Journal of Statistical Computation and Simulation
Vol. 77, No. 8, August 2007, 663–681

Estimating the common parameter of normal models with known coefficients of variation: a sensitivity study of asymptotically efficient estimators

VYTARAS BRAZAUSKAS* and JUGAL GHORAI

Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA

(Revised 24 May 2005; in final form 10 November 2005)

In this article, estimation of the common parameter θ, when data $X_1, \ldots, X_n$ are independent observations where each X_i is normally distributed $N(d_i\theta, \theta^2)$ and coefficients of variation $1/d_1, \ldots, 1/d_n$ are known, is treated. Such a setup is motivated by problems arising in medical, biological, and chemical experiments. We consider maximum likelihood, linear unbiased minimum variance type, linear minimum mean square, Pitman-type, and Bayes estimators of θ. Our results generalize work of previous authors in several ways. First, consideration of known but different coefficients of variation allows more flexibility in designing experiments. Secondly, our treatment can be directly applied to the case of dependent data with known correlation structure. Further, using Monte Carlo simulations, we supplement asymptotic findings with small-sample results. We also investigate the sensitivity of the estimators under various model misspecification scenarios.

Keywords: Curved normal family; Premium-protection plot; Sensitivity; Simulations

1. Introduction

1.1 *Motivation*

In immunoassay problems, the unknown concentration level of a chemical constituent in patient's blood serum is measured indirectly by comparing the absorbance of light through a suitably prepared patient's sample with that of the standard solution. In such studies, a typical relationship between the concentration level and the mean amount of light absorption is developed. A batch of experiments is run at each known concentration level and the amount of light absorption is measured. The variability in light absorption changes according to the concentration level of the chemical. Suppose that at the concentration level l_i, a batch of n_i experiments is run. Let $z_{i1}, \ldots, z_{in_i}$ denote the light absorption measurements. In many cases, the mean, $\bar{z}_i$, and standard deviation, s_i, calculated from these data exhibit a relationship of the form $\bar{z}_i = c_i s_i$. If we assume a normal distribution for the data, then it is equivalent to assuming

*Corresponding author. Email: vytaras@uwm.edu

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/10629360600578221

$Z_{i1}, \ldots, Z_{in_i} \overset{i.i.d.}{\sim} N(c_i\theta, \theta^2)$. Furthermore, coefficients $c_1, \ldots, c_k$ are usually known from the past experiments. Hence, to estimate the mean amount of light absorption at various levels of concentration of the chemical, it is enough to estimate the single parameter θ. Once an estimate of θ is available, the estimates of the mean amount of light absorption can be obtained as $\hat{\mu}_1 = c_1\hat{\theta}, \ldots, \hat{\mu}_k = c_k\hat{\theta}$. To determine the unknown concentration level of chemical in a patient's (or a group of patients') blood serum, light absorption amounts are measured and the averages from these measurements are compared with $\hat{\mu}_1, \ldots, \hat{\mu}_k$.

Problems involving known coefficients of variation are found in other contexts too. For instance, Gerig and Sen [1] analyzed data from the Canadian migratory bird surveys in the various provinces of Canada for the years 1969 and 1970. The individual observations are not available; however, the summary statistics are given in their article. We reproduce the means, variances, and coefficients of variation for the years 1969 and 1970 in the following table.

Region	Mean	Variance	cv69	cv70
1	0.6649	0.2055	0.68	0.64
2	0.8317	0.1998	0.54	0.51
3	0.8150	0.1873	0.53	0.53
4	0.7903	0.1865	0.55	0.56
5	0.9383	0.2355	0.52	0.51
6	0.8022	0.2203	0.59	0.59
7	0.9653	0.1848	0.45	0.42
8	1.0085	0.1663	0.40	0.38
9	1.0817	0.1889	0.40	0.44
10	0.9299	0.2186	0.50	0.51

These data indicate that while the coefficients of variation from year to year remain more or less unchanged, they vary between provinces within each year. Hence, for future years, the coefficient of variation for each province can be assumed to be known, although they can be unequal for various provinces. In this case, the model presented in this article can be used to analyze the data for all the provinces simultaneously instead of analyzing data for each province separately. Further, Gleser and Healy [2], Khan [3], and Searls [4] also considered estimating the common parameter θ in normal samples with known coefficient of variation. Azen and Reed [5] considered estimation of the correlation coefficient in bivariate normal samples with equal coefficient of variation.

1.2 *The problem*

In this article, estimation of the common parameter θ, when data $X_1, \ldots, X_n$ are independent observations where each X_i is normally distributed $N(d_i\theta, \theta^2)$ and coefficients of variation $1/d_1, \ldots, 1/d_n$ are known, is treated. As was illustrated in section 1.1, such problems arise in medical, biological, and chemical experiments. In order to get better understanding of the issues addressed in this article, we reformulate the problem as follows.

> Suppose we observe a realization of n-variate *normal* random vector $(X_1, \ldots, X_n)'$ with the mean vector $\theta(d_1, \ldots, d_n)'$ and the covariance matrix $\theta^2 I$, where I is the $n \times n$ *identity* matrix and $1/d_1, \ldots, 1/d_n$ are known coefficients of variation. The question of interest is how to estimate θ?

Various types of estimators of θ are proposed in the literature on the basis of the assumption that $X_1, \ldots, X_n$ are independent normal variables with known and equal coefficients of variation. These assumptions, however, introduce two potential sources of error: (*i*) data may be independent but not exactly normal and (*ii*) data may be normal but not independent. One of the main objectives of this article is to investigate the sensitivity of the estimators when these violations occur. We consider maximum likelihood, linear unbiased minimum variance type, linear minimum mean square, and Pitman-type and Bayes estimators. The sensitivity of these estimators to the violation of assumptions is studied through simulations. Departure from the normality is studied through the use of various non-normal distributions and the departure from independence is studied by introducing dependence structure among $X_1, \ldots, X_n$. (It is of course possible that the data are not normal and not independent with completely unknown dependence structure at the same time but, as will be seen later, the presence of a single source of error is sufficient to substantially affect the estimators; hence, investigation of the worst case scenario is not considered in this article.) Specific details are discussed later.

The problem formulated above offers more flexibility in designing experiments than that already solved by previous authors. Subsequently, we mention some situations that are special cases of (or can be easily transformed to) our problem and thus can be approached using results of this article.

- Independent measurements with known and equal coefficients of variation. In the special case of $d_1 = \cdots = d_n$, the problem reduces to the one treated by Khan [3], Gleser and Healy [2], and Sinha [6], for example. All types of estimators presented here are also studied by these authors (for the case $d_1 = \cdots = d_n$).
- Independent repeated measurements. In immunoassay problems of section 1.1, for a fixed concentration level l_i, repeated measurements are obtained and it is assumed that data were generated by random variables $Z_{i1}, \ldots, Z_{in_i} \overset{i.i.d.}{\sim} N(c_i\theta, \theta^2)$, where c_i is known. Then, in order to estimate parameter θ, data for all concentration levels $l_1, \ldots, l_k$ can be combined and treated as $(X_1, \ldots, X_n)' \sim N(\theta(d_1, \ldots, d_n)', \theta^2 I)$, where $X_1 = Z_{11}, \ldots, X_{n_1} = Z_{1n_1}, X_{n_1+1} = Z_{21}, \ldots, X_{n_1+n_2} = Z_{2n_2}, \ldots, X_{n_1+\cdots+n_{k-1}+1} = Z_{k1}, \ldots, X_{n_1+\cdots+n_k} = Z_{kn_k}$ and $d_1 = \cdots = d_{n_1} = c_1, \ldots, d_{n_1+\cdots+n_{k-1}+1} = \cdots = d_{n_1+\cdots+n_k} = c_k$, with $n_1 + \cdots + n_k = n$ and I is the $n \times n$ identity matrix.
- Measurements with known dependence structure. Formulas (presented in section 2) for the estimators of θ can also be applied to the case of dependent data with known correlation structure. However, certain data transformations have to be performed first. This is done as follows. Suppose we observe a realization of n-variate normal random vector $(Y_1, \ldots, Y_n)'$ with the mean vector $\theta(d_1, \ldots, d_n)'$ and the covariance matrix $\theta^2\Sigma_*$, where Σ_* is known. Then, as discussed by Rao [7, section 4a.1), the coordinates of the transformed vector $(X_1, \ldots, X_n)' = \Sigma_*^{-1/2}(Y_1, \ldots, Y_n)'$ are independent and each X_i is normally distributed $N(d_i^*\theta, \theta^2)$ with $(d_1^*, \ldots, d_n^*)' = \Sigma_*^{-1/2}(d_1, \ldots, d_n)'$. Hence, the setup for variables $X_1, \ldots, X_n$ is exactly the one described above with d_i's replaced by d_i^*'s. Finally, note that this approach would not work with the estimators of previous authors (designed for $d_1 = \cdots = d_n$) because, except for some trivial choices of Σ_*, in general $d_1^* = \cdots = d_n^*$ does not hold.

The article is organized as follows. In section 2, the estimators under consideration are introduced and their properties are investigated. Specifically, it is first shown that all the estimators are asymptotically equivalent. Then, using Monte Carlo simulations, small-sample performance of these estimators is studied. It is found that they perform quite differently in small samples, *e.g.* mean-square errors for some estimators approach the asymptotic variance from above while for others from below. In section 3, the sensitivity of bias and mean-square error of these procedures under various model misspecification scenarios are explored. The whole study is graphically summarized by applying the premium-protection approach of Anscombe [8]. A final discussion and practical recommendations are provided in section 4.

2. Estimators and their properties

In section 2.1, we present and/or derive the estimators under consideration. In particular, we consider maximum likelihood, linear unbiased minimum variance type, linear minimum mean square, and Pitman-type and Bayes estimators. It is shown in section 2.2 that all these estimators are consistent, asymptotically normal, asymptotically unbiased, and asymptotically efficient (*i.e.* their limiting variances attain the Cramér–Rao lower bound). In section 2.3, we establish – via Monte Carlo simulations – that not all estimators perform equivalently in small- ($n = 5$ and 15) and moderate-size ($n = 25$ and 50) samples and that their convergence to the asymptotic equivalence is from different directions. To bridge between 'moderate' and 'large' sample sizes, we also include the case $n = 100$ and, for completeness, we report the $n \to \infty$ case which corresponds to the asymptotic result of section 2.2.

2.1 *Estimators*

Consider a sample $X_1, \ldots, X_n$ of independent observations, where X_i is normally distributed with mean $d_i\theta$ and variance θ^2, $i = 1, \ldots, n$, $\theta > 0$. Here, constants $1/d_1, \ldots, 1/d_n$ represent known coefficients of variation. Note that information about the sign of θ is incorporated into the prior knowledge of d_i's thus, the assumption of $\theta > 0$ is not restrictive at all.

