Journal of Multivariate Analysis 100 (2009) 1398–1411

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Multivariate limited translation hierarchical Bayes estimators

Malay Ghosh [a,*], Georgios Papageorgiou [a,*], Janet Forrester [b]

[a] University of Florida, Department of Statistics, 103 Griffin/Floyd Hall - P.O. Box 118545, Gainesville, FL 32611-8545, United States

[b] Tufts University, School of Medicine, 136 Harrison Avenue, Stearns 203, Boston, MA 02111-1817, United States

ARTICLE INFO

Article history:
Received 14 September 2007
Available online 25 December 2008

AMS subject classifications:
62Hxx
62H12

Keywords:
Bayes risk
Frequentist risk
g-prior
Influence function
Relevance function

ABSTRACT

Based on the notion of predictive influence functions, the paper develops multivariate limited translation hierarchical Bayes estimators of the normal mean vector which serve as a compromise between the hierarchical Bayes and maximum likelihood estimators. The paper demonstrates the superiority of the limited translation estimators over the usual hierarchical Bayes estimators in terms of the frequentist risks when the true parameter to be estimated departs widely from the grand average of all the parameters.

1. Introduction

Bayesian methods are used quite extensively in the theory and practice of statistics. One appealing feature of the Bayesian procedure is that the posterior inference, based on an approximately elicited prior along with the likelihood, usually leads to more efficient inference than any classical frequentist inferential procedure. This is intuitively true since one is utilizing two sources of information rather than one as in classical analysis. Such elicitation of priors has been possible in the presence of extensive historical data. A very important application is in the medical area where constant updating of information leads to successful prior elicitation. Also, people in Educational Testing Service (ETS) have been using Bayesian methods regularly because they have in store a vast number of test scores from multiple tests that they administered. In particular, IQ test scores are better calibrated when one uses a prior along with the sample data.

We are considering in this paper a hierarchical Bayesian (HB) scenario where the main objective is simultaneous estimation of several multivariate normal means. The HB estimators shrink the individual maximum likelihood estimators (MLEs) towards their grand average (see e.g. Lindley and Smith [1]). If the parameters are 'exchangeable', then the componentwise HB estimators will perform very well both in terms of their Bayes and frequentist risks in comparison with the corresponding MLEs. However, if a certain parameter departs widely from the grand mean, then the corresponding HB estimator may perform poorly in terms of its frequentist risk.

Robust Bayesian methods have been proposed to guard against problems of this type. One such procedure, first introduced by Efron and Morris [2–4], and referred to as 'limited translation estimators', is the subject matter of this paper. The limited translation estimators are compromises between the HB and the ML estimators that slightly increase the Bayes risk but guard against large frequentist risks.

One of the virtues of the limited translation estimators is that they do not fare too badly in their Bayes risk performance, compared to the regular HB estimators, even if the assumed exchangeability among parameters holds true. While

* Corresponding author.
E-mail addresses: ghoshm@stat.ufl.edu (M. Ghosh), gpapageo@stat.ufl.edu (G. Papageorgiou), janet.forrester@tufts.edu (J. Forrester).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2008.12.010

considering frequentist risks, the limited translation estimators do not perform too badly relative to the regular HB estimators if an individual parameter is close to the mean. On the other hand, if the parameter to be estimated is far from the mean, then the limited translation estimators do perform much better than the regular HB estimators.

Efron and Morris [2] developed limited translation estimators for the univariate normal case. The objective here is to develop limited translation estimators for the multivariate normal case. Moreover, the present paper has brought in the notion of predictive influence functions as introduced in Johnson and Geisser [5–7] in motivating the limited translation estimators.

The organization of the remaining sections is as follows. In Section 2 of this paper we review some of the properties of the HB estimators. In Section 3 we introduce the notion of influence functions which is used to develop the limited translation estimators in Section 4. In Section 5 we evaluate the Bayes risk performance under the assumed prior while in Section 6 we evaluate the frequentist risk of these estimators. In Section 6 we consider a specific form of prior, the g-prior originally introduced by Zellner [8]. In Section 7 we apply the proposed inferential procedure to estimate the 'long term average' vitamin intakes of HIV-positive subjects. Some final remarks are made in Section 8. Some of the long algebraic derivations are provided in the Appendix.

2. Hierarchical Bayes estimators

Consider the hierarchical Bayesian model where $\boldsymbol{X}_i|\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n,\boldsymbol{\mu} \overset{\text{ind}}{\sim} N_p(\boldsymbol{\theta}_i,\boldsymbol{\Sigma})$, $i=1,\ldots,n$, where $\boldsymbol{\Sigma}$ (p.d.) is known. Also, let $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n|\boldsymbol{\mu} \overset{\text{iid}}{\sim} \xi \equiv N_p(\boldsymbol{\mu},\boldsymbol{A})$, where $\boldsymbol{A}$ is known. Finally, suppose that $\boldsymbol{\mu}\sim \text{uniform}(\Re^p)$.

Then the HB estimator of $\boldsymbol{\theta}_i$ is

$$\tilde{\boldsymbol{\theta}}_i^B = (\boldsymbol{I}_p-\boldsymbol{B})\boldsymbol{X}_i + \boldsymbol{B}\bar{\boldsymbol{X}}_n = \boldsymbol{X}_i - \boldsymbol{B}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n), \tag{2.1}$$

where $\bar{\boldsymbol{X}}_n = n^{-1}\sum_{i=}^n \boldsymbol{X}_i$ and $\boldsymbol{B}=\boldsymbol{\Sigma}(\boldsymbol{A}+\boldsymbol{\Sigma})^{-1}$. The same estimator can also be given an empirical Bayes (EB) interpretation.

The HB estimator shrinks the MLE $\boldsymbol{X}_i$ of $\boldsymbol{\theta}_i$ towards the grand mean, $\bar{\boldsymbol{X}}_n$. In doing so it attains a lower Bayes risk than the MLE, under the assumed prior. However, the HB estimator introduces high frequentist risk when $\boldsymbol{\theta}_i$ is far from the grand mean, $\bar{\boldsymbol{\theta}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{\theta}_i$. On the other hand, the MLE has constant risk.

We now quantify the statements of the previous paragraph concerning the Bayes and frequentist risks under the matrix loss $\boldsymbol{L}_1(\boldsymbol{\theta}_i,\boldsymbol{a}_i)=(\boldsymbol{\theta}_i-\boldsymbol{a}_i)(\boldsymbol{\theta}_i-\boldsymbol{a}_i)^{\mathrm{T}}$. First, the Bayes risk of $\tilde{\boldsymbol{\theta}}_i^B$ under prior ξ, denoted by $\boldsymbol{r}_1(\xi,\tilde{\boldsymbol{\theta}}_i^B)$, is calculated on the basis of the joint distributions of $\boldsymbol{X}_i$ and $\boldsymbol{\theta}_i$, $i=1,\ldots,n$. Specifically, we have

$$\begin{aligned}\boldsymbol{r}_1(\xi,\tilde{\boldsymbol{\theta}}_i^B) &= E\{(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}_i^B)(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}_i^B)^{\mathrm{T}}\}\\ &= E\{(\boldsymbol{\theta}_i-\hat{\boldsymbol{\theta}}_i^B+\hat{\boldsymbol{\theta}}_i^B-\tilde{\boldsymbol{\theta}}_i^B)(\boldsymbol{\theta}_i-\hat{\boldsymbol{\theta}}_i^B+\hat{\boldsymbol{\theta}}_i^B-\tilde{\boldsymbol{\theta}}_i^B)^{\mathrm{T}}\}\\ &= E\{(\boldsymbol{\theta}_i-\hat{\boldsymbol{\theta}}_i^B)(\boldsymbol{\theta}_i-\hat{\boldsymbol{\theta}}_i^B)^{\mathrm{T}}\} + E\{(\hat{\boldsymbol{\theta}}_i^B-\tilde{\boldsymbol{\theta}}_i^B)(\hat{\boldsymbol{\theta}}_i^B-\tilde{\boldsymbol{\theta}}_i^B)^{\mathrm{T}}\}\\ &= (\boldsymbol{I}_p-\boldsymbol{B})\boldsymbol{\Sigma} + \boldsymbol{B}E\{(\bar{X}_n-\boldsymbol{\mu})(\bar{X}_n-\boldsymbol{\mu})^{\mathrm{T}}\}\boldsymbol{B}^{\mathrm{T}}\\ &= \{\boldsymbol{I}_p-(1-n^{-1})\boldsymbol{B}\}\boldsymbol{\Sigma}.\end{aligned} \tag{2.2}$$

The Bayes risk of $\tilde{\boldsymbol{\theta}}_i^B$ is less than that of the MLE, $\boldsymbol{r}_1(\xi,\boldsymbol{X}_i)=\boldsymbol{\Sigma}$. This can be seen by noting that $\boldsymbol{B}\boldsymbol{\Sigma}=\boldsymbol{\Sigma}(\boldsymbol{A}+\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}$ is positive definite.

