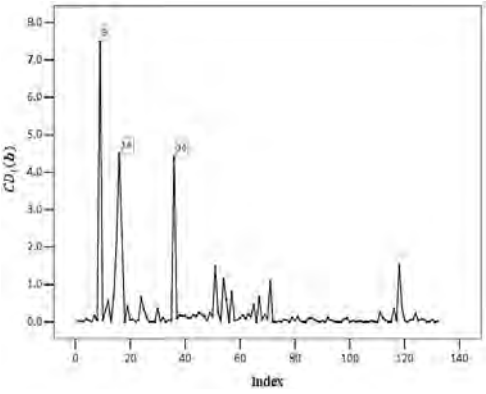


Figure 4: *Index plot of* $CD_i\,(\boldsymbol{b})$.

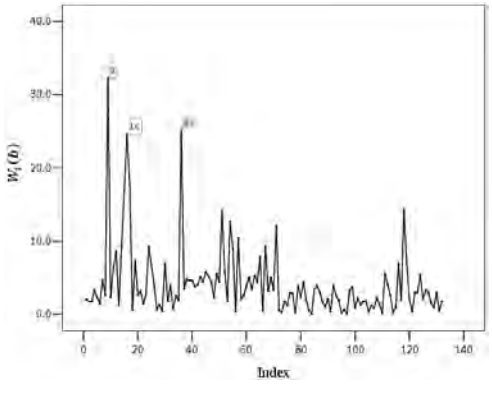


Figure 5: *Index plot of* $W_i\,(\boldsymbol{b})$.

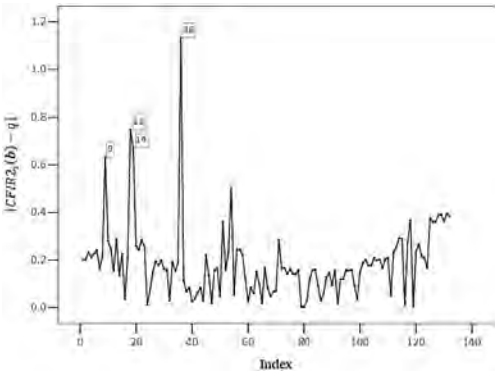


Figure 6: *Index plot of* $|CFIR2_i\,(\boldsymbol{b}) - q|$.

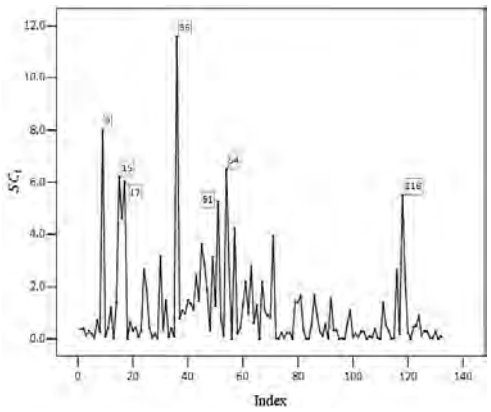


Figure 7: *Index plot of SC_i.*

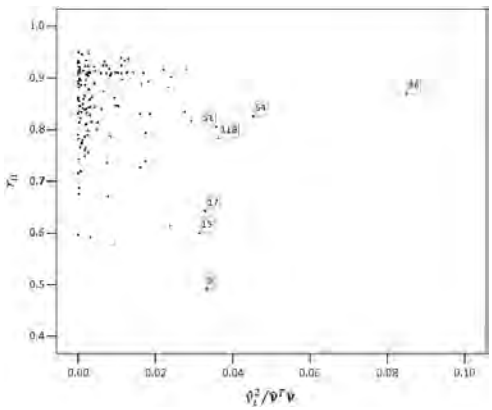


Figure 8: *Scatter plot of r_{ii} versus $\hat{v}_i^2/\hat{\mathbf{v}}^{\mathrm{T}}\hat{\mathbf{v}}$.*

Table 1: *Corrected score estimates for the hedonic housing price data of Boston city. The t-ratios are in parentheses.*

Variable	Full data	Case 9 deleted	% change
Intercept	9.07 (28.72)	8.90 (28.11)	1.9
RM	-1.4×10^{-3} (-0.57)	3.6×10^{-3} (1.26)	352.2
AGE	7.6×10^{-4} (0.4)	-3.7×10^{-4} (-0.2)	150.0
DIS	8.8×10^{-2} (0.59)	1.8×10^{-1} (1.16)	101.2
B	4.6×10^{-1} (2.95)	5.1×10^{-1} (3.43)	12.2
LSTAT	-5.3×10^{-1} (-8.65)	-4.8×10^{-1} (-7.77)	10.5
CRIM	-7.3×10^{-3} (-5.36)	-6.6×10^{-3} (-5.00)	9.7
CHAS	-3.0×10^{-2} (-0.33)	4.6×10^{-3} (0.05)	115.5
NOXSQ	-1.0×10^{-2} (-2.34)	-7.8×10^{-3} (-1.74)	24.1
σ_1^2	4.8×10^{-3}	6.8×10^{-3}	41.7
σ^2	2.8×10^{-3}	2.5×10^{-3}	10.7

Table 2: *Corrected score estimates for the hedonic housing price data of Boston city after deleting pair cases $\{9, 15\}$ and $\{9, 16\}$. The t-ratios are in parentheses.*

Variable	Pair case $\{9, 15\}$ deleted	% change	Pair case $\{9, 16\}$ deleted	% change
Intercept	8.76 (28.05)	3.3	8.86 (28.97)	2.3
RM	3.7×10^{-3} (1.34)	358.1	2.3×10^{-3} (0.81)	264.3
AGE	-1.0×10^{-4} (-0.06)	113.7	-3.4×10^{-4} (-0.19)	144.7
DIS	2.1×10^{-1} (1.4)	144.6	1.4×10^{-1} (0.98)	59.1
B	4.9×10^{-1} (3.41)	7.6	4.6×10^{-1} (3.12)	0.0
LSTAT	-5.2×10^{-1} (-8.35)	2.3	-5.1×10^{-1} (-8.56)	3.8
CRIM	-6.5×10^{-3} (-5.11)	11.6	-6.9×10^{-3} (-5.20)	5.5
CHAS	5.1×10^{-2} (0.60)	270.5	1.1×10^{-1} (1.14)	466.7
NOXSQ	-7.5×10^{-3} (-1.7)	27.3	-6.8×10^{-3} (-1.59)	32.0
σ_1^2	8.4×10^{-3}	75.0	4.3×10^{-3}	10.4
σ^2	2.3×10^{-3}	17.9	2.6×10^{-3}	7.1

Table 2 gives the corrected score estimates after deleting pairs of cases $\{9, 15\}$and $\{9, 16\}$from data set. Deleting these pairs have almost the same effect with deleting case 9 on parameters of model. The only difference is that deleting cases $\{9, 15\}$has more influence on CHAS variable and σ_1^2 while deleting cases $\{9, 16\}$has more influence on CHAS variable. Table 3 indicates the maximum percentage of changes in determinant of corrected Fisher information after deleting case 36. Finally, Figures 7 and 8 indicate that case 36 is also an outlier observation (see Zewotir and Galpin, 2007 for details about plot of Figure 8).

Table 3: *The determinant of the corrected Fisher information (DCFI) for the hedonic housing price data of Boston city.*

DCFI	Full data	Case 36 deleted	% change
Fixed effects	$6.08 \times 10^{+31}$	$1.17 \times 10^{+32}$	91.7
Random effects	$4.52 \times 10^{+36}$	$1.49 \times 10^{+37}$	229.8

6. Concluding remarks

We have presented case deletion and mean shift outlier models for linear mixed measurement error models that appear to be useful and can play important role in data analysis. Also, based on the corrected likelihood, we obtained case deletion diagnostics for detecting influential observations in linear mixed measurement error models. All the diagnostic measures are similar to diagnostics in linear mixed models. They are functions of studentized residuals of model, error contrast matrix $(\mathbf{R})$ and the inverse of the response variable covariance matrix $(\mathbf{C})$. Although no formal cutoff points are presented for these measures, it appears that relative comparisons such as ranking or simple index plots are a promising and practical approach to pinpoint influential observations. Here, the results obtained with the assumption that the γ_i's are known. In practice, we do not know the γ_i's. So, the corrected score estimates of the γ_i's are used and the results are useful as an approximation. In this paper, we fitted a linear mixed model with measurement error in fixed effects (and not in random effects) by specifying the covariance structure of $\mathbf{b}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Delta}$. Here we have assumed that $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are known and $\boldsymbol{\Sigma}$ has diagonal structure with the ith block being $\gamma_i \mathbf{I}$. However, if random effects are also measured with errors and $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are unknown, extending our diagnostics is an area of future research.

Acknowledgments

We wish to thank the Editor and the referee for their valuable comments which enhanced the presentation of this paper.

Appendix

Proof of Theorem 1:

It follows from (2) that the corrected log-likelihood of $\mathbf{y}$ and the conditional corrected log-likelihood of $\mathbf{b}|\mathbf{y}$ for MSOM, respectively, are given by

$$\begin{aligned} l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right) = &-\frac{n}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\mathbf{V}|\right)-\frac{1}{2\sigma^2}\Bigg\{\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\left(\mathbf{V}^{-1}_{[i]}\right. \\ &\left.+\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\right)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+c_{ii}\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)^2 \\ &+2\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\Bigg\}, \end{aligned} \tag{5}$$

$$\begin{aligned} l^*_{\mathbf{b}_{mi}}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right) = &-\frac{q}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(|\boldsymbol{\Sigma}\mathbf{T}|\right)-\frac{1}{2\sigma^2}\Bigg\{\mathbf{b}^{\mathsf{T}}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\mathbf{b} \\ &-2\mathbf{b}^{\mathsf{T}}\left(\boldsymbol{\Sigma}\mathbf{T}\right)^{-1}\boldsymbol{\Sigma}\Big[c_{ii}\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{u}_i+\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{U}^{\mathsf{T}}_{(i)}\mathbf{c}_{i(i)} \\ &+\mathbf{u}_i\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+\mathbf{U}^{\mathsf{T}}_{(i)}\left(\mathbf{V}^{-1}_{[i]}+\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\right)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\ &+F\left(\boldsymbol{\beta},\tau\right)\Big]\Bigg\} \end{aligned} \tag{6}$$

where

$$\begin{aligned} F\left(\boldsymbol{\beta},\tau\right) = &\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\left[\mathbf{I}-\mathbf{V}^{-1}_{[i]}-\mathbf{c}_{i(i)}\mathbf{c}^{\mathsf{T}}_{i(i)}/c_{ii}\right]\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\ &+\left(1-c_{ii}\right)\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)^2-2\left(y_i-\mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}-\tau\right)\mathbf{c}^{\mathsf{T}}_{i(i)}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right) \\ &-\mathrm{tr}\left(\mathbf{I}-\mathbf{V}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta} \end{aligned}$$

The corrected likelihood estimates of $\hat{\boldsymbol{\beta}}_{mi}$, $\hat{\sigma}^2_{mi}$, $\hat{\tau}_{mi}$ and $\tilde{\mathbf{b}}_{mi}$, are derived with differentiating (5) with respect to $\boldsymbol{\beta}$, σ^2 and τ and (6) with respect to $\mathbf{b}$.

Proof of Theorem 2:

Since corrected score estimate is asymptotically normal, the score test can be used (Cox and Hinkley, 1974). Let the corrected Fisher information matrix of $\mathbf{y}$ for $\boldsymbol{\beta}$ and τ be $\mathbf{J}\left(\boldsymbol{\beta},\tau\right)$, then the score statistic under $H_0:\tau=0$ is

$$SC_i = \left[\frac{\partial l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\tau}\right]^{\mathsf{T}} J^{\tau\tau}\left[\frac{\partial l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\tau}\right]\Bigg|_{\left(\hat{\boldsymbol{\beta}},\hat{\sigma}^2\right)},$$

where $J^{\tau\tau}$ is the lower right corner of $\mathbf{J}^{-1}\left(\boldsymbol{\beta},\tau\right)$. It is easily seen that under $H_0:\tau=0$

$$\frac{\partial l^*_{mi}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)}{\partial\tau}=\frac{1}{\sigma^2}\mathbf{c}_i^{\mathsf{T}}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right),$$

$$\mathbf{J}\left(\boldsymbol{\beta},\tau\right)=\frac{1}{\sigma^2}\left[\begin{array}{cc}\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda} & \mathbf{X}^{\mathsf{T}}\mathbf{c}_i\\ \mathbf{c}_i^{\mathsf{T}}\mathbf{X} & c_{ii}\end{array}\right],$$

and $J^{\tau\tau}=\dfrac{\sigma^2}{r_{ii}}$ then under $H_0:\tau=0$

$$SC_i=\left.\frac{\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)^{\mathsf{T}}\mathbf{c}_i\mathbf{c}_i^{\mathsf{T}}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)}{\sigma^2 r_{ii}}\right|_{\left(\hat{\boldsymbol{\beta}},\hat{\sigma}^2\right)}=\frac{\hat{v}_i^2}{\hat{\sigma}^2 r_{ii}}=t_i^2\left(1+\hat{\boldsymbol{\beta}}^{\mathsf{T}}\boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}/\hat{\sigma}^2\right).$$

Proof of Theorem 3:

It follows from (5) that the corrected log-likelihood of $\mathbf{y}$ and the conditional corrected log-likelihood of $\mathbf{b}|\mathbf{y}$ for CDM, respectively, are given by

$$\begin{aligned}l^*_{ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)=&-\frac{n-1}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\log\left(\left|\mathbf{V}_{[i]}\right|\right)-\\&-\frac{1}{2\sigma^2}\left\{\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\mathrm{tr}\left(\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}\right\}\end{aligned}\tag{7}$$

$$\begin{aligned}l^*_{\mathbf{b}ci}\left(\boldsymbol{\beta},\sigma^2;\mathbf{X},\mathbf{y}\right)=&-\frac{q}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\left[\log\left(\left|\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right|\right)\right]-\frac{1}{2\sigma^2}\left\{\mathbf{b}^{\mathsf{T}}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}\right.\right.\\&\left.\left.+\ \boldsymbol{\Sigma}^{-1}\right]\mathbf{b}-2\mathbf{b}^{\mathsf{T}}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{U}_{(i)}+\boldsymbol{\Sigma}^{-1}\right]\boldsymbol{\Sigma}\left[\mathbf{U}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)+F'\left(\boldsymbol{\beta}\right)\right]\right\},\end{aligned}\tag{8}$$

where

$$F'\left(\boldsymbol{\beta}\right)=\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)^{\mathsf{T}}\left(\mathbf{I}-\mathbf{V}_{[i]}^{-1}\right)\left(\mathbf{y}_{(i)}-\mathbf{X}_{(i)}\boldsymbol{\beta}\right)-\mathrm{tr}\left(\mathbf{I}-\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Lambda}\boldsymbol{\beta}.$$