2.1.1 Maximum likelihood estimator. A straightforward maximization of the log-likelihood function,

$$\log L(\theta | X_1, \ldots, X_n) = -\frac{n}{2}\log(2\pi) - n\log(\theta) - \frac{1}{2\theta^2}\sum_{i=1}^{n}(X_i - d_i\theta)^2,$$

yields

$$\hat{\theta}_{\text{MLE}} = \frac{-\sum_{i=1}^{n} d_i X_i + \sqrt{(\sum_{i=1}^{n} d_i X_i)^2 + 4n\sum_{i=1}^{n} X_i^2}}{2n}.$$

2.1.2 Linear unbiased minimum variance type estimator. If unbiasedness is of foremost importance, then there are two obvious (and based on sufficient statistics) choices for estimation of θ, the weighted mean $\bar{X}_u$ and bias-corrected sample standard deviation S_u

$$\bar{X}_u = b_n^{-1}\sum_{i=1}^{n} d_i X_i \quad \text{and} \quad S_u = \alpha_n\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - d_i\bar{X}_u)^2},$$

where $b_n = \sum_{i=1}^{n} d_i^2$ and $\alpha_n = \sqrt{(n-1)/2}\,\Gamma((n-1)/2)/\Gamma(n/2)$. Here $\Gamma(\cdot)$ denotes the gamma function.

As $(\bar{X}_u, S_u)$ is a sufficient statistic (not complete, though) for θ, the Rao–Blackwell theorem implies that search for best unbiased estimators for θ should be confined to functions of the sufficient statistic. Consider the class of *linear unbiased* estimators

$$\mathcal{U} = \{U \mid U = a\bar{X}_u + (1-a)S_u, 0 \le a \le 1\}.$$

Using the fact that $\bar{X}_u$ and S_u^2 are independent (see Lemma 1 in Appendix A), it is not difficult to show that the choice of a, $a_n = v_2/(v_1+v_2)$, where $v_1 = b_n^{-1}$ and $v_2 = \alpha_n^2 - 1$, yields the *minimum variance* estimator

$$\hat{\theta}_{\mathrm{LU}} = \frac{v_2}{v_1+v_2}\bar{X}_u + \frac{v_1}{v_1+v_2}S_u.$$

As the factor $a_n = v_2/(v_1+v_2)$ is independent of θ, the estimator $\hat{\theta}_{\mathrm{LU}}$ is uniformly minimum variance unbiased estimator in the class $\mathcal{U}$.

2.1.3 Linear minimum mean square estimator. Let us consider the class of estimators of θ which are linear in $\bar{X}_u$ and S_u, but not necessarily unbiased. That is, we define

$$\mathcal{V} = \{V \mid V = c_1\bar{X}_u + c_2 S_u\},$$

where we do not assume that $c_1 + c_2 = 1$. (Thus, the class $\mathcal{U}$ is a subset of $\mathcal{V}$.) Our objective is to find an estimator in $\mathcal{V}$ that minimizes the mean-square error $E_\theta(V-\theta)^2$ over all values of θ. It follows from Lemma 2.1 of Gleser and Healy [2] that such an estimator is

$$\hat{\theta}_{\mathrm{LMMS}} = \frac{v_2}{v_1+v_2+v_1v_2}\bar{X}_u + \frac{v_1}{v_1+v_2+v_1v_2}S_u = \frac{v_1+v_2}{v_1+v_2+v_1v_2}\hat{\theta}_{\mathrm{LU}},$$

where v_1 and v_2 are defined as before (see section 2.1.2).

2.1.4 Pitman-type estimator. It is easy to see that the parameter θ is a scale parameter for the family of normal distributions $N(d_i\theta, \theta^2)$. This fact suggests looking at estimators which are equivariant under changes of scale. That is, for data $X_1, \ldots, X_n$, estimators $h(X_1, \ldots, X_n)$ satisfy

$$h(kX_1, \ldots, kX_n) = kh(X_1, \ldots, X_n)$$

for all $k > 0$.

Let $T_1 = \sum_{i=1}^n d_i X_i$ and $T_2 = \sum_{i=1}^n X_i^2$. As T_1 and T_2 are jointly sufficient for θ, the arguments similar to those used by Gleser and Healy [2] imply that any scale equivariant estimator of θ must be of the form $\hat{\theta}_{\mathrm{P}} = \sqrt{T_2}\phi(b)$, where b is defined subsequently. Hence, the minimum risk scale equivariant estimator with respect to squared-error loss can be obtained by minimizing $E_\theta(\theta - \sqrt{T_2}\phi(b))^2$. This gives $\phi(b) = \left[\int_0^\infty t^n \mathrm{e}^{-t^2/2+bt}\,\mathrm{d}t\right]\left[\int_0^\infty t^{n+1}\mathrm{e}^{-t^2/2+bt}\,\mathrm{d}t\right]^{-1} = J_b(n)/J_b(n+1)$. Combining these, we get the scale equivariant (sometimes it is called Pitman-type) estimator as

$$\hat{\theta}_{\mathrm{P}} = \frac{J_b(n)}{J_b(n+1)}\sqrt{\sum_{i=1}^n X_i^2},$$

where function $J_b(\cdot)$ satisfies the following relationships: $J_b(n) = bJ_b(n-1) + (n-1)J_b(n-2)$ for $n \geq 2$, $J_b(1) = bJ_b(0) + 1$, and $J_b(0) = \Phi(b)/\varphi(b)$, with $b = \sum_{i=1}^n d_i X_i/\sqrt{\sum_{i=1}^n X_i^2}$. Here, $\Phi(\cdot)$ and $\varphi(\cdot)$ denote the cdf and pdf of the standard normal distribution, respectively.

2.1.5 Bayes estimator. In this section, we derive the Bayes estimator with inverse gamma prior for θ. As θ^2 represents the variance of the $N(d_i\theta, \theta^2)$ distribution and the inverse gamma distributions have previously been used in estimating variances of normal populations, we consider the prior (inverse gamma) density given by

$$\pi(\theta) = \frac{w^r}{\Gamma(r)}\theta^{-r-1}\mathrm{e}^{-w/\theta}, \quad \theta > 0,$$

where parameters $r > 0$ and $w > 0$ have to be specified in advance.

For the squared-error loss, the Bayes estimator is the mean of the posterior distribution. A straight-forward calculation shows that the Bayes estimator can be expressed as

$$\hat{\theta}_{\mathrm{B}} = \frac{J_{b^*}(n+r-2)}{J_{b^*}(n+r-1)}\sqrt{\sum_{i=1}^{n} X_i^2},$$

where $J_{b^*}(n+r-2) = \int_0^\infty t^{n+r-2}\mathrm{e}^{-t^2/2+b^*t}\,\mathrm{d}t$, with $b^* = (\sum_{i=1}^n d_i X_i - w)/\sqrt{\sum_{i=1}^n X_i^2}$. Obviously, $J_{b^*}(\cdot)$ satisfies the same recurrent relationships as function $J_b(\cdot)$ of section 2.1.4. Computation of $J_{b^*}(\cdot)$, however, is not trivial when both r and w are large. In such situations, we suggest to directly evaluate the ratio $J_{b^*}(n+r-2)/J_{b^*}(n+r-1)$ using numerical integration methods. Note that the ratio can be written as

$$\frac{J_{b^*}(n+r-2)}{J_{b^*}(n+r-1)} = \frac{\int_0^\infty \mathrm{e}^{(n+r-2)\log t - 0.5(t-b^*)^2}\,\mathrm{d}t}{\int_0^\infty \mathrm{e}^{(n+r-1)\log t - 0.5(t-b^*)^2}\,\mathrm{d}t} = \frac{\int_0^\infty \mathrm{e}^{-M+(n+r-2)\log t - 0.5(t-b^*)^2}\,\mathrm{d}t}{\int_0^\infty \mathrm{e}^{-M+(n+r-1)\log t - 0.5(t-b^*)^2}\,\mathrm{d}t}.$$

Here, $M = b^*/2 + \sqrt{(b^*/2)^2 + (n+r-2)}$ is the value of t where function $(n+r-2)\log t - 0.5(t-b^*)^2$ attains its maximum. By subtracting M in the exponent, we can reduce the magnitude of both integrands. (An excellent source of numerical recipes for statistical problems is Lange [9].)

Remark 1 Estimation using uninformative priors. The inverse gamma prior is not the only choice to consider and other alternatives may be pursued. For example, for the problem being considered, the Bayes estimator based on the uninformative (improper) prior, $\pi(\theta) \propto 1/\theta^3, \theta > 0$, coincides with the Pitman estimator of section 2.1.4. Thus, the estimator $\hat{\theta}_{\mathrm{P}}$ serves as an example of Bayes estimation using an uninformative prior.

2.2 *Asymptotic properties*

In the technical derivations of this section (and of Appendix A), the following probabilistic notation is frequently used: '$\xrightarrow{p}$' denotes 'convergence in probability', '$\xrightarrow{\mathcal{D}}$' denotes 'convergence in distribution', '$\overset{\mathcal{D}}{\approx}$' denotes 'approximately equal in distribution', and '$O_p(\cdot)$' denotes 'stochastic $O(\cdot)$'.

For definitions, examples, and some standard approximation arguments involving these concepts in our derivations, see, for example, Serfling (10, chapter 1).

Let $L(\theta \mid X_1, \ldots, X_n) = \prod_{i=1}^n f(X_i \mid \theta)$ denote the likelihood function. Then, for $X_i \sim N(d_i\theta, \theta^2)$, a straightforward calculation shows that the Fisher information, $I(\theta)$, is given by

$$I(\theta) = -E\left(\frac{\partial^2 L(\theta \mid X_1, \ldots, X_n)}{\partial\theta^2}\right) = \frac{2n + b_n}{\theta^2},$$

where $b_n = \sum_{i=1}^n d_i^2$. Next, without the loss of generality, we can assume that the constants d_i are bounded from above. Then, the limit $\lim_{n\to\infty} b_n/n$ exists and let it be denoted by d^2.

The large sample asymptotic properties of the proposed estimators are stated in Theorem 1. In order to have a continuous flow of presentation, a proof of the theorem is postponed to Appendix A.

THEOREM 1 *Let* $\hat{\theta}_{\mathrm{MLE}}, \hat{\theta}_{\mathrm{LU}}, \hat{\theta}_{\mathrm{LMMS}}, \hat{\theta}_{\mathrm{P}}, \hat{\theta}_{\mathrm{B}}$ *be defined as in section* 2.1. *These estimators have the following asymptotic properties*:

(a) $\lim_{n\to\infty} n[E(\hat{\theta}_{\mathrm{MLE}} - \theta)] = -\theta(2+3d^2)(2+d^2)^{-3}$,
(b) $\lim_{n\to\infty} (2n+b_n)[E(\hat{\theta}_{\mathrm{MLE}} - \theta)^2] = \theta^2$,
(c) $\sqrt{n}(\hat{\theta}_{\mathrm{MLE}} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta^2(2+d^2)^{-1})$,
(d) $\sqrt{2n+b_n}(\hat{\theta}_{\mathrm{MLE}} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta^2)$,
(e) $\sqrt{2n+b_n}(\hat{\theta}_{\mathrm{LU}} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta^2)$,
(f) $\lim_{n\to\infty} \sqrt{2n+b_n}[E(\hat{\theta}_{\mathrm{LMMS}} - \theta)] = -\theta$,
(g) $\sqrt{2n+b_n}(\hat{\theta}_{\mathrm{LMMS}} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta^2)$,
(h) *As* $n \to \infty$, $\hat{\theta}_{\mathrm{MLE}} \overset{\mathcal{D}}{\approx} \hat{\theta}_{\mathrm{B}} \overset{\mathcal{D}}{\approx} \hat{\theta}_{\mathrm{P}}$.

