© 2007 Georgios Papageorgiou
To my parents and siblings
ACKNOWLEDGMENTS

First, and foremost, I would like to thank my advisor, Dr. Malay Ghosh, for sharing with me his knowledge, for the invaluable guidance he offered me and for his encouragement. I would also like to thank my committee members, Dr. Ramon Littell, Dr. Ronald Randles, Dr. Andrew Rosalsky and Dr. Myron Chang, for their constructive comments and support. Special thanks also go to the faculty members of the Department of Statistics at the University of Florida as well as to Dr. Janet Forrester of the Tufts University for providing me with data sets and for her collaborative work.

Special thanks are also due to my friends for their support and encouragement. Lastly, and most importantly, I would like to thank my entire family for their love and support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGMENTS</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
</tr>
</tbody>
</table>

## CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW ...................................... 11

   1.1 Introduction .......................................................... 11

   1.2 Overview of Limited Translation Estimators .......................... 12

      1.2.1 The Bayes Case .................................................. 12

      1.2.2 The Empirical Bayes Case ..................................... 15

2 MULTIVARIATE LIMITED TRANSLATION BAYES ESTIMATORS .................... 23

   2.1 Bayes Estimators ..................................................... 23

   2.2 Limited Translation Bayes Estimators ................................ 25

   2.3 Bayes Risk of the Limited Translation Bayes Estimators .......... 26

   2.4 Frequentist Risk of the Limited Translation Bayes Estimators ... 30

   2.5 Robustness of Limited Translation Bayes Estimators .............. 34

3 MULTIVARIATE LIMITED TRANSLATION
   EMPIRICAL BAYES ESTIMATORS: THE CASE OF UNKNOWN PRIOR
   MEAN ............................................................................. 41

   3.1 Bayes, Empirical Bayes Estimators and Influence Functions .... 41

   3.2 Limited Translation Empirical Bayes Estimators .................. 44

   3.3 Bayes Risk of the Limited Translation EB Estimators ............. 45

   3.4 Frequentist Risk of the Limited Translation EB Estimators ...... 48

   3.5 Robustness of Limited Translation EB Estimators ................ 51

4 MULTIVARIATE EMPIRICAL BAYES AND LIMITED TRANSLATION
   EMPIRICAL BAYES ESTIMATORS ............................................ 55

   4.1 Introduction ............................................................. 55

   4.2 Empirical Bayes Estimators ........................................... 57

   4.3 Limited Translation Empirical Bayes Estimators ................... 60

   4.4 Bayes Risk of the Limited Translation EB Estimators ............. 61

   4.5 Frequentist Risk of the Limited Translation EB Estimators ..... 64

   4.6 A Simulation Study .................................................... 65
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Minimum values, $k$, of $\lambda$ and $P(\lambda \geq k)$ for $p = 3$</td>
<td>39</td>
</tr>
<tr>
<td>2-2</td>
<td>Minimum values, $k$, of $\lambda$ and $P(\lambda \geq k)$ for $p = 5$</td>
<td>39</td>
</tr>
<tr>
<td>2-3</td>
<td>Minimum values, $k$, of $\lambda$ and $P(\lambda \geq k)$ for $p = 10$</td>
<td>39</td>
</tr>
<tr>
<td>4-1</td>
<td>Comparison of the risks of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 30$.</td>
<td>70</td>
</tr>
<tr>
<td>4-2</td>
<td>Comparison of the risks of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 30$.</td>
<td>70</td>
</tr>
<tr>
<td>4-3</td>
<td>Comparison of the risks of the estimators under the normal model and the $f_1$ sampling distribution, $n = 30$.</td>
<td>70</td>
</tr>
<tr>
<td>4-4</td>
<td>Comparison of the risks of the estimators under the normal model and the $f_2$ sampling distribution, $n = 30$.</td>
<td>71</td>
</tr>
<tr>
<td>4-5</td>
<td>Comparison of the risk of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 20$.</td>
<td>73</td>
</tr>
<tr>
<td>4-6</td>
<td>Comparison of the risk of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 20$.</td>
<td>73</td>
</tr>
<tr>
<td>4-7</td>
<td>Comparison of the risk of the estimators under the normal model and the $f_1$ sampling distribution, $n = 20$.</td>
<td>73</td>
</tr>
<tr>
<td>4-8</td>
<td>Comparison of the risk of the estimators under the normal model and the $f_2$ sampling distribution, $n = 20$.</td>
<td>74</td>
</tr>
<tr>
<td>5-1</td>
<td>Comparison of the risks of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 30$.</td>
<td>88</td>
</tr>
<tr>
<td>5-2</td>
<td>Comparison of the risks of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 30$.</td>
<td>88</td>
</tr>
<tr>
<td>5-3</td>
<td>Comparison of the risks of the estimators under the normal model and the $f_1$ sampling distribution, $n = 30$.</td>
<td>88</td>
</tr>
<tr>
<td>5-4</td>
<td>Comparison of the risks of the estimators under the normal model and the $f_2$ sampling distribution, $n = 30$.</td>
<td>89</td>
</tr>
<tr>
<td>5-5</td>
<td>Comparison of the risk of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 20$.</td>
<td>89</td>
</tr>
<tr>
<td>5-6</td>
<td>Comparison of the risk of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 20$.</td>
<td>89</td>
</tr>
</tbody>
</table>
5-7 Comparison of the risk of the estimators under the normal model and the $f_1$ sampling distribution, $n = 20$. .................................................. 90

5-8 Comparison of the risk of the estimators under the normal model and the $f_2$ sampling distribution, $n = 20$. .................................................. 90

5-9 MLE, EB, LT Estimates ................................................................. 91
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Plot of $1 - s_c$ as a function of $c$.</td>
<td>38</td>
</tr>
<tr>
<td>2-2</td>
<td>Plot of risk as function of the non-centrality parameter, $\lambda$.</td>
<td>38</td>
</tr>
<tr>
<td>2-3</td>
<td>Bayes risk under misspecified priors.</td>
<td>40</td>
</tr>
<tr>
<td>3-1</td>
<td>Risks plotted against the the non-centrality parameter, $\lambda_i$.</td>
<td>54</td>
</tr>
<tr>
<td>3-2</td>
<td>Bayes risks plotted against the assumed parameter $g$ when the true parameter is taken to be $g^* = 2$.</td>
<td>54</td>
</tr>
<tr>
<td>4-1</td>
<td>Plot of $1 - s_c$ as a function of $c$, for $p = 2, 10$ and $n = 10, 30$.</td>
<td>68</td>
</tr>
<tr>
<td>4-2</td>
<td>The $\theta_i$ generated from (a) the contaminated model and (b) the normal model.</td>
<td>69</td>
</tr>
<tr>
<td>4-3</td>
<td>A sample of $\theta_i$ generated from (a) the contaminated model and (b) the normal model.</td>
<td>72</td>
</tr>
<tr>
<td>5-1</td>
<td>Plot of $RLS = E{1 - \rho_c(aW)}^2$ as function of $c$, for $p = 2, 5$ and $n = 10, 30$.</td>
<td>87</td>
</tr>
<tr>
<td>5-2</td>
<td>Average intakes of vitamins $A$ and $B_1$.</td>
<td>91</td>
</tr>
<tr>
<td>5-3</td>
<td>Estimated intakes of vitamins $A$ and $B_1$.</td>
<td>92</td>
</tr>
</tbody>
</table>
Bayes estimators are used quite often in the theory and practice of statistics. Their popularity is attributed to the more efficient inference that they often lead to compared to classical frequentist procedures. However, the Bayes estimators can result in high Bayes risk when the prior distribution is misspecified. Additionally, they can result in high frequentist risk when they are used for estimating parameters which depart widely from the assumed prior means.

We developed multivariate limited translation Bayes estimators of the normal mean vector which serve as a compromise between the Bayes and the maximum likelihood estimators. We demonstrated the superiority of the limited translation estimators over the usual Bayes estimators under misspecified priors and often also in terms of the frequentist risks under the situation stated in the previous paragraph.

We also extended the above results and developed multivariate limited translation empirical Bayes estimators of the normal mean vector which serve as a compromise between the empirical Bayes estimators and the maximum likelihood estimators. These compromise estimators perform better than the regular empirical Bayes estimators, in a frequentist sense, when there is wide departure of an individual observation from the grand average.
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Bayesian methods are used extensively these days 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 amount 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.

A well known characteristic of the Bayesian methods of estimation for the exponential family of distributions is that the Bayes estimators, namely the posterior means, shrink the maximum likelihood estimators towards the means of the prior distributions, with the extreme values experiencing the most shift. Although the Bayes estimators do, on average, better than any other estimator, they may perform badly in individual cases where true, unobservable means are far away from the assumed prior means. In addition, Bayesian methods can lead to severely wrong conclusions if the assumed prior is widely different from the ‘true’ prior.

Robust Bayesian methods have been proposed to guard against problems of this type. One such procedure, first introduced by Efron & Morris (1971, 1972a, 1975), and referred to as ‘limited translation rules’ by these authors, is the subject matter of this dissertation. The limited translation rules are compromises between the Bayes and the maximum likelihood estimators that slightly increase the Bayes risk but guard against
large frequentist risks and large Bayes risks for cases where the prior distributions are not correctly specified.

One of the virtues of the limited translation rules is that they do not fare too badly in their Bayes risk performance, compared to the regular Bayes estimators, even if the assumed prior is true. On the other hand, if the assumed prior departs widely from the true prior, then the limited translation rules do have much superior Bayes risk performance than the regular Bayes estimators. In a frequentist risk sense, the limited translation rules do not perform too badly relative to the regular Bayes rules if the parameter to be estimated is close to the assumed prior mean. On the other hand, if the parameter to be estimated is far from the prior mean, then the limited translation rules do perform much better than the regular Bayes estimators.

Efron & Morris (1971, 1972a, 1975) developed limited translation rules for the univariate normal case. The objective here is to develop limited translation rules for multivariate normal case. In this chapter we review the literature related to limited translation estimators.

1.2 Overview of Limited Translation Estimators

1.2.1 The Bayes Case

Let $X|\theta \sim N(\theta, 1)$ and $\theta \sim \xi \equiv N(0, A)$. The interest here is in estimating the unobservable quantity $\theta$ based on an observation $x$ on the variable $X$. Under the squared error loss function, $L(\theta, a) = (\theta - a)^2$, or any other increasing function of $|\theta - a|$, the Bayes estimator of $\theta$ given $X$ is

$$\hat{\theta}^B(X) = \frac{A}{A + 1}X = (1 - B)X,$$

where $B = (A + 1)^{-1}$.

The estimating rule in Equation 1–1 is optimal in the sense that it minimizes the expected risk, $r(\xi, \hat{\theta}) = E\{\theta - \hat{\theta}(X)\}^2$, among all choices of estimating rules $\hat{\theta}(X)$, with
the expectation being taken with respect to the joint distribution of $X$ and $\theta$ under the assumed prior $\xi$.

The estimator $\hat{\theta}^B$ shrinks the maximum likelihood estimator (MLE) of $\theta$ towards the mean of the prior distribution, $\mu = 0$, by a fixed proportion, $1 - B$. This type of shrinkage estimators is reasonable if the assumed prior is close to the ‘true’ one. However, shrinkage estimators perform poorly if the assumed and the ‘true’ prior are far apart. Suppose for example that the ‘true’ prior distribution is $\theta \sim \xi^* \equiv N(\mu^*, A^*)$. In this case, the Bayes risk of the estimator in Equation 1–1 is

$$r(\xi^*, \hat{\theta}^B) = B^2(\mu^*)^2 + B^2(A^2 + A^*).$$

(1–2)

Clearly, if the true prior mean, $\mu^*$ is far from the assumed prior mean, $\mu = 0$, then the Bayes risk of $\hat{\theta}^B$ is quite high.

Also, the frequentist risk, that is the risk as a function of $\theta$, denoted by $R(\theta, \hat{\theta}^B)$ and calculated as

$$R(\theta, \hat{\theta}^B) = E_{\theta}(\theta - \hat{\theta}^B)^2 = B^2\theta^2 + (1 - B)^2,$$

(1–3)

will be high when $\theta$ is far from the assumed prior mean, $\mu = 0$, the reason being that $\hat{\theta}^B$ shrinks the maximum likelihood estimator of $\theta$, $\hat{\theta}^0 = X$, towards the prior mean $\mu$.

On the other hand, $\hat{\theta}^0 = X$ has minimax risk equal to $R(\theta, \hat{\theta}^0) = 1$ for all $\theta$, and thus the average risk of $\hat{\theta}^0$, the average taken with respect to prior $\xi$, is $r(\xi, \hat{\theta}^0) = 1$ which, however, is bigger than the average risk of the Bayes estimator, $r(\xi, \hat{\theta}^B) = 1 - B$.

In order to combine the good properties of the Bayes rule with those of the MLE, that is to maintain low Bayes risk and at the same time put a bound to the frequentist risk, Efron & Morris (1971, 1972a, 1975) proposed limited translation rules. These rules are simple compromises between the Bayes rules and the MLEs.

For the estimation problem at hand, they proposed a rule akin to the Bayes rule, which however does not allow deviations from $\hat{\theta}^0 = X$ bigger than a fixed value, say $m$. 

13
That is the rule satisfies the constraint that $|\hat{\theta}^B - X| \leq m$ which holds true if $|X| \leq d$, where $d = mB^{-1}$. The limited translation rule of maximum translation $m$ is defined as

$$\hat{\theta}^{LB} = \begin{cases} X + m & \text{if } X < -d \\ (1 - B)X & \text{if } -d \leq X \leq d \\ X - m & \text{if } X > d. \end{cases} \quad (1-4)$$

The limited translation rule can also be written as

$$\hat{\theta}^{LB} = \{1 - \min(1, \frac{mB^{-\frac{1}{2}}}{|X|B^{\frac{1}{2}}})\} \hat{\theta}^0 + \min(1, \frac{mB^{-\frac{1}{2}}}{|X|B^{\frac{1}{2}}})\hat{\theta}^B$$

$$= \{1 - \rho_c(BX^2)\} \hat{\theta}^0 + \rho_c(BX^2)\hat{\theta}^B$$

$$= \{1 - \rho_c(BX^2)B\}X, \quad (1-5)$$

where $c = mB^{-\frac{1}{2}}$ and $\rho_c(u) = \min(1, c/\sqrt{u})$.

The function $\rho_c(.)$ is termed relevance function. It measures the relevance of the Bayes rule, and thus the relevance of the prior parameters, to the estimation of $\theta$, while $1 - \rho_c(.)$ measures the relevance of the MLE. Its argument, $BX^2$, is the square of the distance, measured in standard deviations, of the variable $X$ from its mean value $0$, under the marginal distribution of $X \sim N(0, B^{-1})$.

The properties of the limited translation rule are examined by calculating its Bayes and frequentist risks. Efron & Morris (1971) showed that

$$r(\xi, \hat{\theta}^{LB}) = r(\xi, \hat{\theta}^0)(1 - s_c) + r(\xi, \hat{\theta}^B)s_c, \quad (1-6)$$

where $1 - s_c$ is a decreasing convex function of $c$. It is calculated as $1 - s_c = E\{1 - \rho_c(U)\}^2$ where $U \sim \chi_3^2$. Also, these authors showed that

$$\sup_{\theta} R(\theta, \hat{\theta}^{LB}) = 1 + m^2. \quad (1-7)$$

The Bayes risk of the limited translation rule is a weighted average of the Bayes risks of the Bayes rule and the MLE, the weights being $s_c$ and $1 - s_c$ respectively. This, of course,
means that the Bayes risk of the limited translation rule is higher than that of the Bayes rule. However, the weight of the risk of the MLE, $1 - s_c$, is a decreasing convex function of $c$. This allows the statistician to choose $c$ by deciding by what proportion it is worth increasing the Bayes risk of the Bayes rule in order to receive protection against large frequentist risks. This protection becomes evident by comparing the frequentist risk of the Bayes rule to the frequentist risk of the limited translation rule. The former is given in Equation 1–3 and it is clearly an unbounded function of $\theta$, while the supremum with respect to $\theta$ of the latter is $1 + m^2$.

1.2.2 The Empirical Bayes Case

We now consider the scenario where the prior variance $A$ is unknown and thus the Bayes rule cannot be used as such. In this case however, $A$ can be estimated from the data and an unbiased estimator of $B = (1 + A)^{-1}$ can be inserted in Equation 1–1, thus resulting in an empirical Bayes (EB) estimator. The EB estimators have become very popular since Efron & Morris (1972a, 1973, 1975) gave the EB interpretation of the celebrated James-Stein estimator, James & Stein (1961). Statisticians have applied these methods to many important problems, in particular for simultaneous estimation of several parameters.

Consider the case where we have $p \geq 3$ independent univariate normal observations

$$X_i | \theta_i \overset{\text{iid}}{\sim} N(\theta_i, \sigma^2), \ i = 1, 2, \ldots, p,$$

(1–8)

where $\sigma^2 = \text{var}(X_i) = 1$. Letting $\mathbf{X} = (X_1, \ldots, X_p)^T$ and $\mathbf{\theta} = (\theta_1, \ldots, \theta_p)^T$, the sampling distributions in 1–8 can more briefly written as $\mathbf{X} | \mathbf{\theta} \sim N_p(\mathbf{\theta}, \mathbf{I}_p)$, where $\mathbf{I}_p$ denotes the identity matrix of order $p$.

Further, suppose that

$$\theta_i \overset{\text{iid}}{\sim} N(\mu, A), \ i = 1, 2, \ldots, p,$$

(1–9)
where the prior mean, for the sake of simplicity, is assumed to be equal to 0, while the
prior variance, \( \text{var}(\theta_i) = A \), is unknown. Equivalently, the prior distributions in 1–9 can be
written as \( \theta \sim \xi \equiv \mathcal{N}_p(0, AI_p) \).

The goal here is to estimate the unknown \( \theta_i, i = 1, 2, \ldots, p \). There are two competing
loss functions that we take into consideration. First, for the estimation of an individual
parameter \( \theta_i \), we consider the individual squared error loss function
\[
L(\theta_i, a_i) = (\theta_i - a_i)^2,
\]
where \( a_i \) is the estimate of the true parameter \( \theta_i \). We also consider the ensemble, or total,
loss function given by adding the individual losses
\[
L(\theta, a) = \sum_{i=1}^{p} L(\theta_i, a_i) = \sum_{i=1}^{p} (\theta_i - a_i)^2,
\]
where \( a = (a_1, \ldots, a_p)^T \) is a vector guess for \( \theta = (\theta_1, \ldots, \theta_p)^T \).

Under the distributional assumptions stated above, the posterior mean of \( \theta_i \) given \( X_i \),
that is the Bayes estimator of \( \theta_i \), is
\[
\hat{\theta}_i^B = (1 - B)X_i.
\]
We may recall that \( B = (1 + A)^{-1} \).

Since \( A \) is unknown so is \( B \), and the Bayes rule cannot be used as such. However,
\( B \) can be estimated from the marginal distribution of the data. Marginally \( \mathbf{X} \sim \mathcal{N}_p(0, B^{-1}I_p) \), and thus \( B\|\mathbf{X}\|^2 \sim \chi^2_p \). It follows that \( E(p - 2/\|\mathbf{X}\|^2) = B \). Thus,
\( \hat{B} = (p - 2)/\|\mathbf{X}\|^2 \) is an unbiased estimator for \( B \). Substituting this expression for \( B \) in
Equation 1–12 results in the celebrated James-Stein estimator for \( \theta_i \), namely
\[
\hat{\theta}_i^{EB} = (1 - \frac{p - 2}{\|\mathbf{X}\|^2})X_i = (1 - \hat{B})X_i.
\]
In vector notations, \( \hat{\theta}^{EB} = (\hat{\theta}_1^{EB}, \ldots, \hat{\theta}_p^{EB})^T \), is written as \( \hat{\theta}^{EB} = (1 - \hat{B})\mathbf{X} \).
Under the ensemble, or total, loss function of Equation 1–11, the risk of the MLE, \( X \), for all \( \theta \), is

\[
R(\theta, X) = E_{\theta} \{ L(\theta, X) \} = \sum_{i=1}^{p} E_{\theta_i} (\theta_i - X_i)^2 = p. \tag{1–14}
\]

Under the same loss function, James & Stein (1961) showed the risk of the James-Stein estimator, \( \hat{\theta}^{EB} \), is

\[
R(\theta, \hat{\theta}^{EB}) = E_{\theta} \{ L(\theta, \hat{\theta}^{EB}) \} = p \left\{ 1 - \frac{p - 2}{p} E_{\lambda} \left( \frac{p - 2}{p - 2 + 2K} \right) \right\}, \tag{1–15}
\]

where \( K \sim \text{Poisson}(\lambda = 2^{-1}||\theta||^2) \). The term curly brackets in the right hand side of Equation 1–15 is a strictly increasing concave function of \( \lambda \). It takes on its minimum value of \( 2/p \) at \( \lambda = 0 \) and it approaches its supremum of 1 as \( \lambda \) increases. Thus, \( \hat{\theta}^{EB} \) has uniformly lower risk than the MLE for every value of \( \theta \) in the sense of the total loss function \( L(\theta, a) \).

Baranchik (1964) considered the risk that occurs when estimating an individual \( \theta_i \) by the James-Stein estimator. He showed that the risk of \( \hat{\theta}_i^{EB} \) is

\[
R(\theta_i, \hat{\theta}_i^{EB}) = E_{\theta_i} (\theta_i - \hat{\theta}_i^{EB})^2
= 1 + 2(p^2 - 4) \frac{\theta_i^2}{||\theta||^2} E_{\lambda} \left\{ \frac{K}{(p - 2 + 2K)(p + 2K)} \right\}
- (p - 2) E_{\lambda} \left\{ \frac{p - 2 + 4K}{(p - 2 + 2K)(p + 2K)} \right\}, \tag{1–16}
\]

where \( K \sim \text{Poisson}(\lambda = 2^{-1}||\theta||^2) \). The risk in Equation 1–16 is maximized for fixed \( ||\theta||^2 \) at \( \theta_i^2 = ||\theta||^2 \), that is when the vector parameter \( \theta \) has all its components except the \( i \)th equal to zero, giving the expression for the maximum risk as

\[
1 + (p - 2) E_{\lambda} \left\{ \frac{2pK - p + 2}{(p - 2 + 2K)(p + 2K)} \right\}. \tag{1–17}
\]

The expression in 1–17 attains its maximum of approximately \( p/4 \) near \( \lambda = p/2 \).

The point of the above is that although the James-Stein estimator has smaller total risk than the MLE, as we saw in Equation 1–15, it may do poorly in estimating individual
\( \theta_i \) that have unusually small or large values. On the other hand, the MLE of \( \theta_i \) has minimax risk equal to 1.

It is worth noting that the Bayes risk of the James-Stein rule, that is the total squared error loss of Equation 1–11 averaged over the joint distribution of \( X \) and \( \theta \), is given by \( r(\xi, \hat{\theta}^{EB}) = p(1 - B) + 2B \). The first term in this expression, \( p(1 - B) \), is the risk of the Bayes rule. We can thus think of the second term as the price for having to estimate the prior variance, \( A \), from the data. The Bayes risk of the James-Stein rule can equivalently be written as \( r(\xi, \hat{\theta}^{EB}) = p\{1 - B(p - 2)/p\} \), where the term in curly brackets is clearly less than one and thus the Bayes risk of \( \hat{\theta}^{EB} \) is less than \( p \), the Bayes risk of the MLE.

**Efron & Morris** (1972a) developed limited translation empirical Bayes rules, a compromise between the James-Stein rule and the MLE, in an effort to lower the maximum risk of the individual components of the James-Stein rule and at the same time maintain low total risk and low Bayes risk. The limited translation rule follows as closely as possible the James-Stein rule provided that it does not deviate from the MLE by more than a fixed value.

The EB, or James-Stein, estimator was obtained as the estimator of the Bayes rule, by replacing \((p - 2)/||X||^2\) for \( B \). Similarly, the limited translation EB rule is obtained as the estimator of the limited translation Bayes rule. Recall that in Equation 1–5, the limited translation Bayes rule was briefly written as \( \hat{\theta}^{LB} = \left\{1 - \rho_c(BX^2)\right\}X \), where \( \rho_c(u) = \min(1, c/\sqrt{u}) \). Replacing \( B \) by its unbiased estimator results in the limited translation EB estimator for \( \theta_i \)

\[
\hat{\theta}_{i}^{LEB} = \left\{1 - \rho_c\left\{\frac{(p - 2)X_i^2}{||X||^2}\right\}\right\} \frac{p - 2}{||X||^2} X_i. \tag{1–18}
\]

In vector notations, \( \hat{\theta}^{LEB} = (\hat{\theta}_{1}^{LEB}, \ldots, \hat{\theta}_{p}^{LEB})^T \).

Since the argument of the function \( \rho_c \) is always less than or equal to \( p - 2 \), the values of \( c \) that need to be considered are those between 0 and \( \sqrt{p - 2} \). For \( c = 0 \) the
resulting estimator is the MLE, while for \( c = \sqrt{p-2} \) the estimator is equivalent to the James-Stein estimator. The value of \( c \) is chosen in such a way that the estimator has the good individual properties of the MLE and the good ensemble properties of the James-Stein estimator.

The relevance function, \( \rho_c(u) = \min(1, c/\sqrt{u}) \), measures the relevance of the whole to the individual. For instance, when \( X_i^2/||X||^2 \) takes on a large value, that is when \( X_i \) is much larger than the rest of the observations in absolute value, then the whole is not considered to be very relevant for the estimation of the corresponding \( \theta_i \). In such a case, the MLE is considered to be more relevant. The relevance function, by decreasing on its domain, \( 0 \leq u \leq p - 2 \), allows the estimator \( \hat{\theta}^{LEB} \) to deviate from the MLE by less as \( X_i^2/||X||^2 \) increases.

The properties of the limited translation EB rule are examined by first calculating its Bayes risk. The result here is similar to the result of the Bayes case. The Bayes risk of the limited translation EB rule can be shown to be equal to a weighted average of the Bayes risks of the MLE and the James-Stein estimator

\[ r(\xi, \hat{\theta}^{LEB}) = r(\xi, X)(1 - s_c) + r(\xi, \hat{\theta}^{EB})s_c. \]

The weight of the Bayes risk of the MLE, \( 1 - s_c \), is calculated as \( 1 - s_c = E[1 - \rho_c((p - 2)W_p)]^2 \), where \( W_p \sim \text{Beta}(3/2, (p-1)/2) \), and for fixed \( p \) it is a decreasing convex function of \( c \). Therefore the inverse function is also defined \( c(s) = c \leftrightarrow s_c(c) = s_c \).

For the two extreme values of \( c \), 0 and \( \sqrt{p-2} \), the resulting estimators, the MLE and the James-Stein estimator respectively, are minimax. It is still not known whether or not any other value of \( c \) results in minimax estimators. However, the limited translation EB rules are shown to have a minimax property in a certain averaged sense. To be precise, let \( u_{p,r} \) denote the uniform distribution on the \( p \)-dimensional sphere of radius \( r \). The estimator \( \hat{\theta} \) of \( \theta \) is said to be spherically minimax if

\[ r(u_{p,r}, \hat{\theta}) = E\{L(\theta, \hat{\theta})\} \leq p \]
for all $r \geq 0$, where the expectation is taken with respect to the conditional distribution of $X$ given $\theta$, $X|\theta \sim N_p(\theta, I_p)$, and the uniform distribution for $\theta$, $\theta \sim u_{p,r}$. If the risk of an estimator as a function of $\theta$ depends on $\theta$ only through $||\theta||^2$, which is the case for the MLE and the James-Stein estimator, then $r(u_{p,||\theta||}, \hat{\theta}) = R(\theta, \hat{\theta})$ so that the ideas of spherically minimax and minimax coincide. For other estimators $r(u_{p,r}, \hat{\theta})$ gives the average risk over $||\theta|| = r$.

The limited translation EB estimator is spherically minimax. Its average risk over a sphere of radius $r = ||\theta||$ is given by

$$r(u_{p,||\theta||}, \hat{\theta}^{LEB}) = p[1 - s_c + s_cS(r^2/2)],$$  \hfill (1–21)

where

$$S(\lambda) = 1 - \frac{p - 2}{p}E_\lambda \left( \frac{p - 2}{p - 2 + 2K} \right),$$  \hfill (1–22)

and $K \sim \text{Poisson}(\lambda)$. The spherical risk of the limited translation rule is less than or equal to $p$ since the bracketed term in Equation 1–21 is less than one.

The risk of $\hat{\theta}_i^{LEB}$, as a function of $\theta$, is now examined. It is shown that the risk of the limited translation estimator of $\theta_i$ depends on $\theta$ through $\theta_i^2/||\theta||^2$ and $||\theta||^2$. We denote this risk by

$$R(\theta_i, \hat{\theta}_i^{LEB}) = E_{\theta}(\theta_i - \hat{\theta}_i^{LEB})^2 = f_{p,c(s)}\left( \frac{\theta_i^2}{||\theta||^2}, ||\theta||^2 \right).$$  \hfill (1–23)

For fixed $p$, $0 \leq c(s) \leq \sqrt{p - 2}$ and $||\theta||^2$, $f_{p,c(s)}(., ||\theta||^2)$ is an increasing function of its first argument. It is therefore maximized for $\theta_i^2/||\theta||^2 = 1$. That is, the most unfavorable case for the estimation of $\theta_i$ occurs when $\theta_j = 0$, $j \neq i$. In this case the vector parameter $\theta$, for real valued $w$, has the form $\theta = w(0, 0, \ldots, 0, 1, 0, \ldots, 0)^T$ and $||\theta|| = |w|$ and thus the risk function becomes $R(\theta_i, \hat{\theta}_i^{LEB}) = f_{p,c(s)}(1, w^2)$. As a function of $|w|$, $f_{p,c(s)}(1, w^2)$ increases from its minimum of $1 - (p - 2)s_c/p$ at $w = 0$ to a maximum exceeding 1 and thereafter decreases asymptotically to 1 as $|w|$ increases. Let $\theta_p(s)$ be the value of $|w|$
which maximizes \( f_{p,c(s)}(1, w^2) \). Then, the supremum over all \( \theta \) of the risk of \( \hat{\theta}^{LEB}_i \), which we will denote by \( \tilde{R}(p, s_c) \), is

\[
\tilde{R}(p, s_c) = \sup_{\theta} R(\theta, \hat{\theta}^{LEB}_i) = f_{p,c(s)}(1, \theta_p(s))^2). \tag{1–24}
\]

Equivalently, \( \tilde{R}(p, s_c) \) is the supremum of the risk of \( \hat{\theta}^{LEB}_i \) over all priors \( \xi \) on \( \theta \)

\[
\tilde{R}(p, s_c) = \sup_{\xi} r(\xi, \hat{\theta}^{LEB}_i). \tag{1–25}
\]

For all values of \( p \), the biggest reductions in \( \tilde{R}(p, s_c) \), as a function of \( s_c \), occur near \( s_c = 1 \). Thus \( \tilde{R}(p, s_c) \) can be considerably reduced by increasing the Bayes risk of the Bayes rule by very little.

We now drop the assumptions that the variance of the sampling distribution, \( \sigma^2 \), and the mean of the prior distribution, \( \mu \), are known. We consider independent normal measurements \( X_{ij} | \theta_i \overset{\text{ind}}{\sim} N(\theta_i, \sigma^2) \), \( j = 1, 2, \ldots, k \), while \( \theta_i \) are themselves normally distributed variables \( \theta_i \overset{\text{iid}}{\sim} N(\mu, \tau^2) \), \( i = 1, 2, \ldots, p \). In this case the Bayes rule for estimating \( \theta_i \) is given by

\[
\hat{\theta}_i^B = \bar{X}_i - \frac{\bar{X}_i - \mu}{1 + (k\tau^2)/\sigma^2} = \bar{X}_i - B(\bar{X}_i - \mu), \tag{1–26}
\]

where \( \bar{X}_i = k^{-1} \sum_{j=1}^{k} X_{ij} \), \( A = (k\tau^2)/\sigma^2 \) and \( B = (1 + A)^{-1} \).

Recall that the limited translation Bayes rule follows as closely as possible the Bayes rule without however allowing deviation from the MLE bigger than a fixed value \( m \), say. That is, we impose the restriction that \( |\bar{X}_i - \hat{\theta}_i^B| \leq m \) which is equivalent to \( |\bar{X}_i - \mu| \leq mB^{-1} \). This rule can be briefly written as

\[
\hat{\theta}_i^{LB} = \bar{X}_i - B\rho_c \left\{ B(\bar{X}_i - \mu)^2 \right\} (\bar{X}_i - \mu) \\
= \mu + \left[ 1 - B\rho_c \left\{ B(\bar{X}_i - \mu)^2 \right\} \right] (\bar{X}_i - \mu). \tag{1–27}
\]

The unknown parameters \( \mu, \sigma^2 \) and \( \tau^2 \) need to be estimated from the data. First, for \( k > 1 \), the conditional distributions of the \( X_{ij} \) given \( \theta_i \), \( j = 1, 2, \ldots, k \) and \( i = 1, 2, \ldots, p \),
give us
\[ \frac{1}{\sigma^2} \sum_{i=1}^{p} \sum_{j=1}^{k} (X_{ij} - \bar{X}_i)^2 \sim \chi^2_{p(k-1)}. \] (1–28)

Thus, letting \( S = \sum_{i=1}^{p} \sum_{j=1}^{k} (X_{ij} - \bar{X}_i)^2 \), it follows that \( \hat{\sigma}^2 = S/[p(k - 1)] \) is unbiased estimator of \( \sigma^2 \).

Marginally, \( \bar{X}_i \sim N(\mu, \sigma^2/k + \tau^2) = N(\mu, \sigma^2/(kB)) \). Hence
\[ V = \sum_{i=1}^{p} (\bar{X}_i - \bar{X})^2 \sim \sigma^2/(kB)\chi^2_{p-1}, \] (1–29)

where \( \bar{X} = (pk)^{-1} \sum_{i=1}^{p} \sum_{j=1}^{k} X_{ij} \). It follows that for \( p \geq 4 \)
\[ E \left( \frac{p - 3}{V} \right) = \frac{B}{\sigma^2/k} = \frac{1}{\sigma^2/k + \tau^2}. \] (1–30)

We now write the unknown \( B \) that appears in the formula of the limited translation Bayes as
\[ B = \frac{\sigma^2}{k} \frac{B}{\sigma^2/k}. \] (1–31)

An unbiased estimator of the first term is \( \hat{\sigma}^2/k \) while an unbiased estimator of the second term is \( (p - 3)/V \). Also note that the unknown prior mean \( \mu \) is estimated by \( \bar{X} \).

Replacing the unknown parameters by their estimators results in the following estimator
\[ \hat{\theta}^{LEB}_i = \bar{X} + \left[ 1 - \rho_c \left( \frac{(\bar{X}_i - \bar{X})^2(p - 3)\hat{\sigma}^2}{kV} \right) \frac{(p - 3)\hat{\sigma}^2}{kV} \right] (\bar{X}_i - \bar{X}). \] (1–32)

The properties of this estimator are examined and it is shown that it has slightly bigger Bayes risk that the regular EB estimator but it protects the statistician against large frequentist risks.
CHAPTER 2
MULTIVARIATE LIMITED TRANSLATION BAYES ESTIMATORS

In this chapter we develop robust estimators of the multivariate normal mean assuming that all the model parameters are known. The cases of some or all parameters unknown will be addressed in later chapters. The organization of the sections of this chapter is as follows. In Section 2.1 we review some facts concerning the multivariate Bayes estimators. In section 2.2 we introduce the limited translation estimators. In Section 2.3 we evaluate their Bayes risk performance under the assumed prior. In Section 2.4 we evaluate the frequentist risk of both the regular Bayes and limited translation rules. In Section 2.5 we compare the Bayes risk performance of the two competing estimators when the assumed prior departs from the true prior. In Sections 2.4 and 2.5 we consider a specific form of prior, the g-prior originally introduced by Zellner (1986). Some of the long algebraic derivations are provided in the Appendix A.