The corrected score estimates of $\boldsymbol{\beta}$, σ^2 and $\mathbf{b}$ will be obtained with differentiating (7) with respect to $\boldsymbol{\beta}$ and σ^2 and (8) with respect to $\mathbf{b}$. Then we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(i)}&=\left[\mathbf{X}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\mathbf{X}_{(i)}-\mathrm{tr}\left(\mathbf{V}_{[i]}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}_{(i)}^{\mathsf{T}}\mathbf{V}_{[i]}^{-1}\mathbf{y}_{(i)}\\&=\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\mathbf{c}_i^{\mathsf{T}}\mathbf{X}/c_{ii}-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}+c_{ii}\boldsymbol{\Lambda}\right]^{-1}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{y}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\mathbf{c}_i^{\mathsf{T}}\mathbf{y}/c_{ii}\right]\\&=\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\mathbf{c}_i^{\mathsf{T}}\mathbf{X}/c_{ii}-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{y}-\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\mathbf{c}_i^{\mathsf{T}}\mathbf{y}/c_{ii}\right]\\&+\mathbf{O}_p\left(n^{-1}\right)\approx\hat{\boldsymbol{\beta}}-\left[\mathbf{X}^{\mathsf{T}}\mathbf{V}^{-1}\mathbf{X}-\mathrm{tr}\left(\mathbf{V}^{-1}\right)\boldsymbol{\Lambda}\right]^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{c}_i\frac{\hat{v}_i}{r_{ii}},\end{aligned}$$

$$
\begin{aligned}
(n-1)\,\hat{\sigma}^2_{(i)} &= \left[\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}_{(i)}\right]^{\mathsf{T}} \mathbf{V}^{-1}_{[i]} \left[\mathbf{y}_{(i)} - \mathbf{X}_{(i)}\hat{\boldsymbol{\beta}}_{(i)}\right] - \operatorname{tr}\left(\mathbf{V}^{-1}_{[i]}\right) \hat{\boldsymbol{\beta}}^{\mathsf{T}}_{(i)} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}_{(i)} \\
&= \mathbf{y}^{\mathsf{T}}_{(i)} \mathbf{V}^{-1}_{[i]} \mathbf{y}_{(i)} - \hat{\boldsymbol{\beta}}^{\mathsf{T}}_{(i)} \mathbf{X}^{\mathsf{T}}_{(i)} \mathbf{V}^{-1}_{[i]} \mathbf{y}_{(i)} \\
&= \mathbf{y}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y} - \mathbf{y}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} / c_{ii} \\
&- \left[\hat{\boldsymbol{\beta}} - \left(\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right)^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \frac{\hat{v}_i}{r_{ii}} + \mathbf{O}_p\left(n^{-1}\right)\right]^{\mathsf{T}} \left(\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y}\right. \\
&\left. - \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} / c_{ii}\right) \\
&= n\hat{\sigma}^2 - \mathbf{y}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} / c_{ii} + \mathbf{c}_i^{\mathsf{T}} \mathbf{X} \left[\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right]^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{y} \frac{\hat{v}_i}{r_{ii}} \\
&- \mathbf{c}_i^{\mathsf{T}} \mathbf{X} \left[\mathbf{X}^{\mathsf{T}} \mathbf{V}^{-1} \mathbf{X} - \operatorname{tr}\left(\mathbf{V}^{-1}\right) \boldsymbol{\Lambda}\right]^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} \frac{\hat{v}_i}{c_{ii} r_{ii}} + \hat{\boldsymbol{\beta}}^{\mathsf{T}} \mathbf{X}^{\mathsf{T}} \mathbf{c}_i \mathbf{c}_i^{\mathsf{T}} \mathbf{y} / c_{ii} \\
&+ O_p(1) = n\hat{\sigma}^2 - \frac{\hat{v}_i^2}{r_{ii}} + O_p(1),
\end{aligned}
$$

and hence, $\hat{\sigma}^2_{(i)} \approx \dfrac{n - t_i^2\left(1 + \hat{\boldsymbol{\beta}}^{\mathsf{T}} \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}} / \hat{\sigma}^2\right)}{n-1} \hat{\sigma}^2,$

$$
\tilde{\mathbf{b}}_{(i)} = \boldsymbol{\Sigma} \mathbf{U}^{\mathsf{T}}_{(i)} \mathbf{V}^{-1}_{[i]} \left(\mathbf{y}_{(i)} - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)}\right) \approx \tilde{\mathbf{b}} - \boldsymbol{\Sigma} \mathbf{U}^{\mathsf{T}} \mathbf{r}_i \frac{\hat{v}_i}{r_{ii}}.
$$

References

Banerjee, M. and Frees, E. W. (1997). Influence diagnostics for linear longitudinal models. *Journal of the American Statistical Association*, 92, 999–1005.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

Chatterjee, S. and Hadi, A. S. (1986). Influential observation, high leverage points and outliers in linear regression (with discussion). *Statistical Science*, 1, 379–416.

Christensen, R., Pearson, L. M. and Johnson, W. (1992). Case deletion diagnostics for mixed models. *Technometrics*, 34, 38–45.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, London.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

Fung, W. K., Zhong, X. P. and Wei, B. C. (2003). On estimation and influence diagnostics in linear mixed measurement errors models. *American Journal of Mathematical and Management Sciences*, 23, 37–59.

Gimenz, P. and Bolfarine, H. (1997). Corrected score functions in classical error-in-variables and incidental parameter models. *The Australian Journal of Statistics*, 39, 325–344.

Hanfelt, J. J. and Liang, K. Y. (1997). Approximate likelihood for generalized linear errors-in-variables models. *Journal of the Royal Statistical Society. Series B*, 59, 627–637.

Harrison, D. and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5, 81–102.

Haslett, J. and Dillane, D. (2004). Application of 'delete=replace' to deletion diagnostics for variance component estimation in linear mixed model. *Journal of the Royal Statistical Society. Series B*, 66, 131–143.

Kelly, G. E. (1984). The influence function in the errors in variables problems, *The Annals of Statistics*, 12, 87–100.

Li, Z., Xu, W., and Zhu, L. (2009). Influence diagnostics and outlier tests for varying coefficient mixed models. *Journal of Multivariate Analysis*, 100, 2002–2017.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77, 127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, 48, 829–838.

Rio, M. (1988). On the potential in the estimation of linear functions in regression. *Communications in Statistics. Theory and Methods*, 17, 729–738.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects (with discussion). *Statistical Science*, 6, 15–51.

Wellman, J. M. and Gunst, R. F. (1991). Influence diagnostics for linear measurement errors models. *Biometrika*, 78, 373–380.

Welsch, R. E. (1982). Influence functions and regression diagnostics. In *Modern Data Analysis* (R.L. Launer and A.F. Siegel, eds.), Academic, New York.

Zare, K., Rasekh, A. and Rasekhi, A. (2011). Estimation of variance components in linear mixed measurement error models. *Statistical Papers*, DOI 10.1007/s00362-011-0387-0.

Zewotir, T. and Galpin, J. S. (2005). Influence diagnostics for linear mixed models. *Journal of Data Science*, 3, 153–177.

Zewotir, T. and Galpin, J. S. (2007). A unified approach on residuals, leverages and outliers in the linear mixed model, *Test*, 16, 58–75.

Zhong, X. P. and Wei, B. C. (1999). Influence analysis on linear models with random effects. *Applied Mathematics. A Journal of Chinese Universities. Ser. B.*, 14, 169–176.

Zhong, X. P., Fung, W. K. and Wei, B. C. (2002). Estimation in linear models with random effects and errors-in-variables. *Annals of the Institute of Statistical Mathematics*, 54, 595–606.

Zhong, X. P., Wei, B. C. and Fung, W. K. (2000). Influence analysis for linear measurement error models. *Annals of the Institute of Statistical Mathematics*, 52, 367–379.

SORT 35 (2) July-December 2011, 145-164

Iterative beam search for simple assembly line balancing with a fixed number of work stations

Christian Blum*

Abstract

The simple assembly line balancing problem (SALBP) concerns the assignment of tasks with pre-defined processing times to work stations that are arranged in a line. Hereby, precedence constraints between the tasks must be respected. The optimization goal of the SALBP-2 variant of the problem concerns the minimization of the so-called cycle time, that is, the time in which the tasks of each work station must be completed. In this work we propose to tackle this problem with an iterative search method based on beam search. The proposed algorithm is able to generate optimal solutions, respectively the best upper bounds, for 283 out of 302 test cases. Moreover, for 9 further test cases the algorithm is able to improve the currently best upper bounds. These numbers indicate that the proposed iterative beam search algorithm is currently a state-of-the-art method for the SALBP-2.

MSC2000 Classification: 90C27 (Combinatorial Optimization)

Keywords: Assembly line balancing, fixed number of work stations, beam search.

1. Introduction

The class of problems known as assembly line balancing problems (ALBPs) concerns the optimization of processes related to the manufacturing of products via assembly lines. Their importance in the industrial world is shown by the fact that much research efforts have been dedicated to many different types of ALBPs during the past 50-60 years Gosh and Gagnon (1989), Salveson (1955). The specific problem considered in this paper is the so-called simple assembly line balancing problem (SALBP) Scholl and Becker (2006), a well-studied scientific test case. An assembly line is composed of a set of work stations arranged in a line, and by a transport system which moves the product to be manufactured along the line. The product is manufactured by executing a given

Address for correspondence: ALBCOM Research Group. Universitat Politècnica de Catalunya. C/ Jordi Girona 1-3, Campus Nord, Omega 112. 08034 Barcelona (Spain). cblum@lsi.upc.edu.
Received: December 2010
Accepted: October 2011

set of tasks. Each of these tasks has a pre-defined processing time. In order to obtain a solution to a given SALBP instance, all tasks must be assigned to work stations subject to precedence constraints between the tasks. In the context of the SALBP, all work stations are considered to be of equal size. Moreover, the assembly line is assumed to move at a constant speed. This implies a maximum of C time units – the so-called *cycle time* – for processing the tasks assigned to each work station. The SALBP has been tackled with several objective functions among which the following ones are the most studied ones in the literature:

- Given a fixed cycle time C, the optimization goal consists in minimizing the number of necessary work stations. This variant of the problem is refered to as SALBP-1.
- Given a fixed number m of work stations, the goal is to minimize the cycle time C. The literature knows this second problem variant as SALBP-2.

The feasibility problem SALBP-F arises when both a cycle time C and a number of work stations m is given and the goal is to find a feasible solution respecting C and m. In this work we will deal with the SALBP-2 variant of the problem.

For what concerns the comparison between SALBP-1 and SALBP-2, much of the scientific work has been dedicated to the SALBP-1. However, also for the SALBP-2 exists a considerable body of research papers. An excellent survey was provided by Scholl and Becker (2006). Approaches for the SALBP-2 can basically be classified as either *iterative solution approaches* or *direct solution approaches*. Iterative approaches tackle the problem by iteratively solving a series of SALBP-F problems that are obtained by fixing the cycle time. This process is started with a cycle time that is set to some calculated upper bound. This cycle time is then decremented during the iterative process, which stops as soon as no solution for the corresponding SALBP-F problem can be found. In contrast to these indirect approaches, direct approaches intend to solve a given SALBP-2 instance directly.

Heuristic as well as exact approaches have been devised for the SALBP-2. Among the existing exact methods we find iterative approaches such as the ones proposed in Hackman et al. (1989), Scholl (1999) but also direct approaches such as the ones described in Klein and Scholl (1996), Scholl (1994). Moreover, the performance of different integer programming formulations of the SALBP-2 have been evaluated in Pastor and Ferrer (2009), Pastor et al. (2007). The currently best-performing exact method is SALOME-2 Klein and Scholl (1996), which is a branch & bound procedure making use of a so-called local lower bound method (LLBM), a bidirectional branching strategy, and several dominance and reduction rules. Surprisingly, exact methods outperform the existing heuristic and metaheuristic approaches for the SALBP-2. While 287 (out of 302) existing problem instances were solved to optimality by exact approaches, the most successful metaheuristic approach to date—a tabu search method proposed in Scholl and Voss (1996) —was only able to find the best upper bounds with respect to 168 prob-

lem instances. Note that SALOME-2 alone was able to solve 217 problem instances to optimality. Apart from the above-mentioned tabu search, another tabu search proposal can be found in Chiang (1998). Other metaheuristic approaches include evolutionary algorithms Anderson and Ferris (1994), Nearchou (2007), Watanabe et al. (1995) and simulated annealing Henrici (1994). Moreover, a two-phase heuristic based on linear programming can be found in Ugurdag et al. (1997), whereas a heuristic based on Petri nets was proposed in Kilincci (2010). Finally, an approach similar to the one proposed in this paper has been presented in Blum and Miralles (2011) for a more general problem, the assembly line worker assignment and balancing problem (ALWABP).

Contribution of this work. Subsequently we propose to tackle the SALBP-2 by means of an iterative approach based on beam search, which is an incomplete variant of branch & bound. The resulting iterative beam search algorithm is inspired by one of the current state-of-the-art methods for the SALBP-1, namely Beam-ACO Blum (2008). Beam-ACO is a hybrid approach that is obtained by combining the metaheuristic ant colony optimization with beam search. In this work we propose to use the beam search component of Beam-ACO in an iterative way for obtaining good SALBP-2 solutions. Our computational results show indeed that the proposed algorithm is currently a state-of-the-art method for the SALBP-2. It is able to generate optimal solutions, respectively the best upper bounds, for 283 out of 302 test cases. Moreover, in further 9 cases the algorithm is able to improve the currently best upper bounds.

Organization of the paper. In Section 2 we present a formal description of the tackled problem. Furthermore, in Section 3 the proposed algorithm is described. Finally, in Section 4 we present a detailed experimental evaluation and in Section 5 we conclude our work and offer an outlook to future work.

2. The SALBP-2

The SALBP-2 can formally be described as follows. An instance (T,G,m) consists of three components. $T = \{1,\ldots,n\}$ is a set of n tasks. Each task $i \in T$ has a pre-defined processing time $t_i > 0$. Without losing generality, the processing times are henceforth assumed to be integer values. Moreover, given is a precedence graph $G = (T,A)$, which is a directed, acyclic graph with T as node set. Finally, m is the pre-defined number of work stations which are ordered from 1 to m. An arc $l_{i,j} \in A$ indicates that $i \in T$ must be processed before $j \in T$. Given a task $j \in T$, $P_j \subset T$ denotes the set of tasks that must be processed before j. A feasible solution is obtained by assigning each task to exactly one work station such that the precedence constraints between the tasks are satisfied. The objective function consists in minimizing the so-called cycle time. The SALBP-2 can be expressed in the following way as an integer programming (IP) problem.

$$\min z \tag{1}$$

subject to:

$$\sum_{s=1}^{m} x_{is} = 1 \quad \forall i \in T \tag{2}$$

$$x_{is} \leq \sum_{s'=1}^{s} x_{js'} \quad \forall i \in T, s = 1,\ldots,m,\ j \in P_i \tag{3}$$

$$\sum_{i \in T} t_i x_{is} \leq z \quad s = 1,\ldots,m \tag{4}$$

$$x_{is} \in \{0,1\} \quad \forall i \in T, s = 1,\ldots,m \tag{5}$$

$$z > 0 \tag{6}$$

This IP model makes use of the following variables and constants: x_{is} is a binary variable which is set to 1 if and only if task $i \in T$ is assigned to work station $1 \leq s \leq m$. The objective function (1) minimizes the cycle time $z > 0$.[1] The constraints (2) ensure that each task $i \in T$ is assigned to a single work station $1 \leq s \leq m$. Constraints (3) reflect the precedence relationships between the tasks. More specifically, if task $i \in T$ is assigned to a work station $1 \leq s \leq m$, all tasks $j \in P_i$ must be assigned to work stations $1 \leq s' \leq m$ with $s' \leq s$. The constraints (4) ensure that the sum of the processing times of the tasks assigned to a work station $1 \leq s \leq m$ do not exceed the cycle time z.

Note that this model was chosen in order to present an easily understandable problem description. An evaluation of alternative models can be found in Pastor and Ferrer (2009), Pastor et al. (2007).

3. Iterative beam search

As mentioned in the introduction, the basic component of our algorithm for the SALBP-2 consists of beam search (BS), which is an incomplete derivative of branch & bound. BS was used for the first time in the context of speech recognition Lowerre (1976). Concerning combinatorial optimization problems, BS has especially been used for solving scheduling problems; see, for example, Ghirardi and Potts (2005), Ow and Morton (1988), Sabuncuoglu and Bayiz (1999), Valente and Alves (2005). To date only very few applications to other types of problems exist. Examples can be found in Akeba et al. (2009), Blum et al. (2009), Lee and Woodruff (2004). In the following we briefly describe how one of the standard variants of BS works. The crucial aspect of BS is the parallel extension of partial solutions in several ways. At all times, the algorithm keeps a set B of at most k_{bw} partial solutions, where B is the so-called *beam*, and k_{bw} is known as the *beam width*. At each step, at most k_{ext} feasible extensions of each partial solution in B are selected on the basis of greedy information. In general, this selection

1. Note that we refer to the variable cycle time of the IP model as z, while fixed cycle times are denoted by C.

is done deterministically. At the end of each step, the algorithm creates a new beam B by choosing up to k_{bw} partial solutions from the set of selected feasible extensions. For that purpose, BS algorithms determine – in the case of minimization – a lower bound value for each extension. Only the maximally k_{bw} best extensions – with respect to these lower bound values – are included in B. Finally, if any complete solution was generated, the algorithm returns the best of those. Note that the underlying constructive heuristic that defines feasible extensions of partial solutions and the lower bound function for evaluating partial solutions are crucial for the working of BS.

In the following, after describing the chosen solution representation, we first present a description of the implementation of the BS component, before we describe the algorithmic scheme in which this BS component is used.

3.1. Solution representation

The following solution representation is used by the proposed BS algorithm. A solution $\mathcal{S}$ is an ordered list $\mathcal{S} = \langle S_1, \ldots, S_m \rangle$ of m sets of tasks, where S_i denotes the set of tasks that are assigned to the i-th work station. Abusing notation we henceforth call S_i a work station. Note that for a solution $\mathcal{S}$ to be valid the following conditions must be fulfilled:

1. $\bigcup_{i=1}^{m} S_i = T = \{1, \ldots, n\}$ and $S_i \cap S_{i'} = \emptyset$ for $i = 1, \ldots, m$ and $i' = i+1, \ldots, m$. That is, each task is assigned to exactly one work station.
2. For each task $j \in S_i$ it is must hold that $P_j \subseteq \bigcup_{k=1}^{i} S_k$. This ensures that the precedence constraints between the tasks are not violated.