Further, the frequentist risk, $\boldsymbol{R}_1(\boldsymbol{\theta}_i,\tilde{\boldsymbol{\theta}}_i^B)$, is calculated on the basis of the conditional distribution of $\boldsymbol{X}_i$ given $\boldsymbol{\theta}_i$, $i=1,\ldots,n$, assuming that $\boldsymbol{\theta}^{\mathrm{T}}=(\boldsymbol{\theta}_1^{\mathrm{T}},\ldots,\boldsymbol{\theta}_n^{\mathrm{T}})$ is fixed. We have

$$\begin{aligned}\boldsymbol{R}_1(\boldsymbol{\theta}_i,\tilde{\boldsymbol{\theta}}_i^B) &= E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}_i^B)(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}_i^B)^{\mathrm{T}}\}\\ &= E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i-\boldsymbol{X}_i)(\boldsymbol{\theta}_i-\boldsymbol{X}_i)^{\mathrm{T}}\} + \boldsymbol{B}E_{\boldsymbol{\theta}}\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\}\boldsymbol{B}^{\mathrm{T}}\\ &\quad + E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i-\boldsymbol{X}_i)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{B}E_{\boldsymbol{\theta}}\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{\theta}_i-\boldsymbol{X}_i)^{\mathrm{T}}\}.\end{aligned} \tag{2.3}$$

Now, $\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n \sim N(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n,(1-1/n)\boldsymbol{\Sigma})$. Thus,

$$E_{\boldsymbol{\theta}}\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\} = \boldsymbol{\Sigma}(1-1/n) + (\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}. \tag{2.4}$$

Also,

$$\begin{aligned}E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i-\boldsymbol{X}_i)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\} &= \boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\mathrm{T}} - n^{-1}\sum_{k=1}^n \boldsymbol{\theta}_i\boldsymbol{\theta}_k^{\mathrm{T}} - \boldsymbol{\theta}_i\boldsymbol{\theta}_i^{\mathrm{T}} - \boldsymbol{\Sigma} + n^{-1}\left(\sum_{k=1}^n \boldsymbol{\theta}_i\boldsymbol{\theta}_k^{\mathrm{T}} + \boldsymbol{\Sigma}\right)\\ &= -(1-1/n)\boldsymbol{\Sigma}.\end{aligned} \tag{2.5}$$

Combining (2.3)–(2.5), we obtain

$$\boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_i^B) = \boldsymbol{\Sigma} + 2(1/n-1)\boldsymbol{B}\boldsymbol{\Sigma} + \boldsymbol{B}\left[\boldsymbol{\Sigma}(1-1/n) + (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}\right]\boldsymbol{B}^{\mathrm{T}}. \tag{2.6}$$

Clearly, when $\boldsymbol{\theta}_i$ is far from $\bar{\boldsymbol{\theta}}_n$, the frequentist risk associated with the HB estimator can be quite high. On the other hand $\boldsymbol{X}_i$ has minimax risk equal to $\boldsymbol{R}_1(\boldsymbol{\theta}_i, \boldsymbol{X}_i) = \boldsymbol{\Sigma}$ for all $\boldsymbol{\theta}$ and thus its Bayes risk is also $\boldsymbol{\Sigma}$ which, however, is bigger than that of the HB estimator, $\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_i^B) = \{\boldsymbol{I}_p - (1-n^{-1})\boldsymbol{B}\}\boldsymbol{\Sigma}$, if the assumed prior ξ is the true prior.

In order to combine the good properties of the HB estimators with those of the MLEs, we develop limited translation HB estimators. To this end, we first introduce the concept of the influence functions.

3. Influence functions

The model assumptions are the same as in the previous section. We find the influence of observations $\boldsymbol{X}_i$, $i = 1, \ldots, n$, on the posterior distribution of $\boldsymbol{\mu}$. The influence is measured using the general divergence formula introduced by Cressie and Read [9].

Let f_1 and f_2 denote two density functions. Then the general divergence measure is given by

$$D_\lambda(f_1, f_2) = \lambda^{-1}(\lambda+1)^{-1}E_{f_1}\left\{(f_1/f_2)^\lambda - 1\right\}. \tag{3.1}$$

Here, f_1 and f_2 denote the posterior densities of $\boldsymbol{\mu}$ given $\boldsymbol{X} = (\boldsymbol{X}_1^{\mathrm{T}}, \ldots, \boldsymbol{X}_n^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{X}^{(-i)} = (\boldsymbol{X}_1^{\mathrm{T}}, \ldots, \boldsymbol{X}_{i-1}^{\mathrm{T}}, \boldsymbol{X}_{i+1}^{\mathrm{T}}, \ldots, \boldsymbol{X}_n^{\mathrm{T}})^{\mathrm{T}}$ respectively. It is easy to show that

$$\boldsymbol{\mu}|\boldsymbol{X} \sim f_1 \equiv N_p(\bar{\boldsymbol{X}}_n, n^{-1}(\boldsymbol{\Sigma} + \boldsymbol{A})), \tag{3.2}$$

$$\boldsymbol{\mu}|\boldsymbol{X}^{(-i)} \sim f_2 \equiv N_p(\bar{\boldsymbol{X}}_{n-1}^{(-i)}, (n-1)^{-1}(\boldsymbol{\Sigma} + \boldsymbol{A})), \tag{3.3}$$

where $\bar{\boldsymbol{X}}_{n-1}^{(-i)}$ is the average of all the random vectors except the ith one.

For $f_1 = N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $f_2 = N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, the $D_\lambda(f_1, f_2)$ divergence measure takes the following form

$$\begin{aligned} D_\lambda(f_1, f_2) = {} & \lambda^{-1}(\lambda+1)^{-1}\Bigg[|\boldsymbol{\Sigma}_2|^{\frac{\lambda+1}{2}}|\boldsymbol{\Sigma}_1|^{-\frac{\lambda}{2}}|(1+\lambda)\boldsymbol{\Sigma}_2 - \lambda\boldsymbol{\Sigma}_1|^{-\frac{1}{2}} \\ & \times \exp\left\{\frac{\lambda(\lambda+1)}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}[(1+\lambda)\boldsymbol{\Sigma}_2 - \lambda\boldsymbol{\Sigma}_1]^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right\} - 1\Bigg]. \end{aligned} \tag{3.4}$$

If the variance–covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are known, which is the case here, the divergence measure is a one-to-one function with $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathrm{T}}\{(1+\lambda)\boldsymbol{\Sigma}_2 - \lambda\boldsymbol{\Sigma}_1\}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

For the special case where the densities f_1 and f_2 are the ones given in (3.2) and (3.3), $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = (n-1)^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)$ and $(1+\lambda)\boldsymbol{\Sigma}_2 - \lambda\boldsymbol{\Sigma}_1 = (\boldsymbol{\Sigma} + \boldsymbol{A})(n+\lambda)n^{-1}(n-1)^{-1}$. It is now easy to see that the divergence measure is a one-to-one function with

$$(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)^{\mathrm{T}}\frac{n(\boldsymbol{\Sigma} + \boldsymbol{A})^{-1}}{(n-1)(n+\lambda)}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n). \tag{3.5}$$

Define $\boldsymbol{D} \equiv \mathrm{var}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n) = (1-1/n)(\boldsymbol{\Sigma} + \boldsymbol{A})$. Then, the expression in (3.5) can be written as

$$(n+\lambda)^{-1}\{\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\}^{\mathrm{T}}\{\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\}, \tag{3.6}$$

which is a quadratic form in the standardized residuals $\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)$. It can be recognized as $(n+\lambda)^{-1}$ times the Mahalanobis [10] distance between $\boldsymbol{X}_i$ and $\bar{\boldsymbol{X}}_n$. Based on this result we will obtain some robust Bayesian estimators in the next section.

4. Limited translation hierarchical Bayes estimators

The limited translation HB estimators are akin to the HB estimators but at the same time they put a limit to the amount of shrinkage of the MLEs towards the grand mean. The goal of these estimators is to maintain low Bayes risk and at the same time put a bound to the frequentist risk.

A modification of the HB estimators will give us the limited translation estimator. Since the influence of the random vectors $\boldsymbol{X}_i$, $i = 1, \ldots, n$, depends on their standardized distance from $\bar{\boldsymbol{X}}_n$, in the HB estimator we want to control the standardized distance of $\boldsymbol{X}_i$ to $\bar{\boldsymbol{X}}_n$. We thus write $\tilde{\boldsymbol{\theta}}_i^B = \boldsymbol{X}_i - \boldsymbol{B}\boldsymbol{D}^{\frac{1}{2}}\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)$.

Definition. For the ith vector $\boldsymbol{\theta}_i$, we define the limited translation HB estimator of maximum translation c as

$$\tilde{\boldsymbol{\theta}}_{c,i}^{LB} = \boldsymbol{X}_i - \boldsymbol{B}\boldsymbol{D}^{\frac{1}{2}} h_c\{\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\}, \tag{4.1}$$

where

$$h_c(\boldsymbol{z}) = \boldsymbol{z}\min(1, c/\|\boldsymbol{z}\|), \quad \boldsymbol{z} \in \Re^p, \tag{4.2}$$

is the multidimensional Huber function, Hampel et al. [11], and c is a known constant.

The proposed estimator can equivalently be written as a weighted average of the MLE and HB estimator since

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_{c,i}^{LB} &= \boldsymbol{X}_i - \boldsymbol{B}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2) \\ &= \boldsymbol{X}_i\{1 - \rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2)\} + \tilde{\boldsymbol{\theta}}_i^{B}\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2),\end{aligned} \tag{4.3}$$

where $\rho_c(u) = \min(1, c/\sqrt{u})$ is termed the relevance function, Efron and Morris [2,3]. It is similar to the Huber function $h_c(u)$ (Huber [12], p. 13, Hampel et al. [11], p. 104) in the sense that $h_c(u) = u\rho_c(u)$.