Remark 2 Asymptotic equivalence of the estimators. The results stated by Theorem 1 indicate that the estimators $\hat{\theta}_{\mathrm{MLE}}, \hat{\theta}_{\mathrm{LU}}, \hat{\theta}_{\mathrm{LMMS}}, \hat{\theta}_{\mathrm{P}}$, and $\hat{\theta}_{\mathrm{B}}$ have the same asymptotic distribution. That is, they are asymptotically normal with mean θ and variance $\theta^2/(2n+b_n)$. Moreover, the limiting variance is exactly $1/I(\theta)$, which means that it attains the Cramér–Rao lower bound.

2.3 *Small-sample performance*

As was seen in the previous section, all the estimators are asymptotically equivalent and, thus, it is of interest to investigate their small-sample performance. Here, we performed simulations of $\hat{\theta}_{\mathrm{MLE}}, \hat{\theta}_{\mathrm{LU}}, \hat{\theta}_{\mathrm{LMMS}}, \hat{\theta}_{\mathrm{P}}$, and three Bayes estimators $\hat{\theta}_{\mathrm{B1}}, \hat{\theta}_{\mathrm{B2}}, \hat{\theta}_{\mathrm{B3}}$. The Bayes estimators are indexed by different prior distribution parameters and are expected to provide additional insight regarding the sensitivity of these estimators to the choice of prior parameters. Parameters $r > 0$ and $w > 0$ (of the inverse gamma distribution) are selected using the following arguments. If θ_0 is the true value of θ used in simulations, then different combinations of r and w lead to different levels of the probability mass concentration around θ_0. Of course, because the inverse gamma family is not symmetric, declaration of 'center' is subjective. One way to define θ_0 as center is to let θ_0 be the average of the $\alpha/2$-quantile and $(1-(\alpha/2))$-quantile of the inverse gamma distribution with parameters r and w, *i.e.* the middle point of the middle $(1-\alpha)\cdot 100\%$ of probability mass. Straightforward simplifications of this condition yield the following relation between w and r

$$\frac{w}{\theta_0} = 2\frac{G_{r,1}^{-1}(\alpha/2)G_{r,1}^{-1}(1-(\alpha/2))}{G_{r,1}^{-1}(\alpha/2) + G_{r,1}^{-1}(1-(\alpha/2))} \tag{1}$$

where $G_{r,1}^{-1}(\alpha)$ denotes the α-quantile of the gamma distribution with parameters r and 1. Further, we add one more condition to control the width of the interval where the middle $(1-\alpha)\cdot 100\%$ of probability mass resides. This is done by deciding on the level of the relative error between the $(1-(\alpha/2))$-quantile and θ_0, which can be expressed as

$$\frac{q_{r,w}(1-(\alpha/2)) - \theta_0}{\theta_0} = \frac{G_{r,1}^{-1}(1-(\alpha/2)) - G_{r,1}^{-1}(\alpha/2)}{G_{r,1}^{-1}(1-(\alpha/2)) + G_{r,1}^{-1}(\alpha/2)}, \tag{2}$$

where $q_{r,w}(\alpha)$ denotes the α-quantile of the inverse gamma distribution with parameters r and w. Finally, for $\alpha = 0.10$ and $[q_{r,w}(1-(\alpha/2)) - \theta_0]/\theta_0 \approx 0.70$, 0.40, and 0.20, equations (1) and (2) lead to prior information, that is,

- significantly off target (estimator $\hat{\theta}_{\mathrm{BI}}$): $r = 4$, $w/\theta_0 = 2.3232$, $q_{r,w}(0.05)/\theta_0 = 0.2996$, $q_{r,w}(0.95)/\theta_0 = 1.7004$; $\mathrm{mode}_{r,w}/\theta_0 = 0.4646$, $\mathrm{median}_{r,w}/\theta_0 = 0.6327$, $\mathrm{mean}_{r,w}/\theta_0 = 0.7744$.
- slightly off target (estimator $\hat{\theta}_{\mathrm{B2}}$): $r = 15$, $w/\theta_0 = 13.0004$, $q_{r,w}(0.05)/\theta_0 = 0.5940$, $q_{r,w}(0.95)/\theta_0 = 1.4060$; $\mathrm{mode}_{r,w}/\theta_0 = 0.8125$, $\mathrm{median}_{r,w}/\theta_0 = 0.8863$, $\mathrm{mean}_{r,w}/\theta_0 = 0.9286$.
- almost on target (estimator $\hat{\theta}_{\mathrm{B3}}$): $r = 66$, $w/\theta_0 = 63.8953$, $q_{r,w}(0.05)/\theta_0 = 0.7996$, $q_{r,w}(0.95)/\theta_0 = 1.2004$; $\mathrm{mode}_{r,w}/\theta_0 = 0.9537$, $\mathrm{median}_{r,w}/\theta_0 = 0.9730$, $\mathrm{mean}_{r,w}/\theta_0 = 0.9830$.

Figure 1 illustrates how on/off target prior densities are for $\theta_0 = 1$.

We used the following design for the Monte Carlo simulation study.

1. A total of 25,000 samples of size n were generated from a normal distribution $N(d_i\theta, \theta^2)$, $i = 1, \ldots, n$, for a fixed choice of $\theta, d_1, \ldots, d_n$. (As our first choice of parameters, we used $\theta = 1$ and $d_i = i^{\log 5/\log 100}$, $i = 1, \ldots, n$. (Rationale: the formula for d_i yields $d_1 = 1$ and $d_{100} = 5$.)
2. For each estimator in the study, these samples were used to compute estimates $\hat{\theta}_1, \ldots, \hat{\theta}_{25,000}$. Then, the mean (denoted $\hat{m}(\hat{\theta})$) and MSE (denoted $\widehat{mse}(\hat{\theta})$) of these $\hat{\theta}_1, \ldots, \hat{\theta}_{25,000}$ were evaluated.
3. Without changing $\theta, d_1, \ldots, d_n$, steps 1 and 2 were repeated 10 times and statistics $\hat{m}_j(\hat{\theta}), \widehat{mse}_j(\hat{\theta})$, $j = 1, \ldots, 10$, were obtained for each estimator. Then, the mean and standard deviation of the standardized quantities, $\hat{m}_1(\hat{\theta})/\theta, \ldots, \hat{m}_{10}(\hat{\theta})/\theta$ and $\widehat{mse}_1(\hat{\theta})/C, \ldots, \widehat{mse}_{10}(\hat{\theta})/C$, where $C = \theta^2/\sum_{i=1}^{n}(d_i^2 + 2)$, were calculated.

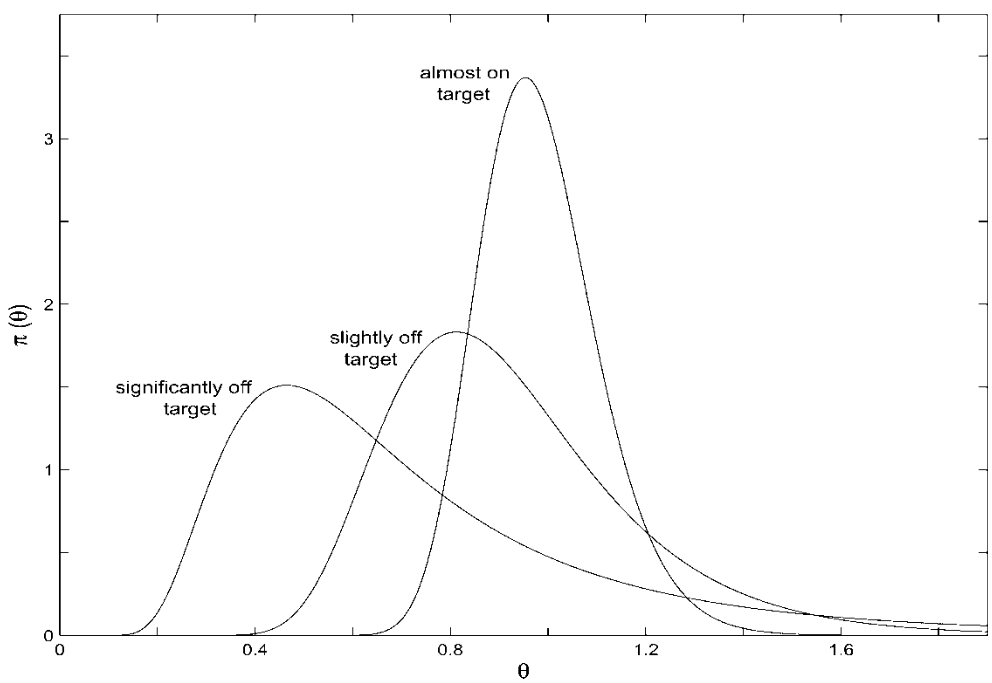

Figure 1. Inverse gamma priors around the target value $\theta_0 = 1$.

Table 1. Values of mean ($\times\theta^{-1}$) and mse ($\times\theta^{-2}\sum_{i=1}^{n}(d_i^2+2)$) for selected n.

Statistic	Estimator	n = 5	15	25	50	100	∞
Mean	$\hat{\theta}_{\mathrm{MLE}}$	$0.95_{(0.001)}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{LU}}$	$1.00_{(0.002)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{LMMS}}$	$0.91_{(0.002)}$	$0.97_{(0.001)}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{P}}$	$0.92_{(0.001)}$	$0.97_{(0.001)}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{B1}}$	$0.95_{(0.001)}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{B2}}$	$0.96_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	1
	$\hat{\theta}_{\mathrm{B3}}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	1
MSE	$\hat{\theta}_{\mathrm{MLE}}$	$0.97_{(0.006)}$	$0.99_{(0.011)}$	$0.99_{(0.009)}$	$0.99_{(0.012)}$	$1.00_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{LU}}$	$1.23_{(0.009)}$	$1.08_{(0.009)}$	$1.05_{(0.009)}$	$1.02_{(0.012)}$	$1.01_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{LMMS}}$	$1.04_{(0.008)}$	$1.02_{(0.009)}$	$1.01_{(0.008)}$	$1.00_{(0.012)}$	$1.00_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{P}}$	$0.92_{(0.006)}$	$0.97_{(0.010)}$	$0.98_{(0.009)}$	$0.99_{(0.012)}$	$0.99_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{B1}}$	$0.68_{(0.004)}$	$0.86_{(0.009)}$	$0.91_{(0.008)}$	$0.95_{(0.012)}$	$0.97_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{B2}}$	$0.23_{(0.001)}$	$0.52_{(0.005)}$	$0.65_{(0.006)}$	$0.79_{(0.011)}$	$0.88_{(0.007)}$	1
	$\hat{\theta}_{\mathrm{B3}}$	$0.03_{(0.001)}$	$0.12_{(0.001)}$	$0.21_{(0.002)}$	$0.39_{(0.005)}$	$0.59_{(0.005)}$	1

Note: The given entries are mean values (and standard deviations in parentheses) based on 10 replications.