In this context it is also convenient to introduce the concept of the *reverse problem instance*. More specifically, the reverse problem instance (T, G^r, m) with respect to an original instance (T, G, m) is obtained by inverting the direction of all arcs of G. It is well-known from the literature Scholl and Becker (2006) that tackling the reverse problem instance may lead an exact algorithm faster to an optimal solution, respectively, may provide a better heuristic solution when tackled with the same heuristic as the original problem instance. Moreover, a solution $\mathcal{S}^r$ to the reverse problem instance (T, G^r, m) can easily be converted into a solution $\mathcal{S}$ to the original problem instance (T, G, m) as follows:

$$S_i := S^r_{m-i+1} \quad \text{for } i = 1, \ldots, m \tag{7}$$

3.2. The beam search component

The BS component described in this section – see Algorithm 1 for the pseudo-code – is the main component of the proposed algorithm for the SALBP-2. The algorithm requires a problem instance (T, G, m), a fixed cycle time C, a beam width k_{bw}, and a maximal

Algorithm 1 Beam search

1: **input:** an instance (T, G, m), a fixed cycle time C, a beam width k_{bw}, and k_{ext}
2: $d := 0$
3: Initialization of an empty solution $\mathcal{S}$
4: $B := \{\mathcal{S}\}$
5: $B_{\text{compl}} := \emptyset$
6: **while** $B \neq \emptyset$ **do**
7: $B_{\text{ext}} := \emptyset$
8: $d := d + 1$
9: **for all** $\mathcal{S} \in B$ **do**
10: **for** $i = 1, \ldots, k_{\text{ext}}$ **do**
11: $\mathcal{S}' := \mathcal{S}$ {copy partial solution $\mathcal{S}$ into $\mathcal{S}'$}
12: $S'_d :=$ ExtendPartialSolution($\mathcal{S}', d, C$) {see Algorithm 2}
13: **if** solution $\mathcal{S}'$ is complete (that is, all tasks are assigned) **then**
14: $B_{\text{compl}} := B_{\text{compl}} \cup \{\mathcal{S}'\}$
15: **else**
16: **if** $d < m$ and S'_d is different to the d-th work station of all other $\mathcal{S} \in B_{\text{ext}}$ **then**
17: $B_{\text{ext}} := B_{\text{ext}} \cup \{\mathcal{S}'\}$
18: **end if**
19: **end if**
20: **end for**
21: **end for**
22: $B \leftarrow$SelectSolutions(B_{ext}, k_{bw})
23: **end while**
24: **output:** If $B_{\text{compl}} \neq \emptyset$ the output is TRUE, otherwise FALSE

number of extensions k_{ext} as input. Given a fixed cycle time C and m (the number of work stations) BS tries to find at least one feasible solution. As mentioned before, the crucial aspect of BS is the extension of partial solutions in several possible ways. At each step the algorithm extends each partial solution from B in a limited number of ways. More specifically, given a partial solution with $d - 1 < m$ work stations already filled, an extension is generated by assigning a set of so-far unassigned tasks to the next work station S_d such that the given cycle time C is not surpassed and the precedence constraints between the tasks are respected (see lines 11–12 of Algorithm 1). The algorithm produces extensions in a (partially) probabilistic way rather than in the usual deterministic manner.[2] Each generated extension (partial solution) is either stored in set B_{compl} in case it is a complete solution, or in set B_{ext} otherwise (see lines 13–19 of

2. This is done for avoiding a *choice-without-replacement* process for which all possible work station fillings must be generated beforehand.

Algorithm 1). However, a partial solution is only stored in set B_{ext} if it uses at most $m-1$ work stations. Moreover, for a partial solution to be stored in set B_{ext} it is required that its d-th work station is different to the d-th work station of all partial solutions that are already in B_{ext}. This criterion can be seen as a computationally cheap, approximate way of checking if two partial solutions are equal. On the downside, with this criterion partial solutions might be excluded from further examination even though they do not belong to B_{ext}. Finally, BS creates a new beam B by selecting up to k_{bw} solutions from set B_{ext} of further extensible partial solutions (see line 22 of Algorithm 1). This is done in function SelectSolutions(B_{ext},k_{bw}) on the basis of a lower bound function LB$(\cdot)$. In the following we describe in detail the extension of partial solutions and the working of function SelectSolutions(B_{ext},k_{bw}).

Extending partial solutions. The generation of an extension of a partial solution $\mathcal{S}'$ with $d-1$ work stations already filled works as follows. Unassigned tasks are iteratively assigned to work station S'_d until the sum of their processing times is such that no other task can be added to S'_d without exceeding the given cycle time C. This procedure is pseudo-coded in Algorithm 2. At each step, T' denotes the set of so-far unassigned tasks that may be added to S'_d without violating any constraints. The definition of this set of *available tasks* is given in line 4, respectively 9, of Algorithm 2.

Algorithm 2 Function ExtendPartialSolution($\mathcal{S}',d,C$) of Algorithm 1

1: **input:** A partial solution $\mathcal{S}'$, the index d of the work station to be filled, and the cycle time C
2: $S'_d := \emptyset$
3: $c_{\text{rem}} := C$
4: $T' := \{i \in T \mid i \notin \bigcup_{r=1}^{d} S'_r, P_i \subseteq \bigcup_{r=1}^{d} S'_r, t_i \leq c_{\text{rem}}\}$
5: **while** $T' \neq \emptyset$ **do**
6: $\quad j :=$ ChooseTask(T',C,c_{rem})
7: $\quad c_{\text{rem}} := c_{\text{rem}} - t_j$
8: $\quad S'_d := S'_d \cup \{j\}$
9: $\quad T' := \{i \in T \mid i \notin \bigcup_{r=1}^{d} S'_r, P_i \subseteq \bigcup_{r=1}^{d} S'_r, t_i \leq c_{\text{rem}}\}$
10: **end while**
11: **output:** Filled work station S'_d

It remains to describe the implementation of function ChooseTask(T',C,c_{rem}) of Algorithm 2. For that purpose let us first define the following subset of T':

$$T^{\text{sat}} := \{i \in T' \mid t_i = c_{\text{rem}}\} \tag{8}$$

This definition is such that T^{sat} contains all tasks that *saturate*, in terms of processing time, the d-th work station S_d. The choice of a task from T' is made on the basis of

greedy information, that is, on the basis of values $\eta_i > 0$ that are assigned to all tasks $i \in T'$ by a greedy function. The first action for choosing a task from T' consists in flipping a coin for deciding if the choice is made deterministically, or probabilistically. In case of a deterministic choice, there are two possibilities. First, if $T^{\text{sat}} \neq \emptyset$, the best task from T^{sat} is chosen, that is, the task with maximal greedy value among all tasks in T^{sat}. Otherwise, we choose the task with maximal greedy value from T'. In case of a probabilistic decision, a task from T' is chosen on the basis of the following probability distribution:

$$\mathbf{p}(i) := \frac{\eta_i}{\sum_{j \in T'} \eta_j} \quad , \forall i \in T' \tag{9}$$

For completing the description of function ChooseTask(T', C, c_{rem}), we must describe the definition of the greedy values η_i, $\forall i \in T$. In a first step a term γ_i is defined as follows:

$$\gamma_i := \kappa_1 \cdot \left(\frac{t_i}{C}\right) + \kappa_2 \cdot \left(\frac{|\text{Suc}_i^{\text{all}}|}{\max_{1 \leq j \leq n} \left|\text{Suc}_j^{\text{all}}\right|}\right) \quad , \forall i \in T \tag{10}$$

Hereby, $\text{Suc}_i^{\text{all}}$ denotes the set of all tasks that can be reached from i in precedence graph G via a directed path. This definition combines two greedy function that are often used in the context of assembly line balancing problems. The first one concerns the task processing times and the second one concerns the size of $\text{Suc}_i^{\text{all}}$. The influence of both heuristics can be adjusted via the setting of weights κ_1 and κ_2. In order to be more flexible we decided to allow for both weights a value from $[-1, 1]$. This means that we consider for each heuristic potentially also its negation. This is motivated by experience from the field of scheduling, where some problem instances are more successfully solved by doing exactly the opposite of what is suggested by certain greedy functions. Given the γ_i-values, the greedy values η_i are then derived as follows:

$$\eta_i := \frac{\gamma_i - \gamma_{\min} + 1}{\gamma_{\max}} \quad \forall\, i \in T \quad , \tag{11}$$

where $\gamma_{\min}$, respectively $\gamma_{\max}$, denote the minimum, respectively maximum, values of all γ_i. Interestingly, for obtaining well-working greedy values, parameters κ_1 and κ_2 have to be chosen in a problem-instance-dependent way. A study concerning the values of parameters κ_1 and κ_2 is presented in Section 4.2.

The lower bound function. The new beam B is – at each step – chosen from B_{ext}. This choice is implemented by function SelectSolutions($B_{\text{ext}}, k_{\text{bw}}$) of Algorithm 2. First, the solutions in B_{ext} are ranked with respect to increasing lower bound values LB($\cdot$). Then, the $\min\{k_{\text{bw}}, |B_{\text{ext}}|\}$ highest ranked partial solutions from B_{ext} are selected. Let us denote

by $\overline{T} \subseteq T$ the set of tasks that have not yet been assigned to work stations in partial solution $\mathcal{S}'$. Then:

$$\mathrm{LB}(\mathcal{S}') = \left\lceil \frac{\sum_{i \in \overline{T}} t_i}{C} \right\rceil \tag{12}$$

Note that this lower bound is inspired by splitting-based bounds for the one-dimensional bin packing problem.

Algorithm 3 Iterative beam search (IBS) for the SALBP-2

input: an instance (T, G, m)
$C :=$ DetermineStartingCycleTime()
$C' := C$
$k_{\text{bw}} := 5$, $k_{\text{ext}} := 2$
success := FALSE
while not *success* **do**
  success := BeamSearch($(T, G, m), C, k_{\text{bw}}, k_{\text{ext}}$) {original instance}
  if not *success* **then**
    success := BeamSearch($(T, G^r, m), C, k_{\text{bw}}, k_{\text{ext}}$) {reverse instance}
    if not *success* **then** $C := C + 1$ **end if**
  end if
end while
if $C > C'$ **then**
  $C := C - 1$
  stop := FALSE
  while not *stop* **do**
    success := FALSE
    while time limit not reached **and not** *success* **do**
      if within 5% of time limit **then** $k_{\text{bw}} := 10$, $k_{\text{ext}} := 5$ **else** $k_{\text{bw}} := 150$, $k_{\text{ext}} := 20$ **end if**
      success := BeamSearch($(T, G, m), C, k_{\text{bw}}, k_{\text{ext}}$) {original instance}
      if not *success* **then**
        success := BeamSearch($(T, G^r, m), C, k_{\text{bw}}, k_{\text{ext}}$) {reverse instance}
      end if
    end while
    if *success* **then** $C := C - 1$ **else** *stop* := TRUE **end if**
  end while
  $C := C + 1$
end if
output: cycle time C

3.3. The algorithmic scheme

The BS component outlined in the previous section is used by an iterative algorithmic scheme that is presented in Algorithm 3. Henceforth this algorithmic scheme is labelled iterated beam search (IBS). The first step consists in determining a starting cycle time C, which is computed in funcion DetermineStartingCycleTime() of Algorithm 3 as

$$C := \max \left\{ \max_{i \in T} \{t_i\}, \left\lceil \frac{\sum_{i \in T} t_i}{m} \right\rceil \right\} \quad . \tag{13}$$

The algorithm works in two phases. In the first phase (see lines 4-12 of Algorithm 3) the algorithm tries to quickly find a first cycle time C for which a valid solution can be found. For this purpose BS is applied with the setting $k_{\text{bw}} = 5$ and $k_{\text{ext}} = 2$. Note that this setting was chosen after tuning by hand. Moreover, note that the first phase only takes a fraction of a second of computation time. This holds for all instances considered in Section 4. The second phase of the algorithm iteratively tries to find a valid solution for the next smaller cycle time. In this phase, the algorithm disposes over a certain time limit for each considered cycle time. Remember that the working of BS is partially probabilistic. Therefore, BS can repeatedly be applied to the same instance with potentially different outcomes. The first five percent of the above-mentioned time limit are spent by BS applications that use the setting $k_{\text{bw}} := 10$ and $k_{\text{ext}} := 5$. This is done with the intention of not wasting too much computation time, if not necessary. However, if BS is not able to solve the given cycle time with this setting, the remaining 95% of the available time are spent by BS applications using the setting $k_{\text{bw}} := 150$ and $k_{\text{ext}} := 20$. With this setting BS is considerably slower. However, the probability of finding feasible solutions is much higher than with the setting described before. The second phase of the algorithm ends when the time limit has passed without having found a feasible solution for the considered cycle time.

4. Experimental evaluation

IBS was implemented in ANSI C++, and GCC 3.4.0 was used for compiling the software. Experimental results were obtained on a PC with an AMD64X2 4400 processor and 4 Gb of memory. In the following we first describe the set of benchmark instances that we used for the experimental evaluation. Subsequently we present a study concerning some of the parameters of the proposed algorithm. Finally, the experimental results are presented.

4.1. Benchmark instances

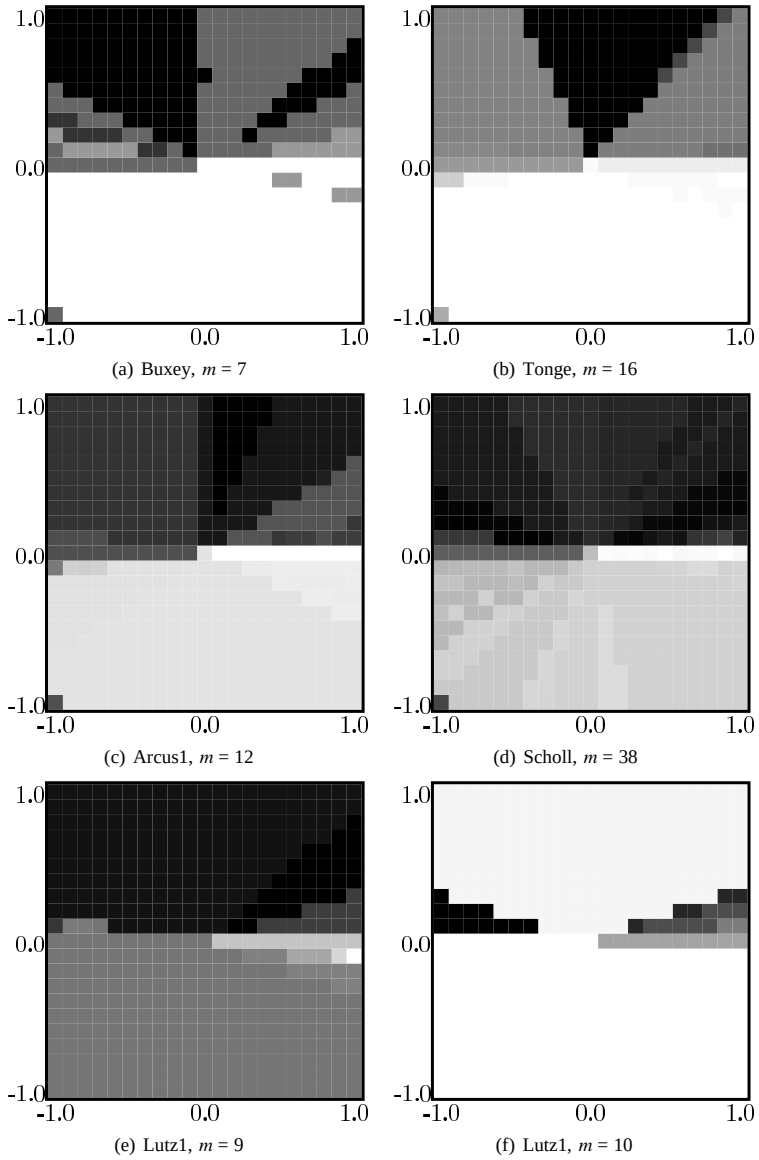
We used the usual set of 302 benchmark instances from the literature. They can be obtained – together with information about optimal solutions, respectively lower and upper bounds – from a website especially dedicated to all kind of assembly line balancing problems maintained by Armin Scholl, `http://www.assembly-line-balancing.de`. Each instance consists of a precedence graph G and a given number m of work stations. The benchmark set is composed of two subsets of instances, henceforth called Dataset1 and Dataset2. Dataset1 consists of 128 instances based on 9 different precedence graphs with a number of tasks between 29 to 111. Dataset2 is composed of 174 instances based on 8 different precedence graphs with a number of tasks varying from 53 to 297.