The limited translation HB estimator follows the HB estimator as closely as possible subject to the constraint that the distance of the observed $\boldsymbol{X}_i$ to the observed mean $\bar{\boldsymbol{X}}_n$, as measured by the standardized norm $\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|$, does not exceed a certain value, c say. When this distance takes on a value bigger than c, the relevance function takes on a value smaller than one, and by the second line of (4.3) we see that the limited translation estimator gives the MLE bigger weight at the expense of the weight of the HB estimator. The idea is that as the distance of $\boldsymbol{X}_i$ to $\bar{\boldsymbol{X}}_n$ increases, the less relevant the HB estimator is considered to be for estimation of the corresponding $\boldsymbol{\theta}_i$. In the next sections we show that this provides the statistician with protection against large values of the frequentist risk, while slightly increasing the Bayes risk.

5. Bayes risk of limited translation estimators

First, it is of interest to know how well the estimator $\tilde{\boldsymbol{\theta}}_{c,i}^{LB}$ performs assuming that the normal prior $N_p(\boldsymbol{\mu}, \boldsymbol{A})$ is the true one. We thus calculate its Bayes risk, $\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) = E\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\}$. The calculations for the most part do not depend on the choice of the relevance function $\rho_c(.)$. The following theorem shows that the Bayes risk of the limited translation estimator can be written as a weighted average of the Bayes risks of the ML and HB estimators.

Theorem 5.1. *For any relevance function $\rho_c(.)$ we have*

$$\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) = \boldsymbol{r}_1(\xi, \boldsymbol{X}_i)(1 - s_c) + \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_i^{B})s_c, \tag{5.1}$$

where $1 - s_c = E\{1 - \rho_c(U)\}^2$ with $U \sim \chi^2_{p+2}$.

Proof. We write

$$\begin{aligned}\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) &= E\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\} \\ &= E\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B} + \tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B} + \tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\} \\ &= \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}_i^{B}) + E\{(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\} + E\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B})(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\} + E\{(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B})^{\mathrm{T}}\}.\end{aligned} \tag{5.2}$$

Noting that $\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB} = \boldsymbol{B}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\{\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2) - 1\}$, from the independence of $\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n$ and $\bar{\boldsymbol{X}}_n$, and the fact that $E(\bar{\boldsymbol{X}}_n) = \boldsymbol{\mu}$, follows that

$$\begin{aligned}E\{(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B})^{\mathrm{T}}\} &= E\{E(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_i^{B})^{\mathrm{T}}|\boldsymbol{X}_i\} \\ &= \boldsymbol{B}E[\{\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2) - 1\}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)(\boldsymbol{\mu} - \bar{\boldsymbol{X}}_n)^{\mathrm{T}}]\boldsymbol{B}^{\mathrm{T}} \\ &= \boldsymbol{0}.\end{aligned} \tag{5.3}$$

Next,

$$\begin{aligned}E\{(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})(\tilde{\boldsymbol{\theta}}_i^{B} - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\} &= \boldsymbol{B}E\{(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)^{\mathrm{T}}[\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2) - 1]^2\}\boldsymbol{B}^{\mathrm{T}} \\ &= (1 - 1/n)\boldsymbol{\Sigma}(\boldsymbol{A} + \boldsymbol{\Sigma})^{-\frac{1}{2}}E\{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}[\rho_c(\|\boldsymbol{Z}\|^2) - 1]^2\}(\boldsymbol{A} + \boldsymbol{\Sigma})^{-\frac{1}{2}}\boldsymbol{\Sigma},\end{aligned} \tag{5.4}$$

where $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{I}_p)$.

The following lemma simplifies the calculation of the Bayes risk.

Lemma 5.2. *Consider the random vector $\boldsymbol{Y} \sim N_p(\mathbf{0}, \tau^2 \boldsymbol{I}_p)$. Then the random scalar $\|\boldsymbol{Y}\|^2$ and the random matrix $\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}/\|\boldsymbol{Y}\|^2$ are independently distributed.*

Proof. First, $\|\boldsymbol{Y}\|^2$ is complete and sufficient for τ^2. Noting that $\tau^{-1}\boldsymbol{Y} \sim N_p(\mathbf{0}, \boldsymbol{I}_p)$, the statistic $\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}/\|\boldsymbol{Y}\|^2$ is ancillary. The independence of $\|\boldsymbol{Y}\|^2$ and $\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}/\|\boldsymbol{Y}\|^2$ follows from the well-known theorem of Basu.

We now continue with the calculation of the expectation that appears in the last line of (5.4). By Lemma 5.2

$$E\{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}[1-\rho_c(\|\boldsymbol{Z}\|^2)]^2\} = E\left\{\frac{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}}{\|\boldsymbol{Z}\|^2}\|\boldsymbol{Z}\|^2[1-\rho_c(\|\boldsymbol{Z}\|^2)]^2\right\}$$
$$= E\left(\frac{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}}{\|\boldsymbol{Z}\|^2}\right)E\{\|\boldsymbol{Z}\|^2[1-\rho_c(\|\boldsymbol{Z}\|^2)]^2\}. \tag{5.5}$$

Again by Lemma 5.2

$$E(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}) = E\left(\frac{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}}{\|\boldsymbol{Z}\|^2}\|\boldsymbol{Z}\|^2\right) = E\left(\frac{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}}{\|\boldsymbol{Z}\|^2}\right)E\left(\|\boldsymbol{Z}\|^2\right), \tag{5.6}$$

and thus

$$E\left(\frac{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}}{\|\boldsymbol{Z}\|^2}\right) = \frac{E(\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}})}{E\left(\|\boldsymbol{Z}\|^2\right)} = p^{-1}\boldsymbol{I}_p. \tag{5.7}$$

Since $\|\boldsymbol{Z}\|^2 \sim \chi^2_p$,

$$E\{\|\boldsymbol{Z}\|^2[1-\rho_c(\|\boldsymbol{Z}\|^2)]^2\} = E\{Y[1-\rho_c(Y)]^2\}$$
$$= \int_0^\infty [1-\rho_c(y)]^2 \exp\left(-\frac{y}{2}\right)\frac{py^{\frac{p+2}{2}-1}}{2^{\frac{p+2}{2}}\Gamma\left(\frac{p+2}{2}\right)}\mathrm{d}y = pE\{[1-\rho_c(U)]^2\}, \tag{5.8}$$

where $U \sim \chi^2_{p+2}$. From (5.5), (5.7) and (5.8) it follows that

$$E\{\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}[1-\rho_c(\|\boldsymbol{Z}\|^2)]^2\} = E\{[1-\rho_c(U)]^2\}\boldsymbol{I}_p. \tag{5.9}$$

Hence, from (5.2), (5.3), (5.4) and (5.9) it follows that

$$\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{LB}_{c,i}) = \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i}) + E\{[1-\rho_c(U)]^2\}(1-1/n)\boldsymbol{B}\boldsymbol{\Sigma}$$
$$= \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i}) + (1-s_c)(1-1/n)\boldsymbol{B}\boldsymbol{\Sigma}, \tag{5.10}$$

where $1-s_c = E\{[1-\rho_c(U)]^2\}$. The second of the two terms can be thought of as the price in terms of increased Bayes risk for, as we will show in the following section, limiting the frequentist risk of the HB estimator. Alternatively we can write

$$\boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{LB}_{c,i}) = \boldsymbol{\Sigma} - \boldsymbol{B}\boldsymbol{\Sigma}s_c(1-1/n)$$
$$= \boldsymbol{\Sigma}(1-s_c) + \{\boldsymbol{\Sigma} - (1-1/n)\boldsymbol{B}\boldsymbol{\Sigma}\}s_c$$
$$= \boldsymbol{r}_1(\xi, \boldsymbol{X}_i)(1-s_c) + \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i})s_c, \tag{5.11}$$

thus completing the proof of the theorem. □ □

Definition. For an estimator $\hat{\boldsymbol{\theta}}_i$ of $\boldsymbol{\theta}_i$ the generalized relative savings loss of $\hat{\boldsymbol{\theta}}$ with respect to $\boldsymbol{X}_i$ is defined as

$$GRSL(\hat{\boldsymbol{\theta}}_i; \boldsymbol{X}_i) = [\boldsymbol{r}_1(\xi, \boldsymbol{X}_i) - \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i})]^{-1}[\boldsymbol{r}_1(\xi, \hat{\boldsymbol{\theta}}_i) - \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i})]. \tag{5.12}$$

The term $\boldsymbol{r}_1(\xi, \boldsymbol{X}_i) - \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i})$ is the savings, in Bayes risk sense, that occurs when using the HB estimator instead of the MLE, while $\boldsymbol{r}_1(\xi, \hat{\boldsymbol{\theta}}_i) - \boldsymbol{r}_1(\xi, \tilde{\boldsymbol{\theta}}^{B}_{i})$ is the loss that occurs when one uses $\hat{\boldsymbol{\theta}}_i$ instead of the HB estimator.