2.1 Bayes Estimators

We begin by considering the following scenario. The \( p \)-dimensional random vector \( X \), conditional on \( \theta \), has the multivariate normal distribution \( X|\theta \sim N_p(\theta, \Sigma) \), where \( \Sigma = E\{(X - \theta)(X - \theta)^T\} \). The mean vector \( \theta \) itself follows a multivariate normal distribution \( \theta \sim \xi \equiv N_p(\mu, A) \). Here we assume that all the model parameters, \( \Sigma \), \( \mu \) and \( A \), are known and the statistician is interested in estimating the unobservable \( \theta \).

The assumptions stated above imply that conditional on \( X = x \), \( \theta \) is distributed as

\[
\theta \mid X = x \sim N(x - \Sigma(A + \Sigma)^{-1}(x - \mu), \Sigma - \Sigma(A + \Sigma)^{-1}\Sigma).
\]

Thus, considering the matrix loss function

\[
L_1(\theta, a) = (\theta - a)(\theta - a)^T,
\]

the Bayes estimator of the unobservable \( \theta \) is given by the posterior mean

\[
\hat{\theta}^B = X - \Sigma(A + \Sigma)^{-1}(X - \mu) = (I_p - B)X + B\mu,
\]
where \( B = \Sigma(A + \Sigma)^{-1} \) and \( I_p \) denotes the identity matrix of order \( p \). The Bayes risk of the Bayes estimator is given by the posterior variance-covariance matrix

\[
\mathbf{r}_1(\xi, \hat{\theta}^B) = E\{(\theta - \hat{\theta}^B)(\theta - \hat{\theta}^B)^T\} = (I_p - B)\Sigma,
\]

(2–4)

which is smaller than the Bayes risk of the maximum likelihood estimator (MLE),

\[
r_1(\xi, \mathbf{X}) = \Sigma.
\]

Much in the spirit of Efron & Morris (1971), we now calculate the Bayes risk of the Bayes estimator \( \hat{\theta}^B \) for a normal distribution with mean vector \( \mu_* \) and variance-covariance matrix \( A_* \). Let \( \xi_* \equiv N_p(\mu_*, A_*) \). Then the true posterior mean is given by

\[
\hat{\theta}_*^B = (I_p - B_*)\mathbf{X} + B_*\mu_*, \text{ where } B_* = \Sigma(A_* + \Sigma)^{-1}.
\]

The Bayes risk of \( \hat{\theta}^B \) under \( \xi_* \) is

\[
r_1(\xi_*, \hat{\theta}^B) = E\{L_1(\theta, \hat{\theta}^B)\} = E\{(\theta - \hat{\theta}^B)(\theta - \hat{\theta}^B)^T\}
= E\{(\theta - \hat{\theta}_*^B + \hat{\theta}_*^B - \hat{\theta}^B)(\theta - \hat{\theta}_*^B + \hat{\theta}_*^B - \hat{\theta}^B)^T\}
= E\{(\theta - \hat{\theta}_*^B)(\theta - \hat{\theta}_*^B)^T\} + E\{(\hat{\theta}_*^B - \hat{\theta}^B)(\hat{\theta}_*^B - \hat{\theta}^B)^T\},
\]

(2–5)

where \( E\{(\theta - \hat{\theta}_*^B)(\theta - \hat{\theta}_*^B)^T\} = (I_p - B_*)\Sigma \). We also have that

\[
E\{(\hat{\theta}_*^B - \hat{\theta}^B)(\hat{\theta}_*^B - \hat{\theta}^B)^T\}
= E\left[\{(I_p - B_*)\mathbf{X} + B_*\mu_* - (I_p - B)\mathbf{X} - B\mu\} \cdot \{(I_p - B_*)\mathbf{X} + B_*\mu_* - (I_p - B)\mathbf{X} - B\mu\}^T\right]
= E\left[\{(B - B_*)(\mathbf{X} - \mu_*) + B(\mu_* - \mu)\} \{(B - B_*)(\mathbf{X} - \mu_*) + B(\mu_* - \mu)\}^T\right]
= (B - B_*)B_*^{-1}\Sigma(B - B_*)^T + B(\mu_* - \mu)(\mu_* - \mu)^T B^T.
\]

(2–6)

Thus, the Bayes risk of the Bayes estimator under misspecified priors can be expressed as

\[
r_1(\xi_*, \hat{\theta}^B) = (I_p - B_*)\Sigma + (B - B_*)B_*^{-1}\Sigma(B - B_*)^T + B(\mu_* - \mu)(\mu_* - \mu)^T B^T. \quad (2–7)
\]

It is clear that when the true prior mean, \( \mu_* \), is far from the assumed prior mean, \( \mu \), the Bayes risk of \( \hat{\theta}^B \) is quite high.
Furthermore, the frequentist risk of $\hat{\theta}^B$, denoted by $R_1(\theta, \hat{\theta}^B)$, is calculated as

$$R_1(\theta, \hat{\theta}^B) = E_\theta\{(\hat{\theta}^B - \theta)(\hat{\theta}^B - \theta)^T\}$$

$$= E_\theta[(I_p - B)(X - \theta) + B(\mu - \theta)\{(I_p - B)(X - \theta) + B(\mu - \theta)\}^T]$$

$$= (I_p - B)\Sigma(I_p - B) + B(\mu - \theta)(\mu - \theta)^TB. \quad (2-8)$$

It is now easily seen that if $\mu$ is far from $\theta$, the frequentist risk of $\hat{\theta}^B$ is quite high, the reason being that $\hat{\theta}^B$ shrinks the MLE of $\theta$, $\hat{\theta}^0 = X$, towards the prior mean $\mu$.

On the other hand $\hat{\theta}^0 = X$ has minimax risk equal to $R_1(\theta, \hat{\theta}^0) = \Sigma$ for all $\theta$ and thus the Bayes risk of $\hat{\theta}^0$ is also equal to $\Sigma$ which, however, is bigger than the Bayes risk of the Bayes estimator, $r_1(\xi, \hat{\theta}^B) = (I_p - B)\Sigma$, if the assumed prior $\xi$ is the ‘true’ prior.

### 2.2 Limited Translation Bayes Estimators

The multivariate limited translation rules are akin to the Bayes rules but at the same time they put a limit to the amount of shrinkage of the maximum likelihood estimator towards the prior mean. The goal of these estimators is to maintain low Bayes risk and at the same time put a bound to the frequentist risk.

Suppose now that $X_i, i = 1, \ldots, n,$ are independently distributed $N_p(\theta_i, \Sigma)$. Also, assume that $\theta_i$ are iid $N_p(\mu, A)$.

**Definition** For the $i$th vector $\theta$, we define the limited translation Bayes estimator of maximum translation $c$ as

$$\hat{\theta}^{LB}_{c,i} = X_i - \Sigma(A + \Sigma)^{-\frac{1}{2}}h_c\{(A + \Sigma)^{-\frac{1}{2}}(X_i - \mu)\}, \quad (2-9)$$

where

$$h_c(z) = z \min(1, c/||z||), z \in \mathbb{R}^p \quad (2-10)$$

is the multidimensional Huber function, Huber (1974), and $c$ is a known constant.
Equivalently, the limited translation estimator can be written as

\[
\hat{\theta}_{c,i}^{LB} = X_i - B(X_i - \mu)\rho_c(||(A + \Sigma)^{-\frac{1}{2}}(X_i - \mu)||^2),
\]

(2–11)

where

\[
\rho_c(u) = \min(1, c/\sqrt{u}), u \in \mathbb{R}^1
\]

(2–12)

is termed the relevance function, Efron & Morris (1971, 1972a). The limited translation rule can also be represented as a weighted average of the maximum likelihood and the Bayes estimators since

\[
\hat{\theta}_{c,i}^{LB} = X_i \{1 - \rho_c(||(A + \Sigma)^{-\frac{1}{2}}(X_i - \mu)||^2)\}
+ \{(I_p - B)X_i + B\mu\} \rho_c(||(A + \Sigma)^{-\frac{1}{2}}(X_i - \mu)||^2).
\]

(2–13)

Marginally the random vectors \(X_i \sim N_p(\mu, A + \Sigma)\). Thus, the argument of the relevance function is the standardized squared norm of \(X_i, ||(A + \Sigma)^{-\frac{1}{2}}(X_i - \mu)||^2\). The value of the relevance function decreases with the increase in the value of the standardized squared norm of \(X_i\), thus reflecting the idea that the relevance of the population parameters, \(\mu\) and \(A\), is not the same for all \(\theta_i\). When the observed \(X_i\) has a high standardized squared norm, the Bayes estimator, and implicitly the prior parameters, is not considered to be very relevant for the corresponding \(\theta_i\). In such a case the MLE is considered to be more relevant and thus the shrinkage towards the prior mean is appropriately controlled.

In the subsequent sections of this chapter we will drop the suffix \(i\) and work with a generic \(\hat{\theta}_c^{LB}\).

2.3 Bayes Risk of the Limited Translation Bayes Estimators

First, it is of interest to know how well the estimator \(\hat{\theta}_c^{LB}\) performs assuming that the normal prior \(N_p(\mu, A)\) is the true one. We thus calculate its Bayes risk, \(r_1(\xi, \hat{\theta}_c^{LB}) = E\{(\theta - \hat{\theta}_c^{LB})(\theta - \hat{\theta}_c^{LB})^T\}\). The calculations for the most part do not depend on the choice of the relevance function \(\rho_c(\cdot)\). 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 MLE and of the Bayes estimator.

**Theorem 2.3.1.** For any relevance function $\rho_c(\cdot)$ we have

$$r_1(\xi, \hat{\theta}_{LB}^c) = r_1(\xi, X)(1 - s_c) + r_1(\xi, \hat{\theta}^B)s_c,$$  \hspace{1cm} (2–14)

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

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

$$GRSL(\hat{\theta}; X) = [r_1(\xi, X) - r_1(\xi, \hat{\theta}^B)]^{-1}[r_1(\xi, \hat{\theta}) - r_1(\xi, \hat{\theta}^B)].$$  \hspace{1cm} (2–15)

The term $r_1(\xi, X) - r_1(\xi, \hat{\theta}^B)$ is the savings, in Bayes risk sense, that occur when using the Bayes estimator instead of the MLE, while $r_1(\xi, \hat{\theta}) - r_1(\xi, \hat{\theta}^B)$ is the loss that occurs when using $\hat{\theta}$ instead of the Bayes estimator.

The generalized relative savings loss of $\hat{\theta}_{LB}^c$ is given by

$$GRSL(\hat{\theta}_{LB}^c; X) = (1 - s_c)I_p,$$  \hspace{1cm} (2–16)

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

$$1 - s_c = P(\chi^2_{p+2} > c^2) - c\sqrt{\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p+2}{2}\right)}}P(\chi^2_{p+1} > c^2) + \frac{c^2}{p}P(\chi^2_{p} > c^2).$$  \hspace{1cm} (2–17)

**Proof.** We first write

$$r_1(\xi, \hat{\theta}_{LB}^c) = E\{(\theta - \hat{\theta}_{LB}^c)(\theta - \hat{\theta}_{LB}^c)^T\}$$

$$= E\{(\theta - \hat{\theta}^B + \hat{\theta}^B - \hat{\theta}_{LB}^c)(\theta - \hat{\theta}^B + \hat{\theta}^B - \hat{\theta}_{LB}^c)^T\}$$

$$= E\{(\theta - \hat{\theta}^B)(\theta - \hat{\theta}^B)^T\} + E\{(\hat{\theta}^B - \hat{\theta}_{LB}^c)(\hat{\theta}^B - \hat{\theta}_{LB}^c)^T\},$$  \hspace{1cm} (2–18)

where the cross product terms do not appear since $E(\theta|X) = \hat{\theta}^B$ and thus

$$E\{(\theta - \hat{\theta}^B)(\hat{\theta}^B - \hat{\theta}_{LB}^c)^T\} = E\{E(\theta - \hat{\theta}^B)(\hat{\theta}^B - \hat{\theta}_{LB}^c)^T|X\} = 0.$$  \hspace{1cm} (2–19)
Also, since \( \hat{\theta}^B \) is the posterior mean, it follows from (2–1) that

\[
E\{(\theta - \hat{\theta}^B)(\theta - \hat{\theta}^B)^T\} = (I_p - B)\Sigma. \tag{2–20}
\]

We now need to calculate \( E\{(\hat{\theta}^B - \hat{\theta}^{LB}_c)(\hat{\theta}^B - \hat{\theta}^{LB}_c)^T\} \). Note that

\[
\hat{\theta}^B - \hat{\theta}^{LB}_c = B(X - \mu)\{1 - \rho_c(||A + \Sigma)^{-\frac{1}{2}}(X - \mu)||^2\}, \tag{2–21}
\]

and we thus have

\[
E\{(\hat{\theta}^B - \hat{\theta}^{LB}_c)(\hat{\theta}^B - \hat{\theta}^{LB}_c)^T\} = BE[(X - \mu)(X - \mu)^T\{1 - \rho_c(||A + \Sigma)^{-\frac{1}{2}}(X - \mu)||^2\}]B^T. \tag{2–22}
\]

Let \( Z = (A + \Sigma)^{-\frac{1}{2}}(X - \mu) \sim N_p(0, I_p) \), and it follows that

\[
E(\hat{\theta}^B - \hat{\theta}^{LB}_c)(\hat{\theta}^B - \hat{\theta}^{LB}_c)^T = \Sigma(A + \Sigma)^{-\frac{1}{2}}E\{|Z|^{2}[1 - \rho_c(||Z||^2)]^2\}(A + \Sigma)^{-\frac{1}{2}}\Sigma. \tag{2–23}
\]

The following lemma simplifies the calculation of the Bayes risk.

**Lemma 2.3.2.** Consider the random vector \( Y \sim N_p(0, \tau^2 I_p) \). Then the random scalar \( ||Y||^2 \) and the random matrix \( YY^T/||Y||^2 \) are independently distributed.

**Proof.** First, \( ||Y||^2 \) is complete and sufficient for \( \tau^2 \). Noting that \( \tau^{-1}Y \sim N_p(0, I_p) \), the statistic \( YY^T/||Y||^2 \) is ancillary. Now the independence of \( ||Y||^2 \) and \( YY^T/||Y||^2 \) follows from the well known theorem of Basu. \( \square \)

We now continue with the calculation of the Bayes risk. By Lemma 2.3.2

\[
E\left\{\frac{ZZ^T}{||Z||^2}[1 - \rho_c(||Z||^2)]^2\right\} = E\left\{\frac{ZZ^T}{||Z||^2}||Z||^2[1 - \rho_c(||Z||^2)]^2\right\}
\]

\[
= E\left(\frac{ZZ^T}{||Z||^2}\right)E\{|Z|^2[1 - \rho_c(||Z||^2)]^2\}. \tag{2–24}
\]

Again by Lemma 2.3.2, we have

\[
E(ZZ^T) = E\left(\frac{ZZ^T}{||Z||^2}||Z||^2\right) = E\left(\frac{ZZ^T}{||Z||^2}\right)E\left(||Z||^2\right), \tag{2–25}
\]

28
and thus
\[ E(\frac{ZZ^T}{||Z||^2}) = \frac{E(ZZ^T)}{E(||Z||^2)} = p^{-1}I_p. \] (2–26)

Since \( ||Z||^2 \sim \chi^2_p \),
\[ E\{||Z||^2[1 - \rho_c(||Z||^2)]^2\} = E\{Y[1 - \rho_c(Y)]^2\} \]
\[ = \int_0^\infty [1 - \rho_c(y)]^2 \exp\left(-\frac{y}{2}\right) \frac{py^{p+2}}{2^{p+2} \Gamma(p+2)} \, dy = pE\{[1 - \rho_c(U)]^2\} \] (2–27)
where \( U \sim \chi^2_{p+2} \). It follows from Equations 2–24, 2–26 and 2–27 that
\[ E\{ZZ^T[1 - \rho_c(||Z||^2)]^2\} = E\{[1 - \rho_c(U)]^2\}I_p. \] (2–28)

Equations 2–18, 2–20, 2–23 and 2–28 show that
\[ r_1(\xi, \hat{\lambda}^{LB}_c) = \Sigma - B\Sigma[1 - E\{1 - \rho_c(U)\}^2] \]
\[ = \Sigma E\{1 - \rho_c(U)\}^2 + (\Sigma - B\Sigma)[1 - E\{1 - \rho_c(U)\}^2] \]
\[ = \Sigma(1 - sc) + (I_p - B)\Sigma sc \]
\[ = r_1(\xi, X)(1 - sc) + r_1(\xi, \hat{\theta}^B)sc, \] (2–29)
where \( 1 - sc = E\{1 - \rho_c(U)\}^2 \).

Now, by choosing \( \rho_c(U) = \min(1, c/\sqrt{U}) \), where \( U \sim \chi^2_{p+2} \), we have
\[ E\{1 - \rho_c(U)\}^2 = E\{1 - \min(1, \frac{c}{\sqrt{U}})\}^2 \]
\[ = E\{(1 - \frac{c}{\sqrt{U}})^2I(U > c^2)\} \]
\[ = E\{I(U > c^2)\} - 2cE\{U^{-\frac{1}{2}}I(U > c^2)\} + c^2E\{U^{-1}I(U > c^2)\} \]
\[ = P(\chi^2_{p+2} > c^2) - c\sqrt{\frac{2\Gamma(p+1)}{\Gamma(p+2)}} P(\chi^2_{p+1} > c^2) + p^{-1}c^2P(\chi^2_p > c^2), \] (2–30)
which, for fixed \( p \) depends only on \( c \) and it is independent of the model parameters.

Further, if we feel that the Bayes rule is irrelevant for observations that have standardized norm bigger than some value \( c_0 \) say, \( c_0 > c \), we can modify the relevance
function in the following manner:

\[
\rho^*_c(U) = \begin{cases} 
\min(1, c/\sqrt{U}) & \text{if } U \leq c_0^2 \\
0 & \text{if } U > c_0^2.
\end{cases}
\] (2–31)

The cost of this modification in terms of increased generalized relative savings loss is

\[
E\{1 - \rho^*_c(U)\}^2 - E\{1 - \rho_c(U)\}^2 = E\{(1 - \frac{c}{\sqrt{U}})^2 I(c^2 < U < c_0^2)\} + E\{I(U > c_0^2)\} - E\{(1 - \frac{c}{\sqrt{U}})^2 I(c^2 < U)\}
\]
\[
= P(U > c_0^2) - E\{(1 - \frac{c}{\sqrt{U}})^2 I(U > c_0^2)\}
\]
\[
= 2cE\{U^{-\frac{1}{2}} I(U > c_0^2)\} - c^2 E\{U^{-1} I(U > c_0^2)\}
\]
\[
= c\sqrt{2} \Gamma\left(\frac{p+1}{2}\right) \Gamma\left(\frac{p+2}{2}\right) P(\chi^2_{p+1} > c_0^2) - p^{-1} c^2 P(\chi^2_p > c_0^2).
\] (2–32)

We have seen that the Bayes risk of the limited translation rule is a weighted average of the Bayes risks of the Bayes rule and the MLE, the weights being \( s_c \) and \( 1 - s_c \) respectively, which causes a generalized relative savings loss of \( (1 - s_c) I_p \). Also, the weight of the Bayes risk of the MLE, \( 1 - s_c \), for fixed \( p \), is a decreasing convex function of \( c \). This allows the statistician to choose \( c \) by deciding by what proportion it is worth increasing the Bayes risk of the Bayes rule in order to receive protection against large frequentist risks.

2.4 Frequentist Risk of the Limited Translation Bayes Estimators

We now turn our attention to the frequentist risk of \( \hat{\theta}_{LB}^c \), a function of \( \theta \) denoted by \( R_1(\theta, \hat{\theta}_{LB}^c) \), to show that the limited translation rule, in return for the increased Bayes risk, does not allow the frequentist risk to be very large, in contrast with the Bayes rule. The calculation of the frequentist risk of the limited translation rule was feasible only under the simplifying assumption that the prior variance-covariance matrix is a multiple of the sampling variance-covariance matrix, that is \( A = g\Sigma \), where \( g > 0 \) is a known positive scalar. That is, we consider the case where \( X|\theta \sim N_p(\theta, \Sigma) \) while the prior distribution is
taken to be \( \theta \sim \xi \equiv N_p(\mu, g\Sigma) \). Such priors, originally introduced by Zellner (1986), are called g-priors.

Under the assumed model, \( B \) reduces to \( B = (1 + g)^{-1}I_p \) and the Bayes estimator is given by \( \hat{\theta}^B = X - (1 + g)^{-1}(X - \mu) \). The frequentist risk associated with it is obtained from Equation 2–8,

\[
R_1(\theta, \hat{\theta}^B) = g^2(1 + g)^{-2}\Sigma + (1 + g)^{-2}(\mu - \theta)(\mu - \theta)^T.
\] (2–33)

Also, the limited translation estimator is given by

\[
\hat{\theta}^{LB}_c = X - (1 + g)^{-1}(X - \mu)\rho_c(||(1 + g)^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}(X - \mu)||^2).
\] (2–34)

We would like to compare the frequentist risk of the Bayes estimator to the frequentist risk of the limited translation estimator. An expression of the latter is provided by the following Theorem.

**Theorem 2.4.1.** Under the assumption that \( A = g\Sigma \), 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 rule is given by

\[
R_1(\theta, \hat{\theta}^{LB}_c) = R_1(\theta, \hat{\theta}^B) + (\theta - \mu)(\theta - \mu)^T\left[ (1 + 2g)(1 + g)^{-2}P[\chi^2_p(\lambda) > c^2(1 + g)]
- 2(1 + g)^{-1}P[\chi^2_{p+2}(\lambda) > c^2(1 + g)]
+ c^2(1 + g)^{-1}E_\lambda\{[\chi^2_{p+4}(\lambda)]^{-1}I[\chi^2_{p+4}(\lambda) > c^2(1 + g)]\}
+ 2c(1 + g)^{-\frac{1}{2}}E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda) > c^2(1 + g)]\}
- 2c(1 + g)^{-\frac{1}{2}}E_\lambda\{[\chi^2_{p+4}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda) > c^2(1 + g)]\}
+ \Sigma\left[ (1 + 2g)(1 + g)^{-2}P[\chi^2_{p+2}(\lambda) > c^2(1 + g)]
+ c^2(1 + g)^{-1}E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-1}I[\chi^2_{p+2}(\lambda) > c^2(1 + g)]\}
- 2c(1 + g)^{-\frac{1}{2}}E_\lambda\{[\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda) > c^2(1 + g)]\}\right].
\] (2–35)
where \( \lambda = (\theta - \mu)^T \Sigma^{-1}(\theta - \mu)/2 \) and \( \chi^2_k(\lambda) \) denotes a non-central chi-square variable with non-centrality parameter \( \lambda \) and degrees of freedom \( k \).

The proof of the theorem is given in section A.2 of Appendix A.

Since the risks in Equations 2–33 and 2–35 involve matrices, in order to graphically compare them, we consider scalar versions of them. Specifically, we consider the quadratic loss function

\[
L_2(\theta, a) = (\theta - a)^T \Sigma^{-1}(\theta - a). \tag{2–36}
\]

It is easy to show using Equation 2–33 that the risk of the Bayes rule under the loss function \( L_2 \) is equal to

\[
R_2(\theta, \hat{\theta}^B) = E_{\theta}\{(\theta - \hat{\theta}^B)^T \Sigma^{-1}(\theta - \hat{\theta}^B)\} = \text{tr}[\Sigma^{-1}R_1(\theta, \hat{\theta}^B)] = \frac{2\lambda + pg^2}{(1+g)^2}. \tag{2–37}
\]

The following Corollary provides an expression for the risk of the limited translation Bayes estimator under the loss function \( L_2 \).

**Corollary 2.4.2.** Under the loss function \( L_2 \) the risk of the limited translation Bayes estimator is given by

\[
R_2(\theta, \hat{\theta}^{LB}_c) = R_2(\theta, \hat{\theta}^B) + P[\chi^2_{p+2}(\lambda) > c^2(1+g)]\left\{\frac{p + 2pg}{(1+g)^2} - \frac{4\lambda}{1+g}\right\}
+ 2\lambda(1 + 2g)(1 + g)^{-2}P[\chi^2_{p+4}(\lambda) > c^2(1+g)] + c^2(1 + g)^{-1}P[\chi^2_p(\lambda) > c^2(1 + g)]
+ 2c(1 + g)^{-\frac{1}{2}}\left[2\lambda E_{\lambda}\{[\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda) > c^2(1+g)]\}
- E_{\lambda}\{[\chi^2_p(\lambda)]^{\frac{1}{2}}I[\chi^2_p(\lambda) > c^2(1+g)]\}\right]. \tag{2–38}
\]

The proof is given in section A.3 of Appendix A.

The bracketed term in the last two lines of Equation 2–38 can be calculated as

\[
\sqrt{2} \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \frac{\Gamma\left(\frac{p+1+2k}{2}\right)}{\Gamma\left(\frac{p+2k}{2}\right)} P[\chi^2_{p+1+2k} > c^2(1 + g)] \left(\frac{2\lambda}{p + 2k} - 1\right). \tag{2–39}
\]

The risk of \( \hat{\theta}^{LB}_c \), for fixed \( p \) and \( g \), quite conveniently, is a function only of the non-centrality parameter \( \lambda = (\theta - \mu)^T \Sigma^{-1}(\theta - \mu)/2 \) and so is the risk in Equation 2–37.
Let us now consider the hypothetical scenario where the statistician is given \( n \) observations of dimension \( p = 3 \). 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 Figure 2.5 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 Figure 2-2 we see how the risks in Equations 2–37 and 2–38 behave as the non-centrality parameter \( \lambda \) increases. For small values of \( \lambda \), i.e. when \( \theta \) is close to the population mean \( \mu \), the Bayes rule has slightly smaller frequentist risk than the limited translation Bayes rule. However, the frequentist risk of the Bayes rule increases linearly with the non-centrality parameter which clearly means that the Bayes rule has high risk when the \( \theta \) is far from \( \mu \). On the contrary, the frequentist risk of the limited translation Bayes rule becomes flat after \( \lambda \) exceeds a certain value. That is, the limited translation rule does not allow large frequentist risks even if the unobservable \( \theta \) is far from the prior mean \( \mu \).

Returning to Equation 2–38, we write \( R_2(\theta, \hat{\theta}_{c}^{LB}) = R_2(\theta, \hat{\theta}^B) + e_{p,c,g}(\lambda) \). The proposed estimator \( \hat{\theta}_{c}^{LB} \), does better than the Bayes estimator when the function \( e_{p,c,g}(\lambda) \) takes on negative values. This, in general, happens when attempting to estimate a random effect \( \theta \) which departs widely from the assumed prior mean \( \mu \), that is, when the non-centrality parameter \( \lambda \) takes on large values. The questions of interest are what values must \( \lambda \) take, for fixed values of \( p, c \) and \( g \), in order for the function \( e_{p,c,g}(\lambda) \) to become negative, and how likely those values are.

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

$$P(\lambda \geq k) = P[(\theta - \mu)^T \Sigma^{-1}(\theta - \mu)/2 \geq k] = P(\chi^2_p \geq 2kg^{-1}).$$

(2-40)

Table 2-1 shows the values $k$ and the corresponding probabilities $P(\lambda \geq k)$ for the case where the dimension is $p = 3$, for five different values of the prior parameter $g$ and for three values of $c$. We may recall that $c$ and $1 - s_c$ are one to one functions and thus the Table provides the generalized relative savings loss, $1 - s_c$, along with the corresponding $c$.

Observing the first row of Table 2-1, it is clear that for all values of $g$, $P(\lambda \geq k)$ is bigger that 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 2-1. 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, $P(\lambda \geq k)$, are calculated assuming that the prior $\xi$ is the true one. We can expect the probabilities $P(\lambda \geq k)$ to increase with the increasing distance of $\xi$ from the true prior.

The results of Tables 2-2 and 2-3, where we have chosen $p = 5$ and $p = 10$ 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 for $1 - s_c = 10\%$ of the Bayes risk are bigger that 10% for $g = 2, 5$ and 10 but they are smaller than 10% for $g = 0.5$ and 1.

2.5 Robustness of Limited Translation Bayes Estimators

In this section we investigate how well the limited translation Bayes rule performs when the assumed prior deviates from the true prior. We consider the same model as in Section 2.4. That is, $X|\theta \sim N_p(\theta, \Sigma)$ and $\theta \sim \xi \equiv N_p(\mu, g\Sigma)$. 

34
Under the assumed model, the Bayes estimator, namely the posterior mean of \( \theta \) given \( X \), is 
\[
\hat{\theta}^B = X - (1 + g)^{-1}(X - \mu),
\]
and under the matrix loss function \( L_1 \), given in 2–2, its Bayes risk is
\[
r_1(\xi, \hat{\theta}^B) = g(1 + g)^{-1}\Sigma.
\] (2–41)

Under \( L_2 \), the quadratic loss function of 2–36, the Bayes risk of the Bayes rule is
\[
r_2(\xi, \hat{\theta}^B) = \text{tr}[g(1 + g)^{-1}\Sigma^{-1}\Sigma] = pg(1 + g)^{-1}.
\] (2–42)

Now suppose that the true prior is \( \theta \sim \xi^* \equiv N(\mu^*, g^*\Sigma) \). Then the Bayes risk of \( \hat{\theta}^B \) under the \( L_1 \) and \( L_2 \) losses is
\[
r_1(\xi^*, \hat{\theta}^B) = (g^2 + g^*)(1 + g)^{-2}\Sigma + (1 + g)^{-2}(\mu^* - \mu)(\mu^* - \mu)^T,
\] (2–43)

and
\[
r_2(\xi^*, \hat{\theta}^B) = p(g^2 + g^*)(1 + g)^{-2} + (1 + g)^{-2}(\mu^* - \mu)^T\Sigma^{-1}(\mu^* - \mu)
\]
\[
= \{p(g^2 + g^*) + 2\lambda(1 + g^*)\}(1 + g)^{-2},
\] (2–44)

respectively, where \( \lambda = 2^{-1}(1 + g^*)^{-1}(\mu^* - \mu)^T\Sigma^{-1}(\mu^* - \mu) \). When \( \lambda = 0 \) and \( g^* \neq g \), that is when the prior mean has been correctly specified but \( g^* \) has been over or under estimated, the Bayes risk \( r_2(\xi, \hat{\theta}^B) \), over or under estimates the true Bayes risk, \( r_2(\xi^*, \hat{\theta}^B) \). Also, \( r_2(\xi^*, \hat{\theta}^B) \) increases linearly with \( \lambda \) and thus \( r_2(\xi, \hat{\theta}^B) \) underestimates the true Bayes risk when the prior mean is misspecified.

We now consider the limited translation Bayes rule. Under the assumed model the limited translation estimator is given by
\[
\hat{\theta}_{LB}^c = X - (1 + g)^{-1}(X - \mu)\rho_c(||(1 + g)^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}(X - \mu)||^2),
\] (2–45)
and using the result of Theorem 2.3.1 it is easy to see that

\[ r_1(\xi, \hat{\theta}_c^{LB}) = \Sigma(1 - s_c) + \Sigma(1 - \frac{1}{1 + g})s_c = (1 - \frac{s_c}{1 + g})\Sigma, \]  \hspace{1cm} (2–46)

which implies that

\[ r_2(\xi, \hat{\theta}_c^{LB}) = p(1 + g)^{-1}(1 + g - s_c) = p\{1 - s_c/(1 + g)\}. \]  \hspace{1cm} (2–47)

The following theorem provides an expression for the Bayes risk under prior \( \xi^* \) of the limited translation Bayes rule obtained under the assumed prior \( \xi \).

**Theorem 2.5.1.** For the relevance function \( \rho_c(u) = \min(1, c/\sqrt{u}) \), the Bayes risk under \( \xi^* \) of the multivariate limited translation rule obtained under the assumed prior \( \xi \) is given by

\[
\begin{align*}
r_1(\xi^*, \hat{\theta}_c^{LB}) &= r_1(\xi^*, \hat{\theta}^B) \\
&= (1 + g)^{-2}(\mu^* - \mu)(\mu^* - \mu)^T\left[ - (g^* - 2g - 1)(1 + g^*)^{-1}P[\chi_{p+4}^2(\lambda) > c^2d] \\
&\quad - 2dP[\chi_{p+2}^2(\lambda) > c^2d] + c^2dE_{\lambda}\{[\chi_{p+4}^2(\lambda)]^{-1}I[\chi_{p+4}^2(\lambda) > c^2d]\} \\
&\quad - 2cd\left\{ E_{\lambda}\{[\chi_{p+4}^2(\lambda)]^{-\frac{1}{2}}I[\chi_{p+4}^2(\lambda) > c^2d]\} - E_{\lambda}\{[\chi_{p+2}^2(\lambda)]^{-\frac{1}{2}}I[\chi_{p+2}^2(\lambda) > c^2d]\}\right\} \\
&\quad + (1 + g)^{-1}\Sigma\left[ - (g^* - 2g - 1)(1 + g)^{-1}P[\chi_{p+2}^2(\lambda) > c^2d] \\
&\quad + cE_{\lambda}\{[\chi_{p+2}^2(\lambda)]^{-1} - 2d^2[\chi_{p+2}^2(\lambda)]^{-\frac{1}{2}}I[\chi_{p+2}^2(\lambda) > c^2d]\}\right], \tag{2–48}
\end{align*}
\]

where \( d = f(g, g^*) = (1 + g)(1 + g^*)^{-1} \).

The proof is given in section A.4 of Appendix A, while in section A.5 we prove the following result.

**Corollary 2.5.2.** For the loss function \( L_2 \), the Bayes risk under \( \xi^* \) of the multivariate limited translation rule obtained under the assumed prior \( \xi \) is given by

\[
\begin{align*}
r_2(\xi^*, \hat{\theta}_c^{LB}) &= r_2(\xi^*, \hat{\theta}^B) - \{4\lambda + p(d^{-1} - 2)\}(1 + g)^{-1}P[\chi_{p+2}^2(\lambda) > c^2d] \\
&\quad - 2\lambda(g^* - 2g - 1)(1 + g)^{-2}P[\chi_{p+4}^2(\lambda) > c^2d] + c^2(1 + g)^{-1}P[\chi_{p}^2(\lambda) > c^2d] \\
&\quad - 2c(1 + g)^{-\frac{1}{2}}(1 + g^*)^{-\frac{1}{2}}E_{\lambda}\{[\chi_{p}^2(\lambda)]^{\frac{1}{2}}I[\chi_{p}^2(\lambda) > c^2d]\} \\
&\quad + 4c\lambda(1 + g)^{-\frac{1}{2}}(1 + g^*)^{-\frac{1}{2}}E_{\lambda}\{[\chi_{p+2}^2(\lambda)]^{-\frac{1}{2}}I[\chi_{p+2}^2(\lambda) > c^2d]\}. \tag{2–49}
\end{align*}
\]

36
The last two lines of the above expression are calculated as

\[
2^3 c(1 + g)^{-\frac{1}{2}}(1 + g^*)^{-\frac{1}{2}} \sum_{k=0}^{\infty} \frac{e^{-\lambda k}}{k!} \frac{\Gamma\left(\frac{b+1+2k}{2}\right)}{\Gamma\left(\frac{b+2k}{2}\right)} P[\chi^2_{p+1+2k} > c^2d] \left\{ \frac{2\lambda}{p + 2k} - 1 \right\}. \tag{2-50}
\]

We now revisit the example of the previous section where we supposed that we are given \(n\) observations of dimension \(p = 3\) and based on our prior beliefs we set \(g = 2\). The choice of \(c = 1.52\) corresponds to generalized relative savings loss of \(1 - s_c = 10\%\). When the prior parameters are correctly specified, the Bayes risk of the Bayes rule and of the limited translation rule are \(r_2(\xi, \hat{\theta}^B) = 2\) and \(r_2(\xi, \hat{\theta}^{LB}_c) = 2.1\) respectively.