4.2. A study of parameters κ_1 and κ_2

During preliminar experiments we realized that parameters k_{bw} and k_{ext} have a rather low impact on the final results of IBS. In other words it is easy to find a reasonable setting for these parameters quite quickly. Their setting dynamically changes during a run of the algorithm as specified in Section 3.3. On the contrary, parameters κ_1 and κ_2 (see Eq. 10) have a rather high impact on the algorithms' performance. Remember that κ_1 is the weight of the greedy function concerning the task processing times, while κ_2 is the weight of the greedy function concerning the number of tasks that have to be processed after the task under consideration. As mentioned before, for both parameters we allowed values from $[-1,1]$. Instead of trying to find a good parameter setting for each single instance, we decided for a process aimed at identifying a single setting of κ_1 and κ_2 for all instances concerning the same precedence graph. For that purpose we applied a specific version of IBS for all combinations of $\kappa_1,\kappa_2 \in \{-1.0,-0.9,\ldots,0.0,\ldots,0.9,1.0\}$ to all 302 instances. This makes a total of 441 different settings for each instance. The specific version of IBS that we apply in the following differs from IBS as outlined in Algorithm 3 in that lines 14-24 are replaced by a single, deterministic, application of beam search with $k_{\text{bw}} = 150$ and $k_{\text{ext}} = 20$. This was done for the purpose of saving computation time. Based on the obtained results we chose the settings presented in Table 1

Table 1: *Values of parameters κ_1 and κ_2 for the final experiments.*

Graph	κ_1	κ_2	**Graph**	κ_1	κ_2
Arcus1	0.0	1.0	Lutz2	−0.1	0.9
Arcus2	−0.5	0.9	Lutz3	0.0	1.0
Barthol2	0.0	1.0	Mukherje	0.0	1.0
Barthold	0.0	1.0	Sawyer	0.0	1.0
Buxey	0.0	1.0	Scholl	0.0	1.0
Gunther	−0.4	0.8	Tonge	−0.1	0.2
Hahn	0.0	1.0	Warnecke	0.0	1.0
Kilbridge	0.0	1.0	Wee−Mag	0.0	1.0
Lutz1	−0.1	0.9			



(a) Buxey, $m = 7$ (b) Tonge, $m = 16$

(c) Arcus1, $m = 12$ (d) Scholl, $m = 38$

(e) Lutz1, $m = 9$ (f) Lutz1, $m = 10$

Figure 1: *Results presented in graphical form for six representative instances. The y-axis ranges over the values of κ_1, while the x-axis ranges over the values of κ_2. The gray levels of the squares indicate the quality of the algorithm when run with the corresponding setting: the lighter a square is painted, the better is the parameter setting.*

for the different precedence graphs. It is interesting to note that, apart from a few exceptions, the greedy heuristic based on task processing times does not seem necessary for obtaining good results.

In Figure 1 a representative sample of the results is provided in graphical form. The y-axis of the presented graphics varies over the different values of κ_1, while the x-axis

ranges over the allowed values of κ_2. Note that each graphic consists of 441 squares representing the 441 different combinations of values for κ_1 and κ_2. The gray level in which each square is painted indicates the quality of the algorithm when run with the corresponding parameter setting. In particular, black color denotes the worst setting, whereas white color indicates the best algorithm setting. In some cases such as (Buxey, $m = 7$) and (Tonge, $m = 16$), as shown in Figures 1(a) and 1(b), there is a wide range of good settings, which are basically all those with $\kappa_1 \leq 0$. In other examples such as (Arcus1, $m = 12$) and (Scholl, $m = 38$) it is strictly required to set κ_1 to 0 and κ_2 to a positive value for obtaining good solutions; see Figures 1(c) and 1(d). Finally, the graphics shown in Figures 1(e) and 1(f) indicate that even for the same precedence graph a good parameter setting might depend strongly on the number of work stations.

Table 2: *Differences in algorithm performance when considering the best and the worst parameter setting.*

Instance	Best setting	Worst setting	Difference (%)
(Buxey, $m = 7$)	47	52	10.64
(Tonge, $m = 16$)	222	265	19.37
(Arcus1, $m = 12$)	12599	13767	9.27
(Scholl, $m = 38$)	1857	2031	9.37
(Lutz1, $m = 9$)	1637	1801	10.02
(Lutz1, $m = 10$)	1525	1619	6.16

It is also interesting to quantify the differences in algorithm performance for different parameter settings. Table 2 shows for the six cases presented in Figure 1 the result of the algorithm with the best setting (column **Best setting**), the result of the algorithm with the worst setting (column **Worst setting**), and the difference (in percent) between these two settings. The results in Table 2 show that there are considerable differences in performance between the best and the worst algorithm setting. This underlines the importance of finding opportune values for κ_1 and κ_2.

4.3. Results

Algorithm IBS was applied 20 times to all 302 instances. Herefore we used a computation time limit of 180 seconds for each cycle time, that is, IBS was given maximally 180 seconds for finding a feasible solution for a given cycle time. In case of success, the algorithm has again 180 seconds for the next smaller cycle time, etc. Detailed results of IBS for all 302 instances are given in Tables 7 and 8 that are to be found in Appendix A. The data are, in both tables, presented as follows. The first two columns provide the name of the precedence graph and the number of work stations (m). The third column (labelled *bub*) provides the values of the optimal solutions in case they are known. If they are unknown the column reports the currently best upper bound. In case a value is not proved to be optimal it is overlined. More in detail, in 15 out of 302 cases optimality

has not been proved yet. The remaining five columns are reserved for the results of IBS. The first of these five columns contains the value of the best solution found by IBS over 20 runs. In case this value is presented with a gray background, a new best upper bound has been found. On the other side, if this value is marked by an asterisk, the obtained result does not reach the value of the best upper bound. In all other cases the values correspond to the values of the best upper bounds. The second column provides the average over 20 runs, while the third column contains the corresponding standard deviation. The fourth column gives the average time (in seconds) at which the best solution of a run was found, averaged over 20 runs. The fifth column provides the corresponding standard deviation. From the results presented in Tables 7 and 8 (see Appendix A) we can observe that IBS obtains optimal solutions, respectively best upper bounds, in 276 out of 302 cases. Moreover, new best upper bounds are obtained in 6 cases. This is remarkable as – despite a considerable amount of ongoing research – in the last 14 years no improved solutions have been reported. Only in 20 cases (all concerning precedence graphs Arcus1, Arcus2, Scholl, and Warnecke) our algorithm was not able to find the best solutions known. However, in most of these cases the deviation from the best upper bound is no more than one unit of cycle time.

In addition to Tables 7 and 8 the results of IBS are presented in a summarized way in Table 3, in comparison to three other algorithms. TABUSEARCH Scholl and Voss (1996), even though already published in 1996, still counts as the current state-of-the-art heuristic method for SALBP-2. DE_RKS is the best version of a differential evolution (DE) algorithm proposed in Nearchou (2007), and PNA-FOR is a Petri net-based heuristic published in Kilincci (2010). The last two methods are, to our knowledge, the most recently published heuristic methods for SALBP-2. Three measures are used in Table 3 for the comparison of IBS with these three algorithms. The row labelled **#opt** provides the number of best upper bounds found by each method (over 302). Moreover, the row labelled **mrd (%)** gives the *mean relative deviation (in percent)* of the results obtained by the four algorithms from the best-known upper bounds for all 302 instances. Finally, row **time** contains the average computation time of the algorithms for all 302 instances. Concerning the quality of the results, we can conclude that IBS clearly outperforms its competitors. For the correct interpretation of the computation times it has to be taken into account that the four algorithms were executed on computers with very different processor speeds. While TABUSEARCH was executed on a 80486 DX2-66 processor, PNA-FOR was run on an Athlon XP 2000+ processor with

Table 3: *Results of* IBS *in comparison to the best (*TABUSEARCH*), respectively most recent (*DE_RKS *and* PNA-FOR*), methods from the literature.*

	DE_RKS	PNA-FOR	TABUSEARCH	**Iterative Beam Search** (IBS)
#opt	n/g	39	168	282
mrd (%)	2.64	2.73	0.40	0.0029
time (s)	10.74	407.28	84.90	31.61

Note: n/g means **not given**

1.67 GHz, and DE_RKS was run on a Pentium IV processor with 1.7 GHz. This means that TABUSEARCH was run by far on the slowest machine, IBS by far on the fastest machine, and PNA-FOR and DE_RKS on comparable machines. Given the computation times in Table 3 we can savely conclude that TABUSEARCH is the fastest algorithm, and PNA-FOR is the slowest one. However, note that assembly line balancing is, in most cases, not a time-critical application. In other words, for most practical purposes it does not matter if an algorithm takes 1 minute or 6 hours of computation time.

In the following we present the results of IBS in comparison to DE_RKS and PNA-FOR in the same way as done in Nearchou (2007) and Kilincci (2010). In these works, results were presented as averages over instances based on the same precedence graph, and also averaged over Dataset1 and Dataset2. The quality of the results is given in terms of the *mean relative deviation (in percent)* from the best-known upper bounds. Tables 4 and 5 clearly show that IBS is largely superior to both competitor algorithms.

Table 4: *Results of* IBS *in comparison to* DE_RKS *and* PNA-FOR *for the 128 instances of* Dataset1 *(averaged over precedence graphs)*

Graph	DE_RKS		PNA-FOR	**Iterative Beam Search** (IBS)	
	mrd (%)	time (s)	mrd (%)	mrd (%)	time (s)
Buxey	1.16	0.80	3.07	0.0	0.06
Sawyer	2.27	1.64	4.00	0.0	0.14
Lutz1	0.32	0.88	4.09	0.0	0.55
Gunther	0.14	1.08	1.27	0.0	0.02
Kilbridge	0.66	1.43	2.19	0.0	0.0067
Tonge	1.88	3.71	2.53	0.0	7.73
Arcus1	0.99	5.29	2.47	0.0287	152.34
Lutz2	3.08	1.00	2.99	0.0	0.035
Arcus2	4.96	19.02	2.06	0.0066	135.83
Average:	**1.72**	**3.87**	**2.57**	**0.0058**	**51.76**

Table 5: *Results of* IBS *in comparison to* DE_RKS *and* PNA-FOR *for the 174 instances of* Dataset2 *(averaged over precedence graphs).*

Graph	DE_RKS		PNA-FOR	**Iterative Beam Search** (IBS)	
	mrd (%)	time (s)	mrd (%)	mrd (%)	time (s)
Hahn	0.0	1.00	2.52	0.0	0.065
Warnecke	3.74	3.53	5.57	0.0579	1.54
Wee-Mag	1.23	3.68	1.56	0.0	0.65
Lutz3	1.68	5.62	2.59	0.0	0.61
Mukherje	n/a	n/a	1.04	0.0	2.52
Barthold	0.26	19.47	1.02	0.0	0.079
Barthol2	6.85	33.08	3.97	0.0	0.96
Scholl	9.51	43.95	3.21	0.0028	98.69
Average:	**3.32**	**15.79**	**2.85**	**0.00071**	**16.79**

Table 6: *Results of a high-performance version of* IBS.

Graph	m	Result	Graph	m	Result
Arcus1 (83)	8	9554	Arcus2 (111)	23	6560
	11	7085*		24	6282
	12	6412		25	6101
	17	4527*		26	5855
	18	4323*	Scholl (297)	31	2247
	19	4071*		36	1936*
	20	3886*		42	1660*
Arcus2 (111)	14	10747		46	1515
	15	10035		47	1484*
	16	9413*		49	1423
	17	8857*	Warnecke (58)	25	64
	18	8377	Wee-Mag (75)	18	87
	19	7922		19	85
	20	7524		23	67
	21	7187		27	65
	22	6856		28	64

4.4. Results of a high-performance version

In an attempt to further improve on the results of our algorithm we decided to apply a high-performance version of IBS to all problem instances for which the optimal solution is unknown and, additionally, to all instances where IBS – with the settings as outlined in the previous section – was not able to find the best known upper bounds. This high-performance version is obtained as follows. First, 1800 seconds are used as a time limit for each cycle time. Second, in line 17 of Algorithm 3 only 1% of the time limit is used (instead of 5%). Third, for each application of beam search in lines 18 and 20 of Algorithm 3 the beam width k_{bw} is randomly chosen from $[150,250]$ and the number of extensions is randomly chosen from $[20,40]$. Moreover, with a probability of 0.5 the heuristic information is – for each application of beam search – calculated using the weight values as outlined in Table 1. Otherwise, the weight values are chosen randomly from $[-1,1]$. With these modifications we applied IBS exactly once to all the instances of Table 6. The results of the algorithm are given in column **Result**: In case the result value is presented with a gray background, a new best upper bound has been found. On the other hand, if this value is marked by an asterisk, the obtained result is inferior to the value of the best known upper bound. Indeed, the number of instances for which the best-known upper bound can not be found before is reduced from 20 to 10 instances. Moreover, the algorithm is now able to find new best-known upper bounds in 9 (instead of only 6) cases. Summarizing, this amounts to 283 best-known upper bounds found and 9 new best upper bounds obtained. In one case, (Scholl, $m = 49$), the new upper bound is provenly optimal, as its value coincides with the best known lower bound.

5. Conclusions and future work

In this work we have proposed an iterative beam search algorithm for the simple assembly line balancing problem with a fixed number of work stations, SALBP-2. The experimental evalution of the algorithm has shown that it is currently a state-of-the-art method for this problem. Appart from producing optimal solutions, respectively best upper bounds, for 283 out of 302 test cases, our algorithm generated new best-known upper bounds in further 9 test cases. Encouraged by the results for the SALBP-1 variant of the problem – as published in Blum (2008) – and the results obtained in this paper for the SALBP-2 we intent to apply similar algorithms based on beam search to other assembly line balancing problems.

Acknowledgements

This work was supported by grant TIN2007-66523 (FORMALISM) of the Spanish government. Moreover, Christian Blum acknowledges support from the *Ramón y Cajal* program of the Spanish Ministry of Science and Innovation.

Many thanks go to Armin Scholl for verifying the new best-known solutions found by the algorithm proposed in this work. Finally, we would also like to express our thanks to Cristóbal Miralles who was involved as a co-author of a similar work for the more general assembly line worker assignment and balancing problem.

References

H. Akeba, M. Hifib and R. MHallah (2009). A beam search algorithm for the circular packing problem. *Computers & Operations Research*, 36(5), 1513–1528.

E. J. Anderson and M. C. Ferris (1994). Genetic algorithms for combinatorial optimization: The assembly line balancing problem. *ORSA Journal on Computing*, 6, 161–173.

C. Blum (2008). Beam-ACO for simple assembly line balancing. *INFORMS Journal on Computing*, 20(4), 618–627.

C. Blum and C. Miralles (2011). On solving the assembly line worker assignment and balancing problem via beam search. *Computers & Operations Research*, 38(1), 328–339.

C. Blum, M. J. Blesa and M. López Ibáñez (2009). Beam search for the longest common subsequence problem. *Computers & Operations Research*, 36(12), 3178–3186.

W.-C. Chiang (1998). The application of a tabu search metaheuristic to the assembly line balancing problem. *Annals of Operations Research*, 77, 209–227.

M. Ghirardi and C. N. Potts (2005). Makespan minimization for scheduling unrelated parallel machines: A recovering beam search approach. *European Journal of Operational Research*, 165(2), 457–467.

S. Gosh and R. J. Gagnon (1989). A comprehensive literature review and analysis of the design, balancing and scheduling of assembly systems. *International Journal of Production Research*, 27, 637–670.

S. T. Hackman, M. J. Magazine and T. S. Wee (1989). Fast, effective algorithms for simple assembly line balancing problems. *Operations Research*, 37, 916–924.

A. Henrici (1994). A comparison between simulated annealing and tabu search with an example from the production planning. In H. Dyckhoff et al., editor, *Operations Research Proceedings*, 498–503. Springer Verlag, Berlin, Germany.

O. Kilincci (2010). A Petri net-based heuristic for simply assembly line balancing problem of type 2. *International Journal of Advanced Manufacturing Technology*, 46, 329–338.

R. Klein and A. Scholl (1996). Maximizing the production rate in simple assembly line balancing – A branch and bound procedure. *European Journal of Operational Research*, 91, 367–385.

G.-C. Lee and D. L. Woodruff (2004). Beam search for peak alignment of NMR signals. *Analytica Chimica Acta*, 513(2), 413–416.

B. Lowerre (1976). *The Harpy Speech Recognition System*. PhD thesis, Carnegie Mellon University.

A. C. Nearchou (2007). Balancing large assembly lines by a new heuristic based on differential evolution. *International Journal of Advanced Manufacturing Technology*, 34, 1016–1029.

P. S. Ow and T. E. Morton (1988). Filtered beam search in scheduling. *International Journal of Production Research*, 26, 297–307.

R. Pastor and L. Ferrer (2009). An improved mathematical program to solve the simple assembly line balancing problem. *International Journal of Production Research*, 47(11), 2943–2959.

R. Pastor, L. Ferrer and A. García (2007). Evaluating optimization models to solve SALBP. In O. Gervasi and M. L. Gavrilova, editors, *Proceedings of ICCSA 2007 – International Conference on Computational Science and Its Applications*, volume 4705 of *Lecture Notes in Computer Science*, 791–803. Springer Verlag, Berlin, Germany.

I. Sabuncuoglu and M. Bayiz (1999). Job shop scheduling with beam search. *European Journal of Operational Research*, 118, 390–412.

M. E. Salveson (1955). The assembly line balancing problem. *Journal of Industrial Engineering*, 6, 18–25.

A. Scholl (1994). Ein B&B-Verfahren zur Abstimmung von Einprodukt-Fließbändern bei gegebener Stationsanzahl. In H. Dyckhoff et al., editor, *Operations Research Proceedings*, 175–181. Springer Verlag, Berlin, Germany.