The results of step 3 are reported in table 1. We also performed additional simulation studies, for other choices of $\theta, d_1, \ldots, d_n$ (in step 1), and found the outcomes to be identical to those of table 1.

The mean of all estimators converges to θ quite fast, underestimating it by no more than 2% for $n \geq 25$. As expected, the unbiased estimator LU beats the competition with respect to this criterion. The MSE of the MLE also converges fast to the asymptotic counterpart, being a bit more favorable in small samples (3% improvement for $n = 5$ and 1% for $n = 15$, 25, 50). Performances of LU and LMMS estimators are inferior to that of MLE in small samples and become comparable only in samples of at least 50 observations. The P estimator is slightly outperforming MLE for $n \leq 25$ and exhibits equivalent performance in larger samples ($n \geq 50$). The Bayes estimator equipped with varying quality prior information can provide from moderate to substantial improvements over MLE in small samples, which decrease as n becomes larger. For example, if prior information is 'significantly off' target, the MSE of the Bayes estimator $\hat{\theta}_{\mathrm{B1}}$ is almost 30% better than that of MLE for $n = 5$ and 13% better for $n = 15$; if prior information is 'slightly off' target (for $\hat{\theta}_{\mathrm{B2}}$), the improvement is even greater: 76% for $n = 5$, 47% for $n = 15$, and 34% for $n = 25$; if prior information is 'almost on' target (for $\hat{\theta}_{\mathrm{B3}}$), the Bayes estimator performs spectacularly: 97% improvement over MLE for $n = 5$, 79% for $n = 25$, and 41% for $n = 100$. Finally, note that the bias contribution to the MSE is negligible when compared with that of variance. Thus, MSE is essentially measuring variability.

3. Sensitivity study

Findings of the previous section demonstrate that the Bayes estimator is sensitive to the choice of the prior distribution parameters. Here, we go further and investigate the sensitivity of all the estimators under various model misspecification scenarios. In section 3.1, we study the case where, instead of a normal distribution, data are generated by other symmetric distributions (but with the same location parameters and the same coefficients of variation). In section 3.2, we treat the problem of undetected dependence, *i.e.* when the assumption of independence is violated, and in section 3.3, we summarize the sensitivity study by applying the premium-protection approach of Anscombe [8].

3.1 *Scenario 1: misspecified distribution*

We fix the coefficients of variation $1/d_1, \ldots, 1/d_n$ to match those used in the study of section 2.3 and, instead of a normal distribution, consider six other similar shape distributions (with the same locations: $d_1\theta, \ldots, d_n\theta$). Specifically, we generate data $X_1, \ldots, X_n$ from the following families:

- *Logistic* with the cdf of X_i given by

$$F_{X_i}^{(\text{Logistic})}(x) = \frac{1}{1 + \mathrm{e}^{-(x-d_i\theta)/(\theta\sqrt{3}/\pi)}}, \quad -\infty < x < \infty,$$

- *Laplace* with the cdf of X_i given by

$$F_{X_i}^{(\text{Laplace})}(x) = \begin{cases} \dfrac{1}{2}\mathrm{e}^{-(d_i\theta - x)/(\theta/\sqrt{2})}, & -\infty < x \le d_i\theta, \\ 1 - \dfrac{1}{2}\mathrm{e}^{-(x-d_i\theta)/(\theta/\sqrt{2})}, & d_i\theta < x < \infty, \end{cases}$$

- *Contaminated normal* with the cdf of X_i given by

$$F_{X_i}^{(\text{contam})}(x) = (1-\varepsilon)\Phi\left(\frac{x - d_i\theta}{\theta}\right) + \varepsilon G(x), \quad -\infty < x < \infty,$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution and $G(\cdot)$ denotes the cdf of a contaminating distribution. We consider four combinations of ε and $G : \varepsilon = 0.20, G = F_{X_i}^{(\text{Logistic})}$ (denoted *'logistic' contamination*, 20%), $\varepsilon = 0.20$, $G = F_{X_i}^{(\text{Laplace})}$ (denoted *'Laplace' contamination*, 20%), $\varepsilon = 0.50$, $G = F_{X_i}^{(\text{Logistic})}$ (denoted *'logistic' contamination*, 50%), $\varepsilon = 0.50$, $G = F_{X_i}^{(\text{Laplace})}$ (denoted *'Laplace' contamination*, 50%). Note that by choosing $\varepsilon = 1$, we get $F_{X_i}^{(\text{contaim})} = G$. Therefore, in figure 2 and table 2, the logistic and Laplace families will be, respectively, denoted as *'logistic' contamination*, 100%, and *'Laplace' contamination*, 100%.

In figure 2, plots of the assumed normal distribution pdf versus various contaminated normal pdf's are presented for $d_i = 10$ and $\theta = 1$. One can see that differences between the curves are small and, thus, the distributional misspecifications are realistic, likely, and dangerous.

Simulations are performed by following the procedure of section 2.3. In this section, however, 25,000 samples are generated from the six non-normal distributions, but parameter θ is estimated using formulas of section 2.1, which are designed for strictly normal data. In table 2, we report the standardized mean's and MSE's of the estimators (with their standard errors in parentheses) under various distributional scenarios, for sample size $n = 5$, 15, 25. Patterns for other choices of n are similar and are graphically summarized in section 3.3. The 'ideal case' where data follow the assumed normal distribution is included as a reference.

The mean of all estimators is virtually unchanged from that in the ideal case, under both types and for all levels of contamination, and for all n. (Thus, similar to table 1, the bias is negligible, and MSE is essentially measuring variability.) However, the MSE's are inflated by contamination for all estimators and that inflation gets larger as sample size increases;

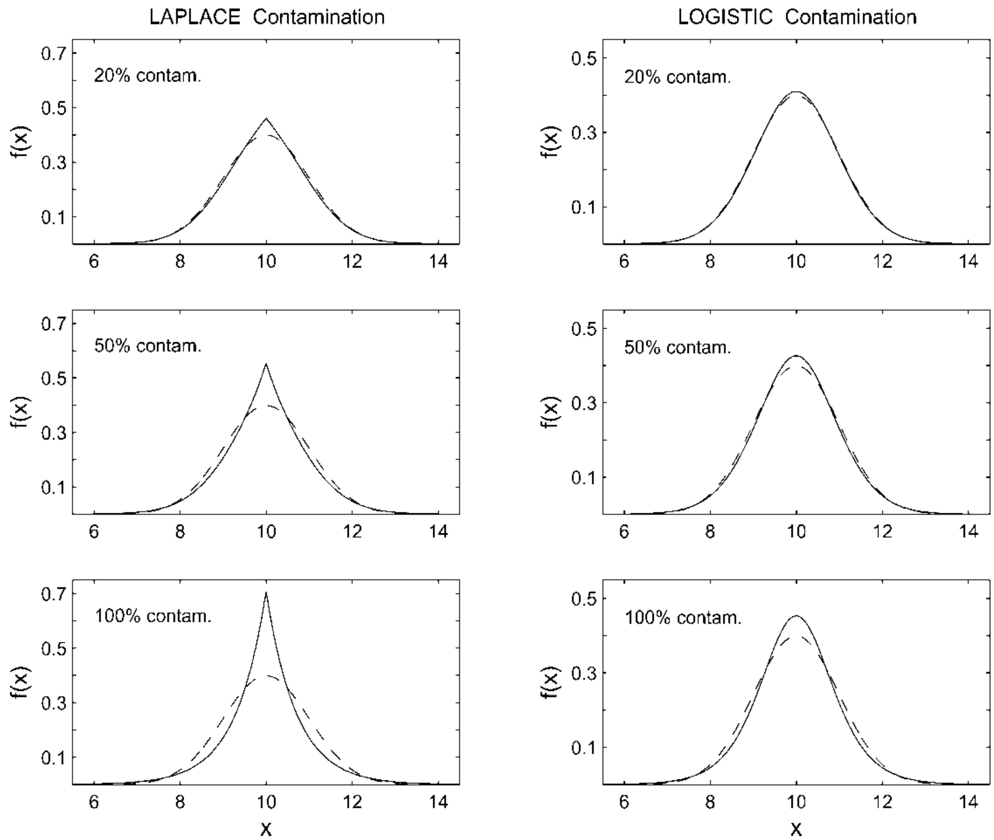

Figure 2. The pdf of $N(d_i\theta, \theta^2)$ with $d_i = 10, \theta = 1$ (dashed line – – –) versus contaminated normal pdf's (solid line ——) for selected levels of Laplace and logistic contaminations.

and, for a fixed level of contamination and a fixed n, 'Laplace' type is more severe than 'logistic'. For example, for $n = 5$, the MSE of the estimators increases by 4–8% (9–19%), 9–19% (24–46%), and 20–40% (49–88%) for the 'logistic' ('Laplace') 20%, 50%, and 100% contamination, respectively; for $n = 15$, the MSE of the estimators increases by 7–8% (18–21%), 19–22% (46–54%), and 39–45% (96–107%) for the 'logistic' ('Laplace') 20%, 50%, and 100% contamination, respectively; for $n = 25$, the MSE of the estimators increases by 8–9% (21–24%), 21–24% (53–57%), 44–48% (110–116%) for the 'logistic' ('Laplace') 20%, 50%, and 100% contamination, respectively. Although not reported in the table, the corresponding percentage ranges for $n = 100$ are 11% (27%), 27–28% (67–68%), and 54% (134–136%) for the 'logistic' ('Laplace') 20%, 50%, and 100% contamination, respectively. Finally, note that computation of these ranges was based on the values of MSE's accurate within four decimal places, and thus for some entries (*e.g.* for $n = 5$ and $\hat{\theta}_{B3}$), there are numerical inconsistencies between the table and this discussion.

3.2 *Scenario 2: undetected dependence*

In this section, we investigate the behavior of the estimators $\hat{\theta}_{MLE}, \hat{\theta}_{LU}, \hat{\theta}_{LMMS}, \hat{\theta}_P, \theta_{B1}, \hat{\theta}_{B2}, \hat{\theta}_{B3}$ when the covariance structure of vector $(X_1, \ldots, X_n)'$ is altered. The alterations are introduced by replacing the identity matrix I with Σ, where $\sigma_{ij} = \rho^{|i-j|}$ for $i, j = 1, \ldots, n, i \neq j$, and $\sigma_{ii} = 1$ for $i = 1, \ldots, n$. As $\rho = 0$ corresponds to the independent case, we choose three

Table 2. Values of mean ($\times\theta^{-1}$) and MSE ($\times\theta^{-2}\sum_{i=1}^{n}(d_i^2+2)$) for $n = 5, 15, 25$.