In Figure 2-3 (a) we plot the risk functions \(r_2(\xi^*, \hat{\theta}^B)\) and \(r_2(\xi^*, \hat{\theta}^{LB}_c)\) for values of the non-centrality parameter \(\lambda\) ranging from 0 to 15 and assuming that \(g^* = g = 2\). In the same graph we plot \(r_2(\xi, \hat{\theta}^B)\) and \(r_2(\xi, \hat{\theta}^{LB}_c)\) which, however, do not depend on \(\lambda\). We see that for very small values of \(\lambda\), the Bayes rule has smaller risk than the limited translation rule. However, the Bayes risk of the Bayes rule increases linearly with \(\lambda\) while the Bayes risk of the limited translation rule increases in a much smaller rate.

In Figure 2-3 (b) we plot the same four risk functions for values of \(g\) ranging from 0.2 to 10 and for \(\lambda = 0\), that is assuming that the prior mean is correctly specified. We see that for values of \(g\) close to the true value, \(g^* = g = 2\), \(r_2(\xi^*, \hat{\theta}^B)\) is less than \(r_2(\xi^*, \hat{\theta}^{LB}_c)\). However, when \(g^*\) is underestimated the limited translation rule does better that the regular Bayes estimator. As the assumed value \(g\), of \(g^*\) becomes bigger than the true value of \(g^*\), the Bayes risk performance of the two estimators becomes similar. As \(g\) increases, the two estimators become closer to the MLE and their Bayes risk tends to the Bayes risk of the MLE, \(r_2(\xi^*, X_i) = p\), and here we have taken \(p = 3\).
Figure 2-1. Plot of $1 - s_c$ as a function of $c$.

Figure 2-2. Plot of risk as function of the non-centrality parameter, $\lambda$. 
Table 2-1. Minimum values, $k$, of $\lambda$ and $P(\lambda \geq k)$ for $p = 3$

<table>
<thead>
<tr>
<th>$g$</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c=2.466$</td>
<td>$1 - s_c=1%$</td>
<td>1.79</td>
<td>3.81</td>
<td>8.27</td>
<td>22.14</td>
</tr>
<tr>
<td>$c=1.840$</td>
<td>$1 - s_c=5%$</td>
<td>1.72</td>
<td>3.32</td>
<td>6.72</td>
<td>16.96</td>
</tr>
<tr>
<td>$c=1.521$</td>
<td>$1 - s_c=10%$</td>
<td>1.76</td>
<td>3.21</td>
<td>6.23</td>
<td>15.21</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>$g$</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c=2.806$</td>
<td>$1 - s_c=1%$</td>
<td>2.49</td>
<td>5.12</td>
<td>10.80</td>
<td>28.31</td>
</tr>
<tr>
<td>$c=2.144$</td>
<td>$1 - s_c=5%$</td>
<td>2.51</td>
<td>4.69</td>
<td>9.24</td>
<td>22.88</td>
</tr>
<tr>
<td>$c=1.797$</td>
<td>$1 - s_c=10%$</td>
<td>2.63</td>
<td>4.68</td>
<td>8.88</td>
<td>21.38</td>
</tr>
</tbody>
</table>

Table 2-3. Minimum values, $k$, of $\lambda$ and $P(\lambda \geq k)$ for $p = 10$

<table>
<thead>
<tr>
<th>$g$</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c=3.490$</td>
<td>$1 - s_c=1%$</td>
<td>4.17</td>
<td>8.24</td>
<td>16.82</td>
<td>42.98</td>
</tr>
<tr>
<td>$c=2.757$</td>
<td>$1 - s_c=5%$</td>
<td>4.41</td>
<td>8.03</td>
<td>15.42</td>
<td>37.57</td>
</tr>
<tr>
<td>$c=2.352$</td>
<td>$1 - s_c=10%$</td>
<td>4.77</td>
<td>8.33</td>
<td>15.53</td>
<td>37.10</td>
</tr>
</tbody>
</table>
Figure 2-3. Bayes risk under misspecified priors. ‘LB True Prior’ refers to $r_2(\xi^*, \hat{\theta}_{LB}^*)$, ‘Bayes True Prior’ refers to $r_2(\xi^*, \hat{\theta}_B^*)$, ‘LB Assumed Prior’ refers to $r_2(\xi, \hat{\theta}_{LB}^*)$, ‘Bayes Assumed Prior’ refers to $r_2(\xi, \hat{\theta}_B^*)$. (a) Bayes risk as function of the non-centrality parameter $\lambda$ and (b) Bayes risk as function of $g^*$. 
CHAPTER 3
MULTIVARIATE LIMITED TRANSLATION
EMPIRICAL BAYES ESTIMATORS: THE CASE OF UNKNOWN PRIOR MEAN

In this chapter we develop limited translation estimators assuming that the prior
mean is unknown but both the sampling and the prior variance-covariance matrices
are known. The organization of the sections of this chapter is as follows. In Section
3.1 we briefly review the Bayes and empirical Bayes estimators as well as the notion of
influence functions. In section 3.2 we introduce the limited translation estimators and in
Section 3.3 we evaluate their Bayes risk performance under the assumed prior. Section
3.4 evaluates their frequentist risk performance while in Section 3.5 we compare the Bayes
risk performance of the two competing estimators assuming misspecification of the prior
distribution. Some of the long algebraic derivations are provided in the Appendix B.

3.1 Bayes, Empirical Bayes Estimators and Influence Functions

We consider the Bayesian example of estimation where the \( n \) random vectors \( X_i, \)
\( i = 1, \ldots, n \), are independently distributed according to the \( p \)-dimensional normal
distribution \( X_i|\theta_i \overset{\text{ind}}{\sim} N_p(\theta_i, \Sigma) \), where \( \Sigma \) is known. Also, the \( \theta_i, i = 1, \ldots, n \), are iid
according to \( \theta_i \overset{\text{iid}}{\sim} \xi \equiv N_p(\mu, A) \), where \( A \) is known. The interest is in estimating the \( \theta_i \)
under the matrix loss function \( L_1(\theta_i, a_i) = (\theta_i - a_i)(\theta_i - a_i)^T \).

First assume that \( \mu \) is known. The Bayes estimator of \( \theta_i \), that is the estimator that
minimizes the posterior risk, is given by the posterior mean of \( \theta_i \) given \( X_i = x_i \).
In order to find this estimator we calculate the posterior distribution of \( \theta_i \) given \( X_i = x_i \)

\[
f(\theta_i|X_i) \propto f(X_i|\theta_i)f(\theta_i)
\]

\[
\propto \exp\{-2^{-1}(X_i - \theta_i)^T\Sigma^{-1}(X_i - \theta_i) - 2^{-1}(\theta_i - \mu)^T A^{-1}(\theta_i - \mu)\}
\]

\[
\propto \exp\{-2^{-1}[\theta_i^T(\Sigma^{-1} + A^{-1})\theta_i - 2\theta_i^T(\Sigma^{-1}X_i + A^{-1}\mu)]\}. \quad (3-1)
\]

It follows that

\[
\theta_i|X_i = x_i \overset{\text{ind}}{\sim} N_p[(\Sigma^{-1} + A^{-1})^{-1}(\Sigma^{-1}x_i + A^{-1}\mu), (\Sigma^{-1} + A^{-1})^{-1}]. \quad (3-2)
\]
Noting that $(\Sigma^{-1} + A^{-1})^{-1} = (\Sigma^{-1}(A + \Sigma)A^{-1})^{-1} = (A^{-1}(A + \Sigma)\Sigma^{-1})^{-1}$, the posterior mean, that is the Bayes estimator of $\theta_i$, can be written as weighted average of $X_i$ and $\mu$ since

$$E(\theta_i|X_i) = (\Sigma^{-1} + A^{-1})^{-1}(\Sigma^{-1}X_i + A^{-1}\mu)$$

$$= A(A + \Sigma)^{-1}X_i + \Sigma(A + \Sigma)^{-1}\mu = (I_p - B)X_i + B\mu,$$

(3–3)

where $B = \Sigma(A + \Sigma)^{-1}$. Further, the posterior variance, also the Bayes risk of the Bayes estimator, can be written as $\text{var}(\theta_i|X_i) = \Sigma - \Sigma(A + \Sigma)^{-1}\Sigma = (I_p - B)\Sigma$.

Since $\mu$ is unknown the Bayes estimator, $\hat{\theta}^B_i = (I_p - B)X_i + B\mu$, cannot be used as such. However, marginally the random vectors $X_i$ have mean $\mu$, and thus the unknown prior mean $\mu$ can be replaced by an unbiased estimator, $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$, thus resulting in an empirical Bayes (EB) estimator

$$\hat{\theta}^{EB}_i = (I_p - B)X_i + B\bar{X}_n,$$

(3–4)

of $\theta_i$. The same estimator can be obtained as a hierarchical Bayes estimator when one assigns a uniform prior distribution on $\mu$.

The EB estimator shrinks every maximum likelihood estimator (MLE), $X_i$, towards the grand mean, $\bar{X}_n$, the MLE of the unknown prior mean, $\mu$. In doing so, it attains a lower Bayes risk than the MLE, under the assumed prior. However, it results in high Bayes risk when the prior distribution is misspecified. It also results in high frequentist risk when attempting to estimate parameters, $\theta_i$, that are far from the grand mean.

On the other hand, the MLE has minimax risk equal to $\Sigma$ for all $\theta_i$. In order to avoid these two problems, we develop robust EB estimators, namely the limited translation EB estimators.

We start by assigning the noninformative prior on $\mu \sim \text{Uniform}(\mathbb{R}^p)$. We then find the influence of observations $X_i$, $i = 1, \ldots, n$, on the posterior distribution of $\mu$. The influence is measured using the general divergence formula introduced by Cressie & Read.
Let \( f_1 \) and \( f_2 \) denote two density functions. Then the general divergence measure is given by

\[
D_\lambda(f_1, f_2) = \lambda^{-1}(\lambda + 1)^{-1} E_{f_1}\{ (f_1/f_2)^\lambda - 1 \}.
\] (3–5)

Here, \( f_1 \) and \( f_2 \) denote the posterior densities of \( \mu \) given \( X = (X_1^T, \ldots, X_n^T)^T \) and \( X^{(-i)} = (X_1^T, \ldots, X_{i-1}^T, X_{i+1}^T, \ldots, X_n^T)^T \) respectively. In order to find \( f_1 \) and \( f_2 \), note that \( X_i|\mu \sim \mathcal{N}_p(\mu, \Sigma + A) \), \( i = 1, \ldots, n \), and \( \mu \sim \text{Uniform}(\mathbb{R}^p) \). It follows that

\[
f(\mu|X_1, \ldots, X_n) \propto \exp\left[ -2^{-1}\{ \mu^T n(\Sigma + A)^{-1}\mu - 2\mu^T n(\Sigma + A)^{-1} \bar{X}_n \} \right]
\]

\[
\propto \exp\left[ -2^{-1}n(\mu - X_n)^T(\Sigma + A)^{-1}(\mu - X_n) \right].
\] (3–6)

We thus have

\[
\mu|X_1, \ldots, X_n \sim \mathcal{N}_p(\bar{X}_n, n^{-1}(\Sigma + A)),
\]

\[
\mu|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n \sim \mathcal{N}_p(\bar{X}_{n-1}^{(-i)}, (n - 1)^{-1}(\Sigma + A)),
\] (3–7)

where \( \bar{X}_{n-1}^{(-i)} \) is the average calculated using all the random vectors except the \( i \)th one.

The following Theorem provides a result concerning the divergence between two multivariate normal densities.

**Theorem 3.1.1.** Let \( f_1 \) denote the \( \mathcal{N}_p(\mu_1, \Sigma_1) \) density and \( f_2 \) denote the \( \mathcal{N}_p(\mu_2, \Sigma_2) \) density. Then

\[
D_\lambda(f_1, f_2) = \lambda^{-1}(\lambda + 1)^{-1}\left[ |\Sigma_2|^{\frac{1}{2}}|\Sigma_1|^{-\frac{1}{2}}|(1 + \lambda)\Sigma_2 - \lambda\Sigma_1|^{-\frac{1}{2}}\right.
\]

\[
\times \exp\left\{ \frac{\lambda(\lambda + 1)}{2}(\mu_1 - \mu_2)^T[(1 + \lambda)\Sigma_2 - \lambda\Sigma_1]^{-1}(\mu_1 - \mu_2) \right\} - 1 \right].
\] (3–8)

If the variance-covariance matrices \( \Sigma_1 \) and \( \Sigma_2 \) are known, which is the case for our estimation problem, the divergence measure is a one to one function with

\[
(\mu_1 - \mu_2)^T[(1 + \lambda)\Sigma_2 - \lambda\Sigma_1]^{-1}(\mu_1 - \mu_2).
\] (3–9)
For the special case where the densities $f_1$ and $f_2$ are the ones given in 3–7 we have

$$\mu_1 - \mu_2 = \bar{X}_n - \bar{X}_{n-1}^{(i)} = (n - 1)^{-1}(X_i - \bar{X}_n)$$

and

$$(1 + \lambda)\Sigma_2 - \lambda\Sigma_1 = (\Sigma + A)(n + \lambda)n^{-1}(n - 1)^{-1}.$$ It is now easy to see that the divergence measure is a one to one function with

$$\begin{align*}
(X_i - \bar{X}_n)^T \frac{n(\Sigma + A)^{-1}}{(n - 1)(n + \lambda)}(X_i - \bar{X}_n)^T,
\end{align*}$$

which is a quadratic form in $(X_i - \bar{X}_n)$. Based on this result we will obtain some robust Bayesian estimators in the following section.

### 3.2 Limited Translation Empirical Bayes Estimators

A modification of the EB estimator will give us the limited translation EB estimator. Since the influence of the random vectors $X_i$, $i = 1, \ldots, n$, depends on their distance from $\bar{X}_n$, in the EB estimator we want to control the standardized distance of $X_i$ to $\bar{X}_n$. We thus define $D \equiv \text{var}(X_i - \bar{X}_n) = (1 - 1/n)(\Sigma + A)$, and we write

$$\hat{\theta}_i^{EB} = X_i - BD^{\frac{1}{2}}D^{-\frac{1}{2}}(X_i - \bar{X}_n).$$

**Definition** For the $i$th vector $\theta_i$, we define the limited translation EB estimator of maximum translation $c$ as

$$\hat{\theta}_{c,i}^{LEB} = X_i - BD^{\frac{1}{2}}h_c(D^{-\frac{1}{2}}(X_i - \bar{X}_n)), \quad (3-11)$$

where

$$h_c(z) = z \min(1, c/\|z\|), z \in \mathbb{R}^p, \quad (3-12)$$

is the multidimensional Huber function, Hampel (1986), and $c$ is a known constant.

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

$$\begin{align*}
\hat{\theta}_{c,i}^{LEB} &= X_i - B(X_i - \bar{X}_n)\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) \\
&= X_i\{1 - \rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2)\} + \hat{\theta}_i^{EB}\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2),
\end{align*}$$

(3–13)
where $\rho_c(u) = \min(1, c/\sqrt{u})$ is termed the relevance function.

The limited translation EB estimator follows the EB estimator as closely as possible subject to the constraint that the distance of the observed $X_i$ to the observed mean $\bar{X}_n$, as measured by the standardized norm $||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||$, 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 3–13 we see that the limited translation EB estimator gives the MLE positive weight at the expense of the weight of the EB estimator and as the distance of $X_i$ to $\bar{X}_n$ increases the less relevant is considered to be the EB rule for the estimation of the corresponding $\theta_i$. 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.

### 3.3 Bayes Risk of the Limited Translation EB Estimators

We now calculate the Bayes risks of the EB estimator, $r_1(\xi, \hat{\theta}_{EB}^i)$, and the limited translation EB estimator, $r_1(\xi, \hat{\theta}_{LEB}^{c,i})$ assuming that the prior $\xi \equiv N_p(\mu,A)$ is the true one. First we calculate

$$r_1(\xi, \hat{\theta}_{EB}^i) = E\{(\theta_i - \hat{\theta}_{EB}^i)(\theta_i - \hat{\theta}_{EB}^i)^T\}$$

$$= E\{(\theta_i - \hat{\theta}_{EB}^i)(\theta_i - \hat{\theta}_{EB}^i)^T\} + E\{(\hat{\theta}_{EB}^i - \hat{\theta}_{EB}^i)(\hat{\theta}_{EB}^i - \hat{\theta}_{EB}^i)^T\}$$

$$= (I_p - B)\Sigma + BE\{(\bar{X}_n - \mu)(\bar{X}_n - \mu)^T\}B^T$$

$$= \{I_p - (1 - n^{-1})B\}\Sigma. \quad (3-14)$$

Comparing $r_1(\xi, \hat{\theta}_{EB}^i)$ to $r_1(\xi, \hat{\theta}_{EB}^i)$, given in Equation 2–4, we see that the price for having to estimate $\mu$ from the data is $n^{-1}B\Sigma$ which converges to a matrix of zeros as $n$ increases, the rate of convergence being $n^{-1}$. This is intuitively clear since as the sample size increases, the sample mean $\bar{X}_n$ converges to the population marginal mean $\mu$.

The following theorem gives an expression for $r_1(\xi, \hat{\theta}_{LEB}^{c,i})$. The calculations do not depend on the special nature of the relevance function $\rho_c(.)$. 

45
Theorem 3.3.1. For any relevance function $\rho(\cdot)$ we have

$$r_1(\xi, \hat{\theta}_i^{LEB}) = r_1(\xi, X_i)(1 - s_c) + r_1(\xi, \hat{\theta}_i^{EB})s_c,$$  \hspace{1cm} (3–15)

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

Hence, the generalized relative savings loss of $\hat{\theta}_i^{LEB}$ with respect to $X_i$, defined by

$$GRSL(\hat{\theta}_i^{LEB}; X_i) = [r_1(\xi, X_i) - r_1(\xi, \hat{\theta}_i^{EB})]^{-1}[r_1(\xi, \hat{\theta}_i^{LEB}) - r_1(\xi, \hat{\theta}_i^{EB})]$$  \hspace{1cm} (3–16)

is calculated as $GRSL(\hat{\theta}_i^{LEB}; X_i) = (1 - s_c)I_p$. If we choose $\rho_c(u) = \min(1, c/\sqrt{u})$, then $1 - s_c$ is given by

$$1 - s_c = P(\chi_{p+2}^2 > c^2) - c\sqrt{2\frac{\Gamma(p+1)}{\Gamma(p+2)}}P(\chi_{p+1}^2 > c^2) + p^{-1}c^2P(\chi_p^2 > c^2),$$  \hspace{1cm} (3–17)

where $\Gamma(.)$ denotes the gamma function.

Proof. We write

$$r_1(\xi, \hat{\theta}_i^{LEB}) = E\{(\theta_i - \hat{\theta}_i^{LEB})(\theta_i - \hat{\theta}_i^{LEB})^T\}$$

$$= E\{(\theta_i - \hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})(\theta_i - \hat{\theta}_i^{EB} + \hat{\theta}_i^{LEB})^T\}$$

$$= r_1(\xi, \hat{\theta}_i^{EB}) + E\{(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})^T\}$$

$$+ E\{(\theta_i - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})^T\} + E\{(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})(\theta_i - \hat{\theta}_i^{EB})^T\}. \hspace{1cm} (3–18)$$

Noting that

$$\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB} = B(X_i - \bar{X}_n)\{\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) - 1\}, \hspace{1cm} (3–19)$$

it follows, from the independence of $X_i - \bar{X}_n$ and $\bar{X}_n$, and the fact that $E(\bar{X}_n) = \mu$, that

$$E\{(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})(\theta_i - \hat{\theta}_i^{EB})^T\} = E\{E(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})(\theta_i - \hat{\theta}_i^{EB})^T|X_i\}$$

$$= BE[\{\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) - 1\}(X_i - \bar{X}_n)(\mu - \bar{X}_n)^T]\mu^T = 0. \hspace{1cm} (3–20)$$
Next,

\[
E\{ (\hat{\theta}_i^{EB} - \hat{\theta}_{c,i}^{LEB})(\hat{\theta}_i^{EB} - \hat{\theta}_{c,i}^{LEB})^T \}
= \mathcal{B} E\{ (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T[\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) - 1]^2\} \mathcal{B}^T
= (1 - 1/n) \Sigma (A + \Sigma)^{-\frac{1}{2}} E\{ Z Z^T[\rho_c(||Z||^2) - 1]^2\} (A + \Sigma)^{-\frac{1}{2}} \Sigma,
\]

(3–21)

where \( Z \sim \mathcal{N}_p(0, I_p) \). It was shown in Equation 2–28 that

\[
E\{ Z Z^T[\rho_c(||Z||^2) - 1]^2\} = E[1 - \rho_c(U)]^2 I_p,
\]

(3–22)

where \( U \sim \chi^2_{p+2} \). Hence, from Equations 3–18, 3–20, 3–21 and 3–22 follows that

\[
r_1(\xi, \hat{\theta}_{c,i}^{LEB}) = r_1(\xi, \hat{\theta}_i^{EB}) + E\{ [1 - \rho_c(U)]^2\} (1 - 1/n) B \Sigma
\]

\[
= r_1(\xi, \hat{\theta}_i^{EB}) + (1 - s_c)(1 - 1/n) B \Sigma,
\]

(3–23)

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 limiting the frequentist risk of the EB estimator. Alternatively we can write

\[
r_1(\xi, \hat{\theta}_{c,i}^{LEB}) = \Sigma - B \Sigma s_c(1 - 1/n) = \Sigma(1 - s_c) + \{ \Sigma - (1 - 1/n) B \Sigma \} s_c
\]

\[
= r_1(\xi, X_i)(1 - s_c) + r_1(\xi, \hat{\theta}_i^{EB}) s_c,
\]

(3–24)

thus completing the proof of the Theorem.

The Bayes risk of the limited translation EB estimator is a weighted average of the Bayes risk of the EB rule and the Bayes risk of the MLE, the weights being \( s_c \) and \( 1 - s_c \) respectively. This causes a loss in the generalized savings of \( (1 - s_c) I_p \). However, the weight of the Bayes risk of the MLE, \( 1 - s_c \), for fixed \( p \), is a decreasing convex function of \( c \). Thus, the choice of \( c \) is equivalent to deciding by what proportion it is worth increasing the Bayes risk of the EB estimator in order to receive protection against large frequentist risks. This protection does not require increasing the Bayes risk by more than 10%.
3.4 Frequentist Risk of the Limited Translation EB Estimators

In this section we focus our attention to the risk of \( \hat{\theta}_{c,i}^{LEB} \) as a function of \( \theta = (\theta_T^1, \ldots, \theta_T^n)^T \), which we denote by \( R_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) \). The purpose is to show that the limited translation EB estimator does not allow to the frequentist risk to take large values. The calculation of the risk of \( \hat{\theta}_{c,i}^{LEB} \) was possible only under the simplifying assumption that the population variance-covariance matrix is a scalar multiple of the sampling variance-covariance matrix, that is \( A = g\Sigma \), where \( g > 0 \) is a known scalar.

It is of interest to compare the frequentist risk of the limited translation EB estimator to the frequentist risk of the EB estimator. To this end, before providing a result concerning the risk of the limited translation estimator, we find an expression for the risk of the regular EB rule. We now write

\[
R_1(\theta_i, \hat{\theta}_{i}^{EB}) = E_{\theta}\{(\theta_i - \hat{\theta}_{i}^{EB})(\theta_i - \hat{\theta}_{i}^{EB})^T\} = E_{\theta}\{(\theta_i - X_i)(\theta_i - X_i)^T\} + BE_{\theta}\{(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\}B^T + E_{\theta}\{(\theta_i - X_i)(X_i - \bar{X}_n)^T\}B + BE_{\theta}\{(X_i - \bar{X}_n)(\theta_i - X_i)^T\}. \tag{3–25}
\]

Now, \( X_i - \bar{X}_n|\theta \sim \mathcal{N}(\theta_i - \bar{\theta}_n, (1 - 1/n)\Sigma) \), where \( \bar{\theta}_n = n^{-1}\sum_{i=1}^n \theta_i \). It follows that

\[
E_{\theta}\{(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} = \Sigma(1 - 1/n) + (\theta_i - \bar{\theta}_n)(\theta_i - \bar{\theta}_n)^T, \tag{3–26}
\]

and that the expectation \( E_{\theta}\{(\theta_i - X_i)(X_i - \bar{X}_n)^T\} \) is equal to

\[
\theta_i\theta_i^T - n^{-1}\sum_{k=1}^n \theta_i \theta_i^T - \Sigma + n^{-1}\sum_{k=1}^n \theta_i \theta_i^T + \Sigma) = -(1 - 1/n)\Sigma. \tag{3–27}
\]

Thus, combining Equations 3–25-3–27, we obtain

\[
R_1(\theta_i, \hat{\theta}_{i}^{EB}) = \Sigma + 2(1/n - 1)B\Sigma + B\{\Sigma(1 - 1/n) + (\theta_i - \bar{\theta}_n)(\theta_i - \bar{\theta}_n)^T\}B^T, \tag{3–28}
\]

which under the assumption that \( A = g\Sigma \) reduces to

\[
R_1(\theta_i, \hat{\theta}_{i}^{EB}) = (1 + g)^{-2}(\theta_i - \bar{\theta}_n)(\theta_i - \bar{\theta}_n)^T + (1 - 1/n)\Sigma[\frac{1}{1 - 1/n} - \frac{(1 + 2g)}{(1 + g)^2}]. \tag{3–29}
\]
The following Theorem provides an expression for the frequentist risk of $\hat{\theta}_{c,i}^{L_{EB}}$.

**Theorem 3.4.1.** Under the assumption that $A = g\Sigma$, where $g > 0$ is a known scalar, and for the relevance function given by $\rho_{c}(u) = \min(1,c/\sqrt{u})$, the frequentist risk of the multivariate limited translation EB rule is given by

$$R_1(\theta_i, \hat{\theta}_{c,i}^{L_{EB}}) = E_\theta\{(\theta_i - \hat{\theta}_{c,i}^{L_{EB}})(\theta_i - \hat{\theta}_{c,i}^{L_{EB}})^T\} = R_1(\theta_i, \hat{\theta}_{i}^{EB})$$

$$+ (\theta_i - \bar{\theta}_n)(\theta_i - \bar{\theta}_n)^T[(1 + 2g)(1 + g)^{-2}P[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]$$

$$+ 2(1 + g)^{-1}P[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]$$

$$+ c^2(1 + g)^{-1}E_{\lambda_i}\{[\chi^2_{p+4}(\lambda_i)]^{-1}I[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]\}$$

$$+ 2c(1 + g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+4}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]\}$$

$$- 2c(1 + g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+4}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]\}]$$

$$+ \Sigma(1 - 1/n)[(1 + 2g)(1 + g)^{-2}P[\chi^2_{p+2}(\lambda_i) > c^2(1 + g)]$$

$$+ c^2(1 + g)^{-1}E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-1}I[\chi^2_{p+2}(\lambda_i) > c^2(1 + g)]\}$$

$$- 2c(1 + g)^{-\frac{1}{2}}E_{\lambda_i}\{[\chi^2_{p+2}(\lambda_i)]^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda_i) > c^2(1 + g)]\}]\}$$

(3–30)

where $\lambda_i = 2^{-1}(1 - 1/n)^{-1}(\theta_i - \bar{\theta}_n)\Sigma^{-1}(\theta_i - \bar{\theta}_n)$ and $\chi^2_{k}^{2}(\lambda)$ denotes the non-central chi-square distribution with non-centrality parameter equal to $\lambda$ and $k$ degrees of freedom.

The proof is given in section B.1 of Appendix B.

For easier comparison of the risks of (3–29) and (3–30), we calculate their scalar versions by considering the $L_2$ loss function

$$L_2(\theta, a) = (\theta - a)^T\Sigma^{-1}(\theta - a).$$

(3–31)

First, it is easy to show that the risk of the EB rule under the loss function $L_2$, is equal to

$$R_2(\theta, \hat{\theta}_i^{EB}) = p + 2(1 - 1/n)(1 + g)^{-2}\lambda_i - (1 - 1/n)(1 + 2g)(1 + g)^{-2}p,$$

(3–32)

which, for fixed $p$ and $g$, is a function only of the non-centrality parameter $\lambda_i$, and so is the risk of the limited translation estimator, as becomes evident in the following Corollary.
Corollary 3.4.2. Under the loss function $L_2$ the frequentist risk of the limited translation EB rule is given by

$$R_2(\theta, \hat{\theta}^{LEB}_{c,i}) = E_{\theta} \{ (\theta_i - \hat{\theta}^{LEB}_{c,i})^T \Sigma^{-1} (\theta_i - \hat{\theta}^{LEB}_{c,i}) \} = tr[\Sigma^{-1} R_1(\theta_i, \hat{\theta}^{LEB}_{c,i})] = tr \left[ \Sigma^{-1} \right] = R_2(\theta, \hat{\theta}^{EB}_{i}) - (1 - 1/n) P[\chi^2_{p+2}(\lambda_i) > c^2(1 + g)] \left\{ \frac{4\lambda_i}{(1 + g)} - \frac{p(1 + 2g)}{(1 + g)^2} \right\}$$

$$+ 2\lambda_i(1 - 1/n)(1 + 2g)(1 + g)^{-2} P[\chi^2_{p+4}(\lambda_i) > c^2(1 + g)]$$

$$+ c^2(1 - 1/n)(1 + g)^{-1} P[\chi^2_{p}(\lambda_i) > c^2(1 + g)]$$

$$+ 2(1 - 1/n)(1 + g)^{-1} \left\{ 2\lambda_i E_{\lambda_i} \left[ \chi^2_{p+2}(\lambda_i) \right]^{1/2} I[\chi^2_{p+2}(\lambda_i) > c^2(1 + g)] \right\}$$

$$- E_{\lambda_i} \left\{ \chi^2_{p}(\lambda_i) \right\}^{1/2} I[\chi^2_{p}(\lambda_i) > c^2(1 + g)] \right\}. \quad (3\text{-}33)$$

The proof is very similar to that of Corollary 2.4.2, given in section A.3 of Appendix A, and thus omitted.

The bracketed term in the last two lines of Equation 3–33 can be calculated as

$$\sqrt{2} \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} \frac{\Gamma(p+1+2k)}{\Gamma(p+2k)} P[\chi^2_{p+1+2k} > c^2(1 + g)] \left( \frac{2\lambda}{p + 2k} - 1 \right). \quad (3\text{-}34)$$

We 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 risk functions in 3–32 and 3–33. 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. For $p = 3$ and $1 - s_c = 10\%$ the corresponding value of $c$ is 1.52.

In Figure 3-1 we see how the risk functions of 3–33 and 3–32 behave as the non-centrality parameter, $\lambda_i$ increases. For small values of $\lambda_i$, i.e. when $\theta_i$ is close to $\theta_n$, the EB estimator has slightly smaller frequentist risk than the limited translation EB estimator. However, the frequentist risk of the EB estimator increases linearly with the non-centrality parameter which clearly means that the EB estimator has high risk when the $\theta_i$ is far from $\theta$. On the contrary, the frequentist risk of the limited translation EB estimator
becomes flat after $\lambda_i$ exceeds a certain value. That is, the limited translation rule does not allow large frequentist risks even if the $\theta_i$ is far from $\hat{\theta}$.

3.5 Robustness of Limited Translation EB Estimators

The purpose of this section is to examine the Bayes risk performance of the proposed estimator and compare it to the performance of the regular EB estimator under misspecified models. This examination proceeds as follows. We first derive EB and limited translation EB estimators assuming that the sampling distributions of the $X_i$ are $X_i|\theta_i \sim N_p(\theta_i, \Sigma)$, while the $\theta_i$ themselves are normally distributed, $\theta_i \sim \xi \equiv N_p(\mu, g\Sigma)$, $i = 1, \ldots, n$. We then calculate the Bayes risk of the two estimators assuming that the true prior is $\xi^* \equiv N_p(\mu^*, g^*\Sigma)$. It should be noted here that the (mis)specification of the prior mean does not really matter because we are assuming it to be unknown.

Assuming that the true prior is $\xi$, the EB estimator of the $i$th vector $\theta_i$ is given as $\hat{\theta}_i^{EB} = X_i - B(X_i - \bar{X}_n)$, where $B = (1 + g)^{-1}$, while its Bayes risk under the assumed prior is given by $r_1(\xi, \hat{\theta}_i^{EB}) = \{1 - (1 - n^{-1})B\}\Sigma$. If, however, the true prior is $\xi^*$, the estimator that we should be using is $\hat{\theta}_i^{EB^*} = X_i - B^*(X_i - \bar{X}_n)$, where $B^* = (1 + g^*)^{-1}$.

We now calculate the Bayes risk associated with the EB estimator derived under the misspecified model. We have that

$$r_1(\xi^*, \hat{\theta}_i^{EB}) = E\{(\theta_i - \hat{\theta}_i^{EB})(\theta_i - \hat{\theta}_i^{EB})^T\}$$

$$= r_1(\xi^*, \hat{\theta}_i^{EB^*}) + E\{(\hat{\theta}_i^{EB^*} - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB^*} - \hat{\theta}_i^{EB})^T\}, \quad (3-35)$$

where

$$E\{(\hat{\theta}_i^{EB^*} - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB^*} - \hat{\theta}_i^{EB})^T\} = (B - B^*)^2E\{(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\}$$

$$= (1 - n^{-1})(B - B^*)^2(B^*)^{-1}\Sigma = (1 - n^{-1})B^*B^2(g^* - g)^2\Sigma, \quad (3-36)$$

and thus

$$r_1(\xi^*, \hat{\theta}_i^{EB}) = r_1(\xi^*, \hat{\theta}_i^{EB^*}) + (1 - n^{-1})B^*B^2(g^* - g)^2\Sigma. \quad (3-37)$$
The Bayes risk of the EB estimator under misspecified priors has two components. The first one, $r_1(\xi^*, \hat{\theta}_i^{EB^*})$, can be thought of as the inevitable risk, the risk due to nature, while the second one can be attributed to the misspecification of the prior parameter, $g^*$. As the distance between the true prior parameter, $g^*$, and the assumed one, $g$, increases so does the second component of the Bayes risk and, of course, the Bayes risk itself.

We now turn our attention to the limited translation estimator. Under prior $\xi$ this estimator is given by

$$\hat{\theta}_{c,i}^{L\text{EB}} = X_i - B(X_i - \bar{X}_n)\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2),$$

(3–38)

where $D = \text{var}(X_i - \bar{X}_n) = (1 - 1/n)B^{-1}\Sigma$. Its Bayes risk under prior $\xi^*$ is given in the following Theorem.