A. Scholl (1999). *Balancing and sequencing assembly lines*. Physica Verlag, Heidelberg, Germany, 2nd edition edition.

A. Scholl and C. Becker (2006). State-of-the-art exact and heuristic solution procedures for simple assembly line balancing. *European Journal of Operational Research*, 168(3), 666–693.

A. Scholl and S. Voss (1996). Simple assembly line balancing – Heuristic approaches. *Journal of Heuristics*, 2, 217–244.

H. F. Ugurdag, R. Rachamadugu and C. A. Papachristou (1997). Designing paced assembly lines with fixed number of stations. *European Journal of Operational Research*, 102, 488–501.

J. M. S. Valente and R. A. F. S. Alves (2005). Filtered and recovering beam search algorithms for the early/tardy scheduling problem with no idle time. *Computers & Industrial Engineering*, 48(2), 363–375.

T. Watanabe, Y. Hashimoto, L. Nishikawa and H. Tokumaru (1995). Line balancing using a genetic evolution model. *Control Engineering Practice*, 3, 69–76.

Appendix A

Table 7: *Detailed results of* IBS *for 302 test instances (Part A).*

Graph	m	bub	Iterative Beam Search (IBS)					Graph	m	bub	Iterative Beam Search (IBS)				
			best	avg	std	time (s)	std				best	avg	std	time (s)	std
Arcus1 (83)	3	25236	25236	25236.00	(0.00)	0.56	(0.29)	Barthold (148)	9	626	626	626.00	(0.00)	0.13	(0.03)
	4	18927	18927	18927.00	(0.00)	6.16	(3.92)		10	564	564	564.00	(0.00)	0.10	(0.02)
	5	15142	15142	15142.00	(0.00)	80.08	(16.69)		11	513	513	513.00	(0.00)	0.09	(0.00)
	6	12620	12620	12620.00	(0.00)	15.52	(5.76)		12	470	470	470.00	(0.00)	0.09	(0.00)
	7	10826	10826	10826.00	(0.00)	59.80	(19.89)		13	434	434	434.00	(0.00)	0.16	(0.02)
	8	9554	*9555	9555.55	(0.51)	120.54	(63.16)		14	403	403	403.00	(0.00)	0.14	(0.01)
	9	8499	8499	8500.70	(0.80)	130.73	(117.23)		15	383	383	383.00	(0.00)	0.05	(0.00)
	10	7580	7580	7580.95	(0.60)	311.89	(109.87)	Buxey (29)	7	47	47	47.00	(0.00)	0.46	(0.51)
	11	7084	*7086	7086.70	(0.47)	71.13	(52.77)		8	41	41	41.00	(0.00)	0.00	(0.00)
	12	6412	*6413	6414.10	(0.72)	362.32	(66.26)		9	37	37	37.00	(0.00)	0.01	(0.00)
	13	5864	5864	5864.00	(0.00)	129.18	(12.99)		10	34	34	34.00	(0.00)	0.00	(0.00)
	14	5441	5441	5441.00	(0.00)	2.02	(0.41)		11	32	32	32.00	(0.00)	0.00	(0.00)
	15	5104	5104	5104.45	(0.60)	222.29	(73.92)		12	28	28	28.00	(0.00)	0.01	(0.00)
	16	4850	4850	4850.00	(0.00)	33.84	(9.59)		13	27	27	27.00	(0.00)	0.01	(0.00)
	17	4516	*4524	4526.20	(0.89)	282.49	(100.96)		14	25	25	25.00	(0.00)	0.00	(0.00)
	18	4317	*4322	4323.20	(0.77)	291.73	(111.23)	Gunther (35)	6	84	84	84.00	(0.00)	0.00	(0.00)
	19	4068	*4073	4074.80	(1.01)	326.78	(125.49)		7	72	72	72.00	(0.00)	0.03	(0.00)
	20	3882	*3886	3889.25	(2.05)	594.21	(204.09)		8	63	63	63.00	(0.00)	0.01	(0.00)
	21	3691	3691	3691.00	(0.00)	5.58	(0.81)		9	54	54	54.00	(0.00)	0.05	(0.04)
	22	3691	3691	3691.00	(0.00)	0.02	(0.00)		10	50	50	50.00	(0.00)	0.01	(0.00)
Arcus2 (111)	3	50133	50133	50133.00	(0.00)	0.09	(0.03)		11	48	48	48.00	(0.00)	0.00	(0.00)
	4	37600	37600	37600.00	(0.00)	0.41	(0.02)		12	44	44	44.00	(0.00)	0.01	(0.00)
	5	30080	30080	30080.00	(0.00)	0.43	(0.22)		13	42	42	42.00	(0.00)	0.01	(0.00)
	6	25067	25067	25067.00	(0.00)	1.13	(0.59)		14	40	40	40.00	(0.00)	0.02	(0.00)
	7	21486	21486	21486.00	(0.00)	1.76	(1.04)		15	40	40	40.00	(0.00)	0.01	(0.00)
	8	18800	18800	18800.00	(0.00)	11.36	(2.86)	Hahn (53)	3	4787	4787	4787.00	(0.00)	0.00	(0.00)
	9	16711	16711	16711.00	(0.00)	28.03	(12.51)		4	3677	3677	3677.00	(0.00)	0.08	(0.01)
	10	15040	15040	15040.00	(0.00)	38.67	(15.08)		5	2823	2823	2823.00	(0.00)	0.01	(0.00)
	11	13673	13673	13673.00	(0.00)	49.47	(12.83)		6	2400	2400	2400.00	(0.00)	0.05	(0.01)
	12	12534	12534	12534.00	(0.00)	43.34	(15.17)		7	2336	2336	2336.00	(0.00)	0.29	(0.01)
	13	11570	11570	11570.00	(0.00)	32.60	(14.63)		8	1907	1907	1907.00	(0.00)	0.09	(0.03)
	14	10747	*10748	10748.00	(0.00)	62.69	(16.17)		9	1827	1827	1827.00	(0.00)	0.00	(0.00)
	15	10035	*10036	10036.40	(0.50)	112.54	(63.37)		10	1775	1775	1775.00	(0.00)	0.00	(0.00)
	16	9412	*9416	9416.60	(0.68)	290.35	(90.83)	Kilbridge (45)	3	184	184	184.00	(0.00)	0.00	(0.00)
	17	8855	*8864	8864.90	(0.31)	87.41	(52.86)		4	138	138	138.00	(0.00)	0.01	(0.00)
	18	$\overline{8377}$	8377	8377.00	(0.00)	8.11	(1.00)		5	111	111	111.00	(0.00)	0.00	(0.00)
	19	$\overline{7928}$	7924	7925.60	(0.60)	205.09	(76.12)		6	92	92	92.00	(0.00)	0.00	(0.00)
	20	$\overline{7526}$	7524	7524.40	(0.50)	159.72	(57.92)		7	79	79	79.00	(0.00)	0.01	(0.00)
	21	$\overline{7188}$	*7192	7193.40	(0.82)	318.78	(89.64)		8	69	69	69.00	(0.00)	0.01	(0.00)
	22	$\overline{6859}$	6858	6858.20	(0.41)	226.94	(75.52)		9	62	62	62.00	(0.00)	0.01	(0.00)
	23	$\overline{6561}$	6560	6563.10	(1.45)	428.68	(168.53)		10	56	56	56.00	(0.00)	0.01	(0.00)
	24	$\overline{6289}$	6284	6285.65	(1.46)	311.77	(144.55)		11	55	55	55.00	(0.00)	0.01	(0.00)
	25	$\overline{6106}$	*6112	6114.15	(0.99)	305.50	(90.99)	Lutz1 (32)	8	1860	1860	1860.00	(0.00)	0.14	(0.00)
	26	$\overline{5856}$	*5858	5860.45	(1.76)	661.48	(192.47)		9	1638	1638	1638.00	(0.00)	2.41	(0.34)
	27	5689	5689	5689.00	(0.00)	9.29	(0.17)		10	1526	1526	1526.00	(0.00)	0.18	(0.01)
Barthol2 (148)	27	157	157	157.00	(0.00)	0.31	(0.03)		11	1400	1400	1400.00	(0.00)	0.01	(0.00)
	28	152	152	152.00	(0.00)	0.44	(0.07)		12	1400	1400	1400.00	(0.00)	0.00	(0.00)
	29	146	146	146.00	(0.00)	0.80	(0.32)	Lutz2 (89)	9	54	54	54.00	(0.00)	0.02	(0.01)
	30	142	142	142.00	(0.00)	0.15	(0.03)		10	49	49	49.00	(0.00)	0.03	(0.00)
	31	137	137	137.00	(0.00)	0.67	(0.09)		11	45	45	45.00	(0.00)	0.04	(0.00)
	32	133	133	133.00	(0.00)	0.57	(0.07)		12	41	41	41.00	(0.00)	0.03	(0.00)
	33	129	129	129.00	(0.00)	0.52	(0.06)		13	38	38	38.00	(0.00)	0.03	(0.00)
	34	125	125	125.00	(0.00)	0.47	(0.05)		14	35	35	35.00	(0.00)	0.03	(0.00)
	35	121	121	121.00	(0.00)	0.94	(0.15)		15	33	33	33.00	(0.00)	0.03	(0.00)
	36	118	118	118.00	(0.00)	0.87	(0.11)		16	31	31	31.00	(0.00)	0.02	(0.00)
	37	115	115	115.00	(0.00)	0.92	(0.19)		17	29	29	29.00	(0.00)	0.04	(0.00)
	38	112	112	112.00	(0.00)	0.76	(0.06)		18	28	28	28.00	(0.00)	0.04	(0.00)
	39	109	109	109.00	(0.00)	0.70	(0.13)		19	26	26	26.00	(0.00)	0.04	(0.01)
	40	106	106	106.00	(0.00)	0.91	(0.32)		20	25	25	25.00	(0.00)	0.04	(0.00)
	41	104	104	104.00	(0.00)	0.72	(0.05)		21	24	24	24.00	(0.00)	0.02	(0.00)
	42	101	101	101.00	(0.00)	2.77	(5.43)		22	23	23	23.00	(0.00)	0.04	(0.00)
	43	99	99	99.00	(0.00)	0.84	(0.23)		23	22	22	22.00	(0.00)	0.05	(0.00)
	44	97	97	97.00	(0.00)	0.59	(0.15)		24	21	21	21.00	(0.00)	0.05	(0.00)
	45	95	95	95.00	(0.00)	0.76	(0.25)		25	20	20	20.00	(0.00)	0.02	(0.00)
	46	93	93	93.00	(0.00)	0.68	(0.29)		26	19	19	19.00	(0.00)	0.05	(0.02)
	47	91	91	91.00	(0.00)	0.67	(0.15)		27	19	19	19.00	(0.00)	0.03	(0.00)
	48	89	89	89.00	(0.00)	1.17	(0.47)		28	18	18	18.00	(0.00)	0.05	(0.00)
	49	87	87	87.00	(0.00)	0.83	(0.26)	Lutz3 (89)	3	548	548	548.00	(0.00)	0.01	(0.00)
	50	85	85	85.90	(0.31)	3.28	(10.33)		4	411	411	411.00	(0.00)	0.03	(0.00)
	51	84	84	84.00	(0.00)	2.60	(2.09)		5	329	329	329.00	(0.00)	0.01	(0.00)
Barthold (148)	3	1878	1878	1878.00	(0.00)	0.01	(0.00)		6	275	275	275.00	(0.00)	0.01	(0.00)
	4	1409	1409	1409.00	(0.00)	0.02	(0.00)		7	236	236	236.00	(0.00)	0.02	(0.00)
	5	1127	1127	1127.00	(0.00)	0.02	(0.00)		8	207	207	207.00	(0.00)	0.04	(0.00)
	6	939	939	939.00	(0.00)	0.06	(0.01)		9	184	184	184.00	(0.00)	5.54	(4.14)
	7	805	805	805.00	(0.00)	0.11	(0.00)		10	165	165	165.00	(0.00)	0.09	(0.01)
	8	705	705	705.00	(0.00)	0.04	(0.01)		11	151	151	151.00	(0.00)	0.06	(0.00)