		Logistic contamination			Ideal case	Laplace contamination		
Statistic	Estimator	100%	50%	20%	Normal	20%	50%	100%
Sample size: $n = 5$								
Mean	$\hat{\theta}_{\mathrm{MLE}}$	$0.94_{(0.002)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.94_{(0.001)}$	$0.93_{(0.002)}$
	$\hat{\theta}_{\mathrm{LU}}$	$0.99_{(0.002)}$	$0.99_{(0.002)}$	$1.00_{(0.001)}$	$1.00_{(0.002)}$	$0.99_{(0.001)}$	$0.98_{(0.001)}$	$0.97_{(0.002)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$0.90_{(0.002)}$	$0.91_{(0.002)}$	$0.91_{(0.001)}$	$0.91_{(0.002)}$	$0.91_{(0.001)}$	$0.90_{(0.001)}$	$0.88_{(0.002)}$
	$\hat{\theta}_{\mathrm{P}}$	$0.91_{(0.002)}$	$0.92_{(0.001)}$	$0.92_{(0.001)}$	$0.92_{(0.001)}$	$0.92_{(0.001)}$	$0.91_{(0.001)}$	$0.90_{(0.002)}$
	$\hat{\theta}_{\mathrm{B1}}$	$0.94_{(0.002)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.94_{(0.001)}$	$0.93_{(0.002)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.95_{(0.001)}$	$0.96_{(0.001)}$	$0.96_{(0.001)}$	$0.96_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$	$0.95_{(0.001)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$
MSE	$\hat{\theta}_{\mathrm{MLE}}$	$1.25_{(0.013)}$	$1.11_{(0.012)}$	$1.02_{(0.014)}$	$0.97_{(0.006)}$	$1.10_{(0.015)}$	$1.31_{(0.014)}$	$1.63_{(0.020)}$
	$\hat{\theta}_{\mathrm{LU}}$	$1.48_{(0.018)}$	$1.35_{(0.016)}$	$1.28_{(0.017)}$	$1.23_{(0.009)}$	$1.35_{(0.018)}$	$1.53_{(0.019)}$	$1.84_{(0.024)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$1.24_{(0.015)}$	$1.13_{(0.013)}$	$1.07_{(0.014)}$	$1.04_{(0.008)}$	$1.13_{(0.015)}$	$1.28_{(0.015)}$	$1.55_{(0.020)}$
	$\hat{\theta}_{\mathrm{P}}$	$1.17_{(0.012)}$	$1.04_{(0.011)}$	$0.96_{(0.013)}$	$0.92_{(0.006)}$	$1.03_{(0.014)}$	$1.22_{(0.013)}$	$1.51_{(0.018)}$
	$\hat{\theta}_{\mathrm{B1}}$	$0.87_{(0.009)}$	$0.78_{(0.008)}$	$0.72_{(0.010)}$	$0.68_{(0.004)}$	$0.77_{(0.011)}$	$0.91_{(0.010)}$	$1.13_{(0.014)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.30_{(0.003)}$	$0.26_{(0.003)}$	$0.24_{(0.003)}$	$0.23_{(0.001)}$	$0.26_{(0.004)}$	$0.31_{(0.004)}$	$0.39_{(0.005)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.04_{(0.001)}$	$0.03_{(0.001)}$	$0.03_{(0.001)}$	$0.03_{(0.001)}$	$0.03_{(0.001)}$	$0.04_{(0.001)}$	$0.05_{(0.001)}$
Sample size: $n = 15$								
Mean	$\hat{\theta}_{\mathrm{MLE}}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.002)}$	$0.97_{(0.001)}$
	$\hat{\theta}_{\mathrm{LU}}$	$0.99_{(0.001)}$	$1.00_{(0.002)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$0.99_{(0.002)}$	$0.98_{(0.001)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$0.96_{(0.001)}$	$0.97_{(0.002)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.96_{(0.002)}$	$0.95_{(0.001)}$
	$\hat{\theta}_{\mathrm{P}}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.002)}$	$0.96_{(0.001)}$
	$\hat{\theta}_{\mathrm{B1}}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.002)}$	$0.97_{(0.001)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.97_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.97_{(0.001)}$	$0.97_{(0.001)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.98_{(0.001)}$
MSE	$\hat{\theta}_{\mathrm{MLE}}$	$1.41_{(0.018)}$	$1.20_{(0.009)}$	$1.07_{(0.013)}$	$0.99_{(0.011)}$	$1.19_{(0.008)}$	$1.49_{(0.010)}$	$2.01_{(0.025)}$
	$\hat{\theta}_{\mathrm{LU}}$	$1.50_{(0.017)}$	$1.28_{(0.009)}$	$1.15_{(0.013)}$	$1.08_{(0.009)}$	$1.27_{(0.010)}$	$1.57_{(0.013)}$	$2.11_{(0.024)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$1.41_{(0.016)}$	$1.21_{(0.009)}$	$1.09_{(0.012)}$	$1.02_{(0.009)}$	$1.20_{(0.009)}$	$1.48_{(0.012)}$	$1.99_{(0.023)}$
	$\hat{\theta}_{\mathrm{P}}$	$1.37_{(0.018)}$	$1.17_{(0.009)}$	$1.04_{(0.013)}$	$0.97_{(0.010)}$	$1.16_{(0.008)}$	$1.45_{(0.010)}$	$1.96_{(0.024)}$
	$\hat{\theta}_{\mathrm{B1}}$	$1.23_{(0.016)}$	$1.04_{(0.008)}$	$0.93_{(0.011)}$	$0.86_{(0.009)}$	$1.04_{(0.007)}$	$1.30_{(0.009)}$	$1.75_{(0.022)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.73_{(0.009)}$	$0.62_{(0.004)}$	$0.56_{(0.007)}$	$0.52_{(0.005)}$	$0.62_{(0.005)}$	$0.78_{(0.005)}$	$1.05_{(0.013)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.17_{(0.002)}$	$0.15_{(0.001)}$	$0.13_{(0.002)}$	$0.12_{(0.001)}$	$0.15_{(0.001)}$	$0.19_{(0.002)}$	$0.25_{(0.003)}$
Sample size: $n = 25$								
Mean	$\hat{\theta}_{\mathrm{MLE}}$	$0.99_{(0.002)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.98_{(0.002)}$
	$\hat{\theta}_{\mathrm{LU}}$	$1.00_{(0.002)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$1.00_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.002)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$0.98_{(0.002)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.97_{(0.002)}$
	$\hat{\theta}_{\mathrm{P}}$	$0.98_{(0.002)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.97_{(0.002)}$
	$\hat{\theta}_{\mathrm{B1}}$	$0.98_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.002)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.001)}$	$0.98_{(0.002)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$	$0.99_{(0.001)}$
MSE	$\hat{\theta}_{\mathrm{MLE}}$	$1.46_{(0.025)}$	$1.22_{(0.011)}$	$1.08_{(0.007)}$	$0.99_{(0.009)}$	$1.22_{(0.013)}$	$1.55_{(0.018)}$	$2.13_{(0.037)}$
	$\hat{\theta}_{\mathrm{LU}}$	$1.51_{(0.026)}$	$1.27_{(0.012)}$	$1.13_{(0.008)}$	$1.05_{(0.009)}$	$1.27_{(0.015)}$	$1.60_{(0.019)}$	$2.19_{(0.040)}$
	$\hat{\theta}_{\mathrm{LMMS}}$	$1.45_{(0.025)}$	$1.22_{(0.011)}$	$1.09_{(0.007)}$	$1.01_{(0.008)}$	$1.22_{(0.014)}$	$1.54_{(0.018)}$	$2.11_{(0.038)}$
	$\hat{\theta}_{\mathrm{P}}$	$1.43_{(0.024)}$	$1.20_{(0.010)}$	$1.06_{(0.007)}$	$0.98_{(0.009)}$	$1.20_{(0.013)}$	$1.52_{(0.018)}$	$2.09_{(0.037)}$
	$\hat{\theta}_{\mathrm{B1}}$	$1.33_{(0.023)}$	$1.12_{(0.010)}$	$0.99_{(0.006)}$	$0.91_{(0.008)}$	$1.12_{(0.012)}$	$1.42_{(0.017)}$	$1.95_{(0.034)}$
	$\hat{\theta}_{\mathrm{B2}}$	$0.95_{(0.016)}$	$0.79_{(0.007)}$	$0.70_{(0.004)}$	$0.65_{(0.006)}$	$0.79_{(0.009)}$	$1.01_{(0.012)}$	$1.39_{(0.024)}$
	$\hat{\theta}_{\mathrm{B3}}$	$0.32_{(0.006)}$	$0.26_{(0.002)}$	$0.23_{(0.002)}$	$0.21_{(0.002)}$	$0.26_{(0.003)}$	$0.34_{(0.004)}$	$0.46_{(0.009)}$

Note: The given entries are mean values (and standard deviations in parentheses) based on 10 replications.

levels of correlation $\rho = 0.75$, 0.95, and 0.999. These respective choices yield positively *weakly*, positively *moderately*, and positively *strongly* correlated data. Such a characterization is motivated by the consideration of the correlation coefficient between two least dependent coordinates, which is given by ρ^{n-1} and some of its numerical values are presented in table 3.

Table 3. Values of ρ^{n-1} for selected n.

	n				
Data dependence	5	15	25	50	100
Weak ($\rho = 0.75$)	0.3164	0.0178	0.0010	$\sim 10^{-6}$	$\sim 10^{-12}$
Moderate ($\rho = 0.95$)	0.8145	0.4877	0.2920	0.0810	0.0062
Strong ($\rho = 0.999$)	0.9960	0.9861	0.9763	0.9522	0.9057

The case of *moderate-mixed* correlation, which corresponds to $\rho = -0.75$, is also included in the study. Here, the term 'mixed correlation' is used because some pairs of the X's coordinates are positively correlated (when $|i - j|$ is even), whereas the remaining ones are negatively correlated (when $|i - j|$ is odd).

Similar to Scenario 1, the simulation study is performed by following the procedure of section 2.3. For each type of dependence, 25,000 samples were generated using Σ but parameter θ was estimated using formulas of section 2.1 which are designed for the independent data. In table 4, the standardized mean's and MSE's are reported for the four types of dependence and for sample size $n = 5, 15, 25$. Results for other choices of n are graphically summarized in section 3.3. The 'zero' correlation case is included as a reference.

In table 4, notation $(+)$ and $(\pm)$ is used to denote scenarios when $\rho > 0$ and $\rho < 0$, respectively. One can see that the presence of correlation in the data has damaging effect on the mean and MSE of the estimators which are designed for the independent data. Indeed, as the level of positive correlation increases, all estimators have increasing negative bias which is even larger for $\rho < 0$. The bias becomes smaller as n increases but it does not vanish (even for sample sizes as large as $n = 100$). The linear estimators, LU and LMMS, are the worst performers with respect to bias. Further examination of the table shows that the effect of correlation on MSE's is catastrophic (measured in *hundreds* of percentage points!), it becomes even larger as $\rho > 0$ increases and is further magnified by increasing n. Comparisons of *moderate* $(+)$ versus *moderate* $(\pm)$ correlations show that the positive scenario has a more severe effect. Interestingly, the linear estimators, LU and LMMS, are least affected by correlation and are even the best performers for $\rho < 0$, with respect to MSE. This fact becomes even more evident when we summarize simulation results in figure 3. Finally, the inconsistencies between the table and this discussion for $n = 5$ disappear when n reaches 15.