**Theorem 3.5.1.** The Bayes risk of $\hat{\theta}_{c,i}^{L\text{EB}}$, derived assuming that the true prior is $\xi$, under prior $\xi^*$ is given by

$$r_1(\xi^*, \hat{\theta}_{c,i}^{L\text{EB}}) = r_1(\xi^*, \hat{\theta}_i^{EB^*}) + (1 - 1/n)B^*E\{(B/B^*)\rho_c(U) - 1\}^2 \Sigma,$$

(3–39)

where, with $c^{**} = (c^*B)/B^*$,

$$E\{(B/B^*)\rho_c(U) - 1\}^2$$

$$= [(B/B^*) - 1]^2P[\chi_{p+2}^2 \leq (c^*)^2] + E\{\left(\frac{c^{**}}{\sqrt{U}} - 1\right)^2I[U > (c^*)^2]\}$$

$$= [(B/B^*) - 1]^2P[\chi_{p+2}^2 \leq (c^*)^2] + P[\chi_{p+2}^2 > (c^*)^2]$$

$$- c^{**}\sqrt{\frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p+2}{2})}}P[\chi_{p+1}^2 > (c^*)^2]$$

$$+ p^{-1}(c^{**})^2P[\chi_{p}^2 > (c^*)^2].$$

(3–40)

The proof is given in section B.2 of Appendix B.

We now consider the $L_2$ loss function, given in 3–31, and provide expressions for the Bayes risks of the two estimators. First, the Bayes risk of the EB estimator is given by

$$r_2(\xi^*, \hat{\theta}_i^{EB}) = \{1 - (1 - n^{-1})B^*\}p + (1 - n^{-1})B^*B^2(g^* - g)^2p,$$

(3–41)
while the Bayes risk of the limited translation estimator is given by

$$r_2(\xi^*, \hat{\theta}_{LEB}^{c,i}) = \{1 - (1 - n^{-1})B^*\}p + (1 - n^{-1})B^*E\{(B/B^*)\rho_{c^*}(U) - 1\}^2 p. \quad (3-42)$$

Figure 3-2 shows how the two functions behave as the assumed prior parameter $g$, varies around the true prior parameter $g^*$, which, for the sake of comparison, we take to be $g^* = 2$. We also take $p = 3$ and $n$ to be large enough to approximate $1 - n^{-1} \approx 1$.

When the true parameter $g^*$ is assumed to take any value smaller than 1.34, $\hat{\theta}_{LEB}^{c,i}$ does much better than $\hat{\theta}_{EB}^{i}$. That is, when $g^*$ is underestimated, the limited translation estimator has much smaller Bayes risk the the EB estimator. When, however, the assumed prior parameter is close to the true parameter, the EB estimator, as one should expect, fares better than the limited translation estimator. As the assumed value $g$, of $g^*$ becomes bigger than the true value of $g^*$, the Bayes risk performance of the two estimators becomes similar. As $g$ increases, the two estimators become closer to the MLE and their Bayes risk tends to the Bayes risk of the MLE, $r_2(\xi^*, X_i) = p$, and here we have taken $p = 3$. 

53
Figure 3-1. Risks plotted against the non-centrality parameter, $\lambda_i$.

Figure 3-2. Bayes risks plotted against the assumed parameter $g$ when the true parameter is taken to be $g^* = 2$. 
4.1 Introduction

Empirical Bayes methods are used quite often in the theory and practice of statistics. Their increasing popularity is attributed to the more efficient inference that they lead to compared to the classical frequentist procedures. Modelling the similarity among the individuals or populations leads to this increased efficiency.

In order to make this point clear, suppose that we observe $n$ random variables, each from a normal distribution with different means but same variance, that is, $X_i | \theta_i \sim N(\theta_i, \sigma^2)$, $i = 1, \ldots, n$, and suppose that the goal is to estimate the unobservable $\theta_i$. The classical approach to this problem is to estimate each of the unknown means using its maximum likelihood (ML) estimator, $\hat{\theta}_i = X_i$. The problem here is that each of the random effects is estimated based on only one observation, and it is clear that estimation based on such small sample sizes cannot be very reliable.

To tackle this problem we model the similarity among the individuals (populations). For instance, suppose that we observe scores on IQ tests of $n$ individuals. The similarity among the individuals, in this case, is that they belong to a population the average IQ score of which is $E(\theta_i) = 100$.

Modelling of the similarity can be achieved by assigning a distribution to the unobservable $\theta_i$ which leads to hierarchical Bayesian models. These models can also be thought of as models for incorporating prior information in the inferential procedure. For the exponential family of distributions, the resulting estimators, namely the posterior means, are weighted averages of the ML estimators and the prior means, the weights being inversely proportional to the sampling and prior variances.

When some or all of the prior parameters are unknown, the Bayes estimators cannot be used as such. However, the prior parameters can be estimated from the marginal distribution of the data. Replacing the unknown parameters that appear in the Bayes
estimators by appropriate estimators, results in the so called empirical Bayes (EB) estimators.

A well known characteristic of the EB estimators, for the exponential family of distributions, is that they shrink the ML estimators towards some synthetic mean. In doing so, they achieve smaller Bayes risk than the ML estimators but they can lead to poor estimation of random effects $\theta_i$, that have unusually small or large values.

Robust EB procedures have been proposed to guard against problems of this type. Efron & Morris (1972a) developed some robust estimators which they referred to as ‘limited translation rules’. These rules are compromises between the EB and the ML estimators that slightly increase the Bayes risk but guard against large frequentist risks. Efron & Morris (1972a) developed limited translation EB rules for the univariate normal case. The objective here is to develop limited translation EB rules for multivariate normal case.

One of the virtues of the limited translation rules is that they do not fare too badly in their Bayes risk performance, compared to the regular EB estimators, even if the assumed prior is close to the true one. In a frequentist risk sense, the limited translation estimators do not perform too badly relative to the regular EB estimators even if the random effect to be estimated is close to the synthetic mean towards which the ML estimators are pulled. On the other hand, if the random effect to be estimated is far from this synthetic mean, then the limited translation estimators do perform much better than the regular EB estimators.

The organization of the remaining sections is as follows. In Section 4.2 we review some results concerning the multivariate EB estimators. In Section 4.3 we introduce the limited translation estimators. Section 4.4 evaluates their Bayes risk under the assumed prior while their frequentist risk is evaluated in Section 4.5. In Section 4.6 we undertake a simulation study to evaluate the effectiveness of the limited translation EB estimators and
compare them with the regular EB and ML estimators. The proofs of the results of this chapter are given in Appendix C.

4.2 Empirical Bayes Estimators

Consider the case where \( n \) random vectors \( X_i, i = 1, \ldots, n \), are independent with \( X_i | \theta_i \overset{\text{iid}}{\sim} N_p(\theta_i, \Sigma) \), where \( \Sigma \) is known. The \( \theta_i, i = 1, \ldots, n \), are iid according to \( \theta_i \overset{\text{iid}}{\sim} \xi \equiv N_p(\mu, A) \). Under the matrix loss function \( L_1(\theta_i, a_i) = (\theta_i - a_i)(\theta_i - a_i)^T \), where \( a_i \) is a vector guess for \( \theta_i \), the Bayes estimator of \( \theta_i \) is

\[
\hat{\theta}_i^B = X_i - \Sigma(A + \Sigma)^{-1}(X_i - \mu) = X_i - B(X_i - \mu) = (I_p - B)X_i + B\mu, \tag{4-1}
\]

where \( B = \Sigma(A + \Sigma)^{-1} \). However, \( \hat{\theta}_i^B \) cannot be used as such when either or both of the prior mean and prior variance-covariance matrix are unknown.

An EB estimator is obtained by observing that marginally \( X_i \overset{\text{iid}}{\sim} N_p(\mu, A + \Sigma) \), \( i = 1, 2, \ldots, n \). Hence, the unknown \( \mu \) is estimated by \( \bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i \). Also, letting \( S = \sum_{i=1}^{n}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \), the inverse of the unknown marginal variance covariance matrix \( (A + \Sigma)^{-1} \), is estimated by \( aS^{-1} \), where \( a \) is a known constant which we set equal to \( n - p - 2 \) if we want an unbiased estimator of \( (A + \Sigma)^{-1} \). Thus, the resulting EB estimator of \( \theta_i \) is

\[
\tilde{\theta}_i^{EB} = X_i - a\Sigma S^{-1}(X_i - \bar{X}_n) = X_i - \hat{B}(X_i - \bar{X}_n) = (I_p - \hat{B})X_i + \hat{B}\bar{X}_n, \tag{4-2}
\]

where \( \hat{B} = a\Sigma S^{-1} \).

The EB estimator shrinks the ML estimator of \( \theta_i \), namely \( X_i \), towards the grand mean, \( \bar{X}_n \), the ML estimator of the unknown prior mean, \( \mu \). In doing so, the EB estimator attains a lower Bayes risk than the ML estimator, under the assumed prior. However, it results in high frequentist risk when the \( \theta_i \) is far from the grand mean. In order to quantify the two preceding statements, we calculate the Bayes and frequentist risks, under the matrix loss function \( L_1 \), of the EB estimator. The following Theorem provides an expression for the Bayes risk of \( \tilde{\theta}_i^{EB} \), which is denoted by \( r_1(\xi, \tilde{\theta}_i^{EB}) \) and calculated as
\[ E(\theta_i - \tilde{\theta}_i^{EB})(\theta_i - \tilde{\theta}_i^{EB})^T, \] with the expectation being taken over the joint distribution of \( \theta^T = (\theta_1^T, \ldots, \theta_n^T) \) and \( X^T = (X_1^T, \ldots, X_n^T) \).

**Theorem 4.2.1.** The Bayes risk of the EB estimator, under the assumed prior, is given by

\[
 r_1(\xi, \tilde{\theta}_i^{EB}) = \Sigma - an^{-1}\{2 - a(n - p - 2)^{-1}\}B\Sigma. \tag{4-3}
\]

The proof is given in section C.1 of Appendix C.

The above Bayes risk is minimal when \( a = n - p - 2 \), and with this choice of \( a \), it becomes

\[
 r_1(\xi, \tilde{\theta}_i^{EB}) = \Sigma - n^{-1}(n - p - 2)B\Sigma, \tag{4-4}
\]

which increases with \( p \) but decreases with \( n \). This is intuitively obvious since as \( p \) increases so does the number of parameters to be estimated. On the other hand, as \( n \) increases so does the information available for the estimation of the prior parameters. Notice that \( B\Sigma = \Sigma(A + \Sigma)^{-1}\Sigma \) is positive definite. Hence, it is clear that the above Bayes risk is smaller than the Bayes risk of the ML estimator \( X_i \), of \( \theta_i \), \( r_1(\xi, X_i) = \Sigma \).

We now turn our attention to the frequentist risk of the EB estimator \( \tilde{\theta}_i^{EB} \) of \( \theta_i \). This risk is denoted by \( R_1(\theta_i, \tilde{\theta}_i^{EB}) \) and it is calculated by averaging the matrix loss function \( L_1 \) over the sampling distribution of \( X \), that is \( R_1(\theta_i, \tilde{\theta}_i^{EB}) = E_{\theta}\{(\theta_i - \tilde{\theta}_i^{EB})(\theta_i - \tilde{\theta}_i^{EB})^T\} \).

In order to obtain an unbiased estimator of the risk of \( \tilde{\theta}_i^{EB} \), we use the multivariate version of Stein’s identity provided in the following Lemma.

**Lemma 4.2.2.** Let \( h : \mathbb{R}^p \to \mathbb{R}^p \) be a vector of differentiable functions. Suppose \( Y \sim N_p(\mu, \Sigma) \). Let \( h_i \) be the \( i \)th element of \( h \) and \( Y_j \) be the \( j \)th element of \( Y \). Then, if \( E\{|\partial h_i(Y)/\partial Y_j|\} \) is a matrix with all elements finite, one has

\[
 \Sigma E\{\partial h(Y)/\partial Y\} = E\{(Y - \mu)h(Y)^T\}. \]

The proof is very similar to the one provided by Stein (1981) for the univariate case and thus omitted.
The frequentist risk of the EB estimator is calculated as

\[
R_1(\theta_i, \tilde{\theta}_i^{EB}) = E_{\theta_i}( (\theta_i - \theta_i^{EB})(\theta_i - \tilde{\theta}_i^{EB})^T ) \\
= E_{\theta_i}( (\theta_i - X_i + \hat{B}(X_i - \bar{X}_n)) (\theta_i - X_i + \hat{B}(X_i - \bar{X}_n))^T ).
\] (4–5)

Using the multivariate version of Stein’s identity, we obtain another expression for the frequentist risk of the EB estimator. This expression is given in the following Theorem.

**Theorem 4.2.3.** The frequentist risk of \(\tilde{\theta}_i^{EB}\) can be expressed as

\[
R_1(\theta_i, \tilde{\theta}_i^{EB}) = \Sigma + a(a + 2) \Sigma E_{\theta_i}(S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}) \Sigma \\
+ 2a \Sigma E_{\theta_i}(tr\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}\}) \Sigma \\
- 2a(1 - 1/n) \Sigma E_{\theta_i}(S^{-1}) \Sigma.
\] (4–6)

The proof is deferred to section C.2 of Appendix C.

Now, let \(\tilde{\theta}^{EB} = (\tilde{\theta}_1^{EB}, \ldots, \tilde{\theta}_n^{EB})^T\) and consider the loss

\[
L_2(\theta, \tilde{\theta}^{EB}) = \sum_{i=1}^{n} L_1(\theta_i, \tilde{\theta}_i^{EB}) = \sum_{i=1}^{n} (\theta_i - \tilde{\theta}_i^{EB})(\theta_i - \tilde{\theta}_i^{EB})^T.
\] (4–7)

Under this loss,

\[
R_2(\theta, \tilde{\theta}^{EB}) = \sum_{i=1}^{n} R_1(\theta_i, \tilde{\theta}_i^{EB}) = n\Sigma - a(2(n - p - 2) - a) \Sigma E_{\theta_i}(S^{-1}) \Sigma.
\] (4–8)

In particular, for \(a = n - p - 2\), which minimizes the above, we have that \(R_2(\theta, \tilde{\theta}^{EB}) = n\Sigma - a\Sigma E_{\theta_i}(S^{-1}) \Sigma\), and it is clear that \(R_2(\theta, \tilde{\theta}^{EB}) < n\Sigma = R_2(\theta, X)\), for all \(\theta\). That is, in the total risk sense of \(L_2\), the EB estimator dominates the ML estimator, a result not surprising in light of the well known univariate analog.

Returning to the result of Theorem 4.2.3, an unbiased estimator \(\hat{R}_1(\theta_i, \tilde{\theta}_i^{EB})\) of \(R_1(\theta_i, \tilde{\theta}_i^{EB})\) is obtained as follows

\[
\hat{R}_1(\theta_i, \tilde{\theta}_i^{EB}) = \Sigma + a(a + 2) \Sigma S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1} \Sigma \\
+ 2a[tr\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} - (1 - 1/n)] \Sigma S^{-1} \Sigma.
\] (4–9)
This quantity will be large when $X_i$ is far from $\bar{X}_n$, that is when $X_i$ is an outlier. Among others, one possible situation when this occurs in a data set is when the true prior is a mixture of two or more normal distributions, with one of them having high probability. Under this scenario, it is not very safe, in light of 4–9, to consider the center of the data $\bar{X}_n$, to be equally relevant for the estimation of all $\theta_i$.

In order to avoid high risks associated with EB estimators corresponding to outlying observations $X_i$, and possibly outlying corresponding random effects $\theta_i$, we develop limited translation EB estimators that serve as compromise between the EB and the ML estimators. Since the ML estimators have minimax risk equal to $\Sigma$ for all $\theta_i$ we expect the limited translation EB estimators also to maintain low frequentist risk and additionally to maintain low Bayes risk.

4.3 Limited Translation Empirical Bayes Estimators

When the prior variance-covariance matrix $A$ is known, the EB estimator is given by

$$\hat{\theta}_i^{EB} = X_i - B(X_i - \bar{X}_n).$$

(4–10)

Based on this estimator, some robust estimators, namely the the limited translation EB estimators, are obtained by modifying it in a way that controls the amount of shrinkage of $X_i$ towards $\bar{X}_n$. This is done by controlling in $\hat{\theta}_i^{EB}$ the standardized version of $X_i - \bar{X}_n$.

Definition The limited translation EB estimator of maximum translation $c$ for the $i$th vector $\theta_i$, is defined as

$$\hat{\theta}_{c,i}^{LEB} = X_i - B(X_i - \bar{X}_n)\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2),$$

where $\rho_c(u) = \min(1, c/\sqrt{u})$ is termed relevance function (Efron & Morris, 1971, 1972a), $c$ is a known constant and $D \equiv \text{var}(X_i - \bar{X}_n) = (1 - 1/n)(A + \Sigma)$.

We may note the similarity between the relevance function, $\rho_c(u)$, of Efron & Morris (1971, 1972a) with the function, $h_c(u)$, of Huber (1974). They are connected through the equality, $h_c(u) = u\rho_c(u)$. 60
Since the prior variance-covariance matrix $A$ is unknown, we replace $(A + \Sigma)^{-1}$ by $aS^{-1}$ to obtain
\[
\tilde{\theta}_{c,i}^{LEB} = X_i - B(X_i - \bar{X}_n)\rho_c(||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2), \quad (4-11)
\]
where $k_1 = (1 - 1/n)^{-\frac{1}{2}}a^{\frac{1}{2}}$.

It can also be written as a weighted average of the ML and EB estimators since
\[
\tilde{\theta}_{c,i}^{LEB} = X_i\{1 - \rho_c(||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2)\} + \hat{\theta}_i^{EB}\rho_c(||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2). \quad (4-12)
\]

The limited translation EB estimator follows the EB estimator as closely as possible subject to the constraint that the distance of the observed $X_i$ to the observed mean $\bar{X}_n$, as measured by $||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)||$, 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 from 4–12 we see that the limited translation EB estimator gives the MLE bigger weight at the expense of the weight of the EB rule. As the distance of $X_i$ to $\bar{X}_n$ increases the less relevant the EB rule is considered as an estimator of $\theta_i$.

### 4.4 Bayes Risk of the Limited Translation EB Estimators

The purpose of this section is to compare the Bayes risk of the regular EB estimator, $r_1(\xi, \tilde{\theta}_i^{EB})$, to the Bayes risk of the limited translation EB estimator, $r_1(\xi, \tilde{\theta}_{c,i}^{LEB})$, assuming that the prior $\xi \equiv N_p(\mu, A)$ is the true one.

The following theorem gives an expression for $r_1(\xi, \tilde{\theta}_{c,i}^{LEB})$ in terms of $r_1(\xi, X_i)$ and $r_1(\xi, \tilde{\theta}_i^{EB})$. The calculations do not depend on the special nature of the relevance function.

**Theorem 4.4.1.** For any relevance function $\rho_c(.)$ we have
\[
\begin{align*}
\ r_1(\xi, \tilde{\theta}_{c,i}^{LEB}) &= \Sigma - an^{-1}B\Sigma \left[ [1 - E\{1 - \rho_c(aW)\}^2]a(n - p - 2)^{-1} \\
&- 2E\{\rho_c(aW)\}\{a(n - p - 2)^{-1} - 1\} \right]. \quad (4-13)
\end{align*}
\]

where, $W \sim Beta((p + 2)/2, (n - p - 1)/2)$. 
The proof of the Theorem is provided in section C.3 of Appendix C.

The Bayes risk of $\hat{\theta}_{c,i}^{LEB}$ is minimal for $a = n - p - 2$, in which case

$$r_1(\xi, \hat{\theta}_{c,i}^{LEB}) = \Sigma - (n - p - 2)n^{-1}BS[1 - E\{1 - \rho_c(aW)\}^2].$$  \hspace{1cm} (4-14)

Letting $1 - s_c = E\{1 - \rho_c(aW)\}^2$ the expression for $r_1(\xi, \hat{\theta}_{c,i}^{LEB})$ becomes

$$r_1(\xi, \hat{\theta}_{c,i}^{LEB}) = r_1(\xi, X_i)(1 - s_c) + r_1(\xi, \hat{\theta}_{i}^{EB})s_c.$$  \hspace{1cm} (4-15)

The generalized relative savings loss of $\hat{\theta}_{c,i}^{LEB}$ with respect to $X_i$ is defined as

$$\text{GRSL}(\hat{\theta}_{c,i}^{LEB}; X_i) = [r_1(\xi, X_i) - r_1(\xi, \hat{\theta}_{c,i}^{EB})]^{-1}[r_1(\xi, \hat{\theta}_{c,i}^{LEB}) - r_1(\xi, \hat{\theta}_{i}^{EB})].$$  \hspace{1cm} (4-16)

The term $r_1(\xi, X_i) - r_1(\xi, \hat{\theta}_{c,i}^{EB})$ is the savings, in Bayes risk sense, that occur when using the EB estimator instead of the MLE, while $r_1(\xi, \hat{\theta}_{c,i}^{LEB}) - r_1(\xi, \hat{\theta}_{i}^{EB})$ is the loss that occurs when using $\hat{\theta}_{c,i}^{LEB}$ instead of the EB estimator.

Hence, for $a = n - p - 2$, the generalized relative savings loss of $\hat{\theta}_{c,i}^{LEB}$ with respect to $X_i$ is is calculated as $\text{GRSL}(\hat{\theta}_{c,i}^{LEB}; X_i) = (1 - s_c)I_p$.

We now give an expression for $1 - s_c = E\{1 - \rho_c(aW)\}^2$ for the choice of relevance function $\rho_c(u) = \min(1, c/\sqrt{u})$. We have that

$$1 - s_c = E\{1 - \rho_c(aW)\}^2 = E\{1 - \min(1, c/\sqrt{aW})\}^2$$
$$= E\{1 - I[c/\sqrt{aW} > 1] - (c/\sqrt{aW})I[c/\sqrt{aW} \leq 1]\}^2$$
$$= E\{(1 - c/\sqrt{aW})^2I[W > a^{-1}c^2]\}$$
$$= P(W_0 > a^{-1}c^2) + c^2(n - 1)(ap)^{-1}P(W_2 > a^{-1}c^2)$$
$$- 2ca^{-1}\Gamma\left(\frac{p + 1}{2}\right)\Gamma\left(\frac{n + 1}{2}\right)\Gamma^{-1}\left(\frac{p + 2}{2}\right)\Gamma^{-1}\left(\frac{n}{2}\right)P(W_1 > a^{-1}c^2),$$  \hspace{1cm} (4-17)

where $W_i, i = 0, 1, 2$, have the Beta((p + 2 - i)/2, (n - p - 1)/2) distributions respectively.

The Bayes risk of the limited translation EB estimator is a weighted average of the Bayes risks of the EB estimator and that of the MLE, the weights being $s_c$ and $1 - s_c$ respectively. This causes a loss in the generalized savings of $(1 - s_c)I_p$. However, the
weight of the Bayes risk of the MLE, \(1 - s_c\), for fixed \(p\) and \(n\), is a decreasing convex function of \(c\). Thus, the choice of \(c\) is equivalent to deciding by what proportion it is worth increasing the Bayes risk of the EB rule, under the assumed prior, in order to receive protection against large frequentist risks.

For \(a = n - p - 2\) and \(\rho_c(u) = \min(1, c/\sqrt{u})\), Figure 4-1 shows how \(1 - s_c\) decreases as \(c\) increases for two values of \(p = 2, 5\) and two values of \(n = 10, 30\). It is interesting to observe that for given values of \(c\) the smallest values of \(1 - s_c\) occur for \(p = 5\) and \(n = 10\) while the biggest ones occur for \(p = 5\) and \(n = 30\). An intuitive interpretation of this, assuming correctly specified priors, would be that when \(n\) is small compared to \(p\), the uncertainty associated with the EB estimator is quite high. Thus, using the limited translation EB estimator instead of the EB estimator does not cause much loss in Bayes risk. On the other hand, when \(n\) is large compared to \(p\) and both \(n\) and \(p\) are large, much information is lost by using the ML estimator instead of the EB estimator, and the limited translation EB estimator is indeed a compromise between the ML and the EB estimators.

Another choice for the relevance function would be

\[
\rho^*_c(u) = \begin{cases} 
\min(1, c/\sqrt{u}) & \text{if } u \leq c^2_0 \\
0 & \text{if } u > c^2_0.
\end{cases} 
\] (4–18)

This relevance function reflects the idea that the EB rule is irrelevant for observations that for which \(||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c_0\) where \(c_0 > c\), and the corresponding GRLS is given by

\[
1 - s_{c,c_0} = E\{1 - \rho^*_{c,c_0}(aW)\}^2 = E\{1 - \min(1, c/\sqrt{aW})I[aW < c^2_0]\}^2 \\
= E[I[aW > c^2_0] + \{1 - \min(1, c/\sqrt{aW})\}I[aW < c^2_0]]^2 \\
= E[I[aW > c^2_0]] + E[\{1 - \min(1, c/\sqrt{aW})\]^2I[aW < c^2_0]\} \\
= P(W > c^2_0/a) + E\{(1 - c/\sqrt{aW})^2I[c^2 < aW < c^2_0]\] \\
= P(W > c^2_0/a) + (c^2/a)E[W^{-1}I\{(c^2/a) < W < (c^2_0/a)\}] \\
- (2c/\sqrt{a})E[W^{-\frac{1}{2}}I\{(c^2/a) < W < (c^2_0/a)\}] \\
\]
Theorem 4.5.1. The frequentist risk of the limited translation EB estimators is given as

\[
\begin{align*}
\hat{R}_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) &= \Sigma + a^2 \Sigma E_\theta \left\{ S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1} \rho_c \right\} \\
&\quad + 2a \Sigma E_\theta \left\{ S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1} \rho_c \right\} \\
&\quad \times \left( (1 - 1/n) \left| S^{-\frac{1}{2}}(X_i - \bar{X}_n) \right|^2 \right)^{I[|k_i S^{-\frac{1}{2}}(X_i - \bar{X}_n)| > c]} \Sigma \\
&\quad + 2a \Sigma E_\theta \left\{ tr\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \} + n^{-1} - 1 \right\} S^{-1} \rho_c \right] \Sigma,
\end{align*}
\]

The proof is given in section C.4 of Appendix C.

Thus, an unbiased estimator of \( R_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) \) is given as

\[
\hat{R}_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) = \Sigma + \rho_c a^2 \Sigma S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1} \Sigma \\
+ 2a \rho_c \left[ tr\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \} - 1 + n^{-1} \right] \Sigma S^{-1} \Sigma
\]
\[ + 2 \alpha \rho_c \left( (1 - 1/n) ||S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^{-2} \right) I_{||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c} \]
\[ \times \Sigma S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \Sigma. \]  

(4–22)

This quantity does not take large values even if \( X_i \) is far from \( \bar{X}_n \). This is because of the presence of the function \( \rho_c \) in each of the terms that depend on \( X_i - \bar{X}_n \). When the distance from \( X_i \) to \( \bar{X}_n \), as measured by \( ||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| \), takes on a large value, bigger than \( c \), \( \rho_c = c/||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| \) takes on a small value. We are thus protected from large frequentist risks, regardless of how well the assumed prior resembles the true prior distribution.

4.6 A Simulation Study

We now undertake a simulation study to evaluate the performance of the proposed estimators and compare them with the ML and EB estimators. Here we focus our attention to the frequentist risk of the three estimators. Thus, in the first step of our simulation study we fix values for the \( \theta_i \), \( i = 1, \ldots, n \). We take \( p \), the dimension of the vectors \( \theta_i \), to be \( p = 2 \) and the sample size to be \( n = 30 \).

In our first scenario, the \( \theta_i \) are obtained by taking a sample of size \( n = 30 \) from the contaminated model \( \theta_i \overset{iid}{\sim} \xi_1 \equiv 0.9N_2(0, A_1) + 0.1N_2(0, A_2) \), where

\[
A_1 = \begin{bmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 36.0 & 7.2 \\ 7.2 & 36.0 \end{bmatrix}.
\]  

(4–23)

Note that the off diagonal elements of \( A_2 \) were so chosen to keep the correlations of \( \theta_{i1} \) and \( \theta_{i2} \) same as those implied by \( A_1 \), \( \text{cor}(\theta_{i1}, \theta_{i2}) = 0.2 \).

In the second scenario, we obtain the \( \theta_i \) from the normal distribution \( \theta_i \overset{iid}{\sim} \xi_2 \equiv N_2(0, A_3) \), where \( A_3 = \begin{bmatrix} 4.5 & 0.9 \\ 0.9 & 4.5 \end{bmatrix} \), is the variance-covariance matrix of the contaminated normal distribution.

Figure 4-2 shows the \( \theta_i \) obtained from the contaminated model as well as those obtained from the normal distribution.
The second step of the simulation study consists of generating the \( X_i, i = 1, \ldots, n \). For each of the two sets of \( \theta_i \), we generate the \( X_i \) firstly as \( X_i|\theta_i \overset{\text{ind}}{\sim} f_1 \equiv \mathcal{N}_2(\theta_i, \Sigma_1) \) and secondly as \( X_i|\theta_i \overset{\text{ind}}{\sim} f_2 \equiv \mathcal{N}_2(\theta_i, \Sigma_2) \), where

\[
\Sigma_1 = \begin{bmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 5.0 & 1.0 \\ 1.0 & 5.0 \end{bmatrix}.
\] (4–24)

At the third step, we estimate each of the \( \theta_i, i = 1, \ldots, n \), of the two sets. Using the data that came from both \( f_1 \) and \( f_2 \), we calculate the estimators \( \hat{\theta}_i^{EB} \) and \( \hat{\theta}_{c,i}^{LEB} \). For the limited translation estimators, we consider three values for the constant \( c \). Corresponding to generalized relative savings loss, \( 1 - s_c = .1 \) the value of \( c \) is \( c = 1.251 \), while for \( 1 - s_c = .05 \) and \( 1 - s_c = .01 \) the values of \( c \) are \( 1.517 \) and \( 2.028 \) respectively.

The second and third steps are repeated \( R = 10,000 \) times. Let \( \hat{\theta}_i \) be any estimator of \( \theta_i \). Then, the frequentist risk of \( \hat{\theta}_i, R_i(\theta_i, \hat{\theta}_i) = E_{\theta}(\theta_i - \hat{\theta}_i)(\theta_i - \hat{\theta}_i)^T \), is approximated by \( \hat{R}_i(\theta_i, \hat{\theta}_i) = R^{-1} \sum_{r=1}^{R}(\theta_i - \hat{\theta}_{i,r})(\theta_i - \hat{\theta}_{i,r})^T \), where \( \hat{\theta}_{i,r} \) is the estimate of \( \theta_i \) from the \( r \)th run. Since the \( \hat{R}_i \) are matrices, we calculate their traces, \( t_i = \text{tr}[\hat{R}_i(\theta_i, \hat{\theta}_i)] \), and their determinants, \( d_i = \text{det}[\hat{R}_i(\theta_i, \hat{\theta}_i)] \), as one number summaries.

Now, for each of the two sets of \( \theta_i \) and for each of the two samplings distributions, \( f_1 \) and \( f_2 \), we summarize the distributions of the resulting \( t_i \) and \( d_i, i = 1, \ldots, 30 \), by reporting the minimum values (\( Q_0 \)), the 25th percentiles (\( Q_{0.25} \)), the medians (\( Q_{0.50} \)), the 75th percentiles (\( Q_{0.75} \)), the maximum values (\( Q_1 \)), the means (Mean) and the standard deviations (Stdev). The results are displayed in Tables 4-1 - 4-4, where in each cell the first entry describes the quantiles of the \( t_i \), while the second one describes the quantiles of the \( d_i \).

Some interesting issues emerge out of Table 4-1. First, even in the case where the assumed prior is not very close to the true prior, the EB estimator performs well. As far as the averages (Mean) of \( t_i \) and \( d_i \) are concerned, it does better than the MLE. On average, the two measures of frequentist risk of the MLE are reduced by \( 1 - 1.843/1.998 = 66 \)
7.76% and $1 - .824/.958 = 13.99\%$ respectively by the EB estimator. Actually, all the entries of columns $Q_0 - Q_{0.75}$ of the row of $\theta_i^{EB}$ are smaller than the ones of the row of $X_i$. However, the EB estimator can result in very large frequentist risks for observation that are far from $\bar{X}_n$, and this becomes obvious by observing the entries of $Q_1$. The two entries in $Q_1$ for the estimator $\tilde{\theta}_i^{EB}$ are $3.708/2.055 = 1.804$ and $2.584/1.015 = 2.546$ times bigger than those of the minimax estimator, $X_i$. Comparing now the EB estimator with the three limited translation estimators, we see that the entries in columns $Q_0 - Q_{0.75}$ are identical or almost identical. The limited translation estimators, however, have much smaller maximum ($Q_1$) risks than the EB estimator, since they become closer to the minimax estimators for outlying observations. They thus have smaller average (Mean) risk than that of the usual EB estimator. For instance, $\tilde{\theta}_{1.251,i}^{LEB}$ reduces the average risks of the EB estimator by $1 - 1.758/1.843 = 4.61\%$ and $1 - .749/.824 = 9.10\%$.

Similarly, in Table 4-2, we see that the EB estimator has maximum risks ($Q_1$) $31.157/10.275 = 3.03$ and $108.290/25.376 = 4.26$ times bigger that those of the MLE. However, the EB estimator reduces the average (Mean) risks of the MLE by $1 - 7.084/9.989 = 29.08\%$ and $1 - 12.997/23.950 = 45.73\%$ respectively. The limited translation estimators compare very favorably to the EB estimators. In particular, the estimator $\tilde{\theta}_{1.251,i}^{LEB}$ decreases the maximum risks ($Q_1$) of the EB estimator by $1 - 12.473/31.157 = 59.96\%$ and $1 - 35.406/108.290 = 67.30\%$ respectively. It also reduces the average (Mean) risks of the EB estimator by $1 - 6.319/7.084 = 10.80\%$ and $1 - 10.249/12.997 = 21.14\%$ respectively.

Continuing to Table 4-3, which displays the results for the set of $\theta_i$ that was obtained from the normal prior $\xi_2$ and the $f_1$ sampling distribution, we see that the EB and the limited translation estimators have very similar performance. The limited translation estimators have slightly bigger Mean risks than the EB estimators but they slightly decrease the maximum risks ($Q_1$) of the EB estimators. Also, compared to the MLE, the limited translation estimators, have smaller Mean risks but bigger maximum risks.
Table 4-4 displays the comparison of the estimators for the case where the $\theta_i$ were obtained from the $\xi_2$ prior, and the $X_i$ from the $f_2$ sampling distribution. This comparison is very similar to the one we have seen for the $\xi_2$ prior and the $f_1$ sampling distribution. Again, the EB and the limited translation estimators perform similarly. Their slight differences are that the limited translation estimators have bigger Mean risks than the EB estimators but they decrease the maximum risks ($Q_1$) of the EB estimators. Also, the limited translation estimators have smaller Mean risks but bigger maximum risks than that of the MLE.

In order to study the effect of the sample size, we took samples of size 20 from each of the two sets of thirty $\theta_i$. Our new sets of $\theta_i$ are displayed in Figure 4-3 (a) and (b). The second and third steps of this study were same as the ones described earlier. The results, shown in Tables 4-5-4-8, are very similar to the ones that we have already seen. It is thus safe to conclude that the sample size does not affect much the performance of the estimators under examination. We may note that for $n = 20$ the values of $c$ corresponding to $1 - s_c = 10\%, 5\%$ and $1\%$ are $c = 1.195, 1.443$ and $1.910$ respectively.