Table 8: *Detailed results of* IBS *for 302 test instances (Part B).*

Graph	m	bub	Iterative Beam Search (IBS)					Graph	m	bub	Iterative Beam Search (IBS)				
			best	avg	std	time (s)	std				best	avg	std	time (s)	std
Lutz3 (89)	12	138	138	138.00	(0.00)	0.02	(0.00)	Tonge (70)	6	585	585	585.00	(0.00)	0.03	(0.02)
	13	128	128	128.00	(0.00)	0.05	(0.03)		7	502	502	502.00	(0.00)	0.06	(0.01)
	14	118	118	118.00	(0.00)	0.67	(0.52)		8	439	439	439.00	(0.00)	0.04	(0.01)
	15	110	110	110.00	(0.00)	0.80	(0.66)		9	391	391	391.00	(0.00)	0.05	(0.00)
	16	105	105	105.00	(0.00)	0.34	(0.30)		10	352	352	352.00	(0.00)	0.06	(0.01)
	17	98	98	98.00	(0.00)	0.19	(0.05)		11	320	320	320.00	(0.00)	0.06	(0.02)
	18	93	93	93.00	(0.00)	0.16	(0.05)		12	294	294	294.00	(0.00)	0.09	(0.01)
	19	89	89	89.00	(0.00)	0.05	(0.03)		13	271	271	271.00	(0.00)	0.06	(0.01)
	20	85	85	85.00	(0.00)	0.14	(0.05)		14	251	251	251.80	(0.41)	20.25	(45.26)
	21	80	80	80.00	(0.00)	0.11	(0.02)		15	235	235	235.00	(0.00)	0.35	(0.16)
	22	76	76	76.00	(0.00)	4.20	(3.30)		16	221	221	221.00	(0.00)	0.97	(0.93)
	23	74	74	74.00	(0.00)	0.29	(0.18)		17	208	208	208.00	(0.00)	1.02	(0.95)
Mukherje (94)	3	1403	1403	1403.00	(0.00)	0.01	(0.00)		18	196	196	196.15	(0.37)	46.76	(39.05)
	4	1052	1052	1052.00	(0.00)	0.10	(0.00)		19	186	186	186.35	(0.49)	87.09	(64.62)
	5	844	844	844.00	(0.00)	0.02	(0.00)		20	177	177	177.90	(0.31)	9.46	(39.39)
	6	704	704	704.00	(0.00)	0.02	(0.00)		21	170	170	170.00	(0.00)	8.71	(9.43)
	7	621	621	621.00	(0.00)	0.12	(0.00)		22	162	162	162.00	(0.00)	2.52	(2.45)
	8	532	532	532.00	(0.00)	0.09	(0.01)		23	156	156	156.00	(0.00)	0.11	(0.04)
	9	471	471	471.00	(0.00)	0.09	(0.01)		24	156	156	156.00	(0.00)	0.02	(0.00)
	10	424	424	424.00	(0.00)	0.02	(0.00)		25	156	156	156.00	(0.00)	0.02	(0.00)
	11	391	391	391.00	(0.00)	0.10	(0.01)	Warnecke (58)	3	516	516	516.00	(0.00)	0.02	(0.00)
	12	358	358	358.00	(0.00)	0.05	(0.01)		4	387	387	387.00	(0.00)	0.08	(0.08)
	13	325	325	325.00	(0.00)	0.17	(0.06)		5	310	310	310.00	(0.00)	0.03	(0.00)
	14	311	311	311.00	(0.00)	0.08	(0.01)		6	258	258	258.00	(0.00)	0.10	(0.06)
	15	288	288	288.00	(0.00)	0.09	(0.02)		7	222	222	222.00	(0.00)	0.03	(0.00)
	16	268	268	268.00	(0.00)	0.20	(0.02)		8	194	194	194.00	(0.00)	0.05	(0.01)
	17	251	251	251.00	(0.00)	0.45	(0.05)		9	172	172	172.00	(0.00)	2.40	(2.28)
	18	239	239	239.00	(0.00)	0.24	(0.02)		10	155	155	155.00	(0.00)	0.05	(0.01)
	19	226	226	226.00	(0.00)	0.03	(0.01)		11	142	142	142.00	(0.00)	0.04	(0.00)
	20	220	220	220.05	(0.22)	57.07	(43.23)		12	130	130	130.00	(0.00)	0.07	(0.01)
	21	208	208	208.00	(0.00)	0.12	(0.00)		13	120	120	120.00	(0.00)	0.08	(0.01)
	22	200	200	200.00	(0.00)	0.13	(0.06)		14	111	111	111.00	(0.00)	0.29	(0.24)
	23	189	189	189.00	(0.00)	0.12	(0.02)		15	104	104	104.00	(0.00)	0.27	(0.13)
	24	179	179	179.00	(0.00)	0.48	(0.41)		16	98	98	98.00	(0.00)	0.08	(0.01)
	25	172	172	172.00	(0.00)	0.43	(0.18)		17	92	92	92.00	(0.00)	0.26	(0.14)
	26	171	171	171.00	(0.00)	0.27	(0.03)		18	87	87	87.00	(0.00)	0.27	(0.14)
Sawyer (30)	7	47	47	47.00	(0.00)	0.14	(0.11)		19	84	84	84.00	(0.00)	0.10	(0.06)
	8	41	41	41.00	(0.00)	0.06	(0.03)		20	79	79	79.00	(0.00)	0.13	(0.03)
	9	37	37	37.00	(0.00)	0.03	(0.01)		21	76	76	76.00	(0.00)	0.08	(0.04)
	10	34	34	34.00	(0.00)	0.02	(0.00)		22	73	73	73.00	(0.00)	0.53	(0.36)
	11	31	31	31.00	(0.00)	0.82	(0.85)		23	69	69	69.00	(0.00)	18.10	(21.25)
	12	28	28	28.00	(0.00)	0.03	(0.01)		24	66	66	66.90	(0.31)	2.88	(9.10)
	13	26	26	26.00	(0.00)	0.02	(0.00)		25	64	*65	65.00	(0.00)	12.30	(11.87)
	14	25	25	25.00	(0.00)	0.02	(0.00)		26	64	64	64.00	(0.00)	0.05	(0.01)
Scholl (297)	25	2787	2787	2787.00	(0.00)	37.11	(6.11)		27	60	60	60.00	(0.00)	1.61	(0.73)
	26	2680	2680	2680.00	(0.00)	32.56	(4.71)		28	59	59	59.00	(0.00)	0.16	(0.06)
	27	2580	2580	2580.00	(0.00)	58.06	(8.65)		29	56	56	56.00	(0.00)	1.43	(1.15)
	28	2488	2488	2488.00	(0.00)	53.78	(12.43)	Wee-Mag (75)	3	500	500	500.00	(0.00)	0.01	(0.00)
	29	2402	2402	2402.00	(0.00)	45.57	(9.93)		4	375	375	375.00	(0.00)	0.03	(0.00)
	30	2322	2322	2322.30	(0.47)	96.86	(53.36)		5	300	300	300.00	(0.00)	0.02	(0.00)
	31	2247	*2248	2248.00	(0.00)	51.23	(4.32)		6	250	250	250.00	(0.00)	0.03	(0.00)
	32	2177	2177	2177.50	(0.51)	91.50	(63.17)		7	215	215	215.00	(0.00)	0.01	(0.00)
	33	2111	2111	2111.80	(0.41)	57.19	(39.95)		8	188	188	188.00	(0.00)	0.04	(0.01)
	34	2049	2049	2049.20	(0.41)	117.47	(53.03)		9	167	167	167.00	(0.00)	0.01	(0.00)
	35	1991	1991	1991.00	(0.00)	95.54	(17.19)		10	150	150	150.00	(0.00)	0.08	(0.02)
	36	1935	*1936	1936.00	(0.00)	61.22	(8.78)		11	137	137	137.00	(0.00)	0.06	(0.01)
	37	1883	1883	1883.05	(0.22)	140.57	(34.46)		12	125	125	125.00	(0.00)	0.10	(0.02)
	38	1834	1834	1834.00	(0.00)	125.92	(34.53)		13	116	116	116.00	(0.00)	0.02	(0.00)
	39	1787	1787	1787.35	(0.49)	192.12	(55.85)		14	108	108	108.00	(0.00)	0.54	(0.60)
	40	1742	1742	1742.00	(0.00)	153.02	(29.66)		15	100	100	100.00	(0.00)	0.24	(0.19)
	41	1700	1700	1700.00	(0.00)	81.51	(18.06)		16	94	94	94.00	(0.00)	0.05	(0.01)
	42	1659	*1660	1660.00	(0.00)	146.58	(32.11)		17	89	89	89.00	(0.00)	0.42	(0.28)
	43	1621	1621	1621.00	(0.00)	149.08	(21.75)		18	$\overline{87}$	87	87.00	(0.00)	0.07	(0.01)
	44	1584	1584	1584.15	(0.37)	127.52	(54.06)		19	$\overline{85}$	85	85.00	(0.00)	0.08	(0.02)
	45	1549	1549	1549.00	(0.00)	150.65	(32.59)		20	77	77	77.00	(0.00)	0.13	(0.04)
	46	1515	*1516	1516.00	(0.00)	121.28	(20.06)		21	72	72	72.00	(0.00)	0.08	(0.01)
	47	1483	*1484	1484.00	(0.00)	165.44	(28.64)		22	69	69	69.00	(0.00)	0.06	(0.01)
	48	1452	1452	1452.70	(0.47)	151.98	(60.76)		23	$\overline{67}$	67	67.00	(0.00)	4.34	(3.58)
	49	$\overline{1427}$	1424	1424.00	(0.00)	113.46	(22.47)		24	66	66	66.00	(0.00)	0.25	(0.14)
	50	1394	1394	1394.85	(0.37)	143.63	(53.48)		25	65	65	65.00	(0.00)	11.36	(26.58)
	51	1386	1386	1386.00	(0.00)	2.23	(0.09)		26	65	65	65.00	(0.00)	0.05	(0.01)
	52	1386	1386	1386.00	(0.00)	0.17	(0.01)		27	$\overline{65}$	65	65.00	(0.00)	0.02	(0.00)
Tonge (70)	3	1170	1170	1170.00	(0.00)	0.01	(0.00)		28	$\overline{64}$	64	64.00	(0.00)	0.02	(0.00)
	4	878	878	878.00	(0.00)	0.02	(0.00)		29	63	63	63.00	(0.00)	0.02	(0.00)
	5	702	702	702.00	(0.00)	0.05	(0.02)		30	56	56	56.00	(0.00)	0.07	(0.00)

SORT 35 (2) July-December 2011, 165-178

Generalized spatio-temporal models

Edilberto Cepeda Cuervo*

Abstract

An important problem in statistics is the study of spatio-temporal data taking into account the effect of explanatory variables such as latitude, longitude and time. In this paper, a new Bayesian approach for analyzing spatial longitudinal data is proposed. It takes into account linear time regression structures on the mean and linear regression structures on the variance-covariance matrix of normal observations. The spatial structure is included in the time regression parameters and also in the regression structure of the variance covariance matrix. Initially, we present a summary of the spatial models and the Bayesian methodology used to fit the models, as a extension of the longitudinal data analysis. Next, the general spatial temporal model is proposed. Finally, this proposal is used to study rainfall data.

MSC: 62F15

Keywords: Bayesian, Fisher scoring, mean-covariance modelling, antedependence, spatial statistics, spatial temporal models, spatial longitudinal data.

1. Introduction

In the context of the parametric multivariate regression model for longitudinal data and under normality, the response variable for each of the m units under study, each having n observations over time, is denoted by $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{in})'$, $i = 1,\ldots,m$. In this case, it is usually assumed that $\mathbf{Y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, where $\mathbf{X}_i$ is a matrix of explanatory variables. Thus, if $nm = n \times m$, it is assumed that the nm-response vector $\mathbf{Y} = (\mathbf{Y}_1,\ldots,\mathbf{Y}_m)'$ follows the model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_i)), \tag{1.1}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_m)'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_m)'$ and $\boldsymbol{\Sigma}_i = \text{Var}(\boldsymbol{\epsilon}_i)$.

*Address for correspondence: Departamento de Estadística. Universidad Nacional de Colombia.
ecepedac@unal.edu.co
Received: December 2010
Accepted: September 2011

In these models, as it is well known, ϵ_{ij} and ϵ_{ik}, $j \neq k$, $i = 1,\ldots,m$, are not independent. Thus, $\text{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i$ is no longer a diagonal matrix and it would be necessary to model and estimate the off-diagonal elements of the covariance matrix. This modelling approach usually requires to impose some constraints on the elements of Σ_i to guarantee its positive definiteness. For example, in stationary Gaussian processes, such as the ones used in Geostatistics, the covariance between two observations is explicitly determined by their correlation function. More specifically, it is modelled as a function of the (Euclidean) distance between these two observations. Moreover, and given that some of the properties of this function are imposed by its spatial structure, only correlation functions belonging to the families where these requirements hold can be considered (see, e.g., Diggle and Verbyla, 1998, or Stein, 1999).

Spatial data consist of several measurements taken on the space in each of the experimental coordinates in the sample. This falls into the framework of correlated observations and requires the specification and estimation of both the mean and the covariance structures. A central idea to be able to efficiently estimate the covariance matrix was first introduced by Macchiavelli and Arnold (1994) and Macchiavelli and Moser (1997) and it is based on its Cholesky decomposition. This approach has been used for several joint modelling proposals for the mean and covariance structures in the context of longitudinal data (see, e.g., Pourahmadi, 1999 and 2000, or Pan and MacKenzie, 2006).

In our work, we apply the modified Cholesky decomposition of the precision matrix proposed in Macchiavelli and Arnold (1994), since it offers a simple unconstrained and statistically meaningful reparametrization of the covariance matrix. It has a statistical interpretation in longitudinal data through consideration of antedependence models (Gabriel, 1962; Macchiavelli and Arnold, 1994). With this reparametrization, the dependence between the components of $\mathbf{Y}$ can be modelled as functions of explanatory variables. In this case the covariance matrix structure does not depend on the ordering of observations, so we can apply this models for the variance-covariance matrix in the analysis of spatial data. The parameters do not have a practical interpretation anymore, but estimation of the covariance matrix may lead to better estimates of the parameters of the mean model. When there are many observational units, this parametrization can be useful to alleviate problems associated with the high dimensionality of the variance-covariance matrix. In this case, other variables beyond distance between observational units may be included in the model for the correlations. A simulation study is presented in section 5.

In this paper, we apply the Bayesian methodology proposed by Cepeda and Gamerman (2004) for the analysis of spatial data. In the regression models of joint regressions for the mean and covariance matrix, we include a spatial structure in the regression mean parameters through the spatial dependence of them and in the regression models of the variance covariance matrix, including spatial variables. We also extend the longitudinal models proposed by Pouramadi (1999) and the Bayesian methodology proposed by Cepeda and Gamerman (2004) for modelling spatio-temporal data sets. A spatio-

temporal structure is included in the mean, that now is a function of the temporal and spatial variables. The spatial-temporal association is captured through $\mathbf{T}$ in the triangular decomposition $\mathbf{\Sigma}^{-1} = \mathbf{T}^{'}\mathbf{D}^{-1}\mathbf{T}$, that now it is not defined as in equation (1.1), since in the spatio-temporal analysis $\mathbf{\Sigma}$ is not a block-diagonal matrix. Initially, it can have all entries different from zero and, thus, $\mathbf{T}$ is a $nm \times nm$ triangular matrix with 1's in the diagonal.

After this introduction, in section 2, a general spatial model is presented. In section 3 a general Bayesian methodology proposed to fit spatial and spatial temporal models is presented. In section 4 the results of a spatial simulation study are presented. In section 5 a general spatial temporal model is proposed. Finally, in section 6, the results of the analysis of rainfall data are presented.

2. The spatial data analysis

In this section, a single observation of each of the n observational units is assumed. Observations are correspondingly arranged in a n dimensional vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, assumed to follow a multivariate normal distribution, so that $\mathbf{Y} \sim (\boldsymbol{\mu}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a non-negative definite matrix. A crucial requirement for the analysis is that the inverse of the covariance matrix can be efficiently computed and that, in addition, it should also be allowed to have a very general and flexible specification, so that its functional specification is not too restrictive. For this reasons, we adopt the models suggested by Pourahmadi (1999) following the general model setting presented in Cepeda (2001) and Cepeda and Gamerman (2004) and considering the general ante-dependence model (Gabriel, 1962, Zimmerman and Núñez-Antón, 1997), where for a given individual having n observations we have that

$$Y_i - \mu_i = \sum_{j=1}^{i-1} \phi_{ij}(Y_j - \mu_j) + \nu_i, \nu_i \sim N(0, \sigma_i^2), \quad i = 1, ..., n, \tag{2.1}$$

where $E(Y_i) = \mu_i$, with $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$ a linear (or nonlinear) function of the vector of parameter $\boldsymbol{\beta}$, $\nu_i \sim N(0, \sigma_i^2)$ are assumed as mutually independent and by convection $\sum_{j=1}^{0} \phi_{ij}(y_i - \mu_j) = 0$. Although i is typically indexed over time (Gabriel, 1962), when a single series of observations is assumed, the covariance matrix structure does not depend on the ordering of observations, and thus, we can apply this model for the variance-covariance matrix in the analysis of spatial data.

Writing (2.1) in matrix form we obtain

$$\boldsymbol{\nu} = \mathbf{T}(\mathbf{Y} - \boldsymbol{\mu}), \ \boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D}) \text{ and } \mathbf{D} = \text{diag}(\sigma_i^2) \tag{2.2}$$

where $\boldsymbol{\nu}' = (\nu_1, \ldots, \nu_n)$, $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_n)$ and $\mathbf{T} = (\tau_{ij})$, with

$$\tau_{ij} = \begin{cases} 1 & \text{if } j = i \\ -\phi_{ij} & \text{if } j < i \\ 0 & \text{elsewhere} \end{cases}$$

and

$$\text{Var}(\boldsymbol{\nu}) = \mathbf{D} = \mathbf{T}\,\text{Var}(\mathbf{Y} - \boldsymbol{\mu})\mathbf{T}' = \mathbf{T}\boldsymbol{\Sigma}\mathbf{T}' \tag{2.3}$$

Thus, as a consequence of (2.3), $\boldsymbol{\Sigma}$ is obtained indirectly by obtaining $\mathbf{D}$ and $\mathbf{T}$. In addition, we should point out that the triangular decomposition in equation (4) is unique. Moreover, given that $\boldsymbol{\Sigma}$ is a symmetric matrix if and only if there exists a unique lower triangular matrix $\mathbf{T}$, with ones in the diagonal, and a unique diagonal matrix $\mathbf{D}$ with positive diagonal entries such that $\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}' = \mathbf{D}$, we also have that $\boldsymbol{\Sigma}$ is positive definite (Pourahmadi, 1999).

From (2.2),

$$\tilde{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{T})\tilde{\mathbf{Y}} + \boldsymbol{\nu}, \tag{2.4}$$

where $\tilde{\mathbf{Y}} = \mathbf{Y} - \boldsymbol{\mu}$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Assuming now that there is a vector of (covariance) explanatory variables $w_{ij} = (w_{ij,1}, \ldots, w_{ij,r})'$, we can write

$$\phi_{ij} = \mathbf{w}_{ij}'\boldsymbol{\lambda}, 1 \leq j < i \leq n \tag{2.5}$$

where λ is a vector of parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)'$. Since $\phi_{ij} = \sum_{l=1}^{r} w_{ij,l}\,\lambda_l$, the matrix $\mathbf{I}_n - \mathbf{T}$ can be expressed as the linear combination

$$\mathbf{I}_n - \mathbf{T} = \lambda_1\mathbf{W}_1 + \cdots + \lambda_r\mathbf{W}_r \tag{2.6}$$

where $\mathbf{W}_l = (w_{ij,l})$, $l = 1, \ldots, r$ are $n \times n$ matrices such that $w_{ij,l} = 0$, if $i \leq j$, and $w_{ij,l} = 1$, if $i > j$ and $l = 1$ to allow for a constant intercept in the covariance model. Note that, given the unrestricted and flexible specification of the ϕ_{ij}'s, there are no particular restrictions imposed on $\boldsymbol{\Sigma}$, indirectly specified in $\mathbf{T}$. Therefore, the covariance matrix is allowed to have any dependence form. Particular structures may be imposed on $\mathbf{T}$, for example by setting some of the non-zero ϕ_{ij}'s to 0. This can be handled by choosing the matrices $\mathbf{W}_l$, $l = 1, \ldots, r$ appropriately. In general, in longitudinal data set analysis the explanatory variables, $w_{ij,k} = (t_i - t_j)^k$ are associated with differences in time measurements. In the case of the spatial data analysis $w_{ij,k}$ can be defined as a function of the spatial variables such as longitude and latitude.

As a consequence of (2.4) and (2.6), model (2.1) can be expressed in the form

$$\begin{aligned}\tilde{\mathbf{Y}} &= \lambda_1 \mathbf{W}_1 \tilde{\mathbf{Y}} + \cdots + \lambda_r \mathbf{W}_r \tilde{\mathbf{Y}} + \boldsymbol{\nu} \\ &= \lambda_1 \mathbf{V}_1 + \cdots + \lambda_r \mathbf{V}_r + \boldsymbol{\nu} \\ &= \mathbf{V}\boldsymbol{\lambda} + \boldsymbol{\nu} \end{aligned} \tag{2.7}$$

where $\boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_r)$ with $\mathbf{V}_l = \mathbf{W}_l \tilde{\mathbf{Y}}$, for $l = 1, \ldots, r$. Note that for a fixed value of $\boldsymbol{\beta}$, the model $\tilde{\mathbf{Y}} \sim N(\mathbf{V}\boldsymbol{\lambda}, \mathbf{D})$ is obtained.