3.3 *Premium–protection plots*

We shall summarize the sensitivity study via the *premium–protection* (PP) approach of Anscombe [8], which was effectively employed by Brazauskas and Serfling [11] to investigate efficiency-robustness trade-offs of robust estimators. In the latter paper, the performance of estimators was summarized using this approach and estimators were displayed as points on so-called premium versus protection plots (PP-*plots*).

The approach works as follows. For each estimator T under consideration, corresponding 'premium' and 'protection' values are defined:

Premium: The relative change, (typically increase) in MSE due to use of T instead of the MLE in the null case V_0 (no violation), *i.e.*

$$\text{Premium}(T, V_0) = \frac{\text{MSE}(T, V_0) - \text{MSE}(\text{MLE}, V_0)}{\text{MSE}(\text{MLE}, V_0)}.$$

Table 4. Values of mean ($\times\theta^{-1}$) and MSE ($\times\theta^{-2}\sum_{i=1}^{n}(d_i^2+2)$) for $n = 5, 15, 25$.

			Correlation				
Sample size	Statistic	Estimator	Strong (+)	Moderate (+)	Weak (+)	Zero	Moderate (±)
$n = 5$	Mean	$\hat{\theta}_{\text{MLE}}$	$0.70_{(0.004)}$	$0.91_{(0.004)}$	$0.97_{(0.003)}$	$0.95_{(0.001)}$	$0.43_{(0.002)}$
		$\hat{\theta}_{\text{LU}}$	$0.32_{(0.002)}$	$0.87_{(0.005)}$	$0.86_{(0.003)}$	$1.00_{(0.002)}$	$0.25_{(0.001)}$
		$\hat{\theta}_{\text{LMMS}}$	$0.32_{(0.002)}$	$0.78_{(0.004)}$	$0.77_{(0.003)}$	$0.91_{(0.002)}$	$0.25_{(0.001)}$
		$\hat{\theta}_{\text{P}}$	$0.68_{(0.004)}$	$0.88_{(0.004)}$	$0.94_{(0.003)}$	$0.92_{(0.001)}$	$0.43_{(0.002)}$
		$\hat{\theta}_{\text{B1}}$	$0.73_{(0.004)}$	$0.92_{(0.004)}$	$0.97_{(0.003)}$	$0.95_{(0.001)}$	$0.45_{(0.002)}$
		$\hat{\theta}_{\text{B2}}$	$0.85_{(0.002)}$	$0.96_{(0.002)}$	$0.98_{(0.002)}$	$0.96_{(0.001)}$	$0.56_{(0.002)}$
		$\hat{\theta}_{\text{B3}}$	$0.96_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	$0.98_{(0.001)}$	$0.83_{(0.001)}$
	MSE	$\hat{\theta}_{\text{MLE}}$	$3.74_{(0.041)}$	$3.90_{(0.034)}$	$2.48_{(0.020)}$	$0.97_{(0.006)}$	$1.46_{(0.014)}$
		$\hat{\theta}_{\text{LU}}$	$1.12_{(0.005)}$	$4.35_{(0.043)}$	$2.07_{(0.022)}$	$1.23_{(0.009)}$	$0.61_{(0.001)}$
		$\hat{\theta}_{\text{LMMS}}$	$1.10_{(0.005)}$	$3.55_{(0.034)}$	$1.70_{(0.017)}$	$1.04_{(0.008)}$	$0.61_{(0.001)}$
		$\hat{\theta}_{\text{P}}$	$3.48_{(0.035)}$	$3.63_{(0.031)}$	$2.32_{(0.018)}$	$0.92_{(0.006)}$	$1.44_{(0.013)}$
		$\hat{\theta}_{\text{B1}}$	$2.74_{(0.026)}$	$2.64_{(0.024)}$	$1.70_{(0.014)}$	$0.68_{(0.004)}$	$1.34_{(0.012)}$
		$\hat{\theta}_{\text{B2}}$	$0.87_{(0.010)}$	$0.84_{(0.009)}$	$0.56_{(0.006)}$	$0.23_{(0.001)}$	$0.82_{(0.006)}$
		$\hat{\theta}_{\text{B3}}$	$0.11_{(0.002)}$	$0.11_{(0.002)}$	$0.07_{(0.001)}$	$0.03_{(0.001)}$	$0.12_{(0.001)}$
$n = 15$	Mean	$\hat{\theta}_{\text{MLE}}$	$0.78_{(0.003)}$	$0.95_{(0.003)}$	$1.02_{(0.001)}$	$0.98_{(0.001)}$	$0.51_{(0.002)}$
		$\hat{\theta}_{\text{LU}}$	$0.51_{(0.002)}$	$0.88_{(0.003)}$	$0.93_{(0.001)}$	$1.00_{(0.001)}$	$0.28_{(0.001)}$
		$\hat{\theta}_{\text{LMMS}}$	$0.50_{(0.002)}$	$0.85_{(0.003)}$	$0.90_{(0.001)}$	$0.97_{(0.001)}$	$0.28_{(0.001)}$
		$\hat{\theta}_{\text{P}}$	$0.77_{(0.003)}$	$0.94_{(0.003)}$	$1.00_{(0.001)}$	$0.97_{(0.001)}$	$0.51_{(0.002)}$
		$\hat{\theta}_{\text{B1}}$	$0.78_{(0.003)}$	$0.95_{(0.003)}$	$1.01_{(0.001)}$	$0.98_{(0.001)}$	$0.52_{(0.002)}$
		$\hat{\theta}_{\text{B2}}$	$0.84_{(0.002)}$	$0.96_{(0.002)}$	$1.00_{(0.001)}$	$0.98_{(0.001)}$	$0.57_{(0.002)}$
		$\hat{\theta}_{\text{B3}}$	$0.94_{(0.001)}$	$0.99_{(0.001)}$	$1.00_{(0.001)}$	$0.99_{(0.001)}$	$0.75_{(0.001)}$
	MSE	$\hat{\theta}_{\text{MLE}}$	$11.39_{(0.105)}$	$8.58_{(0.097)}$	$3.55_{(0.041)}$	$0.99_{(0.011)}$	$3.89_{(0.051)}$
		$\hat{\theta}_{\text{LU}}$	$4.71_{(0.051)}$	$7.57_{(0.099)}$	$2.72_{(0.030)}$	$1.08_{(0.009)}$	$0.70_{(0.002)}$
		$\hat{\theta}_{\text{LMMS}}$	$4.58_{(0.050)}$	$7.08_{(0.093)}$	$2.55_{(0.028)}$	$1.02_{(0.009)}$	$0.70_{(0.002)}$
		$\hat{\theta}_{\text{P}}$	$11.09_{(0.102)}$	$8.35_{(0.094)}$	$3.46_{(0.040)}$	$0.97_{(0.010)}$	$3.84_{(0.051)}$
		$\hat{\theta}_{\text{B1}}$	$10.04_{(0.092)}$	$7.40_{(0.084)}$	$3.07_{(0.036)}$	$0.86_{(0.009)}$	$3.71_{(0.048)}$
		$\hat{\theta}_{\text{B2}}$	$5.81_{(0.055)}$	$4.25_{(0.050)}$	$1.80_{(0.022)}$	$0.52_{(0.005)}$	$2.92_{(0.035)}$
		$\hat{\theta}_{\text{B3}}$	$1.32_{(0.015)}$	$1.02_{(0.014)}$	$0.43_{(0.006)}$	$0.12_{(0.001)}$	$0.96_{(0.012)}$
$n = 25$	Mean	$\hat{\theta}_{\text{MLE}}$	$0.80_{(0.003)}$	$0.96_{(0.002)}$	$1.02_{(0.001)}$	$0.99_{(0.001)}$	$0.57_{(0.002)}$
		$\hat{\theta}_{\text{LU}}$	$0.60_{(0.002)}$	$0.90_{(0.002)}$	$0.96_{(0.001)}$	$1.00_{(0.001)}$	$0.31_{(0.001)}$
		$\hat{\theta}_{\text{LMMS}}$	$0.59_{(0.002)}$	$0.88_{(0.002)}$	$0.94_{(0.001)}$	$0.98_{(0.001)}$	$0.31_{(0.001)}$
		$\hat{\theta}_{\text{P}}$	$0.80_{(0.003)}$	$0.96_{(0.002)}$	$1.01_{(0.001)}$	$0.98_{(0.001)}$	$0.57_{(0.002)}$
		$\hat{\theta}_{\text{B1}}$	$0.81_{(0.003)}$	$0.96_{(0.002)}$	$1.02_{(0.001)}$	$0.99_{(0.001)}$	$0.58_{(0.002)}$
		$\hat{\theta}_{\text{B2}}$	$0.84_{(0.002)}$	$0.97_{(0.001)}$	$1.01_{(0.001)}$	$0.98_{(0.001)}$	$0.61_{(0.002)}$
		$\hat{\theta}_{\text{B3}}$	$0.93_{(0.001)}$	$0.99_{(0.001)}$	$1.01_{(0.001)}$	$0.99_{(0.001)}$	$0.74_{(0.002)}$
	MSE	$\hat{\theta}_{\text{MLE}}$	$18.89_{(0.229)}$	$11.27_{(0.143)}$	$3.81_{(0.033)}$	$0.99_{(0.009)}$	$5.92_{(0.068)}$
		$\hat{\theta}_{\text{LU}}$	$9.89_{(0.102)}$	$9.70_{(0.109)}$	$3.02_{(0.025)}$	$1.05_{(0.009)}$	$0.83_{(0.004)}$
		$\hat{\theta}_{\text{LMMS}}$	$9.66_{(0.100)}$	$9.32_{(0.104)}$	$2.90_{(0.024)}$	$1.01_{(0.008)}$	$0.83_{(0.004)}$
		$\hat{\theta}_{\text{P}}$	$18.57_{(0.224)}$	$11.07_{(0.140)}$	$3.75_{(0.033)}$	$0.98_{(0.009)}$	$5.87_{(0.067)}$
		$\hat{\theta}_{\text{B1}}$	$17.39_{(0.209)}$	$10.28_{(0.130)}$	$3.48_{(0.030)}$	$0.91_{(0.008)}$	$5.71_{(0.065)}$
		$\hat{\theta}_{\text{B2}}$	$12.10_{(0.146)}$	$7.13_{(0.093)}$	$2.44_{(0.022)}$	$0.65_{(0.006)}$	$4.78_{(0.052)}$
		$\hat{\theta}_{\text{B3}}$	$3.76_{(0.054)}$	$2.33_{(0.035)}$	$0.81_{(0.008)}$	$0.21_{(0.002)}$	$2.12_{(0.023)}$

Note: The given entries are mean values (and standard deviations in parentheses) based on 10 replications.