![Figure 4-1. Plot of $1 - s_c$ as a function of $c$, for $p = 2, 10$ and $n = 10, 30$.](image-url)
Figure 4-2. The $\theta_i$ generated from (a) the contaminated model and (b) the normal model.
Table 4-1. Comparison of the risks of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1.961</td>
<td>1.987</td>
<td>1.995</td>
<td>2.011</td>
<td>2.055</td>
<td>1.998</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>0.925</td>
<td>0.948</td>
<td>0.956</td>
<td>0.967</td>
<td>1.015</td>
<td>0.958</td>
<td>0.020</td>
</tr>
<tr>
<td>$\hat{\theta}_{EB}$</td>
<td>1.647</td>
<td>1.687</td>
<td>1.703</td>
<td>1.724</td>
<td>3.708</td>
<td>1.843</td>
<td>0.458</td>
</tr>
<tr>
<td></td>
<td>0.657</td>
<td>0.690</td>
<td>0.701</td>
<td>0.722</td>
<td>2.584</td>
<td>0.824</td>
<td>0.406</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}$</td>
<td>1.647</td>
<td>1.689</td>
<td>1.704</td>
<td>1.729</td>
<td>2.149</td>
<td>1.758</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>0.657</td>
<td>0.691</td>
<td>0.703</td>
<td>0.726</td>
<td>1.109</td>
<td>0.749</td>
<td>0.121</td>
</tr>
<tr>
<td>$\hat{\theta}_{LEB}$</td>
<td>1.647</td>
<td>1.687</td>
<td>1.703</td>
<td>1.725</td>
<td>2.218</td>
<td>1.761</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>0.657</td>
<td>0.690</td>
<td>0.701</td>
<td>0.722</td>
<td>1.176</td>
<td>0.752</td>
<td>0.136</td>
</tr>
<tr>
<td>$\tilde{\theta}_{LEB}$</td>
<td>1.647</td>
<td>1.687</td>
<td>1.703</td>
<td>1.724</td>
<td>2.386</td>
<td>1.772</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.657</td>
<td>0.690</td>
<td>0.701</td>
<td>0.722</td>
<td>1.335</td>
<td>0.761</td>
<td>0.169</td>
</tr>
</tbody>
</table>

Table 4-2. Comparison of the risks of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>9.807</td>
<td>9.935</td>
<td>9.975</td>
<td>10.054</td>
<td>10.275</td>
<td>9.989</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>23.128</td>
<td>23.689</td>
<td>23.889</td>
<td>24.162</td>
<td>25.376</td>
<td>23.950</td>
<td>0.499</td>
</tr>
<tr>
<td>$\hat{\theta}_{EB}$</td>
<td>4.766</td>
<td>5.086</td>
<td>5.227</td>
<td>5.420</td>
<td>31.157</td>
<td>7.084</td>
<td>5.949</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}$</td>
<td>4.953</td>
<td>5.362</td>
<td>5.562</td>
<td>5.840</td>
<td>12.473</td>
<td>6.319</td>
<td>2.016</td>
</tr>
<tr>
<td>$\hat{\theta}_{LEB}$</td>
<td>4.816</td>
<td>5.179</td>
<td>5.352</td>
<td>5.585</td>
<td>13.707</td>
<td>6.223</td>
<td>2.354</td>
</tr>
<tr>
<td></td>
<td>5.703</td>
<td>6.525</td>
<td>6.971</td>
<td>7.747</td>
<td>40.426</td>
<td>10.037</td>
<td>8.478</td>
</tr>
<tr>
<td>$\tilde{\theta}_{LEB}$</td>
<td>4.768</td>
<td>5.092</td>
<td>5.238</td>
<td>5.436</td>
<td>16.809</td>
<td>6.323</td>
<td>3.067</td>
</tr>
<tr>
<td></td>
<td>5.594</td>
<td>6.316</td>
<td>6.694</td>
<td>7.342</td>
<td>52.304</td>
<td>10.422</td>
<td>11.043</td>
</tr>
</tbody>
</table>

Table 4-3. Comparison of the risks of the estimators under the normal model and the $f_1$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1.961</td>
<td>1.987</td>
<td>1.995</td>
<td>2.011</td>
<td>2.055</td>
<td>1.998</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>0.925</td>
<td>0.948</td>
<td>0.956</td>
<td>0.967</td>
<td>1.015</td>
<td>0.958</td>
<td>0.020</td>
</tr>
<tr>
<td>$\hat{\theta}_{EB}$</td>
<td>1.525</td>
<td>1.612</td>
<td>1.704</td>
<td>1.805</td>
<td>2.577</td>
<td>1.767</td>
<td>0.228</td>
</tr>
<tr>
<td></td>
<td>0.564</td>
<td>0.632</td>
<td>0.706</td>
<td>0.793</td>
<td>1.322</td>
<td>0.749</td>
<td>0.172</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}$</td>
<td>1.528</td>
<td>1.634</td>
<td>1.761</td>
<td>1.892</td>
<td>2.211</td>
<td>1.790</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td>0.566</td>
<td>0.646</td>
<td>0.752</td>
<td>0.874</td>
<td>1.099</td>
<td>0.772</td>
<td>0.143</td>
</tr>
<tr>
<td>$\hat{\theta}_{LEB}$</td>
<td>1.526</td>
<td>1.617</td>
<td>1.722</td>
<td>1.871</td>
<td>2.322</td>
<td>1.776</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>0.564</td>
<td>0.635</td>
<td>0.721</td>
<td>0.845</td>
<td>1.173</td>
<td>0.760</td>
<td>0.156</td>
</tr>
<tr>
<td>$\tilde{\theta}_{LEB}$</td>
<td>1.525</td>
<td>1.613</td>
<td>1.704</td>
<td>1.814</td>
<td>2.534</td>
<td>1.769</td>
<td>0.226</td>
</tr>
<tr>
<td></td>
<td>0.564</td>
<td>0.632</td>
<td>0.707</td>
<td>0.797</td>
<td>1.308</td>
<td>0.752</td>
<td>0.174</td>
</tr>
</tbody>
</table>
Table 4.4. Comparison of the risks of the estimators under the normal model and the $f_2$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stddev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>9.831</td>
<td>9.941</td>
<td>9.994</td>
<td>10.081</td>
<td>10.236</td>
<td>10.013</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>23.108</td>
<td>23.730</td>
<td>24.082</td>
<td>24.361</td>
<td>25.150</td>
<td>24.089</td>
<td>0.474</td>
</tr>
<tr>
<td>$\hat{\theta}_{EB}^{i}$</td>
<td>3.752</td>
<td>4.607</td>
<td>5.376</td>
<td>6.748</td>
<td>13.407</td>
<td>6.210</td>
<td>2.260</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}^{1.251,i}$</td>
<td>4.152</td>
<td>5.202</td>
<td>6.067</td>
<td>7.468</td>
<td>11.332</td>
<td>6.605</td>
<td>1.771</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}^{1.517,i}$</td>
<td>3.903</td>
<td>4.874</td>
<td>5.731</td>
<td>7.213</td>
<td>12.041</td>
<td>6.412</td>
<td>1.984</td>
</tr>
<tr>
<td>$\tilde{\theta}_{EB}^{2.028,i}$</td>
<td>3.767</td>
<td>4.647</td>
<td>5.434</td>
<td>6.882</td>
<td>13.019</td>
<td>6.255</td>
<td>2.207</td>
</tr>
<tr>
<td></td>
<td>3.540</td>
<td>5.188</td>
<td>6.870</td>
<td>10.298</td>
<td>25.045</td>
<td>8.888</td>
<td>5.307</td>
</tr>
</tbody>
</table>
Figure 4-3. A sample of $\theta_i$ generated from (a) the contaminated model and (b) the normal model.
Table 4-5. Comparison of the risk of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1.956</td>
<td>1.987</td>
<td>2.002</td>
<td>2.014</td>
<td>2.025</td>
<td>2.000</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>0.922</td>
<td>0.945</td>
<td>0.960</td>
<td>0.970</td>
<td>0.992</td>
<td>0.959</td>
<td>0.017</td>
</tr>
<tr>
<td>$\hat{\theta}^{EB}_i$</td>
<td>1.646</td>
<td>1.665</td>
<td>1.683</td>
<td>1.708</td>
<td>2.825</td>
<td>1.838</td>
<td>0.378</td>
</tr>
<tr>
<td></td>
<td>0.659</td>
<td>0.676</td>
<td>0.689</td>
<td>0.710</td>
<td>1.691</td>
<td>0.822</td>
<td>0.324</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.251,i}$</td>
<td>1.647</td>
<td>1.672</td>
<td>1.685</td>
<td>1.719</td>
<td>2.200</td>
<td>1.764</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>0.660</td>
<td>0.681</td>
<td>0.694</td>
<td>0.722</td>
<td>1.139</td>
<td>0.759</td>
<td>0.154</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.517,i}$</td>
<td>1.646</td>
<td>1.665</td>
<td>1.683</td>
<td>1.710</td>
<td>2.287</td>
<td>1.769</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>0.659</td>
<td>0.677</td>
<td>0.690</td>
<td>0.712</td>
<td>1.212</td>
<td>0.764</td>
<td>0.176</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{2.028,i}$</td>
<td>1.646</td>
<td>1.665</td>
<td>1.683</td>
<td>1.708</td>
<td>2.482</td>
<td>1.792</td>
<td>0.267</td>
</tr>
<tr>
<td></td>
<td>0.659</td>
<td>0.676</td>
<td>0.689</td>
<td>0.710</td>
<td>1.382</td>
<td>0.783</td>
<td>0.226</td>
</tr>
</tbody>
</table>

Table 4-6. Comparison of the risk of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>9.834</td>
<td>9.908</td>
<td>10.041</td>
<td>10.087</td>
<td>10.256</td>
<td>10.012</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>23.176</td>
<td>23.521</td>
<td>24.116</td>
<td>24.234</td>
<td>25.035</td>
<td>24.009</td>
<td>0.515</td>
</tr>
<tr>
<td>$\hat{\theta}^{EB}_i$</td>
<td>4.943</td>
<td>5.145</td>
<td>5.322</td>
<td>5.747</td>
<td>21.543</td>
<td>7.137</td>
<td>4.509</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.251,i}$</td>
<td>5.236</td>
<td>5.479</td>
<td>5.740</td>
<td>6.200</td>
<td>12.124</td>
<td>6.665</td>
<td>2.254</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.517,i}$</td>
<td>5.050</td>
<td>5.271</td>
<td>5.492</td>
<td>5.948</td>
<td>13.079</td>
<td>6.596</td>
<td>2.623</td>
</tr>
<tr>
<td></td>
<td>6.245</td>
<td>6.797</td>
<td>7.293</td>
<td>8.428</td>
<td>34.942</td>
<td>11.358</td>
<td>9.535</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{2.028,i}$</td>
<td>4.954</td>
<td>5.157</td>
<td>5.340</td>
<td>5.773</td>
<td>14.821</td>
<td>6.705</td>
<td>3.231</td>
</tr>
<tr>
<td></td>
<td>5.995</td>
<td>6.496</td>
<td>6.872</td>
<td>7.919</td>
<td>42.474</td>
<td>11.777</td>
<td>11.572</td>
</tr>
</tbody>
</table>

Table 4-7. Comparison of the risk of the estimators under the normal model and the $f_1$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>1.980</td>
<td>1.997</td>
<td>2.003</td>
<td>2.016</td>
<td>2.029</td>
<td>2.006</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>0.945</td>
<td>0.956</td>
<td>0.965</td>
<td>0.976</td>
<td>0.986</td>
<td>0.965</td>
<td>0.011</td>
</tr>
<tr>
<td>$\hat{\theta}^{EB}_i$</td>
<td>1.641</td>
<td>1.744</td>
<td>1.799</td>
<td>1.898</td>
<td>2.182</td>
<td>1.835</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>0.656</td>
<td>0.737</td>
<td>0.780</td>
<td>0.842</td>
<td>1.055</td>
<td>0.807</td>
<td>0.115</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.251,i}$</td>
<td>1.641</td>
<td>1.796</td>
<td>1.861</td>
<td>1.974</td>
<td>2.125</td>
<td>1.871</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>0.656</td>
<td>0.764</td>
<td>0.837</td>
<td>0.905</td>
<td>1.036</td>
<td>0.840</td>
<td>0.115</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.517,i}$</td>
<td>1.641</td>
<td>1.763</td>
<td>1.825</td>
<td>1.939</td>
<td>2.180</td>
<td>1.856</td>
<td>0.152</td>
</tr>
<tr>
<td></td>
<td>0.656</td>
<td>0.744</td>
<td>0.802</td>
<td>0.874</td>
<td>1.069</td>
<td>0.826</td>
<td>0.123</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{2.028,i}$</td>
<td>1.641</td>
<td>1.745</td>
<td>1.800</td>
<td>1.901</td>
<td>2.214</td>
<td>1.841</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>0.656</td>
<td>0.737</td>
<td>0.781</td>
<td>0.844</td>
<td>1.082</td>
<td>0.812</td>
<td>0.124</td>
</tr>
</tbody>
</table>
Table 4-8. Comparison of the risk of the estimators under the normal model and the $f_2$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{\theta}_{i}^{EB}$</td>
<td>4.685</td>
<td>5.964</td>
<td>6.588</td>
<td>7.305</td>
<td>10.600</td>
<td>6.925</td>
<td>1.638</td>
</tr>
<tr>
<td>$\tilde{\theta}_{1.251,i}^{LEB}$</td>
<td>5.006</td>
<td>6.517</td>
<td>7.233</td>
<td>7.842</td>
<td>10.205</td>
<td>7.342</td>
<td>1.430</td>
</tr>
<tr>
<td>$\tilde{\theta}_{1.517,i}^{LEB}$</td>
<td>4.798</td>
<td>6.242</td>
<td>6.947</td>
<td>7.641</td>
<td>10.451</td>
<td>7.164</td>
<td>1.563</td>
</tr>
<tr>
<td>$\tilde{\theta}_{2.028,i}^{LEB}$</td>
<td>4.694</td>
<td>6.012</td>
<td>6.668</td>
<td>7.396</td>
<td>10.648</td>
<td>6.988</td>
<td>1.654</td>
</tr>
</tbody>
</table>
In this chapter we develop estimators assuming that all the model parameters are unknown. The regular EB and robust EB estimators are developed in section 5.1. The Bayes risk of the EB and limited translation EB estimators are evaluated in sections 5.2 and 5.3 respectively. Sections 5.4 and 5.5 examine the frequentist risk of the estimators. In section 5.6 we undertake a simulation study to further evaluate the frequentist risk performance. Further, in section 5.7 we apply the empirical Bayes and limited translation empirical Bayes estimators in order to estimate the average vitamin intakes of HIV-negative drug abusers. The proofs of the results of these chapter are given in Appendix D.

5.1 Development of Estimators

Here we develop empirical Bayes (EB) and some robust empirical Bayes estimators, namely the limited translation estimators, for the case where all parameters are unknown. To this end, consider the following model

\[ X_{ij} | \theta_i \overset{\text{iid}}{\sim} N_p(\theta_i, \Sigma^*), \]

\[ \theta_i \overset{\text{iid}}{\sim} N_p(\mu, A), \]

(5–1)

where \( j = 1, 2, \ldots, k \) and \( i = 1, 2, \ldots, n \). Let \( \Sigma = k^{-1}\Sigma^* \) and \( \bar{X}_i = k^{-1} \sum_{j=1}^{k} X_{ij} \). In order to derive Bayes estimators for the \( \theta_i \), note that for \( i = 1, 2, \ldots, n \),

\[ \bar{X}_i | \theta_i \overset{\text{iid}}{\sim} N_p(\theta_i, \Sigma), \]

\[ \theta_i \overset{\text{iid}}{\sim} N_p(\mu, A). \]

(5–2)

Thus, when all the parameters are known, the Bayes estimator of \( \theta_i \), \( i = 1, 2, \ldots, n \), with respect to the matrix loss function \( L_1(\theta_i, a) = (\theta_i - a)(\theta_i - a)^T \), is given by

\[ \hat{\theta}_i^B = \bar{X}_i - \Sigma(A + \Sigma)^{-1}(\bar{X}_i - \mu) = (I_p - B)\bar{X}_i + B\mu = \bar{X}_i - B(\bar{X}_i - \mu), \]

(5–3)
where \( B = \Sigma(A + \Sigma)^{-1} \).

The assumed model implies that marginally \( X_i \overset{iid}{\sim} N_p(\mu, A + \Sigma), i = 1, 2, \ldots, n \).

First, assume that \( A \) is known, but \( \mu \) is unknown. The latter is estimated by \( \bar{X} = (nk)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k}X_{ij} \) which is its MLE, UMVUE and best equivariant estimator under translation of the sample space. Replacing \( \mu \) by \( \bar{X} \), results in the so called EB estimator,

\[
\hat{\theta}_{i}^{EB} = (I_p - B)\bar{X}_i + B\bar{X} = \bar{X}_i - B(\bar{X}_i - \bar{X}),
\]

which is a weighted average of the MLE of \( \theta_i \), \( \bar{X}_i \), and the sample mean, \( \bar{X} \). This estimator, in contrast with the MLE, uses information included in the whole data set and not just the data corresponding to the \( i \)th individual or population.

Additionally, we have defined the limited translation estimator of maximum translation \( c \) as

\[
\hat{\theta}_{c,i}^{LEB} = \bar{X}_i - B(\bar{X}_i - \bar{X})
\times \rho_c(||(1 - 1/n)^{-\frac{1}{2}}(A + \Sigma)^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^2),
\]

where \( \rho_c(u) = \min(1, c/\sqrt{u}) \) is termed the relevance function. Its argument is the standardized distance of \( \bar{X}_i \) to \( \bar{X} \) and its purpose is to put a bound to the frequentist risk of the EB estimator by controlling the amount of shrinkage of the MLE towards the common mean, \( \bar{X} \).

Suppose now that \( \Sigma \) and \( A \) are also unknown. Let

\[
S = \sum_{i=1}^{n}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T,
\]

\[
V = \sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T.
\]

The inverse of the unknown marginal variance-covariance matrix, \((A + \Sigma)^{-1}\), is estimated by \( aS^{-1} \) and \( \Sigma^* \), the variance-covariance matrix of the sampling distribution, is estimated

76
by \( bV \), where \( a \) and \( b \) are known constants. The resulting estimators are

\[
\begin{align*}
\hat{\theta}_i^{EB} &= \bar{X}_i - \hat{B}(\bar{X}_i - \bar{X}) \\
\hat{\theta}_{c,i}^{LEB} &= \bar{X}_i - \tilde{B}(\bar{X}_i - \bar{X})_c \rho_c(||k_1S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^2),
\end{align*}
\]

(5–8)

where \( \tilde{B} = abk^{-1}VS^{-1} \) and \( k_1 = a^{\frac{1}{2}}(1 - 1/n)^{-\frac{1}{2}} \).

In the following sections we compare the properties of these estimators. We begin by considering their Bayes risk performance. In subsequent sections we also study their frequentist risks by both deriving unbiased estimators of these risks and by undertaking a simulation study. Finally, we apply the proposed inferential procedure in order to estimate the long term average vitamin intakes of HIV-negative drug abusers.

5.2 Bayes Risk of the EB Estimators

We now investigate the Bayes risk performance of the EB estimators assuming the the normal prior \( \xi \equiv N_p(\mu, A) \) is the true one. The following Theorem provides an expression for the Bayes risk of the regular EB estimator.

**Theorem 5.2.1.** The Bayes risk of \( \hat{\theta}^{EB}_i \), under the assumed prior, is given by

\[
\begin{align*}
r_1(\xi, \hat{\theta}^{EB}_i) &= E\{(\theta_i - \hat{\theta}_i^{EB})(\theta_i - \hat{\theta}_i^{EB})^T\} \\
&= \Sigma + ab(k - 1)[ab(n - p - 2)^{-1}\{n(k - 1) + 1\} - 2]B\Sigma \\
&+ a^2b^2(n - p - 2)^{-1}(k - 1)tr(B)\Sigma.
\end{align*}
\]

(5–9)

The proof is given in section D.1 of Appendix D.

Consider now the quadratic loss function, \( L_2 \), given by

\[
L_2(\theta_i, a) = (\theta_i - a)^T\Sigma^{-1}(\theta_i - a).
\]

(5–10)

For fixed \( a = n - p - 2 \), the choice of which results in an unbiased estimator of the marginal variance-covariance matrix, as discussed in the previous chapter, the Bayes risk of the EB estimator, under the \( L_2 \) loss function is minimized for \( b = \{n(k - 1) + p + 1\}^{-1} \). With this
choice of $a$ and $b$, the expression of the Bayes risk under $L_2$ reduces to

$$r_2(\xi, \hat{\theta}_i^{EB}) = E\{(\theta_i - \hat{\theta}_i^{EB})^T \Sigma^{-1}(\theta_i - \hat{\theta}_i^{EB})\} = \text{tr}\{\Sigma^{-1} r_1(\xi, \hat{\theta}_i^{EB})\}$$

$$= p - \frac{(n - p - 2)(k - 1)}{n(k - 1) + p + 1} \text{tr}(B). \tag{5-11}$$

Clearly, $r_2(\xi, \hat{\theta}_i^{EB}) < p = r_2(\xi, \bar{X}_i)$ that is, the EB Bayes estimator has smaller Bayes risk than the ML estimator.

The Bayes risk can also be expressed as

$$r_2(\xi, \hat{\theta}_i^{EB}) = r_2(\xi, \tilde{\theta}_i^{EB}) + (n - p - 2)\{1 - \frac{k - 1}{n(k - 1) + p + 1}\} \text{tr}(B), \tag{5-12}$$

where $\tilde{\theta}_i^{EB}$ is the EB estimator of $\theta_i$ assuming that $\Sigma$ is known, that is, $\tilde{\theta}_i^{EB} = \bar{X}_i - \hat{B}(\bar{X}_i - \bar{X})$ and $\hat{B} = ak^{-1}\Sigma^*S^{-1}$. Thus, the second term in the righthand side of Equation 5–12 can be thought of as the price in terms of Bayes risk for having to estimate $\Sigma$ from the data and it converges to a matrix of zeros as either $n$ or $k$ increase.

### 5.3 Bayes Risk of the Limited Translation EB Estimators

We now obtain an expression for the Bayes risk of the robust estimator. This expression is provided by the following Theorem.

**Theorem 5.3.1.** Let $a_n = n - p - 2$ and $b_n = n^{-1}(k - 1)^{-1}$. Then, the Bayes risk of the limited translation estimator of maximum translation $c$ is given by

$$r_1(\xi, \hat{\theta}_{c,i}^{LEB}) = \Sigma + an^{-1}\left[ -2 + aa_n^{-1} + (aa_n^{-1} - 1)(bb_n^{-1} - 1) + aa_n^{-1}\{1 - 2bb_n^{-1} + b^2b_n^{-1}(b_n^{-1} + 1)\} + aa_n^{-1}b^2n(k - 1)(b_n^{-1} + 1)E\{1 - \rho_c(aW)\}^2 + 2bn(k - 1)\{aa_n^{-1}b(b_n^{-1} + 1) - 1\}E\{\rho_c(aW) - 1\}\right]B\Sigma + a^2b^2a_n^{-1}\left[n^{-1}b_n^{-1} + (k - 1)E\{\rho_c(aW)^2 - 1\}\right] \text{tr}(B)\Sigma \tag{5-13}$$

where $W \sim \text{Beta}((p + 2)/2, (n - p - 1)/2)$.

The proof is given in section D.2 of Appendix D.
For $a = a_n = n - p - 2$ and $b = \{n(k - 1) + p + 1\}^{-1}$, and under the $L_2$ loss function, the Bayes risk of $\hat{\theta}_{c,i}^{LEB}$ is equal to

$$r_2(\xi, \hat{\theta}_{c,i}^{LEB}) = E\{(\theta_i - \hat{\theta}_{c,i}^{LEB})^T\Sigma^{-1}(\theta_i - \hat{\theta}_{c,i}^{LEB})\} = \text{tr}\{\Sigma^{-1}r_1(\xi, \hat{\theta}_{c,i}^{LEB})\}$$

$$= p - \frac{(n - p - 2)(k - 1)}{n(k - 1) + 1}\{2E(\rho_c) - E(\rho_c^2)\}\text{tr}(B). \quad (5-14)$$

It is clear that the proposed estimator has smaller Bayes risk than the MLE. However, it has slightly bigger Bayes risk than the regular EB estimator.

In order to quantify the last statement of the previous paragraph, we use the concept of the relative savings loss of $\hat{\theta}_{c,i}^{LEB}$ with respect to $\bar{X}_i$, defined as

$$\text{RSL}(\hat{\theta}_{c,i}^{LEB}; \bar{X}_i) = [r_2(\xi, \bar{X}_i) - r_2(\xi, \hat{\theta}_{c,i}^{EB})]^{-1}[r_2(\xi, \hat{\theta}_{c,i}^{LEB}) - r_2(\xi, \hat{\theta}_{c,i}^{EB})]. \quad (5-15)$$

The term $r_2(\xi, \bar{X}_i) - r_2(\xi, \hat{\theta}_{c,i}^{EB})$ is the savings, in Bayes risk sense, that occur when using the EB estimator instead of the MLE, while $r_2(\xi, \hat{\theta}_{c,i}^{LEB}) - r_2(\xi, \hat{\theta}_{c,i}^{EB})$ is the loss that occurs when using $\hat{\theta}_{c,i}^{LEB}$ instead of the EB estimator. It can be shown, using Equations 5–11 and 5–14, that $\text{RSL}(\hat{\theta}_{c,i}^{LEB}; \bar{X}_i) = 1 - 2E(\rho_c) + E(\rho_c^2) = E(1 - \rho_c^2) = 1 - s_c$.

In Figure 5-1 we plot RLS against $c$ and it can be seen that RLS is a decreasing convex function of $c$. We can thus choose $c$ by deciding by what percentage is worth increasing the Bayes risk of the EB estimator in order to receive protection against large frequentist risks. The constant $c$ can be so chosen that RLS$= 0.05$ or $0.01$. By sacrificing 5% or 1% of the Bayes risk, we receive considerable protection against large frequentist risks. In order to make the latter point clear, we will examine and compare the frequentist risks of $\hat{\theta}_{c,i}^{EB}$ and $\hat{\theta}_{c,i}^{LEB}$ in the subsequent sections.

Finally, we give another expression for RLS

$$\text{RLS} = P(W_0 > a^{-1}c^2) + c^2(n - 1)(ap)^{-1}P(W_2 > a^{-1}c^2)$$

$$+ ca^{-\frac{1}{2}}\Gamma\left(\frac{p + 1}{2}\right)\Gamma^{-1}\left(\frac{n + 1}{2}\right)P(W_1 > a^{-1}c^2), \quad (5-16)$$

where $W_i, i = 0, 1, 2$, have the Beta($(p + 2 - i)/2, (n - p - 1)/2$) distributions respectively.
5.4 Frequentist Risk of the EB Estimators

In this section we investigate the frequentist risk performance of the EB estimators. We proceed by deriving unbiased estimators of their risks, a task which entails using the multivariate Stein’s identity which was provided in Lemma 4.2.2.

First, the frequentist risk is of EB calculated as

\[
R_1(\theta, \hat{\theta}_{EB}^i) = E_\theta[(\theta_i - \bar{X}_i + \tilde{B}(\bar{X}_i - \bar{X}.))\{(\theta_i - \bar{X}_i + \tilde{B}(\bar{X}_i - \bar{X}.))^T] ,
\]

with the expectation being taken with respect to the sampling distribution of the \(X_{ij}, i = 1, \ldots, n, j = 1, \ldots, k\), and assuming that \(\theta = \{((\theta_1)^T, \ldots, (\theta_n)^T\}^T\) is fixed. Using Stein’s identity we obtain another expression for the risk.

**Theorem 5.4.1.** The frequentist risk of \(\hat{\theta}_{EB}^i\) can be expressed as

\[
R_1(\theta, \hat{\theta}_{EB}^i) = \Sigma + a^2b^2n(k-1)tr\left[E_\theta\{S^{-1}(\bar{X}_i - \bar{X}.)(\bar{X}_i - \bar{X}.)^T\}\Sigma\right] \\
+abn(k-1)[ab\{n(k-1)+1\} + 2]E_\theta\{S^{-1}(\bar{X}_i - \bar{X}.)(\bar{X}_i - \bar{X}.)^T\}S^{-1}\Sigma \\
+2abn(k-1)E_\theta\left[tr\{S^{-1}(\bar{X}_i - \bar{X}.)(\bar{X}_i - \bar{X}.)^T\}\right]S^{-1}\Sigma \\
-2abn(k-1)(1 - 1/n)E_\theta(S^{-1})\Sigma.
\]

The proof is given in section D.3 of Appendix D.

Let \(\tilde{\theta}_{EB} = \{((\tilde{\theta}_1)^T, \ldots, (\tilde{\theta}_n)^T\}^T\) and consider the loss obtained by averaging the individual \(L_2\) losses. The average quadratic loss, using the result of Equation 5–18, and for \(a = n - p - 2\) and \(b = \{n(k-1)+p+1\}^{-1}\), can be shown to be equal to

\[
\bar{R}_2(\theta, \tilde{\theta}_{EB}) = n^{-1}\sum_{i=1}^{n} E_\theta\{(\theta_i - \tilde{\theta}_{EB}^i)^T\Sigma^{-1}(\theta_i - \tilde{\theta}_{EB}^i)\}
\]

\[
= n^{-1}\sum_{i=1}^{n} tr\{\Sigma^{-1}R_1(\theta, \tilde{\theta}_{EB}^i)\} = p - \frac{(n - p - 2)^2(k - 1)}{n(k - 1) + p + 1}tr\{E_\theta(S^{-1})\Sigma\},
\]

which for all \(\theta\) is less than the risk of the MLE. The latter is equal to \(\bar{R}_2(\theta, \bar{X}) = p\) and it is clearly bigger than the risk in Equation 5–19. In other words, the EB estimator dominates the ML estimator, and thus it is a minimax estimator. An interesting approach
for proving the dominance of the EB estimator over the MLE is provided by Efron & Morris (1972b).

Returning to the general result of Theorem 5.4.1, we can obtain an unbiased estimator of the frequentist risk of \( \hat{\theta}_{i}^{EB} \) under the quadratic loss \( L_2 \). This estimator is given by

\[
\hat{R}_2(\theta_i, \hat{\theta}_i^{EB}) = p + pa^2b^3nk^{-1}(k-1)\text{tr}\{S^{-1}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^TS^{-1}V\} + ab^2nk^{-1}(k-1)b[ab\{n(k-1) + 1\} + 2\text{tr}\{S^{-1}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^TS^{-1}\}
\]

\[ + 2ab^2nk^{-1}(k-1)\left[\text{tr}\{S^{-1}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^TS^{-1}\} - 1 + 1/n\right]V S^{-1}. \quad (5–20) \]

The above quantity will be large when \( \bar{X}_i \) is far from \( \bar{X} \). When \( k \) is large and the risk associated with the ML estimators is small, observing an outlying \( \bar{X}_i \), might be an indication that the corresponding \( \theta_i \) is far from the rest of the \( \theta \)'s. Under this scenario, it is not very wise to shrink by a lot the MLE of \( \theta_i \) towards the center of the data. Intuition as well as Equation 5–20 suggest that, in such a case, the risk attached to \( \hat{\theta}_i^{EB} \) is quite high.

### 5.5 Frequentist Risk of the Limited Translation EB Estimators

With the purpose of showing that \( \hat{\theta}_{c,i}^{LEB} \) does not allow high frequentist risks, we obtain an unbiased estimator of its risk. To this end, recall that this estimator is given as \( \hat{\theta}_{c,i}^{LEB} = \bar{X}_i - \bar{B}(\bar{X}_i - \bar{X})\rho_c \), where \( \rho_c \equiv \rho_c(||k_1S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^2) \). First, the risk of \( \hat{\theta}_{c,i}^{LEB} \) is calculated as

\[
R_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) = E_\theta\{(\theta_i - \hat{\theta}_{c,i}^{LEB})(\theta_i - \hat{\theta}_{c,i}^{LEB})^T\}
\]

\[ = E_\theta\{\theta_i - \bar{X}_i + \bar{B}(\bar{X}_i - \bar{X})\rho_c\}\{\theta_i - \bar{X}_i + \bar{B}(\bar{X}_i - \bar{X})\rho_c\}^T. \quad (5–21) \]

Another expression of the above risk is given in the following Theorem.
Theorem 5.5.1. The frequentist risk of $\hat{\theta}_{c,i}^{LEB}$ can be expressed as

$$R_1(\theta, \hat{\theta}_{c,i}^{LEB}) = \Sigma$$

$$+ a^2b^2n(k-1)\{n(k-1) + 1\} \Sigma \mathbb{E}_\theta \{S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}\rho_c^2\} \Sigma$$

$$+ a^2b^2n(k-1)tr \left[ \mathbb{E}_\theta \{S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}\rho_c^2\} \Sigma \right] \Sigma$$

$$+ 2abn(k-1)\Sigma \mathbb{E}_\theta \left[ tr \{S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}\rho_c \} \right] \Sigma$$

$$- 2abn(k-1)(1 - 1/n) \Sigma \mathbb{E}_\theta (S^{-1}\rho_c) \Sigma$$

$$+ 2abn(k-1)\Sigma \mathbb{E}_\theta \left[ S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}\rho_c \right]$$

$$\times \left\{ (1 - 1/n) ||S^{-1/2}(\tilde{X}_i - \bar{X}_i)||^{-2} \right\} I_{[\|k_1S^{-1/2}(X_i - \bar{X})\| > c]} \Sigma. \tag{5-23}$$

The proof is deferred to the Appendix D, section D.4.

Based on the general result of Theorem 5.5.1, we obtain an unbiased estimator of the risk of $\hat{\theta}_{c,i}^{LEB}$ under the quadratic loss $L_2$. This estimator is given by

$$\hat{R}_2(\theta, \hat{\theta}_{c,i}^{LEB}) = p + pa^2b^3 nk^{-1}(k-1) \rho_c^2 \left\{ S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}V \right\}$$

$$+ \left\{ ab\{n(k-1) + 1\}\rho_c + 2\left\{ (1 - 1/n) ||S^{-1/2}(\tilde{X}_i - \bar{X}_i)||^{-2} \right\} I_{[\|k_1S^{-1/2}(X_i - \bar{X})\| > c]} \right\}$$

$$\times ab^2 nk^{-1}(k-1) \rho_c V S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1}$$

$$+ 2ab^2 nk^{-1}(k-1) \rho_c \left[ tr \{S^{-1}(\tilde{X}_i - \bar{X}_i)(\tilde{X}_i - \bar{X}_i)^TS^{-1} \} - 1 + 1/n \right] V S^{-1}. \tag{5-24}$$

The above estimator of the risk suggests that $\hat{\theta}_{c,i}^{LEB}$ does not allow for large frequentist risks. This is because of the presence of the function $\rho_c$ in each of the terms that depend on $\tilde{X}_i - \bar{X}_i$. When the distance from $\tilde{X}_i$ to $\bar{X}$ becomes larger than $c$, then $\rho_c = c/\|k_1S^{-1/2}(\tilde{X}_i - \bar{X}_i)\|$ takes on a small value not allowing the risk to become large.