Given that ϕ_{ij} can be modelled as in (2.5) and that σ_i^2, $i = 1, 2, \ldots, n$, can be modelled in terms of covariates as $g(\sigma_i^2) = \mathbf{z}_i' \boldsymbol{\gamma}$, where g is a real positive function, we summarize the full model for the mean $\boldsymbol{\mu}$ and for the matrices $\mathbf{T}$ and $\mathbf{D}$ by

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta},\ g(\sigma_i^2) = \mathbf{z}_i' \boldsymbol{\gamma},\ h(\phi_{ij}) = \mathbf{w}_{ij}' \boldsymbol{\lambda} \tag{2.8}$$

for some appropriate functions g and h. In (2.8), $\mathbf{x}_i$, $\mathbf{z}_i$, $\mathbf{w}_{ij}$ are $k \times 1$, $s \times 1$ and $r \times 1$ vectors of explanatory variables. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)'$ are the parameters vectors corresponding to the mean, variance and covariance, respectively.

We can apply these models for the variance-covariance matrix in spatial data analysis. We suppose that there are many observational units with a spatial distribution and that we have a variable of interest and the explanatory variables for each observational unit. Then we can propose the mean and covariance models in (2.8) to analyze spatial data. In this situation, we have a random vector, where each component is associated with an observational unit. We consider random vectors $\mathbf{Y} = (Y_1, \ldots, Y_N)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ and concentration matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{T}' \mathbf{D}^{-1} \mathbf{T}$, since the observations Y_i's are not independent.

3. General spatio-temporal model

In section (2) a single observation of each of the n observational units is assumed but, in general, we have several short series, each one associated with one of the observational units, for example, with a meteorological station, where the variable of interest is measured m times through time. This is, we are considering n nonindependent random vectors $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})'$, $i = 1, 2, \ldots, n$, with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im})'$ and variance covariance matrix $\boldsymbol{\Sigma}_i^{-1} = \mathbf{T}_i' \mathbf{D}_i^{-1} \mathbf{T}_i$. Thus, if we assume normal distribution and if $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) = (Y_{11}, \ldots, Y_{1m}, \ldots, Y_{n1}, \ldots, Y_{nm})'$, $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the variance-covariance matrix $\boldsymbol{\Sigma}$ is not a block diagonal matrix. In this case, the variance-covariance matrix $\boldsymbol{\Sigma}$ is $(nm) \times (nm)$ and there is an $(nm) \times (nm)$ triangular matrix $\mathbf{T}$ and an $(nm) \times (nm)$ diagonal matrix $\mathbf{D}$, such that $\boldsymbol{\Sigma}^{-1} = \mathbf{T}' \mathbf{D}^{-1} \mathbf{T}$. Thus, rewriting the vector $\mathbf{Y}$ as $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{nm})'$, we can consider the model given by

$$Y_i - \mu_i = \sum_{j=1}^{i-1} \phi_{ij}(Y_j - \mu_j) + \nu_i, \nu_i \sim N(0, \sigma_i^2), \quad i = 1, \ldots, nm. \tag{3.1}$$

where $E(\mathbf{Y}) = (\mu_1, \mu_2, \ldots, \mu_{nm})'$, ϕ_{ij}s are the entries of $\mathbf{T}$, defined as in (2.2), and $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_{nm})$ is the vector of independent random innovation variables. In this case, the entries ϕ_{ij} of $\mathbf{T}$ can be modelled as a function of the time intervals and spatial differences between the n observational units, as for example, differences between their coordinates or mean temperature, among others. The mean can also be modelled as a spatial temporal function.

In a second approximation, the observation of the interest variable is rearranged as $\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{1m}, \ldots, Y_{k1}, Y_{k2}, \ldots, Y_{km}, \ldots, Y_{n1}, Y_{n2}, \ldots, Y_{nm})'$ and the spatial-temporal dependence through the models is assumed

$$Y_{ki} - \mu_{ki} = \sum_{k=1}^{n} \sum_{j=1}^{i-1} \phi_{ij}{}^{(k)}(Y_{kj} - \mu_{kj}) + \nu_{ki}, \nu_{ki} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, m, \quad k = 1, \ldots, n, \tag{3.2}$$

where $E(Y_{ki}) = \mu_{ki}$, with $\mu_{ki} = \mathbf{x}'_{ki}\boldsymbol{\beta}$, a linear function of parameter $\boldsymbol{\beta}$, $\nu_{ki} \sim N(0, \sigma_{ki}^2)$ are mutually independent and $\sum_{j=1}^{0} \phi_{ij}{}^{(k)}(Y_{kj} - \mu_{kj}) = 0$ is used, and i indexes over time. Writing (2.1) in matrix form, we obtain

$$\boldsymbol{\nu} = \mathbf{T}(\mathbf{Y} - \boldsymbol{\mu}), \ \boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D}) \text{ and } \mathbf{D} = \text{diag}(\sigma_{ki}^2) \tag{3.3}$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{1m}, \ldots, \mu_{k1}, \mu_{k2}, \ldots, \mu_{km}, \ldots, \mu_{n1}, \mu_{11}, \ldots, \mu_{nm})'$ and $\mathbf{T} = (\tau_{ij})$, is an nm lower triangular matrix, with

$$\tau_{ij} = \begin{cases} 1 & \text{if } i = j, \ i = nk_1 + l, \ j = nk_1 + r \\ -\phi_{ij}^{(k)} & \text{if } r < l, \ i = nk_1 + l, \ j = nk_2 + r \\ 0 & \text{elsewhere} \end{cases} \tag{3.4}$$

where $1 \leq l, r \leq m$, $k_1, k_2 = 0, 1, 2 \ldots, n-1$, and $k_1 \leq k_2$. Finally, specifying the mean and the variance-covariance models $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$, $\sigma_i^2 = g(\mathbf{z}_i, \boldsymbol{\Upsilon})$, $\phi_{ij} = h(\mathbf{w}_{ij}, \boldsymbol{\lambda})$, respectively, for some appropriately selected functions h, g, and f, the spatial temporal model is completely defined. These functions can be the same as in equation (2.8), however they can be appropriate nonlinear functions. In the last case it is necessary to build a kernel transition function that can be obtained by defining a normal working variable as proposed in Cepeda and Gamerman (2005). Finally, to fit these spatio-temporal models we propose the Bayesian methodology defined in section (4).

4. Bayesian methodology

In this section we present the Bayesian methodology used to fit spatial and spatial temporal models, following the Bayesian methodology proposed by Cepeda and Gamerman (2004) to fit longitudinal data. Assuming the observational model $\mathbf{Y} = (Y_1, \ldots, Y_n)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ depends on $\boldsymbol{\beta}$ through $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is the matrix of explanatory variables, and $\boldsymbol{\Sigma}$ depends on $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ through (2.8), the likelihood function is given by

$$L(\boldsymbol{\beta},\ \boldsymbol{\gamma},\ \boldsymbol{\lambda}|\mathbf{Y}) \propto |\mathbf{D}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right\},$$

since $|\boldsymbol{\Sigma}| = |\mathbf{T}'|\,|\mathbf{D}|\,|\mathbf{T}| = |\mathbf{D}|$.

Now, a prior distribution $p(\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})'$ must be assigned to obtain the posterior distribution. For simplicity we assume $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\theta}_0 = (\mathbf{b}_0, \mathbf{g}_0, \mathbf{l}_0)'$ as prior distribution. One possible model for $\boldsymbol{\Sigma}_0$ is the diagonal form, implying prior independence between $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. In this case, the full conditional prior distributions for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are given by normal distributions, denoted by $N(\mathbf{b}, \mathbf{B})$, $N(\mathbf{g}, \mathbf{G})$ and $N(\mathbf{l}, \mathbf{L})$, respectively. The values of $(\mathbf{b}, \mathbf{g}, \mathbf{l})$ and $(\mathbf{B}, \mathbf{G}, \mathbf{L})$ are easily evaluated from $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$.

From the Bayes theorem, the posterior distribution for θ is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}|\mathbf{Y}) \propto |\mathbf{D}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'\Sigma^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\Sigma_0^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\} \tag{4.1}$$

The posterior distribution (4.1) is intractable analytically and not easily generated from. However, the posterior full conditional distribution $\pi_{\boldsymbol{\beta}} = \pi(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma}, \boldsymbol{\lambda})$ is given by

$$\pi(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{Y}) \quad \propto \quad \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mathbf{b}^*)'\mathbf{B}^{*-1}(\boldsymbol{\beta}-\mathbf{b}^*)\right\},$$

where $\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y})$ and $\mathbf{B}^* = (\mathbf{B}^{-1} + \mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$. Therefore,

$$(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma}, \boldsymbol{\lambda}, \mathbf{Y}) \sim N(\mathbf{b}^*, \mathbf{B}^*). \tag{4.2}$$

Thus, it is possible to sample $\boldsymbol{\beta}$ directly from $\pi_{\boldsymbol{\beta}}$. Values of $\boldsymbol{\beta}$ can be proposed directly from $\pi_{\boldsymbol{\beta}}$ and accepted with probability 1. This is the Gibbs sampler (Geman and Geman, 1984).

From (2.4) and (2.7), the quadratic form $Q(\mathbf{Y}) = (\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$ appearing in the likelihood can be rewritten as

$$Q(\mathbf{Y}) = \tilde{\mathbf{Y}}'\mathbf{T}'\mathbf{D}^{-1}\mathbf{T}\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}} - \mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}} - \mathbf{V}\boldsymbol{\lambda})$$

Therefore, the full conditional distribution $\pi_{\boldsymbol{\lambda}}$ is given by

$$\begin{aligned}\pi(\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\gamma}) &\propto \exp\left\{-\frac{1}{2}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})-\frac{1}{2}(\boldsymbol{\lambda}-\mathbf{l})'\mathbf{L}^{-1}(\boldsymbol{\lambda}-\mathbf{l})\right\},\\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\mathbf{l}^*)'\mathbf{L}^{*-1}(\boldsymbol{\lambda}-\mathbf{l}^*)\right\}\end{aligned}$$

where $\mathbf{l}^* = \mathbf{L}^*(\mathbf{L}^{-1}\mathbf{l}+\mathbf{V}'\mathbf{D}^{-1}\tilde{\mathbf{Y}})$ and $\mathbf{L}^* = (\mathbf{L}^{-1}+\mathbf{V}'\mathbf{D}^{-1}\mathbf{V})^{-1}$. This is,

$$(\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{Y}) \sim N(\mathbf{l}^*,\mathbf{L}^*). \tag{4.3}$$

Thus, values of $\boldsymbol{\lambda}$ can be proposed directly from $\pi_{\boldsymbol{\lambda}}$ and accepted with probability 1.

Unless full conditional distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are known, the full conditional distribution of $\boldsymbol{\gamma}$, given by

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\lambda}) \propto |\mathbf{D}|^{-1/2}\exp\left\{-\frac{1}{2}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})-\frac{1}{2}(\boldsymbol{\gamma}-\mathbf{g})'\mathbf{G}^{-1}(\boldsymbol{\gamma}-\mathbf{g})\right\}, \tag{4.4}$$

is intractable analytically and not easily generated from. In this case we have to construct suitable proposals for a Metropolis-Hastings step (Hastings, 1970; Gamerman, 1997a).

We used the methodology proposed by Gamerman (1997b) as applied in Cepeda and Gamerman (2001) for modelling heterogeneity in independent normal regression models. The algorithm requires working variables to approximate transformation of the observations around the current parameter estimates. At the $\boldsymbol{\gamma}$ iteration, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are fixed at their current values $\boldsymbol{\beta}^{(c)}$ and $\boldsymbol{\lambda}^{(c)}$ and, given (2.7), the working observation variables are obtained by Fisher scoring process or by Taylor approximation (Cepeda and Gamerman, 2005).

When $g = \log$, the working observation obtained using Fisher scoring process is

$$\tilde{t}_i = \mathbf{z}_i'\boldsymbol{\gamma}^{(c)} + \frac{(\tilde{\mathbf{Y}}_i^{(c)} - \mathbf{v}_i^{(c)'}\boldsymbol{\lambda}^{(c)})^2}{\exp(\mathbf{z}_i'\boldsymbol{\gamma}^{(c)})} - 1, \quad i = 1,\ldots,n.$$

It has $E(\tilde{t}_i) = \mathbf{z}_i'\boldsymbol{\gamma}^{(c)}$ and associated working variances equal to 2. With the process given in Cepeda and Gamerman (2004), the normal transition kernel $q_{\boldsymbol{\gamma}}$ based on Fisher scoring methods is obtained as

$$q_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}^{(c)},\boldsymbol{\gamma}^{(n)}) = N(\mathbf{g}^*,\mathbf{G}^*) \tag{4.5}$$

where

$$\begin{aligned}\mathbf{g}^* &= \mathbf{G}^*(\mathbf{G}^{-1}\mathbf{g}+2^{-1}\mathbf{Z}'\tilde{\mathbf{Y}})\\ \mathbf{G}^* &= (\mathbf{G}^{-1}+2^{-1}\mathbf{Z}'\mathbf{Z})^{-1}.\end{aligned}$$

Given the characteristic of the conditional distributions we will not sample all the components of vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})'$ simultaneously. Explicitly, we sample $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ directly from their full conditionals and $\boldsymbol{\gamma}$ from the proposal given in (4.5), appling Metropolis Hastings algorithm.

5. Spatial data analysis: a simulation study

A simulation study was performed to compare parameter estimates and true values, in a spatial model, assuming that the interest variable $\mathbf{Y}$ has normal distribution. Initially, $n = 50$ values of 5 explanatory variables X_i, $i = 1,2,3$, and W_i, $i = 1,2$, were simulated. Values of X_1, X_2 and X_3 were generated from uniform distributions $U[0,50]$, $U[5,15]$ and $U[0,20]$, respectively, and values of W_i, $i = 1,2$, were generated from uniform distributions $U[0,20]$ and $U[5,15]$, respectively. The values of Y were simulated from a multivariate normal distribution with mean $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ and variance-covariance matrix $\boldsymbol{\Sigma} = \mathbf{T}^{-1}\mathbf{D}(\mathbf{T}')^{-1}$, where $\mathbf{D} = \text{diag}(\sigma_i^2)$, $\mathbf{T} = (-\phi_{ij})$, $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{3i})$ and $\phi_{ij} = \lambda_0 + \lambda_1 w_{ij,1} + \lambda_2 w_{ij,2}$, with $\boldsymbol{\beta} = (20,3,-1.5)'$, $\boldsymbol{\gamma} = (-6,0.05,-0.25)'$ and $\boldsymbol{\lambda} = (-0.5,0.04,-0.02)'$. To apply Bayesian methodology, for simplicity, independent normal prior distributions, $\theta \sim N(0, 10^3\mathbf{I}_9)$ were considered for all the parameters.

The posterior parameter estimates and their respective standard deviation are: $\hat{\beta}_0 = 20.003(6.722 \times 10^{-3})$, $\hat{\beta}_1 = 2.999(1.841 \times 10^{-4})$, $\hat{\beta}_2 = -1.500(6.228 \times 10^{-4})$, $\hat{\lambda}_0 = -0.501(0.004)$, $\hat{\lambda}_1 = 0.040(0.001)$, $\hat{\lambda}_2 = -0.020(4.231 \times 10^{-5})$, $\hat{\gamma}_0 = -5.166(0.619)$, $\hat{\gamma}_1 = 0.025(0.014)$ and $\hat{\gamma}_2 = -0.254(0.041)$. From the comparisons between true values and the corresponding estimates, we conclude that the proposed methodology has excellent performance. The estimates are very close to the true values and, in all cases, they have small standard deviation.

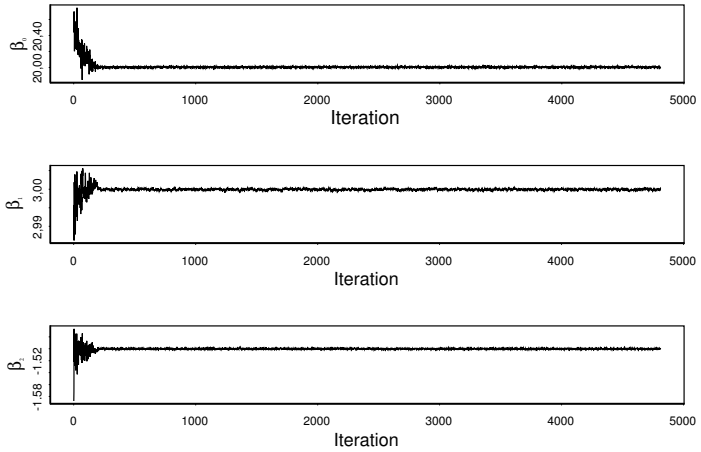


Figure 1: *Posterior chains for the mean parameters: β_0, β_1, β_2.*

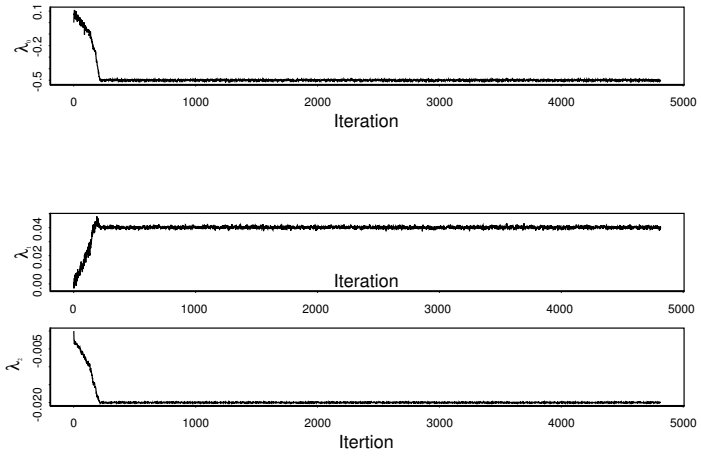


Figure 2: *Posterior samples of the antedependence parameters:* λ_0, λ_1, λ_2.