Protection: The relative change (preferably decrease) in MSE due to use of T instead of the MLE in a non-null case V (violation), *i.e.*

$$\text{Protection}(T, V) = \frac{\text{MSE}(\text{MLE}, V) - \text{MSE}(T, V)}{\text{MSE}(\text{MLE}, V)}.$$

For ease of comparison, estimators are displayed as points on the PP-plots, with favorable estimators being located in the upper left corner (or northwest direction) of the PP-plot, reflecting maximal protection for minimal premium.

In figure 3, the following symbols are used to denote performances of the estimators under the violation scenarios of sections 3.1 and 3.2.

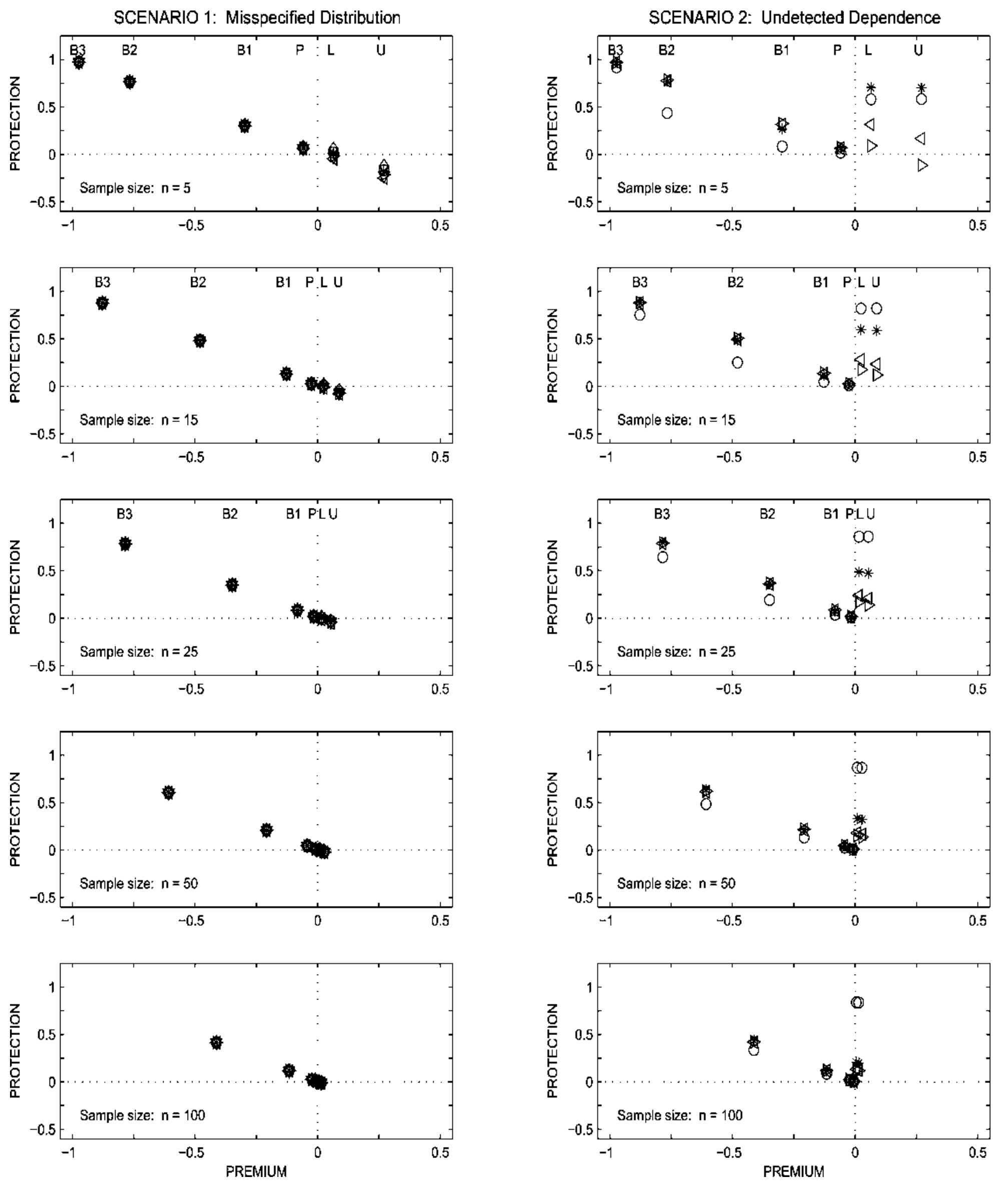

Figure 3. Performance of estimators under various distributional and independence violations for $n = 5$, 15, 25, 50, 100. A benchmark procedure is MLE (coordinates (0, 0)).

Scenario 1: misspecified distribution	Scenario 2: undetected dependence
$\ast$ Logistic contamination, 100%	$\ast$ Weak-positive correlation
$\triangleright$ Logistic contamination, 50%	$\triangleright$ Moderate-positive correlation
$\triangleleft$ Logistic contamination, 20%	$\triangleleft$ Strong-positive correlation
∇ Laplace contamination, 20%	$\circ$ Moderate-mixed correlation
$\circ$ Laplace contamination, 50%	
$\diamond$ Laplace contamination, 100%	

Also, for convenience, the following notation for the estimators is used: $\hat{\theta}_{\mathrm{LU}}$ is denoted as U; $\hat{\theta}_{\mathrm{LMMS}}$ is denoted as L; $\hat{\theta}_{\mathrm{P}}$, $\hat{\theta}_{\mathrm{B1}}$, $\hat{\theta}_{\mathrm{B2}}$, $\hat{\theta}_{\mathrm{B3}}$ are, respectively, denoted as P, B1, B2, B3. In each PP-plot, the estimator's performance under the violations corresponds to a column of symbols below the letter denoting the estimator.

From these plots, the following conclusions emerge. With the exception of estimator B3, premiums for all estimators are very close to 0 (*i.e.* they are almost equivalent to MLE), for sample sizes $n \geq 50$. Protections also follow a similar pattern with one additional exception of L and U estimators under the moderate-mixed correlation scenario. There is no effect of different distributional violations, except for L and U estimators, for $n = 5$. However, these two estimators are consistently outperformed by the other estimators (including the MLE) under Scenario 1. Consideration of Scenario 2 reveals that the effect of different violations is significant for L and U estimators, for $n \leq 25$, and there is almost no effect for other estimators. Among the violations, the moderate-mixed correlation scenario seems to have the least damaging effect on L and U estimators, thus making them very competitive. This is most visible for $n \geq 25$. Overall, estimators B3, B2, B1, and P are located in the favorable 'north-west' territory for all sample sizes and for all scenarios; the L and U estimators demonstrate some strength under Scenario 2, especially in the case of mixed correlation.

4. Final remarks

The estimators considered in this article, $\hat{\theta}_{\mathrm{MLE}}$, $\hat{\theta}_{\mathrm{LU}}$, $\hat{\theta}_{\mathrm{LMMS}}$, $\hat{\theta}_{\mathrm{P}}$, $\hat{\theta}_{\mathrm{B1}}$, $\hat{\theta}_{\mathrm{B2}}$, $\hat{\theta}_{\mathrm{B3}}$, are asymptotically equivalent and their sensitivity to various distributional and independence violations is of a similar extent. The most visible differences among the estimators surface when we examine their small- and moderate-sample behavior. The linear estimators $\hat{\theta}_{\mathrm{LU}}$, $\hat{\theta}_{\mathrm{LMMS}}$ are consistently outperformed by all the other estimators and thus are less competitive, under Scenario 1. However, they show some strength under Scenario 2. In particular, in the case of moderate-mixed correlation, they are very competitive. The Pitman-type estimator $\hat{\theta}_{\mathrm{P}}$ is performing consistently better than $\hat{\theta}_{\mathrm{MLE}}$; however, that improvement is so minimal that, having in mind $\hat{\theta}_{\mathrm{P}}$'s computational complexity, it is probably not worthwhile to implement this estimator in practice. Examination of overall performances of $\hat{\theta}_{\mathrm{P}}$, $\hat{\theta}_{\mathrm{B1}}$, $\hat{\theta}_{\mathrm{B2}}$, $\hat{\theta}_{\mathrm{B3}}$ estimators shows a clear pattern which agrees with the intuition: for a fixed number of sample observations, more accurate prior knowledge about θ should yield a more accurate estimator. (Recall that the Pitman-type estimator is a Bayes estimator based on an uninformative prior (Remark 1) and, thus, can be labeled as 'totally off' target.) Thus, the Bayes estimator enhanced with high quality prior information can provide remarkable improvements over the MLE and should be included in practical use. In practice, however, formulating the prior information in as precise a manner as in θ_{B3} may not be an easy task.

In summary, for most practical situations, the MLE offers a reasonable trade-off between performance (measured in terms of MSE) and computational simplicity. The Bayes estimator based on even uninformative priors can still provide improvements over the MLE but is significantly more complex computationally. Finally, we emphasize again that all these estimators are sensitive to model violations and, thus, they are not robust. Development of robust estimators for θ (*e.g.* M-estimators) is an open but more complicated problem and, therefore, is postponed to future projects.

Acknowledgement

We are very appreciative of valuable insights and comments provided by an anonymous referee, leading to many improvements in the article.

References

[1] Gerig, T.M. and Sen, A.R., 1980, MLE in two normal samples with equal but unknown population coefficients of variation. *Journal of the American Statistical Association*, **75**, 704–708.
[2] Gleser, L.J. and Healy, J.D., 1976, Estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **71**, 977–981.
[3] Khan, R.A., 1968, A note on estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association*, **63**, 1039–1041.
[4] Searls, D.T., 1964, The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, **59**, 1225–1226.
[5] Azen, S.P. and Reed, A.H., 1973, Maximum likelihood estimation of correlation between variates having equal coefficients of variation. *Technometrics*, **15**(3), 457–462.
[6] Sinha, S.K., 1983, Bayesian estimation of the mean of a normal distribution when the coefficient of variation is known. *The Statistician*, **32**, 339–345.
[7] Rao, C.R., 1973, *Linear Statistical Inference and Its Applications* (2nd edn) (New York: Wiley).
[8] Anscombe, F.J., 1960, Rejection of outliers. *Technometrics*, **2**, 123–147.
[9] Lange, K., 1999, *Numerical Analysis for Statisticians* (New York: Springer).
[10] Serfling, R.J., 1980, *Approximation Theorems of Mathematical Statistics* (New York: Wiley).
[11] Brazauskas, V. and Serfling, R., 2001, Small sample performance of robust estimators of tail parameters for Pareto and exponential models. *Journal of Statistical Computation and Simulation*, **70**(1), 1–19.