From Theorem 5.1 it follows that the generalized relative savings loss of $\tilde{\boldsymbol{\theta}}^{LB}_{c,i}$ is given by

$$GRSL(\tilde{\boldsymbol{\theta}}^{LB}_{c,i}; \boldsymbol{X}_i) = (1-s_c)\boldsymbol{I}_p, \tag{5.13}$$

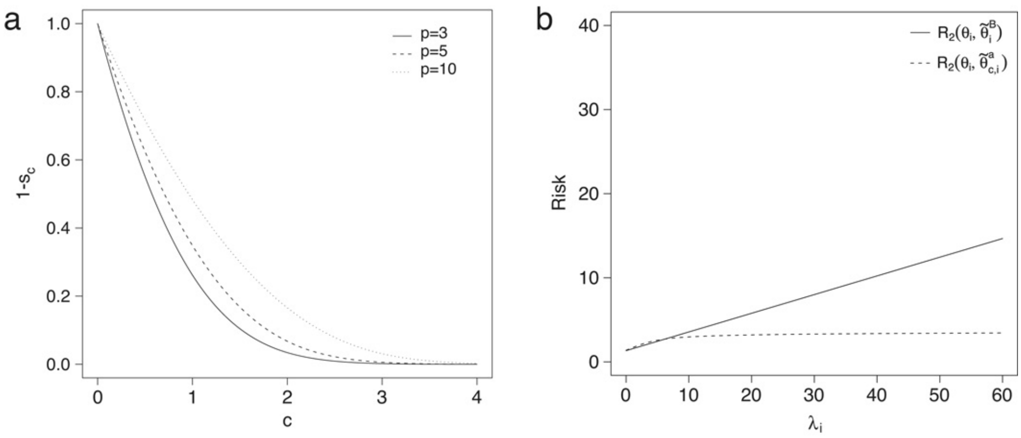

Fig. 1. (a) Plot of $1 - s_c$ as a function of c and (b) Risk as function of the non-centrality parameter.

and for the special case where $\rho_c(u) = \min(1, c/\sqrt{u})$,

$$\begin{aligned} 1 - s_c &= E\{1 - \rho_c(U)\}^2 = E[(1 - c/\sqrt{U})^2 I(U > c^2)] \\ &= E[I(U > c^2)] - 2cE[U^{-\frac{1}{2}} I(U > c^2)] + c^2 E[U^{-1} I(U > c^2)] \\ &= P(\chi^2_{p+2} > c^2) - c\sqrt{2}\Gamma\left(\frac{p+1}{2}\right)\Gamma^{-1}\left(\frac{p+2}{2}\right) P(\chi^2_{p+1} > c^2) + p^{-1}c^2 P(\chi^2_p > c^2), \end{aligned} \tag{5.14}$$

which, for fixed p, depends only on c and it is independent of the model parameters.

The Bayes risk of the limited translation estimator is a weighted average of the Bayes risk of the HB and the ML estimators, the weights being s_c and $1 - s_c$ respectively. This causes a loss in the generalized savings of $(1 - s_c)\boldsymbol{I}_p$. However, the weight of the Bayes risk of the MLE, $1 - s_c$, for fixed p, is a decreasing convex function of c, see Fig. 1(a). Thus, the choice of c is equivalent to deciding by what proportion it is worth increasing the Bayes risk of the HB estimator in order to receive protection against large frequentist risks.

6. Frequentist risk of limited translation estimators

We now turn our attention to the frequentist risk of $\tilde{\boldsymbol{\theta}}^{LB}_{c,i}$, a function of $\boldsymbol{\theta}$ denoted by $\boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}^{LB}_{c,i})$, to show that the limited translation estimator, in return for the increased Bayes risk, does not allow the frequentist risk to be very large, in contrast with the HB estimator. The calculation of the frequentist risk of the limited translation estimator was feasible only under the simplifying assumption that the population variance–covariance matrix is a multiple of the sampling variance–covariance matrix, that is $\boldsymbol{A} = g\boldsymbol{\Sigma}$, where $g > 0$ is a known positive scalar. We thus consider the case where $\boldsymbol{X}_i|\boldsymbol{\theta}_i \sim N_p(\boldsymbol{\theta}_i, \boldsymbol{\Sigma})$ while the prior distributions are taken to be $\boldsymbol{\theta}_i \sim \xi \equiv N_p(\boldsymbol{\mu}, g\boldsymbol{\Sigma})$, $i = 1, \ldots, n$. Such priors, originally introduced by Zellner [8], are called g-priors.

Under the assumed model, the HB estimator of $\boldsymbol{\theta}_i$ is given by $\tilde{\boldsymbol{\theta}}^B_i = \boldsymbol{X}_i - (1+g)^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)$, and the frequentist risk associated with it is obtained by simplifying (2.6),

$$\boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}^B_i) = (1+g)^{-2}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)^{\mathrm{T}} + \boldsymbol{\Sigma}\{1 - (1 - 1/n)(1+2g)(1+g)^{-2}\}. \tag{6.1}$$

Also, the limited translation estimator is given by $\tilde{\boldsymbol{\theta}}^{LB}_{c,i} = \boldsymbol{X}_i - (1+g)^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\rho_c(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i - \bar{\boldsymbol{X}}_n)\|^2)$, where $\boldsymbol{D} = (1 - n^{-1})(1+g)\boldsymbol{\Sigma}$.

We would like to compare the frequentist risk of the HB estimator to the frequentist risk of the limited translation estimator. An expression of the latter is provided in the following theorem.

Theorem 6.1. *Under the assumption that $\boldsymbol{A} = g\boldsymbol{\Sigma}$, where $g > 0$ is a known scalar and for the relevance function $\rho_c(u) = \min(1, c/\sqrt{u})$, the frequentist risk of the multivariate limited translation estimator is given by*

$$\begin{aligned} \boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}^{LB}_{c,i}) &= E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}^{LB}_{c,i})(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}^{LB}_{c,i})^{\mathrm{T}}\} = \boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}^B_i) \\ &\quad + (\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}[(1+2g)(1+g)^{-2}P[\chi^2_{p+4}(\lambda_i) > c^2(1+g)] \\ &\quad + 2(1+g)^{-1}P[\chi^2_{p+2}(\lambda_i) > c^2(1+g)] + c^2(1+g)^{-1}E_{\lambda_i}\{[\chi^2_{p+4}(\lambda_i)]^{-1} I[\chi^2_{p+4}(\lambda_i) > c^2(1+g)]\} \end{aligned}$$

$$
\begin{aligned}
&+2c(1+g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda_i)>c^2(1+g)]\}\\
&-2c(1+g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+4}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda_i)>c^2(1+g)]\}]\\
&+\boldsymbol{\Sigma}(1-1/n)[(1+2g)(1+g)^{-2}P[\chi^2_{p+2}(\lambda_i)>c^2(1+g)]\\
&+c^2(1+g)^{-1}E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-1}I[\chi^2_{p+2}(\lambda_i)>c^2(1+g)]\}\\
&-2c(1+g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda_i)>c^2(1+g)]\}] \qquad (6.2)
\end{aligned}
$$

where $\lambda_i = 2^{-1}(1-1/n)^{-1}(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)$ and $\chi^2_k(\lambda_i)$ denotes the non-central chi-square distribution with non-centrality parameter λ_i and k degrees of freedom.

The proof of the theorem is given in the Appendix.

From the general matrix valued risks in (6.1) and (6.2), we obtain scalar versions of them. Specifically, we consider the quadratic loss function $L_2(\boldsymbol{\theta}_i, \boldsymbol{a}_i) = (\boldsymbol{\theta}_i - \boldsymbol{a}_i)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{a}_i)$.

First, it is easy to show that the risk of the HB estimator, under the loss function L_2, is equal to

$$R_2(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}_i^B) = p + 2(1-1/n)(1+g)^{-2}\lambda_i - (1-1/n)(1+2g)(1+g)^{-2}p. \qquad (6.3)$$

Corollary 6.2. *Under the loss function L_2, the risk of the limited translation estimator is given by*

$$
\begin{aligned}
R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) &= E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \tilde{\boldsymbol{\theta}}_{c,i}^{LB})\} = tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB})]\\
&= R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_i^B) - (1-1/n)P[\chi^2_{p+2}(\lambda_i) > c^2(1+g)]\left\{\frac{4\lambda_i}{(1+g)} - \frac{p(1+2g)}{(1+g)^2}\right\}\\
&\quad + 2\lambda_i(1-1/n)(1+2g)(1+g)^{-2}P[\chi^2_{p+4}(\lambda_i) > c^2(1+g)]\\
&\quad + c^2(1-1/n)(1+g)^{-1}P[\chi^2_p(\lambda_i) > c^2(1+g)] + 2c(1-1/n)(1+g)^{-\frac{1}{2}}\\
&\quad \times\left[2\lambda_i E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda_i) > c^2(1+g)]\} - E_{\lambda_i}\{[\chi^2_p(\lambda_i)]^{\frac{1}{2}}I[\chi^2_p(\lambda_i) > c^2(1+g)]\}\right]. \qquad (6.4)
\end{aligned}
$$

The proof is provided in the Appendix.

The bracketed term in the last line of (6.4) can be calculated as

$$\sqrt{2}\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}\frac{\Gamma\left(\frac{p+1+2k}{2}\right)}{\Gamma\left(\frac{p+2k}{2}\right)}P[\chi^2_{p+1+2k} > c^2(1+g)]\left(\frac{2\lambda}{p+2k} - 1\right). \qquad (6.5)$$

The risk of $\tilde{\boldsymbol{\theta}}_{c,i}^{LB}$, for fixed p, n and g, quite conveniently, is a function only of the non-centrality parameter $\lambda_i = 2^{-1}(1-1/n)^{-1}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)$ and so is the risk in (6.3).

Let us now consider the hypothetical scenario where the statistician is given n observations of dimension $p = 3$. We also suppose that n is large enough to ignore the $1/n$ terms in the risks in (6.3) and (6.4). Also, suppose that $g = 2$ and that the statistician is willing to have a generalized relative savings loss of $1 - s_c = 10\%$ in order to receive protection against large frequentist risks.