5.6 A Simulation Study

We now undertake a simulation study to further evaluate the performance of the EB and limited translation EB estimators. Since in previous sections we obtained closed
form formulas for the Bayes risk of these estimators, here we focus our attention to their frequentist risks.

Our simulation study here is very similar to the one of section 4.6. It consists of three steps which we briefly describe. At the first step we obtain values for the \( \theta_i \), \( i = 1, \ldots, n \), and keep them fixed. We take \( n = 30 \) and the dimension of the \( \theta_i \) to be \( p = 2 \). The method of obtaining \( \theta_i \) is exactly the same as the one described in section 4.6. We may recall that the two sets of \( \theta_i \) are shown in Figure 4-2.

At the second step of the simulation study we generate the observations \( X_{ij} \), \( i = 1, \ldots, n \), \( j = 1, \ldots, k \), where \( k \) was selected to be \( k = 3 \). For each of the two sets of \( \theta_i \), we generate the \( X_{ij} \) firstly as \( X_{ij} \mid \theta_i \text{ind} \sim f_1 \equiv N_2(\theta_i, \Sigma_1) \) and secondly as \( X_{ij} \mid \theta_i \text{ind} \sim f_2 \equiv N_2(\theta_i, \Sigma_2) \), where \( \Sigma_1 \) and \( \Sigma_2 \) are given in 4–24.

At the third step, we estimate each of the \( \theta_i \), \( i = 1, \ldots, n \), of the two sets using the observations that came from both \( f_1 \) and \( f_2 \). We calculate the estimators \( \hat{X}_i \), \( \hat{\theta}_i^{EB} \) and \( \hat{\theta}_i^{LEB} \). For the limited translation estimators, we consider three values for the constant \( c \). Corresponding to \( 1 - s_c = .1 \) the value of \( c \) is \( c = 1.251 \), while for \( 1 - s_c = .05 \) and \( 1 - s_c = .01 \) the values of \( c \) are 1.517 and 2.028 respectively.

The second and third steps are repeated \( R = 10,000 \) times. Let \( \hat{\theta}_i \) be any estimator of \( \theta_i \). Then, the frequentist risk of \( \hat{\theta}_i \), \( R_1(\theta_i, \hat{\theta}_i) = E_{\theta_i}( (\theta_i - \hat{\theta}_i)(\theta_i - \hat{\theta}_i)^T ) \), is approximated by \( \hat{R}_1(\theta_i, \hat{\theta}_i) = R^{-1} \sum_{r=1}^R (\theta_i - \hat{\theta}_{i,r})(\theta_i - \hat{\theta}_{i,r})^T \), where \( \hat{\theta}_{i,r} \) is the estimate of \( \theta_i \) from the \( r \)th run. Since the \( \hat{R}_1 \) are matrices, we calculate their traces, \( t_i = \text{tr}[\hat{R}_1(\theta_i, \hat{\theta}_i)] \), and their determinants, \( d_i = \text{det}[\hat{R}_1(\theta_i, \hat{\theta}_i)] \), as one number summaries.

For each of the two sets of \( \theta_i \) and for each of the two samplings distributions, \( f_1 \) and \( f_2 \), we summarize the distributions of the resulting \( t_i \) and \( d_i \), \( i = 1, \ldots, 30 \), by reporting the minimum values (\( Q_0 \)), the 25th percentiles (\( Q_{0.25} \)), the medians (\( Q_{0.50} \)), the 75th percentiles (\( Q_{0.75} \)), the maximum values (\( Q_1 \)), the means (Mean) and the standard deviations (Stdev). The results are displayed in Tables 5-1-5-4, where in each cell the two entries describe the quantiles of the \( t_i \) and \( d_i \) respectively.
Observing Table 5-1, we first notice that the entries of the row that corresponds to the MLE are very close to the theoretical values, \( \text{tr}(k^{-1}\Sigma_1) = 0.667 \) and \( \text{det}(k^{-1}\Sigma_1) = 0.107 \), with very small variability. Now, even in the case where the assumed prior is not very close to the true prior, the EB estimator performs well. The averages (Mean) of \( ti \) and \( di \) for the EB estimator are smaller than those of the MLE. On average, the two measures of frequentist risk of the MLE are reduced by \( 1 - 0.649/0.666 = 2.55\% \) and \( 1 - 0.101/0.107 = 5.94\% \) respectively by the EB estimator. Not only the means but all the entries of columns \( Q_0 - Q_{0.75} \) corresponding to \( \hat{\theta}^{EB}_i \) are smaller than the ones corresponding to \( \hat{X}_i \). The bad property of the EB estimator is that it can result in large frequentist risks for observation that are far from \( \hat{X}_i \). This becomes obvious by observing the entries of \( Q_1 \). The two entries in \( Q_1 \) for the estimator \( \hat{\theta}^{EB}_i \) are 0.863/0.680 = 1.27 and 0.173/0.110 = 1.57 times bigger than those of the minimax estimator, \( \hat{X}_i \). We now compare the three limited translation estimators to the EB estimator. Firstly, we observe that the entries in columns \( Q_0 - Q_{0.75} \) are identical. The limited translation estimators, however, have smaller maximum (\( Q_1 \)) risks than the EB estimator, since they become closer to the minimax estimators for outlying observations. They thus have smaller average (Mean) risk than that of the usual EB estimator. For instance, \( \hat{\theta}^{LEB}_{1.251,i} \) reduces the maximum risks of the EB estimator by \( 1 - 0.686/0.863 = 22.05\% \) and \( 1 - 0.113/0.173 = 34.68\% \).

In Table 5-2, we see that the EB estimator has maximum risks (\( Q_1 \)) 7.282/3.332 = 2.19 and 8.953/2.780 = 3.22 times bigger that those of the MLE. However, it reduces the average (Mean) risks of the MLE by \( 1 - 2.938/3.332 = 11.82\% \) and \( 1 - 2.127/2.666 = 20.22\% \) respectively. The limited translation estimators compare very favorably to the EB estimators. In particular, the estimator \( \hat{\theta}^{LEB}_{1.251,i} \) decreases the maximum risks (\( Q_1 \)) of the EB estimator by \( 1 - 3.746/7.282 = 48.56\% \) and \( 1 - 3.300/8.953 = 63.14\% \) respectively. It also reduces the average (Mean) risks of the EB estimator by \( 1 - 2.752/2.938 = 6.33\% \) and \( 1 - 1.852/2.127 = 12.92\% \) respectively.
Tables 5-3 and 5-4 display the comparison of the estimators, for the case where the \( \theta_i \) were obtained from the normal prior \( \xi_2 \), and the \( X_i \) from the \( f_1 \) and \( f_2 \) sampling distributions respectively. In both Tables we see that the EB and the limited translation estimators perform very similarly. Their slight differences are that the limited translation estimators have bigger average risks (Mean) than the EB estimators but they decrease the maximum risks \( (Q_1) \) of the EB estimators. Further, the limited translation estimators have smaller Mean risks but bigger maximum risks than that of the MLE.

In order to study the effect of the sample size, we took samples of size 20 from each of the two sets of thirty \( \theta_i \) that were obtained in the first step of the simulation study, see Figure 4-3 (a) and (b). The second and third steps of this study were same as the ones described earlier. Note that for \( n = 20 \), the values of \( c \) that correspond to RLS= 10%, 5% and 1% are \( c = 1.195, 1.443 \) and 1.910 respectively. The results were very similar to the ones that we have already seen.

5.7 Application

In this section we apply the proposed inferential procedure in order to estimate the ‘long term average’ vitamin intakes of HIV-negative drug abusers. In this application, we will be using the baseline data from a prospective cohort study of the role of drug abuse in HIV/AIDS weight loss and malnutrition conducted in Boston, Massachusetts, USA.

Each of the \( n = 54 \) 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 intake was determined. The intakes of several nutrients were determined but here, for the sake if simplicity, selected for analysis only two of those nutrients, specifically, vitamin A and Thiamin (also known as vitamin \( B_1 \)).

The observed distribution of the intakes of the two vitamins 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 distributions. We thus start our
analysis by considering a bivariate Box-Cox, Box & Cox (1964), transformation. It turns out that the values of \( \lambda_i, i = 1, 2 \), for the transformation are \( \lambda_1 = 0.10 \) and \( \lambda_2 = 0.35 \).

The average intakes of the two vitamins of the \( n = 54 \) subjects, after the transformation, are displayed in Figure 5-2, and it is clear that even after the transformation the assumptions of normality are not exactly met. For this reason, a robust procedure, like the limited translation estimators, would be more appropriate than the regular EB estimators.

Let \( X_{ij1} \) denote the intake of vitamin A of person \( i \) in day \( j \), and, likewise, \( X_{ij2} \) the intake of vitamin \( B_1 \) of person \( i \) in day \( j \). Further, \( X_{ij} = (X_{ij1}, X_{ij2})^T \) is the response vector of person \( i \) in day \( j \). Additionally, \( \theta_{i1} \) and \( \theta_{i2} \) denote the average daily intake of vitamin A and \( B_1 \), respectively, of person \( i \). The vector \( \theta_i = (\theta_{i1}, \theta_{i2})^T \) is accordingly defined. Now, the EB and limited translation EB estimators are derived based on the assumed model

\[
X_{ij} | \theta_i \overset{\text{ind}}{\sim} N_2(\theta_i, \Sigma) \\
\theta_i \overset{\text{iid}}{\sim} N_2(\mu, A), \ i = 1, 2, \ldots, 54, \ j = 1, 2, 3,
\]

which can equivalently be written as a mixed linear model since,

\[
X_{ij} = \mu + \theta_i + \epsilon_{ij} \\
\theta_i \overset{\text{iid}}{\sim} N_2(0, A), \ \epsilon_{ij} \overset{\text{iid}}{\sim} N_2(0, \Sigma), \ i = 1, 2, \ldots, 54, \ j = 1, 2, 3.
\]

Table 5.7 displays the estimated long term average intakes of the two vitamins for the first ten patients in our sample. The estimates were obtained using the ML, EB and limited translation EB estimators. The first column under each heading refers to the estimated vitamin A intakes while the second one refers to the intakes of vitamin \( B_1 \).

Note that the average vitamin A and \( B_1 \) intakes, after transformation, are \( \bar{X} = (9.096, 0.721)^T \) while \( (n - 1)^{-1} S \), the estimated marginal variance-covariance matrix is
given by \((n - 1)^{-1} \mathbf{S} = \begin{bmatrix} 2.14 & 0.61 \\ 0.61 & 0.31 \end{bmatrix}\). Also note that for the limited translation estimator we have chosen \(c = 1.301\) which corresponds to a relative savings loss of \(1 - s_c = 0.10\).

From Table 5.7, we see that the EB estimator pools the ML estimates towards the grand mean. For those ML estimates that are close to the grand mean, the EB and the limited translation EB estimates are identical while for those that are far from the grand mean, the limited translation estimates are somewhere between the ML and EB estimates.

Similarly, in Figure 5-3, we see the ML, EB and limited translation EB estimates for all \(n = 54\) subjects.

![Figure 5-1. Plot of RLS=E\(\{1 - \rho_c(aW)\}\)\(^2\) as function of \(c\), for \(p = 2, 5\) and \(n = 10, 30\).](image-url)
Table 5-1. Comparison of the risks of the estimators under the contaminated model and the $f_1$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stddev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>0.654</td>
<td>0.664</td>
<td>0.666</td>
<td>0.669</td>
<td>0.680</td>
<td>0.666</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>0.103</td>
<td>0.106</td>
<td>0.106</td>
<td>0.108</td>
<td>0.110</td>
<td>0.107</td>
<td>0.002</td>
</tr>
<tr>
<td>$\bar{\theta}_{EB}^i$</td>
<td>0.617</td>
<td>0.629</td>
<td>0.634</td>
<td>0.641</td>
<td>0.863</td>
<td>0.649</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.095</td>
<td>0.097</td>
<td>0.099</td>
<td>0.173</td>
<td>0.101</td>
<td>0.017</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.251,i}$</td>
<td>0.617</td>
<td>0.629</td>
<td>0.634</td>
<td>0.641</td>
<td>0.686</td>
<td>0.638</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.095</td>
<td>0.097</td>
<td>0.099</td>
<td>0.113</td>
<td>0.098</td>
<td>0.005</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.517,i}$</td>
<td>0.617</td>
<td>0.629</td>
<td>0.634</td>
<td>0.641</td>
<td>0.695</td>
<td>0.639</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.095</td>
<td>0.097</td>
<td>0.099</td>
<td>0.113</td>
<td>0.098</td>
<td>0.006</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{2.028,i}$</td>
<td>0.617</td>
<td>0.629</td>
<td>0.634</td>
<td>0.641</td>
<td>0.716</td>
<td>0.640</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>0.092</td>
<td>0.095</td>
<td>0.097</td>
<td>0.099</td>
<td>0.121</td>
<td>0.099</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Table 5-2. Comparison of the risks of the estimators under the contaminated model and the $f_2$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stddev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>3.280</td>
<td>3.314</td>
<td>3.328</td>
<td>3.353</td>
<td>3.401</td>
<td>3.332</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>2.584</td>
<td>2.632</td>
<td>2.659</td>
<td>2.699</td>
<td>2.780</td>
<td>2.666</td>
<td>0.048</td>
</tr>
<tr>
<td>$\bar{\theta}_{EB}^i$</td>
<td>2.542</td>
<td>2.568</td>
<td>2.610</td>
<td>2.652</td>
<td>7.282</td>
<td>2.938</td>
<td>1.073</td>
</tr>
<tr>
<td></td>
<td>1.566</td>
<td>1.602</td>
<td>1.660</td>
<td>1.724</td>
<td>8.953</td>
<td>2.127</td>
<td>1.583</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.251,i}$</td>
<td>2.546</td>
<td>2.577</td>
<td>2.622</td>
<td>2.680</td>
<td>3.746</td>
<td>2.752</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td>1.574</td>
<td>1.610</td>
<td>1.679</td>
<td>1.764</td>
<td>3.300</td>
<td>1.852</td>
<td>0.469</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.517,i}$</td>
<td>2.543</td>
<td>2.570</td>
<td>2.612</td>
<td>2.656</td>
<td>3.926</td>
<td>2.753</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>1.567</td>
<td>1.603</td>
<td>1.664</td>
<td>1.729</td>
<td>3.563</td>
<td>1.855</td>
<td>0.531</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{2.028,i}$</td>
<td>2.542</td>
<td>2.568</td>
<td>2.610</td>
<td>2.652</td>
<td>4.367</td>
<td>2.778</td>
<td>0.474</td>
</tr>
<tr>
<td></td>
<td>1.566</td>
<td>1.602</td>
<td>1.660</td>
<td>1.724</td>
<td>4.200</td>
<td>1.889</td>
<td>0.669</td>
</tr>
</tbody>
</table>

Table 5-3. Comparison of the risks of the estimators under the normal model and the $f_1$ sampling distribution, $n = 30$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stddev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_i$</td>
<td>0.646</td>
<td>0.660</td>
<td>0.666</td>
<td>0.672</td>
<td>0.690</td>
<td>0.666</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>0.101</td>
<td>0.104</td>
<td>0.106</td>
<td>0.108</td>
<td>0.114</td>
<td>0.106</td>
<td>0.003</td>
</tr>
<tr>
<td>$\bar{\theta}_{EB}^i$</td>
<td>0.592</td>
<td>0.619</td>
<td>0.632</td>
<td>0.650</td>
<td>0.745</td>
<td>0.640</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>0.085</td>
<td>0.092</td>
<td>0.096</td>
<td>0.102</td>
<td>0.125</td>
<td>0.098</td>
<td>0.009</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.251,i}$</td>
<td>0.592</td>
<td>0.621</td>
<td>0.636</td>
<td>0.666</td>
<td>0.696</td>
<td>0.643</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>0.085</td>
<td>0.093</td>
<td>0.097</td>
<td>0.106</td>
<td>0.116</td>
<td>0.099</td>
<td>0.008</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{1.517,i}$</td>
<td>0.592</td>
<td>0.619</td>
<td>0.633</td>
<td>0.659</td>
<td>0.708</td>
<td>0.641</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>0.085</td>
<td>0.092</td>
<td>0.096</td>
<td>0.104</td>
<td>0.118</td>
<td>0.099</td>
<td>0.009</td>
</tr>
<tr>
<td>$\bar{\theta}_{LEB}^{2.028,i}$</td>
<td>0.592</td>
<td>0.619</td>
<td>0.632</td>
<td>0.651</td>
<td>0.740</td>
<td>0.640</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>0.085</td>
<td>0.092</td>
<td>0.096</td>
<td>0.102</td>
<td>0.124</td>
<td>0.099</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Table 5-4. Comparison of the risks of the estimators under the normal model and the \( f_2 \) sampling distribution, \( n = 30 \).

<table>
<thead>
<tr>
<th></th>
<th>( Q_{0.0} )</th>
<th>( Q_{0.25} )</th>
<th>( Q_{0.50} )</th>
<th>( Q_{0.75} )</th>
<th>( Q_{1} )</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_i )</td>
<td>3.233</td>
<td>3.314</td>
<td>3.329</td>
<td>3.358</td>
<td>3.418</td>
<td>3.338</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>2.518</td>
<td>2.637</td>
<td>2.661</td>
<td>2.706</td>
<td>2.806</td>
<td>2.674</td>
<td>0.061</td>
</tr>
<tr>
<td>( \hat{\theta}_{EB}^i )</td>
<td>2.213</td>
<td>2.406</td>
<td>2.628</td>
<td>2.902</td>
<td>4.458</td>
<td>2.772</td>
<td>0.506</td>
</tr>
<tr>
<td></td>
<td>1.205</td>
<td>1.412</td>
<td>1.668</td>
<td>1.992</td>
<td>3.721</td>
<td>1.838</td>
<td>0.594</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.251,i} )</td>
<td>2.232</td>
<td>2.471</td>
<td>2.758</td>
<td>3.075</td>
<td>3.779</td>
<td>2.832</td>
<td>0.403</td>
</tr>
<tr>
<td></td>
<td>1.223</td>
<td>1.486</td>
<td>1.827</td>
<td>2.264</td>
<td>3.099</td>
<td>1.933</td>
<td>0.507</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.517,i} )</td>
<td>2.217</td>
<td>2.425</td>
<td>2.682</td>
<td>3.013</td>
<td>4.004</td>
<td>2.800</td>
<td>0.445</td>
</tr>
<tr>
<td></td>
<td>1.208</td>
<td>1.433</td>
<td>1.732</td>
<td>2.165</td>
<td>3.328</td>
<td>1.887</td>
<td>0.551</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{2.028,i} )</td>
<td>2.213</td>
<td>2.407</td>
<td>2.632</td>
<td>2.919</td>
<td>4.375</td>
<td>2.779</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>1.205</td>
<td>1.413</td>
<td>1.672</td>
<td>2.022</td>
<td>3.678</td>
<td>1.852</td>
<td>0.601</td>
</tr>
</tbody>
</table>

Table 5-5. Comparison of the risk of the estimators under the contaminated model and the \( f_1 \) sampling distribution, \( n = 20 \).

<table>
<thead>
<tr>
<th></th>
<th>( Q_{0.0} )</th>
<th>( Q_{0.25} )</th>
<th>( Q_{0.50} )</th>
<th>( Q_{0.75} )</th>
<th>( Q_{1} )</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_i )</td>
<td>0.653</td>
<td>0.660</td>
<td>0.666</td>
<td>0.671</td>
<td>0.678</td>
<td>0.666</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.102</td>
<td>0.105</td>
<td>0.106</td>
<td>0.108</td>
<td>0.109</td>
<td>0.106</td>
<td>0.002</td>
</tr>
<tr>
<td>( \hat{\theta}_{EB}^i )</td>
<td>0.618</td>
<td>0.626</td>
<td>0.633</td>
<td>0.640</td>
<td>0.750</td>
<td>0.647</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.093</td>
<td>0.095</td>
<td>0.096</td>
<td>0.099</td>
<td>0.133</td>
<td>0.101</td>
<td>0.012</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.251,i} )</td>
<td>0.618</td>
<td>0.626</td>
<td>0.633</td>
<td>0.640</td>
<td>0.682</td>
<td>0.638</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>0.093</td>
<td>0.095</td>
<td>0.096</td>
<td>0.099</td>
<td>0.111</td>
<td>0.098</td>
<td>0.005</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.517,i} )</td>
<td>0.618</td>
<td>0.626</td>
<td>0.633</td>
<td>0.640</td>
<td>0.693</td>
<td>0.638</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>0.093</td>
<td>0.095</td>
<td>0.096</td>
<td>0.099</td>
<td>0.114</td>
<td>0.098</td>
<td>0.006</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{2.028,i} )</td>
<td>0.618</td>
<td>0.626</td>
<td>0.633</td>
<td>0.640</td>
<td>0.718</td>
<td>0.641</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>0.093</td>
<td>0.095</td>
<td>0.096</td>
<td>0.099</td>
<td>0.120</td>
<td>0.099</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table 5-6. Comparison of the risk of the estimators under the contaminated model and the \( f_2 \) sampling distribution, \( n = 20 \).

<table>
<thead>
<tr>
<th></th>
<th>( Q_{0.0} )</th>
<th>( Q_{0.25} )</th>
<th>( Q_{0.50} )</th>
<th>( Q_{0.75} )</th>
<th>( Q_{1} )</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{X}_i )</td>
<td>3.244</td>
<td>3.309</td>
<td>3.324</td>
<td>3.346</td>
<td>3.374</td>
<td>3.322</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>2.527</td>
<td>2.628</td>
<td>2.656</td>
<td>2.684</td>
<td>2.722</td>
<td>2.648</td>
<td>0.048</td>
</tr>
<tr>
<td>( \hat{\theta}_{EB}^i )</td>
<td>2.475</td>
<td>2.547</td>
<td>2.575</td>
<td>2.680</td>
<td>5.275</td>
<td>2.928</td>
<td>0.837</td>
</tr>
<tr>
<td></td>
<td>1.499</td>
<td>1.594</td>
<td>1.627</td>
<td>1.744</td>
<td>5.606</td>
<td>2.119</td>
<td>1.190</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.251,i} )</td>
<td>2.487</td>
<td>2.559</td>
<td>2.599</td>
<td>2.725</td>
<td>3.759</td>
<td>2.781</td>
<td>0.402</td>
</tr>
<tr>
<td></td>
<td>1.512</td>
<td>1.610</td>
<td>1.653</td>
<td>1.807</td>
<td>3.286</td>
<td>1.904</td>
<td>0.558</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{1.517,i} )</td>
<td>2.477</td>
<td>2.549</td>
<td>2.580</td>
<td>2.693</td>
<td>3.951</td>
<td>2.787</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td>1.501</td>
<td>1.596</td>
<td>1.632</td>
<td>1.762</td>
<td>3.557</td>
<td>1.915</td>
<td>0.645</td>
</tr>
<tr>
<td>( \tilde{\theta}_{EB}^{2.028,i} )</td>
<td>2.475</td>
<td>2.547</td>
<td>2.575</td>
<td>2.6817</td>
<td>4.388</td>
<td>2.831</td>
<td>0.590</td>
</tr>
<tr>
<td></td>
<td>1.499</td>
<td>1.594</td>
<td>1.627</td>
<td>1.745</td>
<td>4.165</td>
<td>1.977</td>
<td>0.819</td>
</tr>
</tbody>
</table>
Table 5-7. Comparison of the risk of the estimators under the normal model and the $f_1$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_i$</td>
<td>0.655</td>
<td>0.661</td>
<td>0.663</td>
<td>0.665</td>
<td>0.680</td>
<td>0.665</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>0.103</td>
<td>0.105</td>
<td>0.106</td>
<td>0.107</td>
<td>0.111</td>
<td>0.106</td>
<td>0.002</td>
</tr>
<tr>
<td>$\tilde{\theta}^{EB}_i$</td>
<td>0.622</td>
<td>0.633</td>
<td>0.643</td>
<td>0.653</td>
<td>0.683</td>
<td>0.646</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>0.096</td>
<td>0.100</td>
<td>0.102</td>
<td>0.112</td>
<td>0.100</td>
<td>0.006</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.251,i}$</td>
<td>0.622</td>
<td>0.636</td>
<td>0.653</td>
<td>0.662</td>
<td>0.694</td>
<td>0.651</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>0.097</td>
<td>0.102</td>
<td>0.105</td>
<td>0.113</td>
<td>0.102</td>
<td>0.006</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.517,i}$</td>
<td>0.622</td>
<td>0.633</td>
<td>0.646</td>
<td>0.657</td>
<td>0.695</td>
<td>0.648</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>0.096</td>
<td>0.100</td>
<td>0.103</td>
<td>0.113</td>
<td>0.101</td>
<td>0.006</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{2.028,i}$</td>
<td>0.622</td>
<td>0.633</td>
<td>0.643</td>
<td>0.654</td>
<td>0.688</td>
<td>0.647</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>0.094</td>
<td>0.096</td>
<td>0.100</td>
<td>0.102</td>
<td>0.114</td>
<td>0.101</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 5-8. Comparison of the risk of the estimators under the normal model and the $f_2$ sampling distribution, $n = 20$.

<table>
<thead>
<tr>
<th></th>
<th>$Q_0$</th>
<th>$Q_{0.25}$</th>
<th>$Q_{0.50}$</th>
<th>$Q_{0.75}$</th>
<th>$Q_1$</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_i$</td>
<td>3.249</td>
<td>3.302</td>
<td>3.324</td>
<td>3.349</td>
<td>3.380</td>
<td>3.327</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>2.530</td>
<td>2.623</td>
<td>2.662</td>
<td>2.694</td>
<td>2.758</td>
<td>2.660</td>
<td>0.056</td>
</tr>
<tr>
<td>$\tilde{\theta}^{EB}_i$</td>
<td>2.460</td>
<td>2.747</td>
<td>2.850</td>
<td>3.015</td>
<td>3.668</td>
<td>2.917</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>1.477</td>
<td>1.794</td>
<td>2.001</td>
<td>2.147</td>
<td>2.884</td>
<td>2.043</td>
<td>0.401</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.251,i}$</td>
<td>2.469</td>
<td>2.838</td>
<td>2.989</td>
<td>3.156</td>
<td>3.562</td>
<td>2.995</td>
<td>0.301</td>
</tr>
<tr>
<td></td>
<td>1.487</td>
<td>1.900</td>
<td>2.205</td>
<td>2.339</td>
<td>2.845</td>
<td>2.157</td>
<td>0.394</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{1.517,i}$</td>
<td>2.461</td>
<td>2.782</td>
<td>2.918</td>
<td>3.096</td>
<td>3.660</td>
<td>2.964</td>
<td>0.327</td>
</tr>
<tr>
<td></td>
<td>1.478</td>
<td>1.833</td>
<td>2.103</td>
<td>2.252</td>
<td>2.932</td>
<td>2.109</td>
<td>0.421</td>
</tr>
<tr>
<td>$\tilde{\theta}^{LEB}_{2.028,i}$</td>
<td>2.460</td>
<td>2.750</td>
<td>2.856</td>
<td>3.025</td>
<td>3.714</td>
<td>2.930</td>
<td>0.337</td>
</tr>
<tr>
<td></td>
<td>1.477</td>
<td>1.797</td>
<td>2.010</td>
<td>2.160</td>
<td>2.951</td>
<td>2.060</td>
<td>0.425</td>
</tr>
</tbody>
</table>
Figure 5-2. Average intakes of vitamins A and $B_1$.

Table 5-9. MLE, EB, LT Estimates

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>EB</th>
<th>LT</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.143</td>
<td>-0.171</td>
<td>8.451</td>
<td>0.086</td>
</tr>
<tr>
<td>10.415</td>
<td>1.059</td>
<td>9.973</td>
<td>1.017</td>
</tr>
<tr>
<td>10.774</td>
<td>1.215</td>
<td>10.260</td>
<td>1.143</td>
</tr>
<tr>
<td>8.134</td>
<td>0.759</td>
<td>8.670</td>
<td>0.708</td>
</tr>
<tr>
<td>9.798</td>
<td>1.038</td>
<td>9.665</td>
<td>0.976</td>
</tr>
<tr>
<td>5.992</td>
<td>-0.521</td>
<td>6.700</td>
<td>-0.295</td>
</tr>
<tr>
<td>11.537</td>
<td>2.178</td>
<td>11.340</td>
<td>1.862</td>
</tr>
<tr>
<td>9.319</td>
<td>1.430</td>
<td>9.732</td>
<td>1.236</td>
</tr>
<tr>
<td>10.412</td>
<td>1.054</td>
<td>9.968</td>
<td>1.013</td>
</tr>
<tr>
<td>8.890</td>
<td>0.518</td>
<td>8.847</td>
<td>0.567</td>
</tr>
</tbody>
</table>
Figure 5-3. Estimated intakes of vitamins A and $B_1$. 
CHAPTER 6
SUMMARY AND CONCLUSIONS

In this dissertation, we developed some robust estimators of the multivariate normal mean, namely the limited translation Bayes and empirical Bayes estimators, by extending the work of Efron & Morris (1971, 1972a).

The multivariate limited translation Bayes estimators serve as a compromise between the Bayes and the maximum likelihood estimators. We have demonstrated the usefulness of such estimators over the usual Bayes estimators, in a Bayes risk sense, under misspecified priors. From the criteria of frequentist risks, we have demonstrated the usefulness of such estimators, when they are used for estimating parameters which depart widely from the assumed prior means.

Additionally, we developed multivariate limited translation empirical Bayes estimators of the normal mean vector which serve as a compromise between the empirical Bayes estimators and the maximum likelihood estimators. We examined the properties of such estimators and demonstrated their usefulness from the frequentist risks criteria, when there is wide departure of an individual observation from the grand average.

Future work will develop similar estimators using, however, shrinkage estimators, as those suggested by Efron & Morris (1976), for the unknown variance-covariance matrices. In the process, we will address directly estimation of the variance-covariance matrix of a multivariate normal distribution in a very general setup.
Hence,

\[ \text{We now differentiate both sides of Equation } k \frac{X}{\lambda} (A–2) \]

Proof. We write \[ 2.5.1 \]

\[ \sim \]

We first prove two basic lemmas useful to the proof of Theorems 2.4.1 and 2.5.1.

Lemma A.1.1. Let \( Y \sim N_p(\eta, a\Sigma) \) where \( a > 0 \). Then, for any fixed scalars \( b \) and \( d \) and any fixed \( p \)-dimensional vector \( \phi \), we have

\[
E\{ \frac{Y - \eta}{||\lambda^{-\frac{1}{2}}(Y - \phi)||^2} I[||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 \leq b] \} = (\eta - \phi) \left[ E\lambda\{[\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \leq b] \} - E\lambda\{[\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \leq b] \} \right].
\] (A–1)

Proof. We write \[ Q = a^{-1}(Y - \phi)^T\Sigma^{-1}(Y - \phi) \] and observe that \( Q \sim \chi^2_p(\lambda) \) where \( \lambda = (\eta - \phi)^T(a\Sigma)^{-1}(\eta - \phi)/2 \). 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 \( \chi^2_{p+2k} \) variables, \( k = 0, 1, 2, \ldots \), with Poisson weights \((e^{-\lambda}\lambda^k)/k!\).