Appendix A

LEMMA 1 *Let* $X_1, \ldots, X_n$ *be independent random variables with* $X_i \sim N(d_i\theta, \theta^2)$ *and let* $\bar{X}_u$ *and* S_u *be defined as in section* 2.1.2. *Then* (a) $\bar{X}_u \sim N(\theta, b_n^{-1}\theta^2)$, $(b)(n-1)S_u^2/\theta^2 \sim \chi^2_{n-1}$, *and* $(c)\bar{X}_u$ *and* S_u^2 *are independent.*

Proof First, notice that the joint pdf of $X_1, \ldots, X_n$ is given by

$$f(x_1, x_2, \ldots, x_n \mid \theta, d_1, \ldots, d_n) = (\theta\sqrt{2\pi})^{-n} \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - d_i\theta)^2\right\}.$$

Next, consider the transformation $Y = AX$, where $X = (X_1, \ldots, X_n)'$, $Y = (Y_1, \ldots, Y_n)'$, and A is any orthonormal matrix with the first row $(d_1/\sqrt{b_n}, \ldots, d_n/\sqrt{b_n})$. The following facts are easily verified: $\sum_{i=1}^n Y_i^2 = \sum_{i=1}^n X_i^2$, $Y_1 = \sqrt{b_n}\bar{X}_u$, $\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n X_i^2 - Y_1^2 = (n-1)S_u^2/\alpha_n$, and $\sum_{i=1}^n (X_i - d_i\theta)^2 = b_n(\bar{X}_u - \theta)^2 + \sum_{i=1}^n (X_i - d_i\bar{X}_u)^2 = (Y_1 - \sqrt{b_n}\theta)^2 + Y_2^2 + \cdots + Y_n^2$. As the Jacobian of the transformation is 1, the joint pdf of $Y_1, \ldots, Y_n$ is given by

$$f(y_1, y_2, \ldots, y_n \mid \theta, d_1, \ldots, d_n) = (\theta\sqrt{2\pi})^{-n} \times \exp\left\{-\frac{1}{2\theta^2}\left[\left(y_1 - \sqrt{b_n}\theta\right)^2 + y_2^2 + \cdots + y_n^2\right]\right\},$$

which implies that $Y_1, \ldots, Y_n$ are independent normal random variables. The conclusions (a), (b), (c) now follow. ■

Proof of Theorem 1 The proof is based on the asymptotic expansion of the maximum-likelihood estimator. Define $T_1 = \sum_{i=1}^n d_i X_i$, $T_2 = \sum_{i=1}^n X_i^2$, $\lambda_1 = b_n\theta$, and $\lambda_2 = (n+b_n)\theta^2$. The following results are easy to establish: $E(T_1) = \lambda_1$, $E(T_2) = \lambda_2$, $E(T_1-\lambda_1)^{2k} = (2k)!b_n^k\theta^{2k}/k!2^k$, $E(T_2-\lambda_2)^2 = 2(n+2b_n)\theta^4$, $E(T_2-\lambda_2)^{2k} = O(n^k)$, $E(T_1-\lambda_1)(T_2-\lambda_2) = 2b_n\theta^3$. Note that the MLE can be expressed as $\hat{\theta}_{\text{MLE}} = g(T_1,T_2) = 1/2n(-T_1 + \sqrt{T_1^2 + 4nT_2})$. Then, the first- and second-order expansions of the MLE are given by

$$g(T_1,T_2) = g(\lambda_1,\lambda_2) + (T_1-\lambda_1)g_{10} + (T_2-\lambda_2)g_{01} + R_{n1}(\tilde{\lambda}_1,\tilde{\lambda}_2)$$

and

$$\begin{aligned} g(T_1,T_2) = g(\lambda_1,\lambda_2) &+ (T_1-\lambda_1)g_{10} + (T_2-\lambda_2)g_{01} + 1/2[(T_1-\lambda_1)^2 g_{20} \\ &+ 2(T_1-\lambda_1)(T_2-\lambda_2)g_{11} + (T_2-\lambda_2)^2 g_{02}] + R_{n2}(\lambda_1^*,\lambda_2^*), \end{aligned}$$

where constants $\tilde{\lambda}_1, \lambda_1^*$ are between T_1 and λ_1 and constants $\tilde{\lambda}_2, \lambda_2^*$ are between T_2 and λ_2, and

$$g_{ij}(\lambda_1,\lambda_2) = \left.\frac{\partial^{i+j} g(u,v)}{\partial u^i \partial v^j}\right|_{\lambda_1,\lambda_2}.$$

Further, as $b_n^{-1}T_1 \xrightarrow{p} \theta$ and $(n+b_n)^{-1}T_2 \xrightarrow{p} \theta^2$, it follows that $b_n^{-1}\tilde{\lambda}_1 \xrightarrow{p} \theta$, $b_n^{-1}\lambda_1^* \xrightarrow{p} \theta$, $(n+b_n)^{-1}\tilde{\lambda}_2 \xrightarrow{p} \theta^2$, and $(n+b_n)^{-1}\lambda_2^* \xrightarrow{p} \theta^2$. Using these and the standard convergence arguments, it can be shown that $E(|nR_{n2}(\lambda_1^*,\lambda_2^*)|) = O(n^{-1/2})$ and $\sqrt{n}R_{n1}(\tilde{\lambda}_1,\tilde{\lambda}_2) = O_p(n^{-1/2})$.

Then, as $g(\lambda_1,\lambda_2) = \theta$, part ($a$) follows from the second-order expansion of $g(T_1,T_2)$ upon substituting the values of g_{20}, g_{11}, g_{02}, expected values, and taking limits as $n \to \infty$. Similarly, part (b) follows from the first-order expansion of $g(T_1,T_2)$.

In order to show parts (c) and (d), we observe that

$$\begin{aligned} \sqrt{n}(T_1-\lambda_1)g_{10} + \sqrt{n}(T_2-\lambda_2)g_{01} &= \left[\frac{\theta\sqrt{nb_n}}{2n+b_n}\right] Z_n + \left[\frac{\theta\sqrt{2n(n-1)}}{2n+b_n}\right] W_n \\ &\quad + \left[\frac{\theta\sqrt{n}}{2n+b_n}\right] Z_n^2 + \sqrt{n}R_n - \left[\frac{\theta\sqrt{n}}{2n+b_n}\right], \end{aligned}$$

where $Z_n = \sqrt{b_n}(\bar{X}_u - \theta)/\theta \sim N(0,1)$ and $W_n = (\sum_{i=1}^n (X_i - d_i\bar{X}_u)^2 - (n-1)\theta^2)/\sqrt{2(n-1)\theta^4}$ are independent random variables. Also, as $n \to \infty$, we have the following: $W_n \xrightarrow{\mathcal{D}} N(0,1)$ (by central limit theorem), the third and fourth terms $\xrightarrow{p} 0$, and the last term $\to 0$. As a consequence of the above arguments, $\sqrt{n}(T_1-\lambda_1)g_{10} + \sqrt{n}(T_2-\lambda_2)g_{01} \xrightarrow{\mathcal{D}} [\theta d/(2+d^2)]N_1 + [\theta\sqrt{2}/(2+d^2)]N_2$, where N_1 and N_2 are independent $N(0,\,1)$ random variables. Hence, the right-hand side is distributed as $N(0,\theta^2(2+d^2)^{-1})$.

Further, for part (e), we decompose $\sqrt{n}(\hat{\theta}_{\text{LU}} - \theta)$ into

$$\left[\frac{\theta a_n}{\sqrt{b_n/n}}\right] Z_n + \left[\frac{(1-a_n)\sqrt{2n(n-1)}\alpha_n^2\theta^2}{(n-1)(S_u+\theta)}\right] W_n + \left[\frac{\theta^2(1-a_n)(\alpha_n^2-1)\sqrt{n}}{S_u+\theta}\right],$$

where Z_n and W_n are same as defined above. As $n \to \infty$, the third term is $O_p(n^{-1/2})$ (because $S_u \xrightarrow{p} \theta$). Hence, $\sqrt{n}(\hat{\theta}_{\text{LU}} - \theta) \xrightarrow{\mathcal{D}} [\theta d/(2+d^2)]N_1 + [\theta\sqrt{2}/(2+d^2)]N_2 \overset{\mathcal{D}}{\approx} N(0,\theta^2(2+d^2)^{-1})$.

For parts (f) and (g), we observe that $\hat{\theta}_{\mathrm{LMMS}} - \hat{\theta}_{\mathrm{LU}} = -v_1 v_2 (v_1 + v_2 + v_1 v_2)^{-1} \hat{\theta}_{\mathrm{LU}}$. As $v_1 v_2 (v_1 + v_2 + v_1 v_2)^{-1} = O((2n + b_n)^{-1})$, it follows that $\lim_{n\to\infty}(2n + b_n)\ E(\hat{\theta}_{\mathrm{LMMS}} - \theta) = -\theta$. Also, $\sqrt{n}(\hat{\theta}_{\mathrm{LMMS}} - \hat{\theta}_{\mathrm{LU}}) = \sqrt{n}(\hat{\theta}_{\mathrm{LMMS}} - \theta) - \sqrt{n}(\hat{\theta}_{\mathrm{LU}} - \theta) = O_p(n^{-1/2})$ implies that $\sqrt{n}(\hat{\theta}_{\mathrm{LMMS}} - \theta) \overset{\mathcal{D}}{\approx} \sqrt{n}(\hat{\theta}_{\mathrm{LU}} - \theta) \overset{\mathcal{D}}{\to} N(0, \theta^2(2 + d^2)^{-1})$.

To show part (h), we use the inequality (4.12) in Gleser and Healy [2]. Applying the arguments similar to those used to derive (4.12) by these authors, it can be shown that $\hat{\theta}_{\mathrm{B}}^{-} \le \hat{\theta}_{\mathrm{B}} \le \hat{\theta}_{\mathrm{B}}^{+}$, where $\hat{\theta}_{\mathrm{B}}^{-}$ is given by

$$\hat{\theta}_{\mathrm{B}}^{-} = \frac{\sqrt{T_2}}{2}\left(\frac{-b^*}{n+r-1} + \sqrt{\left(\frac{b^*}{n+r-1}\right)^2 + \frac{4}{n+r-1}}\right)$$

and $\hat{\theta}_{\mathrm{B}}^{+}$ is obtained from $\hat{\theta}_{\mathrm{B}}^{-}$ by replacing $r-1$ by $r-2$. Then, it is not difficult to show that $\hat{\theta}_{\mathrm{B}}^{+} - \hat{\theta}_{\mathrm{B}}^{-} = O_p(n^{-1})$. As $\hat{\theta}_{\mathrm{B}}^{-}$ reduces to $\hat{\theta}_{\mathrm{MLE}}$ when $r = 1$ and $w = 0$, the asymptotic properties of $\hat{\theta}_{\mathrm{B}}^{-}$ can be derived in exactly the same way as those of $\hat{\theta}_{\mathrm{MLE}}$. Finally, as $\hat{\theta}_{\mathrm{P}}$ is a special case of $\hat{\theta}_{\mathrm{B}}$, with $r = 2$ and $w = 0$, the technique used for $\hat{\theta}_{\mathrm{B}}$ can be used to show $\hat{\theta}_{\mathrm{P}} = \hat{\theta}_{\mathrm{MLE}} + O_p(n^{-1})$. ■