In Fig. 1(a) we see how $1 - s_c$ decreases as c increases for three different values of $p = 3, 5$ and 10 and for fixed $g = 2$. For $p = 3$ and $1 - s_c = 10\%$ the corresponding value of c is 1.52. In Fig. 1(b) we see how the risks in (6.3) and (6.4) behave as the non-centrality parameter, λ_i increases. For small values of λ_i, i.e. when $\boldsymbol{\theta}_i$ is close to $\bar{\boldsymbol{\theta}}_n$, the HB estimator has slightly smaller frequentist risk than the limited translation HB estimator. However, the frequentist risk of the HB estimator increases linearly with the non-centrality parameter which clearly means that the HB estimator has high risk when the $\boldsymbol{\theta}_i$ is far from $\bar{\boldsymbol{\theta}}_n$. On the contrary, the frequentist risk of the limited translation HB estimator becomes flat after λ_i exceeds a certain value. That is, the limited translation estimator does not allow large frequentist risks even if $\boldsymbol{\theta}_i$ is far from $\bar{\boldsymbol{\theta}}_n$.

Returning to (6.4), we write $R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) = R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_i^B) + e_{p,c,g,n}(\lambda_i)$. The proposed estimator, $\tilde{\boldsymbol{\theta}}_{c,i}^{LB}$, does better than the HB estimator when the function $e_{p,c,g,n}(\lambda_i)$ takes on negative values. This, in general, happens when attempting to estimate a parameter $\boldsymbol{\theta}_i$ which departs widely from the mean $\bar{\boldsymbol{\theta}}_n$, that is, when the non-centrality parameter λ_i takes on large values. The question of interest is what values must λ_i take, for fixed values of p, c, g and n, in order for the function $e_{p,c,g,n}(\lambda_i)$ to become negative, and how likely these values are.

We attempt to partly answer this question by providing in Tables 1, 2 and 3 the minimum values, k, of λ_i needed in order for $e_{p,c,g,n}(\lambda_i)$ to take negative values, for fixed p, c, n and g. We also provide the probabilities that λ_i takes a value as big or bigger than k. These probabilities are calculated assuming that prior $\xi = N_p(\boldsymbol{\mu}, g\boldsymbol{\Sigma})$ is the true one, i.e.

$$
\begin{aligned}
P(\lambda_i \geq k) &= P\{2^{-1}(1-1/n)^{-1}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \bar{\boldsymbol{\theta}}_n) \geq k\}\\
&= P(\chi^2_p \geq 2\,kg^{-1}). \qquad (6.6)
\end{aligned}
$$

Table 1
Minimum values, k, of λ_i and $P(\lambda_i \geq k)$ for $p = 3$.

	g				
	0.5	1	2	5	10
$c = 2.466$	1.79	3.81	8.27	22.14	44.94
$1 - s_c = 1\%$	6.70%	5.45%	4.07%	3.13%	2.95%
$c = 1.840$	1.72	3.32	6.72	16.96	33.60
$1 - s_c = 5\%$	7.58%	8.43%	8.14%	7.91%	8.14%
$c = 1.521$	1.76	3.21	6.23	15.21	29.84
$1 - s_c = 10\%$	7.06%	9.29%	10.09%	10.76%	11.32%

Table 2
Minimum values, k, of λ_i and $P(\lambda_i \geq k)$ for $p = 5$.

	g				
	0.5	1	2	5	10
$c = 2.806$	2.49	5.12	10.80	28.31	57.11
$1 - s_c = 1\%$	7.64%	6.87%	5.55%	4.53%	4.36%
$c = 2.144$	2.51	4.69	9.24	22.88	45.19
$1 - s_c = 5\%$	7.41%	9.48%	9.99%	10.32%	10.76%
$c = 1.797$	2.63	4.68	8.88	21.38	41.98
$1 - s_c = 10\%$	6.17%	9.55%	11.39%	12.83%	13.57%

Table 3
Minimum values, k, of λ_i and $P(\lambda_i \geq k)$ for $p = 10$.

	g				
	0.5	1	2	5	10
$c = 3.490$	4.17	8.24	16.82	42.98	86.05
$1 - s_c = 1\%$	8.18%	8.70%	7.84%	7.02%	6.98%
$c = 2.757$	4.41	8.03	15.42	37.57	74.21
$1 - s_c = 5\%$	6.13%	9.79%	11.75%	13.10%	13.79%
$c = 2.352$	4.77	8.33	15.53	37.10	73.01
$1 - s_c = 10\%$	3.93%	8.22%	11.39%	13.80%	15.72%

Table 1 shows the values k and the corresponding probabilities $P(\lambda_i \geq k)$ for the case where the dimension is $p = 3$, for five different values of the prior parameter g and for three values of c. For the sake of simplicity, we take n to be large enough to be able to ignore the n^{-1} terms in (6.4). We may recall that c and $1 - s_c$ are one-to-one functions and thus Table 1 provides the generalized relative savings loss, $1 - s_c$, along with the corresponding c.

From the first row of Table 1, it is clear that for all values of g, $P(\lambda_i \geq k)$ is bigger than 1%, the generalized relative savings loss. That is, by sacrificing 1% of the Bayes risk, we have fairly big returns in terms of the frequentist risk. Similar are the results displayed on the second row of Table 1. The generalized relative savings loss is 5% while the returns in frequentist risk are bigger than 5% for all values of g. For the case where $1 - s_c = 10\%$, the returns in frequentist risk are bigger than 10% for $g = 2$, 5 and 10 and smaller than 10% for $g = 0.5$ and 1. This, however, is not discouraging because the reported percentages, $P(\lambda_i \geq k)$, are calculated assuming that the prior ξ is the true one. We can expect the probabilities $P(\lambda_i \geq k)$ to increase with the increasing distance of ξ from the true prior.

The results of Tables 2 and 3, where $p = 5$ and 10 respectively, are similar. We have fairly big returns in frequentist risk when sacrificing $1 - s_c = 1\%$ and 5% of the Bayes risk. The returns in frequentist risk when sacrificing $1 - s_c = 10\%$ of the Bayes risk are bigger than 10% for $g = 2$, 5 and 10 but they are smaller than 10% for $g = 0.5$ and 1.

7. Application

In this section we apply the proposed inferential procedure in order to estimate the 'long term average' vitamin intakes of HIV-positive subjects. We will be using the baseline data from a prospective study of the role of drug abuse in HIV/AIDS weight loss and malnutrition conducted in Boston, Massachusetts, USA. This study gathered data on both HIV-positive and HIV-negative subjects. However, here we will only be concerned with the 127 HIV-positive subjects. A similar data-set, that includes information on vitamin intakes on 70 HIV-positive subjects, will be used in order to elicit the prior distribution.

Each of the 197 subjects completed 3-day food records, recording type and amount of food, including supplements and vitamins. Dietary analysis was performed on the 3-day food records and daily nutrient intakes were determined. The intakes of several nutrients were determined but here, for the sake of simplicity, we will be analyzing only two of those nutrients, specifically, vitamins B_6 and B_{12}.

The observed distribution of the intakes of the two vitamins, on all 197 subjects, is not close to a realization from a bivariate normal distribution. This indicates the need of transforming the data before applying methods that require normal

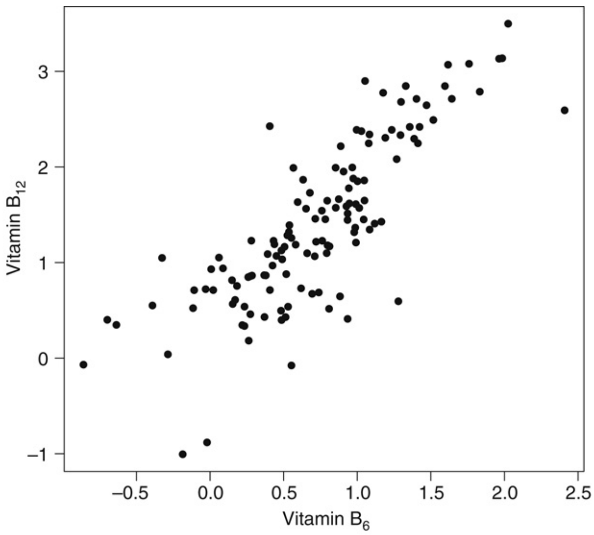

Fig. 2. Average of the transformed intakes of vitamins B_6 and B_{12}.

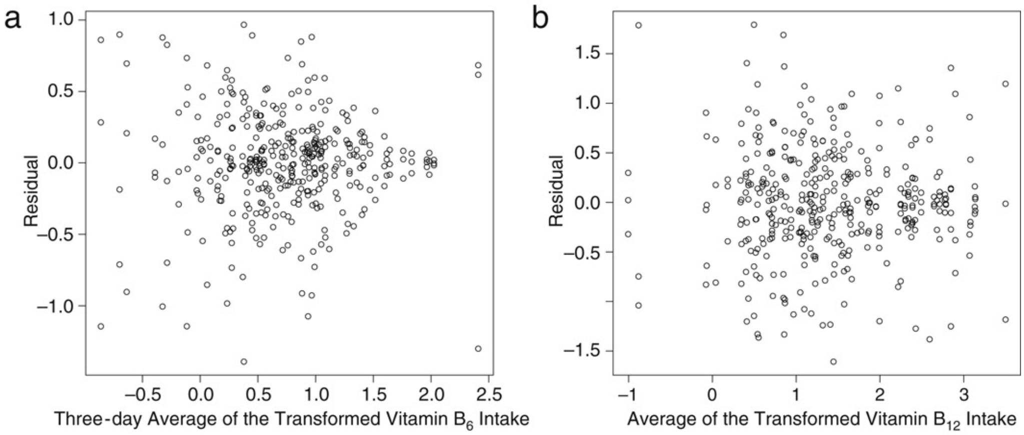

Fig. 3. Plot of the residuals $X_{ijk} - \bar{X}_{i.k}$ versus $\bar{X}_{i.k}$ of the transformed intakes of vitamins (a) $B_6 (k = 1)$ and (b) $B_{12} (k = 2)$.

distributions. We thus start our analysis by considering a bivariate Box–Cox [13,14] transformation. It turns out that the values of λ_i, $i = 1, 2$, for the transformation are $\lambda_1 = 0.017$ and $\lambda_2 = 0.023$ for vitamins B_6 and B_{12} respectively.