We begin with the equality

\[
\int_0^b \frac{e^{-\frac{1}{2}||\lambda^{-\frac{1}{2}}(Y - \phi)||^2} \lambda^{-\frac{1}{2}}(Y - \eta)}{||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 I[||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 \leq b]} dY = E\{\lambda^{-\frac{1}{2}}(Y - \phi)^{-2} I[||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 \leq b] \} \]

\[ \text{We now differentiate both sides of Equation A–2 with respect to } \eta. \text{ First note that} \]

\[
\frac{\partial}{\partial \eta} (Y - \eta)^T(a\Sigma)^{-1}(Y - \eta) = 2(a\Sigma)^{-1}(\eta - Y).
\] (A–3)

Hence,

\[
\int_0^b \frac{(a\Sigma)^{-1}(Y - \eta)e^{-\frac{1}{2}||\lambda^{-\frac{1}{2}}(Y - \phi)||^2} \lambda^{-\frac{1}{2}}(Y - \eta)}{||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 I[||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 \leq b]} dY = (a\Sigma)^{-1}E\lambda\{\lambda^{-\frac{1}{2}}(Y - \eta)^T I[||\lambda^{-\frac{1}{2}}(Y - \phi)||^2 \leq b] \}
\]

\[ = \sum_{k=0}^\infty (k!)^{-1}E\lambda\{((\chi^2_{p+2k})^{-d} I[\chi^2_{p+2k} \leq b]) \frac{\partial(e^{-\lambda}\lambda^k)}{\partial \eta} \}. \] (A–4)
Note that

\[
\frac{\partial (e^{-\lambda} \chi^k)}{\partial \eta} = -\frac{\partial \lambda}{\partial \eta} e^{-\lambda} \chi^k + e^{-\lambda} \chi^{k-1} \frac{\partial \lambda}{\partial \eta} = (a\Sigma)^{-1}(\eta - \phi)e^{-\lambda} \chi^{k-1}(k - \lambda). \tag{A-5}
\]

Combining Equations A–4 and A–5 we obtain

\[
E\left\{ \frac{Y - \eta}{\| (a\Sigma)^{-\frac{1}{2}}(Y - \phi) \|^2} I[\| (a\Sigma)^{-\frac{1}{2}}(Y - \phi) \|^2 \leq b] \right\} = (\eta - \phi) \sum_{k=0}^{\infty} \frac{e^{-\lambda} \chi^{k-1}(k - \lambda)}{k!} E\left\{ (\chi^2)_{p+2k}^{-d} I[\chi^2_{p+2k} \leq b] \right\}
\]

\[
= (\eta - \phi) \left[ \sum_{k=0}^{\infty} \frac{e^{-\lambda} \chi^k}{k!} E\left\{ (\chi^2)_{p+2k}^{-d} I[\chi^2_{p+2k} \leq b] \right\} - \sum_{k=0}^{\infty} \frac{e^{-\lambda} \chi^k}{k!} E\left\{ (\chi^2)_{p+2k}^{-d} I[\chi^2_{p+2k} \leq b] \right\} \right]
\]

\[
= (\eta - \phi) \left[ E\{ [\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \leq b] \} - E\{ [\chi^2_{p}(\lambda)]^{-d} I[\chi^2_{p}(\lambda) \leq b] \} \right]. \tag{A-6}
\]

This completes the proof of Lemma A.1.1. \hfill \Box

**Lemma A.1.2.** Consider the same setting as in Lemma A.1.1. Then

\[
E\left\{ \frac{(Y - \eta)(Y - \eta)^T}{\| (a\Sigma)^{-\frac{1}{2}}(Y - \phi) \|^2} I[\| (a\Sigma)^{-\frac{1}{2}}(Y - \phi) \|^2 \leq b] \right\} = (\eta - \phi)(\eta - \phi)^T \left[ E\{ [\chi^2_{p}(\lambda)]^{-d} I[\chi^2_{p}(\lambda) \leq b] \} + E\{ [\chi^2_{p+4}(\lambda)]^{-d} I[\chi^2_{p+4}(\lambda) \leq b] \}
\]

\[
-2E\{ [\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \leq b] \} + a\Sigma E\{ [\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \leq b] \}. \tag{A-7}
\]

**Proof.** We start by differentiating twice both sides of Equation A–2 with respect to $\eta$.

Note that

\[
\frac{\partial^2}{\partial \eta \partial \eta^T} \exp\{ -(Y - \eta)^T(a\Sigma)^{-1}(Y - \eta)/2 \} = (a\Sigma)^{-1}(Y - \eta)(Y - \eta)^T(a\Sigma)^{-1} \exp\{ -(Y - \eta)^T(a\Sigma)^{-1}(Y - \eta)/2 \}
\]

\[
- (a\Sigma)^{-1} \exp\{ -(Y - \eta)^T(a\Sigma)^{-1}(Y - \eta)/2 \}. \tag{A-8}
\]
Thus,

\[
(a \Sigma)^{-1} E \left\{ \frac{(Y - \eta)(Y - \eta)^T}{||((a \Sigma)^{-\frac{1}{2}}(Y - \phi)||^2 I[||((a \Sigma)^{-\frac{1}{2}}(Y - \phi)||^2 \leq b]](a \Sigma)^{-1} \right.
\]

\[
- (a \Sigma)^{-1} E \left\{ ||((a \Sigma)^{-\frac{1}{2}}(Y - \phi)||^{-2d} I[||((a \Sigma)^{-\frac{1}{2}}(Y - \phi)||^2 \leq b]] \right. 
\]

\[
= \sum_{k=0}^{\infty} (k!)^{-1} E \left\{ (\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2 \leq b] \right\} \frac{\partial^2 (e^{-\lambda} \lambda^k)}{\partial \eta \partial \eta^T}, \quad (A-9)
\]

where

\[
\frac{\partial^2 (e^{-\lambda} \lambda^k)}{\partial \eta \partial \eta^T} = \frac{\partial}{\partial \eta} \left\{(a \Sigma)^{-1}(\eta - \phi)e^{-\lambda} \lambda^{k-1}(k - \lambda)\right\}
\]

\[
= (a \Sigma)^{-1} e^{-\lambda} \lambda^{k-1}(k - \lambda)
\]

\[
+ (a \Sigma)^{-1}(\eta - \phi)(\eta - \phi)^T (a \Sigma)^{-1} e^{-\lambda} \lambda^{k-2} \{\lambda^2 + k(k - 1) - 2k\lambda\}. \quad (A-10)
\]

Substituting the last expression in Equation A–9, we obtain

\[
\sum_{k=0}^{\infty} (k!)^{-1} E \left\{ (\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2 \leq b] \right\} \frac{\partial^2 (e^{-\lambda} \lambda^k)}{\partial \eta \partial \eta^T}
\]

\[
= (a \Sigma)^{-1} (\eta - \phi)(\eta - \phi)^T (a \Sigma)^{-1}
\]

\[
\times \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^{k-2}}{k!} \{\lambda^2 + k(k - 1) - 2k\lambda\} E \left\{ (\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2 \leq b] \right\}
\]

\[
+ (a \Sigma)^{-1} \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^{k-1}}{k!} (k - \lambda) E \left\{ (\chi_{p+2k}^2)^{-d} I[\chi_{p+2k}^2 \leq b] \right\}
\]

\[
= (a \Sigma)^{-1} (\eta - \phi)(\eta - \phi)^T (a \Sigma)^{-1}
\]

\[
\times \left[ E_{\lambda}\{[\chi_p^2(\lambda)]^{-d} I[\chi_p^2(\lambda) \leq b]\} + E_{\lambda}\{[\chi_{p+4}^2(\lambda)]^{-d} I[\chi_{p+4}^2(\lambda) \leq b]\}
\]

\[
- 2E_{\lambda}\{[\chi_{p+2}^2(\lambda)]^{-d} I[\chi_{p+2}^2(\lambda) \leq b]\} \right]
\]

\[
+ (a \Sigma)^{-1} \left[ E_{\lambda}\{[\chi_{p+2}^2(\lambda)]^{-d} I[\chi_{p+2}^2(\lambda) \leq b]\} - E_{\lambda}\{[\chi_p^2(\lambda)]^{-d} I[\chi_p^2(\lambda) \leq b]\} \right]. \quad (A-11)
\]

Combining Equations A–9 and A–11 and collecting terms we obtain Lemma A.1.2. \qed

**Remark** The results of Lemmas A.1.1 and A.1.2 hold even if we change the inequalities from \( \leq b \) to \( > b \) with obvious modifications.
A.2 Proof of Theorem 2.4.1

Proof. Let \( Q_1 = (X - \mu)^T \Sigma^{-1} (X - \mu) \) and observe that \( Q_1 | \theta \sim \chi^2_p(\lambda) \), where \( \lambda = (\theta - \mu)^T \Sigma^{-1} (\theta - \mu) / 2. \) We thus have the following equality

\[
\int_{||\Sigma^{-\frac{1}{2}}(X - \mu)||^2 \leq c^2(1 + g)} e^{-\frac{1}{2}||\Sigma^{-\frac{1}{2}}(X - \theta)||^2} \sum_{d} I[||\Sigma^{-\frac{1}{2}}(X - \mu)||^2 \leq c^2(1 + g)] dX = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} E\{ (\chi^2_{p+2k})^{-d} I[\chi^2_{p+2k} \leq c^2(1 + g)] \},
\]

(A-12)

which is the same as Equation A-2 with

\[
a = 1, \ b = c^2(1 + g), \ \eta = \theta \ \text{and} \ \phi = \mu.
\]

(A-13)

Since Equation A-2 was the basis of proving Lemmas A.1.1 and A.1.2, the results of these two Lemmas hold for the special case defined by the Equations in A-13. This gives us the following two equalities

\[
E\{ \frac{X - \theta}{||\Sigma^{-\frac{1}{2}}(X - \mu)||^2} I[||\Sigma^{-\frac{1}{2}}(X - \mu)||^2 \leq c^2(1 + g)] \} = (\theta - \mu) \left[ E\{ [\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \leq c^2(1 + g)] \} - E\{ [\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \geq c^2(1 + g)] \} \right],
\]

(A-14)

and

\[
\begin{align*}
E\{ \frac{(X - \theta)(X - \theta)^T}{||\Sigma^{-\frac{1}{2}}(X - \mu)||^2} I[||\Sigma^{-\frac{1}{2}}(X - \mu)||^2 \leq c^2(1 + g)] \} &= (\theta - \mu)(\theta - \mu)^T \left[ E\{ [\chi^2_p(\lambda)]^{-d} I[\chi^2_p(\lambda) \leq c^2(1 + g)] \} + E\{ [\chi^2_{p+4}(\lambda)]^{-d} I[\chi^2_{p+4}(\lambda) \leq c^2(1 + g)] \} - 2E\{ [\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \leq c^2(1 + g)] \} \right] \\
&+ \Sigma E\{ [\chi^2_{p+2}(\lambda)]^{-d} I[\chi^2_{p+2}(\lambda) \leq c^2(1 + g)] \}.
\end{align*}
\]

(A-15)

Also, note that for any \( k \)

\[
\rho^k_e(||(1 + g)^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}}(X - \mu)||^2) = \rho^k_e((1 + g)^{-1} Q_1)) =
\]

97
\[ I[Q_1 \leq c^2(1 + g)] + c^k(1 + g)^{\frac{k}{2}}Q_1^{\frac{k}{2}}I[Q_1 > c^2(1 + g)]. \quad (A-16) \]

We now write

\[ R_1(\theta, \hat{\theta}^{LB}) = E_{\theta}\{ (\hat{\theta}^{LB} - \theta)(\hat{\theta}^{LB} - \theta)^T \} \]

\[ = E_{\theta}\{ [X - (1 + g)^{-1}(X - \mu)\rho_c((1 + g)^{-1}Q_1) - \theta] \]

\[ \times [X - (1 + g)^{-1}(X - \mu)\rho_c((1 + g)^{-1}Q_1) - \theta]^T \}

\[ = E_{\theta}\{ (X - \theta)(X - \theta)^T \}

\[ \quad \quad \quad - (1 + g)^{-1}E_{\theta}\{ [(X - \theta)(X - \mu)^T + (X - \mu)(X - \theta)^T]\rho_c((1 + g)^{-1}Q_1) \}

\[ \quad \quad \quad \quad + (1 + g)^{-2}E_{\theta}\{ (X - \mu)(X - \mu)^T\rho_c^2((1 + g)^{-1}Q_1) \}. \quad (A-17) \]

Because of the form of the function \( \rho_c(\cdot) \), we need to calculate

\[ E_{\theta}\{ [(X - \theta)(X - \mu)^T + (X - \mu)(X - \theta)^T]I[Q_1 \leq c^2(1 + g)] \}

\[ = E_{\theta}\{ [2(X - \theta)(X - \mu)^T + (X - \theta)(\theta - \mu)^T

\[ \quad \quad \quad \quad \quad + (\theta - \mu)(X - \theta)^T]I[Q_1 \leq c^2(1 + g)] \}, \quad (A-18) \]

and

\[ E_{\theta}\{ (X - \mu)(X - \mu)^T I[Q_1 \leq c^2(1 + g)] \}

\[ = E_{\theta}\{ [(X - \theta)(X - \theta)^T + (X - \theta)(\theta - \mu)^T + (\theta - \mu)(X - \theta)^T

\[ \quad \quad \quad \quad \quad + (\theta - \mu)(\theta - \mu)^T]I[Q_1 \leq c^2(1 + g)] \}. \quad (A-19) \]

We thus apply Equation A–14 with \( d = 0 \) to obtain

\[ E_{\theta}\{ (X - \theta) I[Q_1 \leq c^2(1 + g)] \}

\[ = (\theta - \mu)\{ P[\chi^2_{p+2}(\lambda) \leq (1 + g)c^2] - P[\chi^2_{p}(\lambda) \leq (1 + g)c^2] \}, \quad (A-20) \]

and Equation A–15, again with \( d = 0 \), to obtain

\[ E_{\theta}\{ (X - \theta)(X - \theta)^T I[Q_1 \leq c^2(1 + g)] \} = \]
\[ = \Sigma P[\chi^2_{p+2}(\lambda) \leq (1 + g)c^2] + (\theta - \mu)(\theta - \mu)^T \]

\[
\times \left\{ P[\chi^2_{p+4}(\lambda) \leq (1 + g)c^2] + P[\chi^2_p(\lambda) \leq (1 + g)c^2] 
- 2P[\chi^2_{p+2}(\lambda) \leq (1 + g)c^2] \right\}. \quad (A-21)
\]

Furthermore, the form of the function \( \rho_c(.) \), given in Equation A–16, and Equation A–17 indicate that we also need to find an expression for

\[
G \equiv E_{\theta} \left\{ (X - \theta)(X - \mu)^T + (X - \mu)(X - \theta)^T \right\} Q_1^{-\frac{1}{2}} I[Q_1 > c^2(1 + g)]
= 2E_{\theta} \left\{ (X - \theta)(X - \theta)^T Q_1^{-\frac{1}{2}} I[Q_1 > c^2(1 + g)] \right\}
+ E_{\theta} \left\{ (X - \theta)(\theta - \mu)^T + (\theta - \mu)(X - \theta)^T \right\} Q_1^{-\frac{1}{2}} I[Q_1 > c^2(1 + g)] \right\}. \quad (A-22)
\]

The results of Equations A–14 and A–15 with \( d = 1/2 \) and reversed inequalities show that

\[
G = 2(\theta - \mu)(\theta - \mu)^T E_{\lambda} \left\{ [\chi^2_{p+4}(\lambda)]^{-\frac{1}{2}} I[\chi^2_{p+4}(\lambda) > c^2(1 + g)] \right\}
- E_{\lambda} \left\{ [\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}} I[\chi^2_{p+2}(\lambda) > c^2(1 + g)] \right\}
+ 2\Sigma E_{\lambda} \left\{ [\chi^2_{p+2}(\lambda)]^{-\frac{1}{2}} I[\chi^2_{p+2}(\lambda) > c^2(1 + g)] \right\} \quad (A-23)
\]

Finally, we rewrite

\[
E_{\theta} \left\{ (X - \mu)(X - \mu)^T Q_1^{-1} I[Q_1 > c^2(1 + g)] \right\}
= E_{\theta} \left\{ (X - \theta)(X - \theta)^T Q_1^{-1} I[Q_1 > c^2(1 + g)] \right\}
+ E_{\theta} \left\{ ((X - \theta)(\theta - \mu)^T + (\theta - \mu)(X - \theta)^T) Q_1^{-1} I[Q_1 > c^2(1 + g)] \right\}
+ E_{\theta} \left\{ (\theta - \mu)(\theta - \mu)^T Q_1^{-1} I[Q_1 > c^2(1 + g)] \right\}. \quad (A-24)
\]

Again the results of Equations A–14 and A–15 with \( d = 1 \) and reversed inequalities show that

\[
E_{\theta} \left\{ (X - \mu)(X - \mu)^T Q_1^{-1} I[Q_1 > c^2(1 + g)] \right\}
= (\theta - \mu)(\theta - \mu)^T E_{\lambda} \left\{ [\chi^2_{p+4}(\lambda)]^{-1} I[\chi^2_{p+4}(\lambda) > c^2(1 + g)] \right\}
+ E_{\lambda} \left\{ [\chi^2_{p+2}(\lambda)]^{-1} I[\chi^2_{p+2}(\lambda) > c^2(1 + g)] \right\}. \quad (A-25)
\]

99
Combining Equations A–16–A–25 and collecting the coefficients of $\Sigma$ and $(\theta - \mu)(\theta - \mu)^T$ separately, the result follows.

A.3 Proof of Corollary 2.4.2

Proof. First, we obtain an expression for $R_2(\theta, \hat{\theta}^{LB}_c)$ by directly using the result of Theorem 2.4.1. That is, we calculate $R_2(\theta, \hat{\theta}^{LB}_c) = \text{tr}[\Sigma^{-1}R_1(\theta, \hat{\theta}^{LB}_c)]$. The resulting expression can be simplified by making use of the two equalities that follow. First,

$$
pe_\lambda\{(\chi^2_{p+2}(\lambda))^{-1}I[\chi^2_{p+2}(\lambda) > c^2(1 + g)]
$$

$$
+ \ 2\lambda e_\lambda\{(\chi^2_{p+4}(\lambda))^{-1}I[\chi^2_{p+4}(\lambda) > c^2(1 + g)]
$$

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

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

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

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

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

$$
= \ P[\chi^2_{p}(\lambda) > c^2(1 + g)],
$$

(A–26)

and similarly

$$
pe_\lambda\{(\chi^2_{p+2}(\lambda))^{-\frac{1}{2}}I[\chi^2_{p+2}(\lambda) > c^2(1 + g)]
$$

$$
+ \ 2\lambda e_\lambda\{(\chi^2_{p+4}(\lambda))^{-\frac{1}{2}}I[\chi^2_{p+4}(\lambda) > c^2(1 + g)]
$$

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

$$
+ \ 2\lambda \sum_{k=0}^{\infty} \frac{e^{-\lambda^2}}{k!} \frac{\Gamma(p+3+2k)}{2\Gamma(p+3+2k+2)} P[\chi^2_{p+3+2k} > (1 + g)c^2]
$$

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

\[ = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^{k}}{k!} \frac{\Gamma\left(\frac{p+1+2k}{2}\right)}{\sqrt{2\Gamma\left(\frac{p+2+2k}{2}\right)}} (p + 2k) P[\chi_{p+1+2k}^2 > (1 + g)c^2] \]

\[ = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^{k}}{k!} \frac{\Gamma\left(\frac{p+1+2k}{2}\right)}{\Gamma\left(\frac{p+2k}{2}\right)} \int_{c^2(1+g)}^{\infty} x^{\frac{p+2k+1}{2}} \exp\left(-\frac{c^2}{2}x\right) dx \]

\[ = E_{\lambda}\{|\chi_{p}^{2}(\lambda)|^{\frac{1}{2}} I[\chi_{p}^{2}(\lambda) > c^2(1 + g)]\}. \quad (A-27) \]

The proof of Corollary 2.4.2 is now complete. \[\square\]

### A.4 Proof of Theorem 2.5.1

**Proof.** Write \(B^* = (1+g)^{-1}, B = (1+g)^{-1}\) and \(Q_2 = B^*Q_1 = B^*(X - \mu)^T\Sigma^{-1}(X - \mu)\). Note that under the \(N_p(\mu^*,\Sigma^*)\) prior, the posterior is

\[ \theta|X \sim N_p(\hat{\theta}_{c}^B = (1-B^*)X + B^*\mu^*, (1-B^*)\Sigma). \tag{A-28} \]

Also, marginally \(X \sim N_p(\mu^*, (1+g^*)\Sigma)\) so that \(Q_2 \sim \chi_{p}^2(\lambda)\), where \(\lambda = B^*(\mu^* - \mu)^T\Sigma^{-1}(\mu^* - \mu)/2\). The risk under prior \(\xi^*\) of the limited translation estimator is calculated as

\[ r_1(\xi^*, \hat{\theta}_{c}^B) = E\{(\theta - \hat{\theta}_{c}^B)(\theta - \hat{\theta}_{c}^B)^T\} \]

\[ = E\{(\theta - \hat{\theta}_{c}^B + \hat{\theta}_{c}^B - \hat{\theta}_{c}^L)(\theta - \hat{\theta}_{c}^B + \hat{\theta}_{c}^B - \hat{\theta}_{c}^L)^T\} \]

\[ = r_1(\xi^*, \hat{\theta}_{c}^B) + E\{(\hat{\theta}_{c}^L - \hat{\theta}_{c}^B)(\hat{\theta}_{c}^L - \hat{\theta}_{c}^L)^T\}, \tag{A-29} \]

with the last equality following from the fact that \(E\{(\theta - \hat{\theta}_{c}^B)(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L)^T\} = 0\). We also have that \(r_1(\xi^*, \hat{\theta}_{c}^B) = (1-B^*)\Sigma\), and we need only to calculate \(E\{(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L)(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L)^T\}\). Note that \(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L = -B^*(X - \mu^*) + B(X - \mu)\rho_c(BQ_1)\). Hence,

\[ E\{(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L)(\hat{\theta}_{c}^B - \hat{\theta}_{c}^L)^T\} \]

\[ = (B^*)^2 E\{(X - \mu^*)(X - \mu)^T\} + B^2 E\{(X - \mu)(X - \mu)^T\rho_c^2(BQ_1)\} \]

\[ - BB^* E\{(X - \mu^*)(X - \mu)^T + (X - \mu)(X - \mu)^T\rho_c(BQ_1)\}. \tag{A-30} \]
Rewriting

\[(X - \mu)(X - \mu^*)^T = (X - \mu^*)(X - \mu^*)^T + (\mu^* - \mu)(X - \mu^*)^T \quad (A-31)\]

and

\[(X - \mu)(X - \mu)^T = (X - \mu^*)(X - \mu^*)^T + (\mu^* - \mu)(\mu^* - \mu)^T + (X - \mu^*)(\mu^* - \mu)^T + (\mu^* - \mu)(X - \mu^*)^T, \quad (A-32)\]

it follows from Equation \(A-30\) and some simplification

\[
E\{ \hat{\theta}_c^B - \hat{\theta}_c^{LB} \} (\hat{\theta}_c^B - \hat{\theta}_c^{LB})^T = B^2(\mu^* - \mu)(\mu^* - \mu)^T E\{ \rho_e^2(BQ_1) \}
+ E\{ (X - \mu^*)(X - \mu)^T \} [B^* - B \rho_e^2(BQ_1)]^2
- BE \left[ \{(X - \mu^*)(\mu^* - \mu)^T + (\mu^* - \mu)(X - \mu^*)^T \right.
\times \rho_e(BQ_1) \{ B^* - B \rho_e(BQ_1) \} \right]. \quad (A-33)
\]

For any \(k > 0\), we write

\[
E\{ (X - \mu^*)(\mu^* - \mu)^T \rho_e^k(BQ_1) \}
= E\{ (X - \mu^*)(\mu^* - \mu)^T I[B^*Q_1 \leq c^2B^*B^{-1}] \}
+ c^k (B/B^*)^{-\frac{k}{2}} E\{ (B^*Q_1)^{-\frac{k}{2}} (X - \mu^*)(\mu^* - \mu)^T I[B^*Q_1 > c^2B^*B^{-1}] \}, \quad (A-34)
\]

and the two components of this expectations are calculated by applying Lemma A.1.2 and the remark following it with \(Y = X\), \(a = 1 + g^*\), \(b = c^2(1 + g)(1 + g^*)^{-1}\), \(\eta = \mu^*\), \(\phi = \mu\), and \(d = 0\) and \(k/2\) respectively.

We also have that

\[
E\{ (X - \mu^*)^k \rho_e^k(BQ_1) \} = E\{ (X - \mu^*) I[B^*Q_1 \leq c^2B^*B^{-1}] \}
+ c^k (B/B^*)^{-\frac{k}{2}} E\{ (B^*Q_1)^{-\frac{k}{2}} (X - \mu^*) I[B^*Q_1 > c^2B^*B^{-1}] \}. \quad (A-35)
\]
Here, the two components of this expectations are calculated by applying Lemma A.1.1 and the remark following it with $Y = X$, $a = 1 + g^*$, $b = c^2(1 + g)(1 + g^*)^{-1}$, $\eta = \mu^*$, $\phi = \mu$, and $d = 0$ and $k/2$ respectively.

Finally, we collecting the coefficients $(\mu^* - \mu)(\mu^* - \mu)^T$ and $\Sigma$ separately and the result follows.

\[ \Box \]

### A.5 Proof of Corollary 2.5.2

**Proof.** We first calculate $r_2(\xi^*, \hat{\theta}_c^{LB}) = \text{tr}\{\Sigma^{-1}r_1(\xi^*, \hat{\theta}_e^{LB})\}$ and we simplify the resulting expression using Equations A–26 and A–27.

\[ \Box \]
APPENDIX B
PROOF OF THEOREM 3 SERIES

B.1 Proof of Theorem 3.4.1

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

\[
\rho_c^k(\|D^{-\frac{1}{2}}(X_i - \bar{X}_n)\|^2) = \rho_c^k\{B(1 - n^{-1})Q\} = I[(1 - n^{-1})Q \leq c^2(1 + g)]
\]

We write

\[
R_1(\theta_i, \hat{\theta}_{LEB}^{ci}) = E_{\theta}\{(\theta_i - \hat{\theta}_{LEB}^{ci})(\theta_i - \hat{\theta}_{LEB}^{ci})^T\} = E_{\theta}\{(\theta_i - X_i)(\theta_i - X_i)^T\} - BE_{\theta}\{((X_i - \theta_i)(X_i - \bar{X}_n)^T + (X_i - \bar{X}_n)(X_i - \theta_i)^T)\} \rho_c\{(1 - n^{-1})BQ\} + B^2E_{\theta}\{(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} \rho_c^2\{(1 + n^{-1})BQ\}. \tag{B-2}
\]

Now,

\[
E_{\theta}\{(X_i - \theta_i)(X_i - \bar{X}_n)^T\} \rho_c\{(1 + n^{-1})BQ\} = E_{\theta}\{((X_i - \bar{X}_n) - (\theta_i - \bar{\theta}_n))(X_i - \bar{X}_n)^T\} \rho_c\{(1 + n^{-1})BQ\} = E_{\theta}\{((X_i - \bar{X}_n) - (\theta_i - \bar{\theta}_n))(X_i - \bar{X}_n)^T\} I[(1 - 1/n)^{-1}Q \leq c^2(1 + g)] + c(1 + g)^{\frac{1}{2}}E_{\theta}\{((X_i - \bar{X}_n) - (\theta_i - \bar{\theta}_n))(X_i - \bar{X}_n)^T\} \times \{(1 - 1/n)^{-1}Q\}^{-\frac{1}{2}} I[(1 - 1/n)^{-1}Q > c^2(1 + g)], \tag{B-3}
\]

and the first of the two expectations in the last three lines of the above equation is calculated by applying Lemma A.1.2 with \( Y = X_i - \bar{X}_n, \eta = \theta_i - \bar{\theta}_n, a = 1 - 1/n, \)
\[ b = c^2(1 + g), \ \phi = 0 \text{ and } d = 0, \text{ while the second one is calculated by setting } d = 1/2, \]
keeping the rest of the specifications same as before and reversing inequalities.

Similarly,

\[
E_\theta[(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \rho_c^2((1 + n^{-1})BQ)]
\]
\[= E_\theta[(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T I[(1 - 1/n)^{-1}Q \leq c^2(1 + g)]]
\]
\[+ c^2(1 + g)E_\theta[(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \times \{(1 - 1/n)^{-1}Q \leq c^2(1 + g)\}] \quad (B-4)
\]
and these two expectations are calculated using Lemma A.1.2 exactly as we did in Equation B–3, 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 Equations B–2-B–4, and collecting the coefficients of \( \Sigma \) and \( (\theta_i - \bar{\theta}_n)(\theta_i - \bar{\theta}_n)^T \) separately.

\section*{B.2 Proof of Theorem 3.5.1}

We first write

\[
r_1(\xi^*, \hat{\theta}_{LEB}^{c,i}) = E\{(\theta_i - \hat{\theta}_{LEB}^{c,i})(\theta_i - \hat{\theta}_{LEB}^{c,i})^T\}
\]
\[= E\{(\theta_i - \hat{\theta}_i^{EB*} + \hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})(\theta_i - \hat{\theta}_i^{EB*} + \hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})^T\}
\]
\[= r_1(\xi^*, \hat{\theta}_i^{EB*}) + E\{(\hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})(\hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})^T\}. \quad (B-5)
\]

Now,

\[
\hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB} = (X_i - \bar{X}_n)\{B\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) - B^*\}. \quad (B-6)
\]

Therefore

\[
M \equiv E\{(\hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})(\hat{\theta}_i^{EB*} - \hat{\theta}_{c,i}^{LEB})^T\}
\]
\[= E[(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \{B\rho_c(||D^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2) - B^*\}^2]. \quad (B-7)
\]
Note that $X_i - \bar{X}_n \overset{d}{=} (1 - 1/n)^{1/2}(B^*)^{-1/2}\Sigma^{1/2}Z$, where $Z \sim N_p(0, I_p)$. Also that $\|D^{-1/2}(X_i - \bar{X}_n)\|^2 = (B/B^*)\|Z\|^2$. Using the last equality, it is easy to show that $\rho_c(||D^{-1/2}(X_i - \bar{X}_n)||^2) = \rho_{c^*}(||Z||^2)$ where $c^* = c(B^*/B)^{1/2}$. Returning to Equation B–7, we write $M$ as

$$M = (1 - 1/n)(B^*)^{-1}\Sigma^{1/2}E[ZZ^T\{B\rho_{c^*}(||Z||^2) - B^*\}^2]\Sigma^{1/2}, \quad (B-8)$$

and it follows from Equation 3–22 that

$$M = (1 - 1/n)B^*E\{(B/B^*)\rho_{c^*}(U) - 1\}^2\Sigma, \quad (B-9)$$

where $U \sim \chi^2_{p+2}$.

Combining Equations B–5 and B–9, we obtain the result.
APPENDIX C
PROOF OF THEOREM 4 SERIES

C.1 Proof of Theorem 4.2.1

Proof. We first write

\[ r_1(\xi, \hat{\theta}_{i}^{EB}) = E\{(\theta_i - \hat{\theta}_{i}^{EB})(\theta_i - \hat{\theta}_{i}^{EB})^T\} \]

\[ = E\{(\theta_i - \hat{\theta}_i^B + \hat{\theta}_i^B - \hat{\theta}_i^{EB})(\theta_i - \hat{\theta}_i^B + \hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\} \]

\[ = (I_p - B)\Sigma + E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\}. \quad (C-1) \]

Next, with \( \hat{\theta}_i^{EB} = (I_p - B)X_i + B\bar{X}_n \), we write

\[ E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\} \]

\[ = E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB} + \hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB} + \hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})^T\}. \quad (C-2) \]

It is easy to show that

\[ E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\} = n^{-1}B\Sigma, \quad (C-3) \]

and also that

\[ E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})^T\} = BE\{\mu - \bar{X}_n\}(X_i - \bar{X}_n)^T(\hat{B}^T - B^T)\} = 0, \quad (C-4) \]

with the last equality following from the independence of \( X_n \) and \( (X_i - \bar{X}_n, S) \).

Finding an expression for \( E\{(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})^T\} \) completes the task. First,

\[ \hat{\theta}_i^{EB} - \hat{\theta}_i^{EB} = \hat{B}(X_i - \bar{X}_n) - B(X_i - \bar{X}_n) = (\hat{B} - B)(X_i - \bar{X}_n), \quad (C-5) \]

and thus

\[ C_i \equiv E\{(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{EB})^T\} \]

\[ = E\{(\hat{B} - B)(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T(\hat{B} - B)^T\}. \quad (C-6) \]
Notice that

\[ C_1 = C_2 = \cdots = C_n = n^{-1}\sum_{j=1}^{n} C_j = C \equiv n^{-1}E\{(\hat{B} - B)S(\hat{B} - B)^T\}. \quad (C-7) \]

Also, \( E(S) = (n - 1)(A + \Sigma) \) and \( E(S^{-1}) = (n - p - 2)^{-1}(A + \Sigma)^{-1} \). Thus, by expansion,

\[ E\{(\hat{\theta}^{EB}_i - \bar{\theta}^{EB}_i)(\hat{\theta}^{EB}_i - \bar{\theta}^{EB}_i)^T\} = n^{-1}\{n - 1 - 2a + a^2(n - p - 2)^{-1}\}B\Sigma. \quad (C-8) \]

The proof is completed by combining Equations C–1-C–4 and C–8.

\[ \square \]

**C.2 Proof of Theorem 4.2.3**

**Proof.** First, the frequentist risk of the \( \bar{\theta}^{EB}_i \) is expressed as

\[
R_1(\theta_i, \bar{\theta}^{EB}_i) = E_\theta\{(\theta_i - \bar{\theta}^{EB}_i)(\theta_i - \bar{\theta}^{EB}_i)^T\} \\
= E_\theta[(\theta_i - X_i + \hat{B}(X_i - \bar{X}_n))(\theta_i - X_i + \hat{B}(X_i - \bar{X}_n))^T] \\
= \Sigma + a^2\Sigma E_\theta\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^TS^{-1}\}\Sigma \\
+ aE_\theta\{(\theta_i - X_i)(X_i - \bar{X}_n)^TS^{-1}\}\Sigma + a\Sigma E_\theta\{S^{-1}(X_i - \bar{X}_n)(\theta_i - X_i)^T\}. \quad (C-9)
\]

We write \( X^{(-i)} = (X_1^T, \ldots, X_{i-1}^T, X_{i+1}^T, \ldots, X_n^T)^T \) and using the result of Lemma 4.2.2, we can see that

\[
E_\theta\{(\theta_i - X_i)(X_i - \bar{X}_n)^TS^{-1}\} = E\left[E_\theta\{(\theta_i - X_i)(X_i - \bar{X}_n)^TS^{-1}|X^{(-i)}\}\right] \\
= -\Sigma E\left(\partial(S^{-1}(X_i - \bar{X}_n))/\partial X_i\right), \quad (C-10)
\]

which when combined with Equation C–9 gives us

\[
R_1(\theta_i, \bar{\theta}^{EB}_i) = \Sigma + a^2\Sigma E_\theta\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^TS^{-1}\}\Sigma \\
- a\Sigma E_\theta\left\{\partial(S^{-1}(X_i - \bar{X}_n))/\partial X_i + \left(\partial(S^{-1}(X_i - \bar{X}_n))/\partial X_i\right)^T\right\}\Sigma. \quad (C-11)
\]

Next, observe that

\[
\partial(S^{-1}(X_i - \bar{X}_n))/\partial X_i = \{\partial S^{-1}/\partial X_i\}(X_i - \bar{X}_n) + S^{-1}\{\partial(X_i - \bar{X}_n)/\partial X_i\}. \quad (C-12)
\]
Now, it is easy to see that
\[
\frac{\partial (X_i - \bar{X}_n)}{\partial X_i} = (1 - 1/n)I_p. \tag{C-13}
\]

Also, it is true that
\[
\frac{\partial S^{-1}}{\partial X_{ij}} = -S^{-1}(\partial S/\partial X_{ij})S^{-1}, \tag{C-14}
\]
where \(X_{ij}\) is the \(j\)th element of vector \(X_i\). Since \(S = \sum_{k=1}^{n} X_kX_k^T - n\bar{X}_n\bar{X}_n^T\),
\[
\frac{\partial S}{\partial X_{ij}} = \frac{\partial X_i}{\partial X_{ij}}X_i^T + X_i\frac{\partial X_i^T}{\partial X_{ij}} - \frac{\partial X_i}{\partial X_{ij}}\bar{X}_n - \bar{X}_n\frac{\partial X_i^T}{\partial X_{ij}}
\]
\[
= (X_i - \bar{X}_n)\frac{\partial X_i^T}{\partial X_{ij}} + \frac{\partial X_i}{\partial X_{ij}}(X_i - \bar{X}_n)^T. \tag{C-15}
\]

From Equations (C–14) and (C–15),
\[
\frac{\partial S^{-1}}{\partial X_{ij}} = -S^{-1}(X_i - \bar{X}_n)f_j^T - f_j(X_i - \bar{X}_n)^T S^{-1}, \tag{C-16}
\]
where \(f_j\) is the \(j\)th column of matrix \(S^{-1} = (f_1, \ldots, f_p)\). Now, from Equation (C–16), we see that
\[
(\frac{\partial S^{-1}}{\partial X_{ij}})(X_i - \bar{X}_n) = -S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T f_j
\]
\[-\text{tr}\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} f_j, \tag{C-17}
\]
and thus
\[
(\frac{\partial S^{-1}}{\partial X_i})(X_i - \bar{X}_n) = -S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}
\]
\[-\text{tr}\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} S^{-1}. \tag{C-18}
\]

The result of Equation (C–18), along with Equations (C–12) and (C–13), shows that
\[
\frac{\partial}{\partial X_i} S^{-1}(X_i - \bar{X}_n) = -S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}
\]
\[-\text{tr}\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\} S^{-1} + (1 - 1/n)S^{-1}. \tag{C-19}
\]

109
The result of Theorem 4.2.3 follows from combining Equations C–11 and C–19.

C.3 Proof of Theorem 4.4.1

Proof. Without loss of generality, we derive the Bayes risk of the $n$th limited translation estimator, $\hat{\theta}_{c,n}^{LEB}$. We begin with the Helmert orthogonal transformation

\[
Y_1 = n^{-\frac{1}{2}}(X_1 + \cdots + X_n) = \sqrt{n} \bar{X}_n
\]

\[
Y_2 = 2^{-\frac{1}{2}}(X_1 - X_2)
\]

\[\vdots\]

\[
Y_n = \{n(n-1)\}^{-\frac{1}{2}}(X_1 + \cdots + X_{n-1} - (n-1)X_n)
\]

\[= -(1 - 1/n)^{-\frac{1}{2}}(X_n - \bar{X}_n). \tag{C–20}\]

Then,

\[
S = \sum_{i=2}^{n} Y_i Y_i^T,
\]

\[
X_n - \bar{X}_n = -(1 - 1/n)^{\frac{1}{2}} Y_n, \tag{C–21}\]

and $Y_i \sim N_p(0, A + \Sigma), i = 2, \ldots, n$. Accordingly,

\[
\tilde{\theta}_{c,n}^{LEB} = X_n + (1 - 1/n)^{\frac{1}{2}} \hat{B}Y_n \rho_c(||a^{\frac{1}{2}}S^{-\frac{1}{2}}Y_n||^2), \tag{C–22}\]

\[
\tilde{\theta}_{n}^{EB} = X_n + (1 - 1/n)^{\frac{1}{2}} \hat{B}Y_n. \tag{C–23}\]