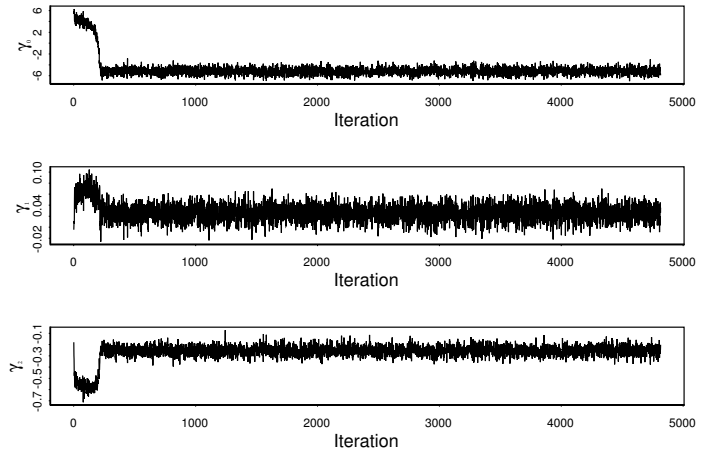


Figure 3: *Posterior samples of the covariance parameters.* γ_1, γ_1, γ_2.

Figures 1, 2 and 3 show the behavior of the chain for the sample simulated for each parameter, where each one has a small transient stage, indicating the speed convergence of simulation for the algorithm. The chain samples are given for the first 4800 iterations. The other results reported in this section are based on a sample of 4000 draws after a burn-in of 800 draws to eliminate the effect of initial values.

The posterior marginal distributions for all the parameters are approximately normal. The p-values of the Kolmogorov-Smirnov test are all larger than 0.05. The posterior sample shows large correlation between parameters of mean models, large correlations between parameters of variance-covariance models, and small but non-negligible correlation between parameters of mean models and parameters of variance models.

6. Application

In this application we consider the precipitation index in the Guajira, a department of Colombia. This department is located in the northeast region of Colombia in a peninsula bounded on the west and north by the Caribbean Sea, on the east by the Gulf of Venezuela and on the south by the Sierra Nevada de Santa Marta. There are six measurement centers distributed along the region. Although it is a flat region, there are differences between the levels of precipitation in each of its points and between seasons and years. For example, in January, February, March and July the level of precipitation is small but in April and May or in October and November all measurements centers report the highest levels of precipitation. In this application we analyze the mean sample of the cumulative monthly precipitation level for 10 years, from 1995 to 2005. The general behavior of the cumulative mean precipitation can be seen in figure 4. From the figure we can infer that the hydrological stations are located in regions with two different regimes of rainfall, characterized by very low levels of precipitation in the first three months of the year. As of April, it is clear that there are three stations located in a region whith high rates of precipitation and three stations where precipitation levels are much lower.

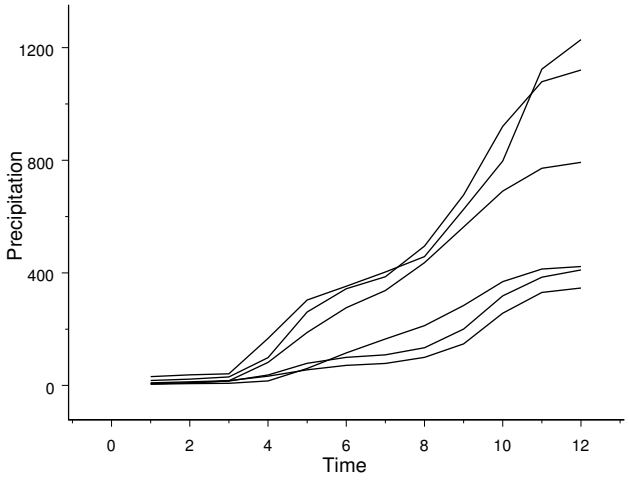


Figure 4: *Annual mean of precipitation.*

In the first analysis of this data we consider the spatio-temporal time model given by

$$\begin{aligned}
\mu_{trs} &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (st)^3 + \beta_5 (rt)^3 \\
\log\left(\sigma_t^2\right) &= \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 \\
\phi_{ij} &= \lambda_0 + \lambda_1 (i-j) + \lambda_2 (i-j)^2 + \lambda_3 (i-j)^3 + \lambda_4 (s_i - s_j)^3 + \lambda_5 (r_i - r_j)^3
\end{aligned}$$

where t is the time, s is the latitude with respect to the mean value of the latitude of the observational units, r is the longitude with respect to the mean value of the longitude of the observational units, st is the random variable given by the product of the time by spatial latitude, and rt is the explanatory variable given by the product of the time and spatial longitude. Thus, assuming normal flat prior distribution for the mean, innovation variance and covariance parameters, the posterior estimates and the correspondent standard deviation for the parameters of the model are given by the following values.

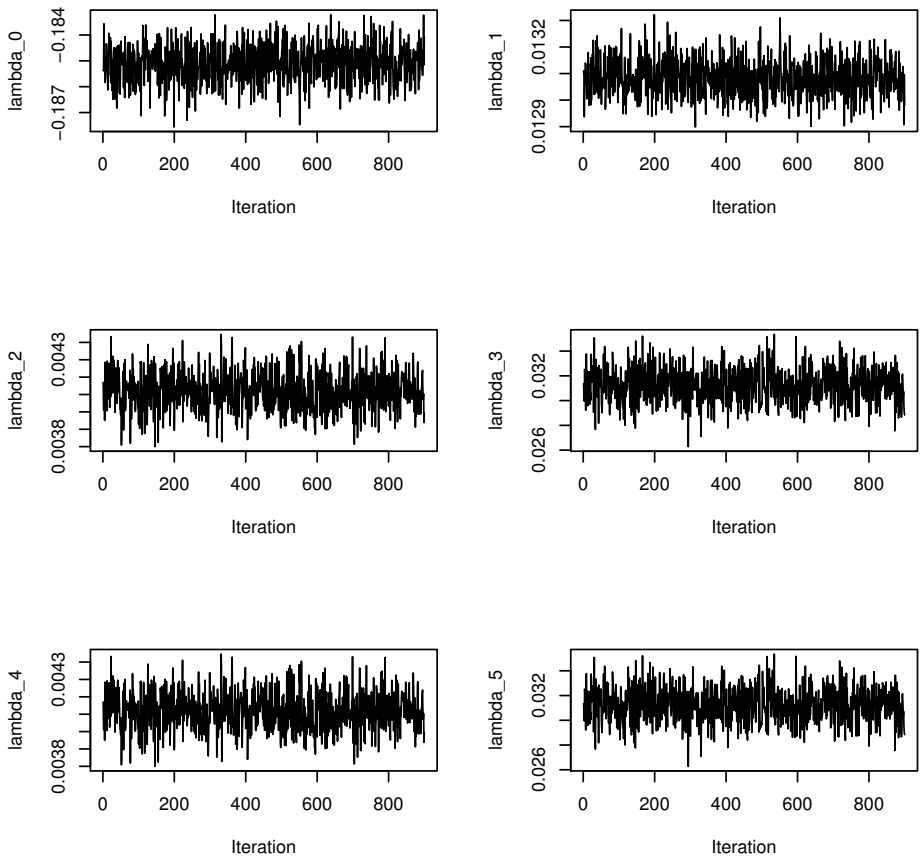


Figure 5: *Posterior samples of the covariance parameters.*

1. For the mean parameters:

$$\hat{\beta}_0 = -32.635(3.877),\ \hat{\beta}_1 = 26.213(2.788),\ \hat{\beta}_2 = -0.548(0.094),$$
$$\hat{\beta}_3 = 0.434(0.001),\ \hat{\beta}_4 = -17.461(0.330),\ \hat{\beta}_5 = 0.184(0.000).$$

2. For the innovation variance:

$$\hat{\gamma}_0 = -9.722(0.357),\ \hat{\gamma}_1 = 0.566(0.000),$$
$$\hat{\gamma}_2 = -0.185(0.000),\ \hat{\gamma}_3 = 0.013(0.000).$$

3. For the covariance parameters:

$$\hat{\lambda}_0 = 0.004(0.000),\ \hat{\lambda}_1 = 1.031(0.000),\ \hat{\lambda}_2 = 53.379(6.611),$$
$$\hat{\lambda}_3 = -45.967(4.604),\ \hat{\lambda}_4 = 9.723(0.201),\ \hat{\lambda}_5 = -0.573(0.001).$$

In all cases, for all the parameters in the model, the chains have a small transient period showing the performance of the proposed algorithm. To illustrate this behavior the chain of the posterior samples for the covariance parameters are included in figure 5. As in the simulation study, in this case, the histograms show that the posterior marginal distributions for all the parameters are approximately normal. The p-values of the Kolmogorov-Smirnov test are all larger than 0.05.

7. Conclusions

In this paper, a new Bayesian flexible and unrestricted approach for analyzing spatial longitudinal data is proposed. We illustrate the usefulness of our proposed methodology by including a Monte Carlo analysis, a simulation study and an application to a real data set. Our proposal puts forward a new way to explore the QR decomposition of the covariance matrix in a longitudinal data setting, by generalizing its application within this context.

Natural extensions of the research presented in this paper are also possible. Classical methodologies applying the Newton Raphson or the Fisher scoring algorithms to fit the proposed models can be easily introduced in the same way as in Pourahmadi (1999) or Cepeda and Gamerman (2004). Nonlinear spatio-temporal models also can be defined by assuming a nonlinear regression models in the mean and covariance models, so that it would be a generalization of the nonlinear longitudinal models proposed in Cepeda-Cuervo and Núñez-Antón (2009), which included a classic and Bayesian generalization that allowed to fit these new models.

Acknowledgements

Cepeda's work was supported by a grant from The Research Division of the National University of Colombia (Universidad Nacional de Colombia). The author wishes to thank the anonymous referees for their careful reading of this manuscript and also for their suggestions that led to very important improvements.

References

Cepeda, E. C. (2001). Variability modelling in generalized linear models. Unpublished Ph.D. thesis, Mathematics Institute, Universidade Federal do Rio de Janeiro.

Cepeda, E. C. and Gamerman D. (2004). Bayesian modelling of joint regression for the mean and covariance matrix. *Biometrical Journal*, 14, 430–440.

Cepeda-Cuervo, E. and Núñez-Antón, V. (2009). Bayesian modelling of the mean and covariance matrix in normal nonlinear models. *Journal of Statistical Computation and Simulation*, 79(6), 837–853.

Cepeda-Cuervo, E. (2005). Bayesian methodology for modelling parameters in the two parameter exponential family. *Revista Estadística*, 57(168-169).

Diggle, P. and Verbyla, A. (1999). Nonparametric estimation of covariance structure in longitudinal date. *Biometrics*, 54, 401–415.

Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, 33, 201–212.

Gamerman, D. (1997a). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall: London.

Gamerman, D. (1997b). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7, 57–68.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97–109.

Macchiavelli, R. E. and Arnold, S. F. (1994). Variable order antedependence models. *Communications in Statistics*, 23, 2683–2699.

Macchiavelli, R. and E. B. Moser (1997). Analysis of repeated measures with ante-dependence models. *Biometrical Journal*, 39(3), 339–350.

Pan, J. and MacKenzie G (2006). Regression models for covariance structures in longitudinal studies. *Statistical Modelling*, 6(1), 43–57.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterizations. *Biometrika*, 86, 667–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, 87, 425–435.

Stein, M. (1999). *Interpolation of Spatial Data*. Springer: New York.

Zimmerman, D. and Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data, in *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*, T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. Wolnger (eds.), 63–76. Lecture Notes in Statistics No. 122. Springer: New York.

Information for authors and subscribers

Information for authors and subscribers

Submitting articles to SORT

Guidelines for submitting articles

SORT accepts for publication only original articles that have not been submitted to any other journal in the areas of Statistics, Operations Research, Official Statistics and Biometrics. Articles may be either theoretical or applied, and may include computational or educational elements. Publication will be exclusively in English, with an abstract in Catalan (abstract translation into Catalan can be done by the journal's staff).

All articles submitted to thematic sections will be forwarded for systematic peer review by independent specialists and/or members of the Editorial Board, except for those articles specifically invited by the journal or reprinted with permission. Reviewer comments will be sent to the primary author if changes are requested in form or content.

When an original article is received, the journal sends the author a dated receipt; this date will appear in the publication as "reception date". The date in which its final version is received will be the "acceptance date".

To submit an article, the author must send it in PDF format to the electronic address of the journal (sort@idescat.cat) especifying his e-mail address and/or a phone number. Once the article has been approved, the author must send a final electronic version, following the instructions of the editor responsible for that article. It is recommended that this final version be submitted using a LaTeX 2_ε .

In any case, upon request the journal secretary will provide authors with LaTeX 2_ε templates and appropriate references to the MSC2000 classification of the American Mathematical Society. For more information about the criteria for using this classification system, authors may consult directly the Mathematics Subject Classification 2000 (MSC2000) at the website of the Institut d'Estadística de Catalunya (http://www.idescat.es/sort/Normes.stm).

Publishing rights and authors' opinions

Authored articles represent the opinions of the authors; the journal does not necessarily agree with the opinions expressed in authored articles.

Formal guidelines for articles

The first page of the article must contain the following items: title, name of the author(s), professional affiliation and complete address, and an abstract (100-200 words) followed by MSC2000 classification of the American Mathematical Society and the keywords. Before submitting an article, the author(s) would be well advised to ensure that the text uses correct English.

Bibliographic references have to be listed alphabetically at the end of the article, according to the following examples:

Citations
Mahalanobis (1936), Rao (1982b)

Journal articles
Pfeffermann, D. and Barnard, C. H. (1991). Some new estimators for small-area means with application to the assessment of farmland values. *Journal of Business & Economic Statistics*, 9, 73-84.

Books
Platek, R., Rao, J. N. K., Särndal, C. E. and Singh, M. P. (eds.) (1987). *Small Area Statistics: an International Symposium*. New York: John Wiley and Sons.

Parts of books
Lee, E.W., Wei, L.J. and Amato, D.A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival Analysis: State of the Art*, J.P. Klein and P.K. Goel (eds.), Kluwer Academic Publishers, 237-247.

Web files or "pages"
Nielsen, S. F. (2001). *Proper and improper multiple imputation*
http://www.stat.ku.dk/~feodor/publications/ (10th May 2003).

Multiple publications by a single author are to be listed chronologically. Explanatory notes should be numbered sequentially and placed at the bottom of the corresponding page. Tables and figures should also be numbered sequentially and will be reproduced directly from the submitted originals if it is impossible to include them in the electronic text.

SORT *(Statistics and Operations Research Transactions)*
Institut d'Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona (Spain)
Tel.: +34-93 412 09 24 or +34-93 412 00 88
Fax: +34-93 412 31 45
E-mail: sort@idescat.cat

How to cite articles published in SORT

Commenges, D. (2003). Likelihood for interval-censored observations from multi-state models. *SORT*, 27 (1), 1-12.

Subscription form
SORT *(Statistics and Operations Research Transactions)*

Name _____

_____

Organisation _____

_____

Street Address _____

_____

Zip/Postal code _____ City _____

State/Country _____ Tel. _____

Fax _____ NIF/VAT Registration Number _____

E-mail _____

Date _____

Signature

I wish to subscribe to **SORT *(Statistics and Operations Research Transactions)*** for the year 2011 (volume 35)

Annual subscription rates:

– Spain: €22 (4% VAT included)
– Other countries: €25 (4% VAT included)

Price for individual issues (current and back issues):

– Spain: €15/issue (4% VAT included)
– Other countries: €17/issue (4% VAT included)

Method of payment:

☐ Bank transfer to account number 2013-0100-53-0200698577

☐ Automatic bank withdrawal from the following account number

☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

☐ Check made payable to the Institut d´Estadística de Catalunya

Please send this subscription form (or a photocopy) to:

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45

Bank copy

Authorisation for automatic bank withdrawal in payment for

SORT *(Statistics and Operations Research Transactions)*

The undersigned _____

authorises Bank/Financial institution _____

located at (Street Address) _____

Zip/postal code _____ City _____

Country _____

to draft the subscription to **SORT *(Statistics and Operations Research Transactions)*** from my account

number ☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

Date _____

Signature

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45