The average intakes, based on the 3-day food records, of the two vitamins of the 127 subjects in the main data set, after the transformation, are displayed in Fig. 2. Notice that some of these averages are negative. Further, it is clear that even after the transformation the assumption of normality is not exactly met. The presence of outliers indicates that a robust procedure, like the limited translation estimators, would be more appropriate than the regular HB estimators.

For $i = 1, \dots, n = 127$ and $j = 1, 2, 3$, let X_{ij1} and X_{ij2} denote the intake of vitamins B_6 and B_{12} of the ith subject in day j respectively. Further, $\boldsymbol{X}_{ij} = (X_{ij1}, X_{ij2})^{\mathrm{T}}$ is the response vector of subject i on day j. Additionally, θ_{i1} and θ_{i2} denote the 'long term average' daily intakes of vitamins B_6 and B_{12}, respectively, of subject i. The vector $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2})^{\mathrm{T}}$ is accordingly defined.

There are reasonable grounds to believe that patients with higher average vitamin intake have higher variability. In order to test this assumption, we plot the residuals $X_{ijk} - \bar{X}_{i.k}$, $i = 1, \dots, n = 127$ and $k = 1, 2$, where $\bar{X}_{i.k}$ denotes each patients average 3-day intake, after transformation, of vitamins B_6 and B_{12} for $k = 1$ and $k = 2$, respectively. These plots are shown in Fig. 3. It appears as if the bivariate Box–Cox transformations has solved the suspected problem of unequal variances.

Note that the average intakes, of the $n = 127$ subjects, after transformation, are $\bar{\boldsymbol{X}}_n = (0.73, 1.40)^{\mathrm{T}}$. The prior variance–covariance matrix, $\boldsymbol{A}$, is obtained from the 70 observations in the second data set, $\boldsymbol{A} = \begin{bmatrix} 0.744 & 0.740 \\ 0.740 & 1.290 \end{bmatrix}$. Also, to get an idea about the entries of the sampling variance–covariance matrix $\boldsymbol{\Sigma}$, we use the main data set, $\boldsymbol{\Sigma} = \begin{bmatrix} 0.189 & 0.140 \\ 0.140 & 0.452 \end{bmatrix}$. Finally, note that for the limited translation estimator, we have chosen $c = 1.36$ which corresponds to a generalized relative savings loss of $1 - s_c = 0.10$.

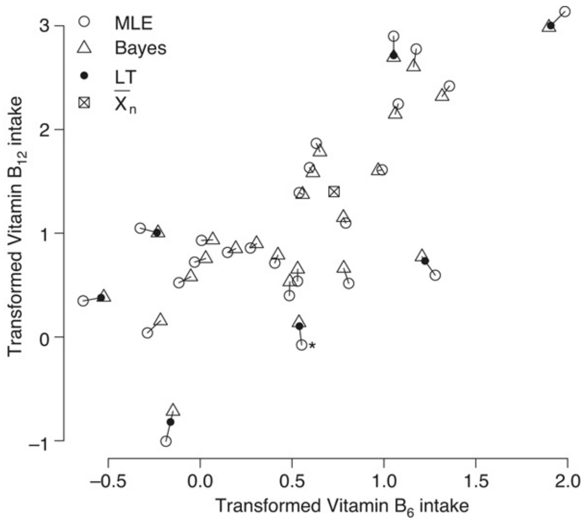

Fig. 4. Estimated 'long term average' intakes of vitamins B_6 and B_{12} on the transformed scale.

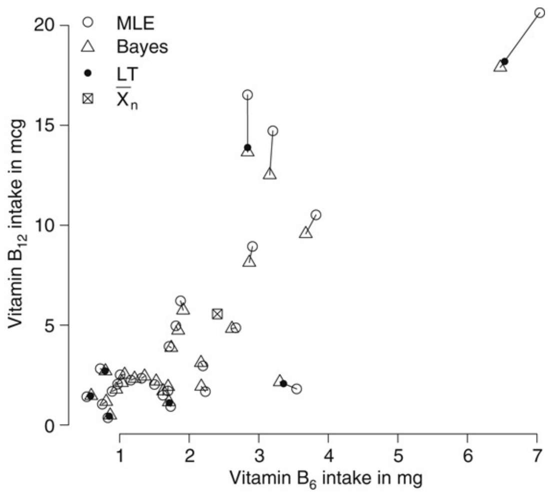

Fig. 5. Estimated 'long term average' intakes of vitamins B_6 and B_{12} on the original scale.

Fig. 4 displays the estimated transformed 'long term average' intakes of the two vitamins for 25 of the 127 subjects in our sample. The estimates were obtained using the ML, HB and limited translation estimators. The estimates that correspond to the same subject are connected by a line. For those subjects that the HB and LT estimators provide the same estimate, the LT estimates are not displayed. A few points are worth noting. The HB estimator pulls the ML estimates towards the grand mean. For those ML estimates that are close to the grand mean, the HB and the limited translation estimates are identical while for those that are far from the grand mean, the limited translation estimates are somewhere between the ML and HB estimates. The difference between the Euclidean and Mahalanobis distance is well displayed in this graph. For instance, the ML estimator at approximately $(-0.3, 0)$ is far from the mean in the Euclidean but not in the Mahalanobis sense, and for this reason the limiting translation does not kick in. The opposite holds true for the ML estimate that is marked with an asterisk. The latter displays another interesting point. Componentwise speaking, the HB (or the LT) estimator does not necessarily pull the ML estimator closer to the grand mean. The bivariate point under consideration corresponds to to an ML estimate of the 'long term average' intake of vitamin B_6 of 0.55, while the HB estimate is 0.54. As noted earlier, however, the grand mean that corresponds to the intake of vitamin B_6 is 0.73.

In order to complete the proposed purpose of the application, we back-transform the estimates to the original vitamin intake scale. These estimates, for the same 25 subjects as those that we saw in Fig. 4, are displayed in Fig. 5. The comparison of the estimates in the original scale is very similar to the one we saw for the estimates that were obtained based on the transformed data.

Table 4
Comparison of the estimates from the bivariate and univariate approaches.

	$Q_{0.05}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.95}$	Mean	Stdev
HB B_6 (mg)	−0.029	−0.007	0.004	0.031	0.087	0.015	0.040
LT B_6 (mg)	−0.028	−0.007	0.004	0.030	0.086	0.015	0.036
HB B_{12} (mcg)	−0.113	−0.026	0.039	0.116	0.449	0.089	0.245
LT B_{12} (mcg)	−0.112	−0.026	0.039	0.113	0.319	0.076	0.200

It should be pointed out that the limiting of translation kicked in for 17 subjects, that is for $13.4\% = 17/127$ of the subjects. We may also note that For $c = 1.67$ which corresponds to a generalized relative savings loss of 5%, the limiting of translations kicks in for $5.5\% = 7/127$ of the subjects.

When treating the problem of the estimation of the 'long term average' intake of the two vitamins as two univariate problems and for $1 - s_c = 0.10$ (corresponding $c = 1.18$) the limiting of translation kicks in for $7.9\% = 10/127$ of the subjects for vitamin B_6 and for $10.2\% = 13/127$ of the subjects for vitamin B_{12}. Again for the two univariate problems and for $1 - s_c = 0.05$ (corresponding $c = 1.46$) the limiting of translation kicks in for $3.1\% = 4/127$ and $2.4\% = 3/127$ of the subjects for vitamins B_6 and B_{12}, respectively. It should also be noted that in the marginal approach, we used two univariate Box–Cox transformations and it turned out that the values of λ_i, $i = 1, 2$, were $\lambda_1 = -0.021$ and $\lambda_2 = -0.029$ for vitamins B_6 and B_{12} respectively.

In order to compare the marginal HB and LT estimates from the multivariate and univariate treatments, we begin by transforming the obtained estimates back to the original scale. We then calculate the differences $\hat{\theta}_i^m - \hat{\theta}_i^u$, where $\hat{\theta}_i^m$ is the estimate of the 'long term average' intake of subject i for any of the two vitamins, using any of the two estimators, that results from the multivariate treatment of the estimation problem. Similarly, $\hat{\theta}_i^u$ is the estimate that results from the univariate treatment. We then summarize these differences by reporting the 5th, 25th, 50th, 75th and 95th percentiles, denoted by $Q_{0.05}$, $Q_{0.25}$, $Q_{0.50}$, $Q_{0.75}$ and $Q_{0.95}$, as well as the means (Mean) and the standard deviations (Stdev). These are shown in Table 4.

Noting that the average of the observed intakes of vitamin B_6 is 2.4 mg, while for vitamin B_{12} is 5.6 mcg, we see that the differences between the estimates are relatively small. The means, however, of these differences are always positive, indicating that the multivariate treatment of the problem provides, on average, bigger estimates that the univariate treatment. The medians of the differences are also positive. By inspecting the provided quantiles, it can be seen that the distributions of these differences are skewed to the right.

The reported results for the LT estimates are based on a generalized relative savings loss of $1 - s_c = 10\%$. We have also compared the LT estimates that resulted by setting $1 - s_c = 5\%$ but this comparison was very similar to the one presented, that is for $1 - s_c = 10\%$, and was thus omitted.

8. Summary and conclusions

The paper has developed limited translation HB estimators of the multivariate normal mean by extending the work of Efron and Morris [2] and by utilizing the notion of influence functions of Johnson and Geisser [5–7]. We have demonstrated the usefulness of such estimators from the criteria of frequentist risks when the true parameter to be estimated departs widely from the grand average of all the parameters.

Appendix. Proof of Theorem 6.1

We first prove two basic lemmas useful to the proof of Theorem 6.1.

Lemma A.1. *Let $\boldsymbol{Y} \sim N_p(\boldsymbol{\eta}, a\boldsymbol{\Sigma})$ where $a > 0$. Then, for any fixed scalars b and d and any fixed p-dimensional vector $\boldsymbol{\phi}$, we have*

$$E\left\{\frac{\boldsymbol{Y}-\boldsymbol{\eta}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2 \le b]\right\} = (\boldsymbol{\eta}-\boldsymbol{\phi}) \times \left[E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \le b]\} - E_\lambda\{[\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \le b]\}\right]. \quad (A.1)$$

Proof. We write $Q = a^{-1}(\boldsymbol{Y}-\boldsymbol{\phi})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{\phi})$ and observe that $Q \sim \chi^2_p(\lambda)$ where $\lambda = (\boldsymbol{\eta}-\boldsymbol{\phi})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})/2$. In what follows we repeatedly use the result that if $X \sim \chi^2_p(\lambda)$, then the density function of X is an infinite sum of χ^2_{p+2k} variables, $k = 0, 1, 2, \ldots$, with Poisson weights $(\mathrm{e}^{-\lambda}\lambda^k)/(k!)$.

We begin with the equality

$$\int_{[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]} \frac{\mathrm{e}^{-\frac{1}{2}\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\eta})\|^2}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}(2\pi)^{\frac{p}{2}}|a\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{d}\boldsymbol{Y} = E\{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{-2d} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\}$$
$$= E_\lambda\{[\chi_p^2(\lambda)]^{-d} I[\chi_p^2(\lambda)\le b]\}$$
$$= \sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}E\{(\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2\le b]\}. \tag{A.2}$$

We now differentiate both sides of (A.2) with respect to $\boldsymbol{\eta}$. First note that

$$\frac{\partial}{\partial\boldsymbol{\eta}}(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta}) = 2(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{Y}). \tag{A.3}$$

Hence,

$$\int_{[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]} \frac{(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta})\mathrm{e}^{-\frac{1}{2}\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\eta})\|^2}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}(2\pi)^{\frac{p}{2}}|a\boldsymbol{\Sigma}|^{\frac{1}{2}}}\mathrm{d}\boldsymbol{Y}$$
$$= (a\boldsymbol{\Sigma})^{-1}E\left\{\frac{\boldsymbol{Y}-\boldsymbol{\eta}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\right\}$$
$$= \sum_{k=0}^{\infty}(k!)^{-1}E_\lambda\{(\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2\le b]\}\frac{\partial(\mathrm{e}^{-\lambda}\lambda^k)}{\partial\boldsymbol{\eta}}. \tag{A.4}$$

Now,

$$\frac{\partial(\mathrm{e}^{-\lambda}\lambda^k)}{\partial\boldsymbol{\eta}} = -\frac{\partial\lambda}{\partial\boldsymbol{\eta}}\mathrm{e}^{-\lambda}\lambda^k + \mathrm{e}^{-\lambda}k\lambda^{k-1}\frac{\partial\lambda}{\partial\boldsymbol{\eta}}$$
$$= (a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})\mathrm{e}^{-\lambda}\lambda^{k-1}(k-\lambda). \tag{A.5}$$

Combining (A.4) and (A.5) we obtain

$$E\left\{\frac{\boldsymbol{Y}-\boldsymbol{\eta}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\right\} = (\boldsymbol{\eta}-\boldsymbol{\phi})\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^{k-1}(k-\lambda)}{k!}E\{(\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2\le b]\}$$
$$= (\boldsymbol{\eta}-\boldsymbol{\phi})\left[\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}E\{(\chi_{p+2+2k}^2)^{-d} I[\chi_{p+2+2k}^2\le b]\} - \sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}E\{(\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2\le b]\}\right]$$
$$= (\boldsymbol{\eta}-\boldsymbol{\phi})\left[E_\lambda\{[\chi_{p+2}^2(\lambda)]^{-d} I[\chi_{p+2}^2(\lambda)\le b]\} - E_\lambda\{[\chi_p^2(\lambda)]^{-d} I[\chi_p^2(\lambda)\le b]\}\right]. \tag{A.6}$$

This completes the proof of Lemma A.1. □

Lemma A.2. *Consider the same settings as in Lemma A.1. Then*

$$E\left\{\frac{(\boldsymbol{Y}-\boldsymbol{\eta})(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\right\} = (\boldsymbol{\eta}-\boldsymbol{\phi})(\boldsymbol{\eta}-\boldsymbol{\phi})^{\mathrm{T}}$$
$$\times\left[E_\lambda\{[\chi_p^2(\lambda)]^{-d} I[\chi_p^2(\lambda)\le b]\} + E_\lambda\{[\chi_{p+4}^2(\lambda)]^{-d} I[\chi_{p+4}^2(\lambda)\le b]\} - 2E_\lambda\{[\chi_{p+2}^2(\lambda)]^{-d} I[\chi_{p+2}^2(\lambda)\le b]\}\right]$$
$$+ a\boldsymbol{\Sigma}E_\lambda\{[\chi_{p+2}^2(\lambda)]^{-d} I[\chi_{p+2}^2(\lambda)\le b]\}. \tag{A.7}$$

Proof. We start by differentiating twice both sides of (A.2) with respect to $\boldsymbol{\eta}$. Note that

$$\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}}}\exp\{-(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta})/2\} = (a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta})(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}$$
$$\times\exp\{-(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta})/2\} - (a\boldsymbol{\Sigma})^{-1}\exp\{-(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{Y}-\boldsymbol{\eta})/2\}. \tag{A.8}$$

Thus,

$$(a\boldsymbol{\Sigma})^{-1}E\left\{\frac{(\boldsymbol{Y}-\boldsymbol{\eta})(\boldsymbol{Y}-\boldsymbol{\eta})^{\mathrm{T}}}{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{2d}} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\right\}(a\boldsymbol{\Sigma})^{-1}$$
$$-(a\boldsymbol{\Sigma})^{-1}E\{\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^{-2d} I[\|(a\boldsymbol{\Sigma})^{-\frac{1}{2}}(\boldsymbol{Y}-\boldsymbol{\phi})\|^2\le b]\}$$
$$= \sum_{k=0}^{\infty}(k!)^{-1}E\{(\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2\le b]\}\frac{\partial^2(\mathrm{e}^{-\lambda}\lambda^k)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}}}, \tag{A.9}$$

where

$$\frac{\partial^2(\mathrm{e}^{-\lambda}\lambda^k)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}}} = \frac{\partial}{\partial\boldsymbol{\eta}^{\mathrm{T}}}\{(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})\mathrm{e}^{-\lambda}\lambda^{k-1}(k-\lambda)\} = (a\boldsymbol{\Sigma})^{-1}\mathrm{e}^{-\lambda}\lambda^{k-1}(k-\lambda)$$
$$+ (a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})(\boldsymbol{\eta}-\boldsymbol{\phi})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}\mathrm{e}^{-\lambda}\lambda^{k-2}\{\lambda^2+k(k-1)-2k\lambda\}. \tag{A.10}$$

Substituting the last expression in (A.9), we obtain

$$\sum_{k=0}^{\infty}(k!)^{-1}E\{(\chi^2_{p+2k})^{-d}I[\chi^2_{p+2k}\le b]\}\frac{\partial^2(\mathrm{e}^{-\lambda}\lambda^k)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\mathrm{T}}} = (a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})(\boldsymbol{\eta}-\boldsymbol{\phi})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}$$
$$\times\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^{k-2}}{k!}\{\lambda^2+k(k-1)-2k\lambda\}E\{(\chi^2_{p+2k})^{-d}I[\chi^2_{p+2k}\le b]\}$$
$$+(a\boldsymbol{\Sigma})^{-1}\sum_{k=0}^{\infty}\frac{\mathrm{e}^{-\lambda}\lambda^{k-1}}{k!}(k-\lambda)E\{(\chi^2_{p+2k})^{-d}I[\chi^2_{p+2k}\le b]\}$$
$$=(a\boldsymbol{\Sigma})^{-1}(\boldsymbol{\eta}-\boldsymbol{\phi})(\boldsymbol{\eta}-\boldsymbol{\phi})^{\mathrm{T}}(a\boldsymbol{\Sigma})^{-1}$$
$$\times\left[E_\lambda\{[\chi^2_p(\lambda)]^{-d}I[\chi^2_p(\lambda)\le b]\}+E_\lambda\{[\chi^2_{p+4}(\lambda)]^{-d}I[\chi^2_{p+4}(\lambda)\le b]\}-2E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-d}I[\chi^2_{p+2}(\lambda)\le b]\}\right]$$
$$+(a\boldsymbol{\Sigma})^{-1}\left[E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-d}I[\chi^2_{p+2}(\lambda)\le b]\}-E_\lambda\{[\chi^2_p(\lambda)]^{-d}I[\chi^2_p(\lambda)\le b]\}\right]. \tag{A.11}$$

Combining (A.9) and (A.11) and collecting terms we obtain Lemma A.2. □

Remark. The results of Lemmas A.1 and A.2 hold even if we change the inequalities from $\le b$ to $>b$ with obvious modifications.

Proof of Theorem 6.1. Let $Q = (\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)$ and recall that $B=(1+g)^{-1}$. Also, recall that $\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n\sim N_p(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n,(1-1/n)\boldsymbol{\Sigma})$ and thus $(1-1/n)^{-1}Q\sim\chi^2_p(\lambda_i)$ where $\lambda_i=2^{-1}(1-1/n)^{-1}(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)$. Further, for any $k>0$, we write

$$\rho_c^k(\|\boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)\|^2)=\rho_c^k\{B(1-n^{-1})Q\}=I[(1-n^{-1})Q\le c^2(1+g)]$$
$$+c^kB^{-\frac{k}{2}}(1-n^{-1})^{-\frac{k}{2}}Q^{-\frac{k}{2}}I[(1-n^{-1})Q>c^2(1+g)]. \tag{A.12}$$

We write

$$\boldsymbol{R}_1(\boldsymbol{\theta}_i,\tilde{\boldsymbol{\theta}}^{LB}_{c,i})=E_{\boldsymbol{\theta}}(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}^{LB}_{c,i})(\boldsymbol{\theta}_i-\tilde{\boldsymbol{\theta}}^{LB}_{c,i})^{\mathrm{T}}$$
$$=E_{\boldsymbol{\theta}}\left[\boldsymbol{\theta}_i-\boldsymbol{X}_i+B(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)\rho_c\{(1-n^{-1})BQ\}\right]\times\left[\boldsymbol{\theta}_i-\boldsymbol{X}_i+B(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)\rho_c\{(1-n^{-1})BQ\}\right]^{\mathrm{T}}$$
$$=E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta}_i-\boldsymbol{X}_i)(\boldsymbol{\theta}_i-\boldsymbol{X}_i)^{\mathrm{T}}\}-BE_{\boldsymbol{\theta}}\left[\{(\boldsymbol{X}_i-\boldsymbol{\theta}_i)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}+(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\boldsymbol{\theta}_i)^{\mathrm{T}}\}\rho_c\{(1+n^{-1})BQ\}\right]$$
$$+B^2E_{\boldsymbol{\theta}}[(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\rho_c^2\{(1+n^{-1})BQ\}]. \tag{A.13}$$

Now,

$$E_{\boldsymbol{\theta}}\left[(\boldsymbol{X}_i-\boldsymbol{\theta}_i)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\rho_c\{(1+n^{-1})BQ\}\right]=E_{\boldsymbol{\theta}}\left[\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)-(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)\}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\rho_c\{(1+n^{-1})BQ\}\right]$$
$$=E_{\boldsymbol{\theta}}\left[\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)-(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)\}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}I[(1-1/n)^{-1}Q\le c^2(1+g)]\right]$$
$$+c(1+g)^{\frac{1}{2}}E_{\boldsymbol{\theta}}\left[\{(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)-(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)\}(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\times\{(1-1/n)^{-1}Q\}^{-\frac{1}{2}}I[(1-1/n)^{-1}Q>c^2(1+g)]\right], \tag{A.14}$$

and the first of the two expectations in the last three lines of the above equation is calculated by applying Lemmas A.1 and A.2 with $\boldsymbol{Y}=\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n$, $\boldsymbol{\eta}=\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n$, $a=1-1/n$, $b=c^2(1+g)$, $\boldsymbol{\phi}=\boldsymbol{0}$ and $d=0$, while the second one is calculated by setting $d=1/2$, keeping the rest of the specifications same as before and reversing inequalities.

Similarly,

$$E_{\boldsymbol{\theta}}[(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\rho_c^2\{(1+n^{-1})BQ\}]=E_{\boldsymbol{\theta}}[(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}I[(1-1/n)^{-1}Q\le c^2(1+g)]]$$
$$+c^2(1+g)E_{\boldsymbol{\theta}}\left[(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)(\boldsymbol{X}_i-\bar{\boldsymbol{X}}_n)^{\mathrm{T}}\times\{(1-1/n)^{-1}Q\}^{-1}I[(1-1/n)^{-1}Q>c^2(1+g)]\right], \tag{A.15}$$

and these two expectations are calculated using Lemmas A.1 and A.2 exactly as we did in Eq. (A.14), with the only difference being that in the second of the two expectations of the above equation we set $d=1$ instead of $d=1/2$. The result follows from combining Eqs. (A.13)–(A.15), and collecting the coefficients of $\boldsymbol{\Sigma}$ and $(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)(\boldsymbol{\theta}_i-\bar{\boldsymbol{\theta}}_n)^{\mathrm{T}}$ separately. □

Proof of Corollary 6.2. We obtain an expression for $R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB})$ by directly using the result of Theorem 6.1 to calculate $R_2(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB}) = \mathrm{tr}[\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_1(\boldsymbol{\theta}_i, \tilde{\boldsymbol{\theta}}_{c,i}^{LB})]$. The resulting expression is simplified by making use of the two equalities that follow. First,

$$
\begin{aligned}
&pE_\lambda\{[\chi^2_{p+2}(\lambda)]^{-1}I[\chi^2_{p+2}(\lambda) > c^2(1+g)]\} + 2\lambda E_\lambda\{[\chi^2_{p+4}(\lambda)]^{-1}I[\chi^2_{p+4}(\lambda) > c^2(1+g)]\} \\
&= p\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\frac{\Gamma\left(\frac{p+2k}{2}\right)}{2\Gamma\left(\frac{p+2+2k}{2}\right)}P[\chi^2_{p+2k} > (1+g)c^2] + 2\lambda\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\frac{\Gamma\left(\frac{p+2+2k}{2}\right)}{2\Gamma\left(\frac{p+4+2k}{2}\right)}P[\chi^2_{p+2+2k} > (1+g)c^2] \\
&= p\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\frac{\Gamma\left(\frac{p+2k}{2}\right)}{2\Gamma\left(\frac{p+2+2k}{2}\right)}P[\chi^2_{p+2k} > (1+g)c^2] + 2\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k k}{k!}\frac{\Gamma\left(\frac{p+2k}{2}\right)}{2\Gamma\left(\frac{p+2+2k}{2}\right)}P[\chi^2_{p+2k} > (1+g)c^2] \\
&= \sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}\frac{\Gamma\left(\frac{p+2k}{2}\right)}{2\Gamma\left(\frac{p+2+2k}{2}\right)}(p+2k)P[\chi^2_{p+2k} > (1+g)c^2] \\
&= P[\chi^2_p(\lambda) > c^2(1+g)],
\end{aligned} \tag{A.16}
$$

and similarly

$$
\begin{aligned}
&pE_\lambda\{[\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda) > c^2(1+g)]\} + 2\lambda E_\lambda\{[\chi^2_{p+4}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda) > c^2(1+g)]\} \\
&= E_\lambda\{[\chi^2_p(\lambda)]^{\frac{1}{2}}I[\chi^2_p(\lambda) > c^2(1+g)]\}.
\end{aligned} \tag{A.17}
$$

This completes the proof of the corollary. □

References

[1] D.V. Lindley, A.F.M. Smith, Bayes estimates for the linear model, J. R. Statist. Soc. B 32 (1972) 1–41.
[2] B. Efron, C. Moris, Limiting the risk of Bayes and empirical Bayes estimators–part I: The Bayes case, J. Ann. Statist. Assoc. 66 (1971) 807–815.
[3] B. Efron, C. Moris, Limiting the risk of Bayes and empirical Bayes estimators–part II: The empirical Bayes case, J. Ann. Statist. Assoc. 67 (1972) 130–139.
[4] B. Efron, C. Moris, Data analysis using Stein's estimator and its generalizations, J. Ann. Statist. Assoc. 70 (1975) 311–319.
[5] W. Johnson, S. Geisser, Assessing the predictive influence of observations, in: G. Kallianpur, P.R. Krishnaiah, J.K. Ghosh (Eds.), Statistics and Probability: Essays in Honor of C.R. Rao, North-Holland Publishing Co., Amsterdam, 1982, pp. 343–358.
[6] W. Johnson, S. Geisser, A predictive view of the detection and characterization of influential observations in regression analysis, J. Amer. Statist. Assoc. 78 (1983) 137–144.
[7] W. Johnson, S. Geisser, Estimative influence measures for the multivariate general linear model, J. Statist. Plann. Inference 11 (1985) 33–56.
[8] A. Zellner, On assessing prior distributions and Bayesian regression analysis with g-prior distributions, in: P.K. Goel, A. Zellner (Eds.), Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti, North-Holland Publishing Co., Amsterdam, 1986, pp. 233–243.
[9] N. Cressie, T.R.C. Read, Multinomial goodness-of-fit tests, J. R. Statist. Soc. B 46 (1984) 440–464.
[10] P. Mahalanobis, On the generalized distance in statistics, Proc. Nat. Inst. Sci. India (Calcutta) 2 (1936) 49–55.
[11] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust statistics, in: The Approach Based on Influence Functions, Wiley, New York, 1986.
[12] P.J. Huber, Robust Statistical Procedures, SIAM, Philadelphia, 1977.
[13] F. Hernandez, R.A. Johnson, The large-sample behavior of transformations to normality, J. Amer. Statist. Assoc. 75 (1980) 855–861.
[14] R.A. Rode, V.M. Chinchilli, The use of Box–Cox transformations in the development of multivariate tolerance regions with applications to clinical chemistry, Amer. Statistician 42 (1988) 23–30.