Next we calculate $r_1(\xi, \tilde{\theta}_{c,n}^{LEB})$,

\[
r_1(\xi, \tilde{\theta}_{c,n}^{LEB}) = E\{(\theta_n - \tilde{\theta}_{c,n}^{LEB})(\theta_n - \tilde{\theta}_{c,n}^{LEB})^T\}
\]

\[= E\{(\theta_n - \tilde{\theta}_{n}^{EB} + \tilde{\theta}_{n}^{LEB} - \tilde{\theta}_{c,n}^{LEB})(\theta_n - \tilde{\theta}_{n}^{EB} + \tilde{\theta}_{n}^{EB} - \tilde{\theta}_{c,n}^{LEB})^T\}. \tag{C–24}\]

The Bayes risk of $\tilde{\theta}_{c,n}^{LEB}$ is now written as

\[
r_1(\xi, \tilde{\theta}_{c,n}^{LEB}) = r_1(\xi, \tilde{\theta}_{n}^{EB}) + E\{(\tilde{\theta}_{n}^{EB} - \tilde{\theta}_{c,n}^{LEB})(\tilde{\theta}_{n}^{EB} - \tilde{\theta}_{c,n}^{LEB})^T\}
\]

\[+ E\{(\theta_n - \tilde{\theta}_{n}^{EB})(\tilde{\theta}_{n}^{EB} - \tilde{\theta}_{c,n}^{LEB})^T\} + E\{(\tilde{\theta}_{n}^{EB} - \tilde{\theta}_{c,n}^{LEB})(\theta_n - \tilde{\theta}_{n}^{EB})^T\}. \tag{C–25}\]
We continue by considering \(E\left\{ (\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n})(\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n})^T \right\} \). Note that

\[
\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n} = \hat{B}(X_n - X_n)\{\rho_c(||k_1S^{-\frac{1}{2}}(X_n - X_n)||^2) - 1\}, \quad (C-26)
\]

where \(k_1 = (1 - 1/n)^{-\frac{1}{2}}a^\frac{1}{2} \). Also, writing \(d \) as equal in distribution

\[
S^{-\frac{1}{2}}(X_n - X_n) = -(1 - 1/n)^{\frac{1}{2}}(\sum_{i=2}^{n} Y_i Y_i^T)^{-1}Y_n
\]

\[
= -(1 - 1/n)^{\frac{1}{2}}(\sum_{i=2}^{n} Y_i Y_i^T + Y_n Y_n^T)^{-1}Y_n
\]

\[
d \equiv -(1 - 1/n)^{\frac{1}{2}}(A + \Sigma)^{-\frac{1}{2}}(W_1 + ZZ)^{-1}Z, \quad (C-27)
\]

where \(W_1 \sim W(I_p, n - 2, \Sigma) \) independently of \(Z \sim N(0, I_p) \). Further,

\[
||S^{-\frac{1}{2}}(X_n - X_n)||^2 = (X_n - X_n)^T S^{-\frac{1}{2}}(X_n - X_n)
\]

\[
\overset{d}{=} (1 - 1/n)Z^T(W_1 + ZZ)^{-1}Z = (1 - 1/n)||Z||^2. \quad (C-28)
\]

From Equations \(C-26-C-28 \), it follows that

\[
E\left\{ (\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n})(\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n})^T \right\} 
\]

\[
= a^2(1 - 1/n)\Sigma(A + \Sigma)^{-\frac{1}{2}}E\left[ (W_1 + ZZ)^{-1}ZZ^T(W_1 + ZZ)^{-1} \right]
\]

\[
\times\{\rho_c(a||Z||^2 - 1)^2\} (A + \Sigma)^{-\frac{1}{2}} \Sigma. \quad (C-29)
\]

We now continue with the calculation of

\[
M \equiv E\left[ (W_1 + ZZ)^{-1}ZZ^T(W_1 + ZZ)^{-1} \right]
\]

\[
\times\{\rho_c(a||Z||^2 - 1)^2\} = M_{1,0} - 2M_{2,0} + M_{3,0}, \quad (C-30)
\]

where, for \(k = 1, 2, 3 \) and \(l = 0, 1 \), \(M_{k,l} \) are defined as

\[
M_{k,l} \equiv E\left\{ (W_1 + ZZ)^{-\frac{l-1}{2}}ZZ^T(W_1 + ZZ)^{-\frac{l}{2}}\rho_c^{k-1}(a||W_1 + ZZ||^2) \right\}. \quad (C-31)
\]

For the time being, we need only \(M_{k,0}, k=1, 2, 3 \), but we will need \(M_{k,1}, k=1, 2 \), in order to calculate the cross product, \(E\left\{ (\theta_n - \hat{\theta}^{EB}_n)(\hat{\theta}^{EB}_n - \tilde{\theta}^{EB}_{c,n})^T \right\} \). Now, let \(U = \)
Then the matrices $UU^T$ and $W_1 + ZZ^T$ are independently distributed (Srivastava & Khatri (1979), p. 95). This independence result allows us to simplify matters. We first rewrite $M_{k,l}$ as

$$M_{k,l} = E\left\{ (W_1 + ZZ^T)^{-\frac{1}{2}}UU^T(W_1 + ZZ^T)^{-\frac{1}{2}}\rho_{c}^{k-1}(a||U||^2) \right\}. \quad (C-32)$$

Notice that $||U||^2 = \text{tr}(UU^T)$. Thus, alternatively, $M_{k,l}$ is written as

$$M_{k,l} = E\left\{ (W_1 + ZZ^T)^{-\frac{1}{2}}H(W_1 + ZZ^T)^{-\frac{1}{2}} \right\}, \quad (C-33)$$

where $H_k = E\{UU^T\rho_{c}^{k-1}(a||U||^2)\}$. Now, the density of the random vector $U$ is given by

$$f_U(u) = c^*|I_p - uu^T|^{n-p-3}I[u^T u \leq 1] = c^*(1 - u^T u)^{\frac{n-p-3}{2}}I[u^T u \leq 1], \quad (C-34)$$

(Srivastava & Khatri (1979), p. 95), where $I[.]$ is the indicator function and $c^*$ is the normalizing constant. It follows that for $i \neq j,$

$$E\{U_iU_j\rho_{c}^{k-1}(a||U||^2)\} = E\{U_i(-U_j)\rho_{c}^{k-1}(a||U||^2)\}. \quad (C-35)$$

The above equality holds true if and only if $E\{U_iU_j\rho_{c}^{k-1}(a||U||^2)\} = 0, i \neq j$. Further, we have that

$$E\{U_1^2\rho_{c}^{k-1}(a||U||^2)\} = \cdots = E\{U_p^2\rho_{c}^{k-1}(a||U||^2)\}$$

$$= p^{-1}E\left\{ (\sum_{i=1}^{p} U_i^2)\rho_{c}^{k-1}(a||U||^2) \right\}$$

$$= p^{-1} \int_{-1}^{1} \cdots \int_{-1}^{1} (u'u)\rho_{c}^{k-1}(au'u)f(u)du_1 \cdots du_p$$

$$= p^{-1}\frac{\int_{0}^{1} \cdots \int_{0}^{1} (u'u)\rho_{c}^{k-1}(au'u)(1 - u'u)^{\frac{n-p-3}{2}}du_1 \cdots du_p}{\int_{0}^{1} \cdots \int_{0}^{1} (1 - u'u)^{\frac{n-p-3}{2}}du_1 \cdots du_p}. \quad (C-36)$$

In order to evaluate the above integral we consider the polar transformation

$$u_1 = r \sin \theta_1, u_2 = r \cos \theta_1 \sin \theta_2, u_3 = r \cos \theta_1 \cos \theta_2 \sin \theta_3,$$

$$\ldots, u_{p-1} = r \cos \theta_1 \ldots \cos \theta_{p-2} \sin \theta_{p-1}, u_p = r \cos \theta_1 \ldots \cos \theta_{p-2} \cos \theta_{p-1}. \quad (C-37)$$
Finally, we need to calculate which when combined with Equation (C–33) gives

\[ E\{U_i^2 \rho^{-1}_c(a||U||^2)\} = p^{-1} \int_0^1 \rho^{-1}_c(r)^{p+1}(1 - r^{n-p/2}) dr \]

\[ = p^{-1} \int_0^1 \rho^{-1}_c(r)^{p+1}(1 - r^{n-p/2}) dr = p^{-1} \frac{\text{Beta}(\frac{p+2}{2}, \frac{n-p-1}{2})}{\text{Beta}(\frac{p}{2}, \frac{n-p-1}{2})} E\{\rho^{-1}_c(aW)\} \]

\[ = (n-1)^{-1} E\{\rho^{-1}_c(aW)\}, \quad (C-38) \]

where \( W \sim \text{Beta}((p+2)/2, (n-p-1)/2) \). Thus, \( H_k \) is a diagonal matrix with entries \((n-1)^{-1} E\{\rho^{-1}_c(aW)\}\) in its main diagonal, that is \( H_k = (n-1)^{-1} E\{\rho^{-1}_c(aW)\} I_p \).

This, along with Equation C–33, implies that, for \( k = 1, 2, 3 \) and \( l = 0, 1 \),

\[ M_{k,l} = (n-1)^{-1} E\{\rho^{-1}_c(aW)\} E(W_1 + ZZ^T)^{l-1}. \quad (C-39) \]

Since \((W_1 + ZZ^T) \sim W_p(I_p, n-1)\),

\[ M_{k,0} = (n-1)^{-1}(n-p-2)^{-1} E\{\rho^{-1}_c(aW)\} I_p, \quad (C-40) \]

\[ M_{k,1} = (n-1)^{-1} E\{\rho^{-1}_c(aW)\} I_p = (n-p-2)M_{k,0}. \quad (C-41) \]

Thus, returning to Equation C–30, we see that

\[ M = (n-1)^{-1}(n-p-2)^{-1}[1 - 2E\{\rho_c(aW)\} + E\{\rho^2_c(aW)\}]I_p \]

\[ = (n-1)^{-1}(n-p-2)^{-1}E\{1 - \rho_c(aW)\}^2 I_p, \quad (C-42) \]

which when combined with Equation C–29 gives

\[ E\{(\hat{\theta}^E_{n} - \bar{\theta}^L_{c,n})(\bar{\theta}^E_{n} - \bar{\theta}^L_{c,n})^T\} = a^2 n^{-1}(n-p-2)^{-1}E\{1 - \rho_c(aW)\}^2 B \Sigma. \quad (C-43) \]

Finally, we need to calculate

\[ E\{(\theta_n - \bar{\theta}^E_{n})(\bar{\theta}^E_{n} - \bar{\theta}^L_{c,n})^T\} = E\{E(\theta_n - \bar{\theta}^E_{n})(\bar{\theta}^E_{n} - \bar{\theta}^L_{c,n})^T | X\} = \]

113
\[ E\left[ \{ \hat{B}(\mathbf{X}_n - \bar{\mathbf{X}}_n) - B(\mathbf{X}_n - \mu + \bar{\mathbf{X}}_n) \} \times (\mathbf{X}_n - \bar{\mathbf{X}}_n)^T \hat{B}^T (\rho_c(||k_1\mathbf{S}^{-\frac{1}{2}}(\mathbf{X}_n - \bar{\mathbf{X}}_n)||^2) - 1) \right]. \]  \hspace{1cm} (C-44)

Using the results of the orthogonal transformation, shown in the Equations C–21, we write the above expectation as

\[
E\{ (\theta_n - \hat{\theta}_n^{EB})(\hat{\theta}_n^{EB} - \hat{\theta}_c,n^{EB})^T \} \\
= a(1 - 1/n) \Sigma E\left[ \{ a(\sum_{i=2}^{n} \mathbf{Y}_i \mathbf{Y}_i^T)^{-1} - (\mathbf{A} + \Sigma)^{-1} \} \mathbf{Y}_n \mathbf{Y}_n^T (\sum_{i=2}^{n} \mathbf{Y}_i \mathbf{Y}_i^T)^{-1} \right] \\
\times \{ \rho_c(||a^{\frac{1}{2}}(\sum_{i=2}^{n} \mathbf{Y}_i \mathbf{Y}_i^T)^{-\frac{1}{2}} \mathbf{Y}_n||^2) - 1 \} \Sigma \\
+ a(1 - 1/n)^{\frac{1}{2}} B E\left[ (n^{-\frac{1}{2}} \mathbf{Y}_1 - \mu) \mathbf{Y}_n^T (\sum_{i=2}^{n} \mathbf{Y}_i \mathbf{Y}_i^T)^{-1} \right] \\
\times \{ \rho_c(||a^{\frac{1}{2}}(\sum_{i=2}^{n} \mathbf{Y}_i \mathbf{Y}_i^T)^{-\frac{1}{2}} \mathbf{Y}_n||^2) - 1 \} \Sigma, \hspace{1cm} (C-45)
\]

and due to the independence of the \( \mathbf{Y}_i, i = 1, \ldots, n, \) and \( n^{-\frac{1}{2}} E(\mathbf{Y}_1) = \mu, \) the second of the two terms of the last expression is equal to a matrix of zeros. We now continue with the calculation of the first expectation in Equation C–45. Writing \( \mathbf{Z}_i = (\mathbf{A} + \Sigma)^{-\frac{1}{2}} \mathbf{Y}_i, \) \( i = 2, \ldots, n, \) yields

\[
(A + \Sigma)^{-\frac{1}{2}} E\left[ \{ a(\sum_{i=2}^{n} \mathbf{Z}_i \mathbf{Z}_i^T)^{-1} - I_p \} \mathbf{Z}_n \mathbf{Z}_n^T (\sum_{i=2}^{n} \mathbf{Z}_i \mathbf{Z}_i^T)^{-1} \right] \\
\times \{ \rho_c(a(||\sum_{i=2}^{n} \mathbf{Z}_i \mathbf{Z}_i^T)^{-\frac{1}{2}} \mathbf{Z}_n||^2) - 1 \} \right](A + \Sigma)^{-\frac{1}{2}} \\
= (A + \Sigma)^{-\frac{1}{2}} (a\mathbf{M}_{2,0} - a\mathbf{M}_{1,0} - \mathbf{M}_{2,1} + \mathbf{M}_{1,1})(A + \Sigma)^{-\frac{1}{2}}. \hspace{1cm} (C-46)
\]

Recalling the results of Equations C–40 and C–41 and combining Equations C–45 and C–46 yields

\[
E\{ (\theta_n - \hat{\theta}_n^{EB})(\hat{\theta}_n^{EB} - \hat{\theta}_c,n^{EB})^T \} = an^{-1}\{1 - a(n - p - 2)^{-1}\}\{1 - E\rho_c(aW)\} B \Sigma, \hspace{1cm} (C-47)
\]

which for the choice of \( a = n - p - 2 \) is equal to a matrix of zeros.
The result of the Theorem follows from combining the result of Theorem 4.2.1 with Equations C–25, C–43 and C–47.

C.4 Proof of Theorem 4.5.1

Proof. Starting from Equation 4–20 and using the multivariate version of Stein’s identity, we write

\[
E_{\theta} \{(\theta_i - X_i)(X_i - \bar{X}_n)^T S^{-1} \rho_c \} = -\Sigma E_{\theta} \left[ \partial \{S^{-1}(X_i - \bar{X}_n)\rho_c\}/\partial X_i \right]. \tag{C–48}
\]

Now, using the differentiation product rule, we obtain that

\[
\frac{\partial [S^{-1}(X_i - \bar{X}_n)\rho_c]}{\partial X_i} = \frac{\partial [S^{-1}(X_i - \bar{X}_n)]}{\partial X_i} \rho_c + S^{-1}(X_i - \bar{X}_n) \frac{\partial \rho_c}{\partial X_i^T}. \tag{C–49}
\]

We have provided an expression for \(\frac{\partial [S^{-1}(X_i - \bar{X}_n)]}{\partial X_i}\) in Equation C–19 and we now obtain an expression for \(\frac{\partial \rho_c}{\partial X_i}\). Since the function \(\rho_c\) is given as

\[
\rho_c = I[||k_1 S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| \leq c] + \frac{cI[||k_1 S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c]}{||k_1 S^{-\frac{1}{2}}(X_i - \bar{X}_n)||}, \tag{C–50}
\]

it follows that

\[
\frac{\partial \rho_c}{\partial X_i} = \rho_d I[||k_1 S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c], \tag{C–51}
\]

where

\[
\rho_d = c \partial \{k_1^2 (X_i - \bar{X}_n)^T S^{-1}(X_i - \bar{X}_n)\}^{-\frac{1}{2}} / \partial X_i \nonumber
\]

\[
= -\frac{c}{2k_1} ||S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^{-3} \frac{\partial}{\partial X_i} \{ (X_i - \bar{X}_n)^T S^{-1}(X_i - \bar{X}_n) \}. \tag{C–52}
\]

Using the differentiation product rule and the result of Equation C–16, we can show that

\[
\frac{\partial}{\partial X_{ij}} (X_i - \bar{X}_n) S^{-1}(X_i - \bar{X}_n)^T = 2(1 - 1/n) f_j^T (X_i - \bar{X}_n) \nonumber
\]

\[
-2f_j^T (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}(X_i - \bar{X}_n). \tag{C–53}
\]
where $f_j$ is the $j^{th}$ column of $S^{-1}$. Therefore

$$
\frac{\partial}{\partial X_i}(X_i - \bar{X}_n)S^{-1}(X_i - \bar{X}_n)^T = 2(1 - 1/n)S^{-1}(X_i - \bar{X}_n)
- 2S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}(X_i - \bar{X}_n).
$$

(C-54)

Thus, by combining Equations C–51, C–52 and C–54, we obtain

$$
\frac{\partial \rho_c}{\partial X_i^T} = -c \frac{\hat{I}[||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c]1 - n^{-1} - ||S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^2}{||S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^3}
\times (X_i - \bar{X}_n)^T S^{-1}.
$$

(C-55)

Further, we combine the Equations C–55, C–49 and C–19 to obtain

$$
\begin{align*}
\frac{\partial}{\partial X_i}[S^{-1}(X_i - \bar{X}_n)\rho_c]/\partial X_i & \\
& = -\text{tr}\{S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T\}S^{-1}\rho_c + (1 - 1/n)S^{-1}\rho_c \\
& - S^{-1}(X_i - \bar{X}_n)(X_i - \bar{X}_n)^T S^{-1}\rho_c \\
& \times \left(1 - 1/n\right)||S^{-\frac{1}{2}}(X_i - \bar{X}_n)||^{-2}\hat{I}[||k_1S^{-\frac{1}{2}}(X_i - \bar{X}_n)|| > c].
\end{align*}
$$

(C-56)

Combining Equations C–56 with 4–20 and C–48 completes the proof of Theorem 4.5.1. □
APPENDIX D
PROOF OF THEOREM 5 SERIES

D.1 Proof of Theorem 5.2.1

Proof. First we write

\[ r_1(\xi, \hat{\theta}_i^{EB}) = E\{(\theta_i - \hat{\theta}_i^{EB})(\theta_i - \hat{\theta}_i^{EB})^T\} \]

\[ = E\{(\theta_i - \hat{\theta}_i^B + \hat{\theta}_i^B - \hat{\theta}_i^{EB})(\theta_i - \hat{\theta}_i^B + \hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\} \]

\[ = (I_p - B)\Sigma + E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\}. \quad \text{(D-1)} \]

Now,

\[ \hat{\theta}_i^B - \hat{\theta}_i^{EB} = \tilde{B}(\bar{X}_i - \bar{X}) - B(\bar{X}_i - \mu) = (\tilde{B} - B)(\bar{X}_i - \bar{X}) - B(\bar{X} - \mu), \quad \text{(D-2)} \]

and it follows that

\[ E\{(\hat{\theta}_i^B - \hat{\theta}_i^{EB})(\hat{\theta}_i^B - \hat{\theta}_i^{EB})^T\} \]

\[ = E\{(\tilde{B} - B)(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T(\tilde{B} - B)^T\} \]

\[ + E\{B(\bar{X} - \mu)(\bar{X} - \mu)^T B^T\}. \quad \text{(D-3)} \]

Marginally, \( \bar{X}_i \overset{\text{iid}}{\sim} N_p(\mu, \Sigma + A) \), \( i = 1, \ldots, n \), and thus

\[ E\{B(\bar{X} - \mu)(\bar{X} - \mu)^T B^T\} = n^{-1}BS. \quad \text{(D-4)} \]

We now calculate \( W_i \equiv E\{(\tilde{B} - B)(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T(\tilde{B} - B)^T\} \). Notice that

\[ W_i = W \equiv n^{-1}\sum_{i=1}^n W_i = n^{-1}E\{(\tilde{B} - B)S(\tilde{B} - B)^T\} \]

Now,

\[ E(\tilde{B}S\tilde{B}^T) = a^2b^2k^{-2}E(VS^{-1}V), \quad \text{(D-5)} \]

and

\[ E(VS^{-1}V) = E\{E(VS^{-1}V|V)\} = E\{VE(S^{-1}V)\} \]

\[ = (n - p - 2)^{-1}E\{V(A + \Sigma)^{-1}V\}. \quad \text{(D-6)} \]
Recall that $V \sim W_p(k\Sigma, df = n(k - 1))$, that is $V \overset{d}{=} \sum_{i=1}^{n(k-1)} Y_i Y_i^T$ where $Y_i \overset{iid}{\sim} N_p(0, k\Sigma)$. Also, note that $Y_i \overset{d}{=} (k\Sigma)^{\frac{1}{2}} Z_i$ where $Z_i \overset{iid}{\sim} N_p(0, I_p)$. We thus write
\[
E\{V(A + \Sigma)^{-1}V\} = k^{2}\Sigma^{\frac{1}{2}} E\left\{ \sum_{i=1}^{n(k-1)} (Z_i Z_i^T) \Sigma^{\frac{1}{2}} (A + \Sigma)^{-1} \Sigma^{\frac{1}{2}} \sum_{j=1}^{n(k-1)} (Z_j Z_j^T) \right\} \Sigma^{\frac{1}{2}}
\]
\[
= k^{2}\Sigma^{\frac{1}{2}} E\left\{ \sum_{i\neq j} Z_i Z_i^T \Sigma^{\frac{1}{2}} (A + \Sigma)^{-1} \Sigma^{\frac{1}{2}} Z_j Z_j^T \right\} \Sigma^{\frac{1}{2}}
\]
\[
+ k^{2}\Sigma^{\frac{1}{2}} E\left\{ \sum_{i=1}^{n(k-1)} Z_i Z_i^T \Sigma^{\frac{1}{2}} (A + \Sigma)^{-1} \Sigma^{\frac{1}{2}} Z_i Z_i^T \right\} \Sigma^{\frac{1}{2}}, \tag{D–7}
\]
and it is now easy to see that
\[
E\{V(A + \Sigma)^{-1}V\} = k^{2} n(k - 1)\{n(k - 1) - 1\} B \Sigma
\]
\[+ k^{2} n(k - 1) \Sigma^{\frac{1}{2}} E\{ZZ^T \Sigma^{\frac{1}{2}} (A + \Sigma)^{-1} \Sigma^{\frac{1}{2}} ZZ^T \} \Sigma^{\frac{1}{2}}, \tag{D–8}
\]
where $Z$ has the standard normal distribution. Let $D \equiv \Sigma^{\frac{1}{2}} (A + \Sigma^{*})^{-1} \Sigma^{\frac{1}{2}}$. Now the expectation in the last line of D–8 is written as
\[
Q \equiv E\{ZZ^T D ZZ^T\} = E\{(ZZ^T D) ZZ^T\} = E(\sum_{i,j=1}^{p} z_i z_j d_{ij} ZZ^T), \tag{D–9}
\]
where $d_{ij}$ is the $(i, j)^{th}$ element of the matrix $D$, $i, j = 1, \ldots, p$, and $z_i$ is the $i^{th}$ element of vector $Z$, $i = 1, \ldots, p$. The $k^{th}$ diagonal element of matrix $Q$ is calculated as
\[
E\{\sum_{i,j=1}^{p} (z_i z_j d_{ij}) z_k^2\} = E(\sum_{i=1}^{p} z_i^2 d_{ii} z_k^2) = \text{tr}(D) + 2d_{kk} \tag{D–10}
\]
while the expression of the $(k, l)^{th}$, $k \neq l$, element of matrix $Q$ is obtained as
\[
E\{\sum_{i,j=1}^{p} (z_i z_j d_{ij}) z_k z_l\} = 2d_{kl}, \tag{D–11}
\]
since the matrix $D$ is symmetric.

Combining Equations D–9-D–11, we obtain that
\[
Q = 2D + \text{tr}(D)I_p. \tag{D–12}
\]
Next, combining Equations D–8, D–9 and D–12 follows that

\[
E(\hat{B}SB^T) = a^2b^2(n - p - 2)^{-1}n(k - 1)\left\{\{n(k - 1) + 1\}B + \text{tr}(B)I_p\right\}\Sigma. \tag{D–13}
\]

Now, from Equations D–5, D–6 and D–13 follows that

\[
E(\hat{B}SB^T) = a^2b^2(n - p - 2)^{-1}n(k - 1)\left\{\{n(k - 1) + 1\}B + \text{tr}(B)I_p\right\}\Sigma. \tag{D–14}
\]

Further, we have the following two results

\[
E(BSB^T) = (n - 1)B\Sigma, \tag{D–15}
\]

\[
E(\hat{B}SB^T) = abn(k - 1)B\Sigma. \tag{D–15}
\]

From Equations D–14-D–15 follows that

\[
E\{(\tilde{B} - B)(\bar{X}_i - \bar{X}_c)(\bar{X}_i - \bar{X}_c)^T(\tilde{B} - B)^T\}
\]

\[
= n^{-1}\left[a^2b^2(n - p - 2)^{-1}n(k - 1)\{n(k - 1) + 1\} + n - 1 - 2abn(k - 1)\right]B\Sigma
\]

\[
+ n^{-1}a^2b^2(n - p - 2)^{-1}n(k - 1)\text{tr}(B)\Sigma \tag{D–16}
\]

The result of the Theorem follows from combining Equations D–1, D–3, D–4 and D–16.

D.2 Proof of Theorem 5.3.1

Proof. Without loss of generality we calculate the Bayes risk of the estimator of \(\theta_n, \hat{\theta}_{c,n}^{\text{LEB}}\).

\[
r_1(\xi, \hat{\theta}_{c,n}^{\text{LEB}}) = E\{(\theta_n - \hat{\theta}_{c,n}^{\text{LEB}})(\theta_n - \hat{\theta}_{c,n}^{\text{LEB}})^T\}
\]

\[
= E\{(\theta_n - \hat{\theta}_n^{\text{EB}} + \hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})(\theta_n - \hat{\theta}_n^{\text{EB}} + \hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})^T\}
\]

\[
= r_1(\xi, \hat{\theta}_n^{\text{EB}}) + E\{(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})^T\}
\]

\[
+ E\{(\theta_n - \hat{\theta}_n^{\text{EB}})(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})^T\} + E\{(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})(\theta_n - \hat{\theta}_n^{\text{EB}})^T\}. \tag{D–17}
\]

In order to calculate \(E\{(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})(\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}})^T\}\) we write

\[
\hat{\theta}_n^{\text{EB}} - \hat{\theta}_{c,n}^{\text{LEB}} = \tilde{B}(\bar{X}_n - \bar{X}_c)\{\rho_c(||k_1S^{-\frac{1}{2}}(\bar{X}_n - \bar{X}_c)||^2) - 1\}. \tag{D–18}
\]
where \(\rho_c(u) = \min(1, c/\sqrt{u})\), and \(k_1 = (1 - 1/n)^{-\frac{1}{2}}a^{\frac{1}{2}}\).

Note that marginally \(\bar{X}_i \sim \mathcal{N}_p(\mu, A + \Sigma)\), \(i = 1, \ldots, n\). Consider now the Helmert orthogonal transformation

\[
\begin{align*}
Y_1 &= n^{-\frac{1}{2}}(X_1 + \cdots + X_n) = n^{\frac{1}{2}} \bar{X}.
Y_2 &= 2^{\frac{1}{2}}(X_1 - \bar{X}_2)
\vdots
Y_n &= \{n(n-1)\}^{-\frac{1}{2}}(X_1 + \cdots + X_{n-1} - (n-1)\bar{X}_n)
&= -(1 - 1/n)^{-\frac{1}{2}}(\bar{X}_n - \bar{X}).
\end{align*}
\]

Then

\[
\begin{align*}
S &= \sum_{i=2}^{n} Y_i Y_i^T,
\bar{X}_n - \bar{X} &= -(1 - 1/n)^{\frac{1}{2}} Y_n.
\end{align*}
\]

Also, \(Y_i \sim \mathcal{N}_p(0, A + \Sigma), i = 2, \ldots, n\). Recall that \(\bar{B} = abk^{-1}VS^{-1}\). Then,

\[
\begin{align*}
\hat{\theta}_n^{EB} - \hat{\theta}_n^{LEB} &= -(1 - 1/n)^{\frac{1}{2}}abk^{-1}VS^{-1}Y_n\{\rho_c(||a^{\frac{1}{2}}S^{-\frac{1}{2}}Y_n||^2) - 1\},
\end{align*}
\]

and thus

\[
E\{(\hat{\theta}_n^{EB} - \hat{\theta}_n^{LEB}) (\hat{\theta}_n^{EB} - \hat{\theta}_n^{LEB})^T\}
= (1 - 1/n)a^{2}b^{2}k^{-2}E\left[VS^{-1}Y_n Y_n^T S^{-1}V \{\rho_c(||a^{\frac{1}{2}}S^{-\frac{1}{2}}Y_n||^2) - 1\}^2 \right].
\]

Since \(V\) is independent of \((S, Y_n)\), the last expectation can be calculated as

\[
E\left[VE[S^{-1}Y_n Y_n^T S^{-1} \{\rho_c(||a^{\frac{1}{2}}S^{-\frac{1}{2}}Y_n||^2) - 1\}]^2] \right].
\]

In order to evaluate the inner expectation notice that

\[
S^{-1} Y_n = \sum_{i=2}^{n-1} Y_n Y_n^T Y_n \overset{d}{=} (A + \Sigma)^{-\frac{1}{2}}(W_1 + ZZ)^{-1}Z.
\]
Also, $\|S^{-\frac{1}{2}} Y_n\| \overset{d}{=} \|(W_1 + ZZ^T)^{-\frac{1}{2}} Z\|$, where $W_1 \sim W_p(I_p, n - 2)$ independently of $Z \sim N_p(0, I_p)$. Therefore

$$E[S^{-1} Y_n Y_n^T S^{-1}\{\rho_c(||a^2 S^{-\frac{1}{2}} Y_n||^2) - 1\}^2] = (A + \Sigma)^{-\frac{1}{2}} E[(W_1 + ZZ^T)^{-1} ZZ^T (W_1 + ZZ^T)^{-1} \times \{\rho_c(||a^2 (W_1 + ZZ^T)^{-\frac{1}{2}} Z||^2) - 1\}^2] (A + \Sigma)^{-\frac{1}{2}}. \quad (D-25)$$

In Equation C–39 we showed that

$$M_{k,l} \equiv E\{(W_1 + ZZ^T)^{l-1} ZZ^T (W_1 + ZZ^T)^{-1} \rho_c^{k-1}(a||(W_1 + ZZ^T)^{-\frac{1}{2}} Z||^2)\} = (n - 1)^{-1} E\{\rho_c^{k-1}(aW)\} E(W_1 + ZZ^T)^{l-1}, \quad (D-26)$$

where $W \sim \text{Beta}((p+2)/2, (n-p-1)/2)$, $k = 1, 2, 3$ and $l = 0, 1$. Hence, combining Equations D–22, D–23, D–25 and D–26, we obtain that

$$E\{\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB}\} \{\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB}\}^T = a^2 b^2 k^{-2} n^{-1} (n - p - 2)^{-1} E\{1 - \rho_c(aW)\}^2 E\{V (A + \Sigma)^{-1} V\}. \quad (D-27)$$

Using the result about $E\{V (A + \Sigma)^{-1} V\}$, given in Equation D–13, we obtain

$$E\{\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB}\} \{\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB}\}^T = a^2 b^2 (n - p - 2)^{-1} (k - 1) \times E\{1 - \rho_c(aW)\}^2 \{n(k - 1) + 1\} B + \text{tr}(B) \Sigma. \quad (D-28)$$

The final step for calculating the Bayes risk is to provide an expression for the cross product

$$L_n \equiv E\{(\theta_n - \hat{\Theta}_n^{EB})(\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB})^T\} = E\{E(\theta_n | \tilde{X}_n) - \hat{\Theta}_n^{EB}\} (\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB})^T. \quad (D-29)$$

To this end, we calculate

$$E(\theta_n | \tilde{X}_n) - \hat{\Theta}_n^{EB} = \hat{B}(\tilde{X}_n - \tilde{X}) - B(\tilde{X}_n - \mu - \tilde{X} + \tilde{X}),$$

$$\hat{\Theta}_n^{EB} - \hat{\Theta}_c,n^{EB} = \hat{B}(\tilde{X}_n - \tilde{X}) \{\rho_c(||k_1 S^{-\frac{1}{2}}(\tilde{X}_n - \tilde{X})||^2) - 1\}. \quad (D-30)$$
Thus, in light of Equation D–20 and writing $\rho_c$ for $\rho_c(||k_1S^{-\frac{1}{2}}(X_i - \bar{X})||^2)$, $L_n$ becomes

\[
L_n = E \left\{ -(1 - 1/n)^{\frac{1}{2}}\hat{B} Y_n + (1 - 1/n)^{\frac{1}{2}}B Y_n + B(\mu - n^{-\frac{1}{2}} Y_1) \right\} \times \left\{ -(1 - 1/n)^{\frac{1}{2}} Y_n^T(\hat{B})^T(\rho_c - 1) \right\}. \tag{D–31}
\]

Now, $n^{-\frac{1}{2}}E(Y_1) = \mu$ and $Y_1$ is independent of $(Y_n, S, V)$. Thus, $L_n$ is equal to

\[
L_n = (1 - 1/n)E\{ (\hat{B} - B) Y_n \} Y_n^T(\hat{B})^T(\rho_c - 1) = L_{n,1} + L_{n,2}, \tag{D–32}
\]

where

\[
L_{n,1} \equiv a^2b^2k^{-2}(1 - 1/n)E\{ VS^{-1} Y_n \} Y_n^T S^{-1} V(\rho_c - 1),
\]

\[
L_{n,2} \equiv -abk^{-1}(1 - 1/n)BE\{ Y_n \} Y_n^T S^{-1} V(\rho_c - 1). \tag{D–33}
\]

First,

\[
L_{n,1} = a^2b^2k^{-2}(1 - 1/n)E\{ V E\{ S^{-1} Y_n \} Y_n^T S^{-1}(\rho_c - 1) \} V \}. \tag{D–34}
\]

Now, the inner expectation is written as

\[
(A + \Sigma)^{-\frac{1}{2}}E\left[ \left( \sum_{i=2}^{n} Z_i Z_i^T \right)^{-1} Z_n \left( \sum_{i=2}^{n} Z_i Z_i^T \right)^{-1} \right] \times \rho_c(||a(\sum_{i=2}^{n} Z_i Z_i^T)^{-\frac{1}{2}}Z_n||^2 - 1) (A + \Sigma)^{-\frac{1}{2}} \tag{D–35}
\]

where $Z_i \sim N_p(0, I_p)$. Thus, combining Equations D–34 and D–35 with D–26 we obtain

\[
L_{n,1} = a^2b^2k^{-2}(1 - 1/n)E\{ V (A + \Sigma)^{-\frac{1}{2}} (M_{2,0} - M_{1,0})(A + \Sigma)^{-\frac{1}{2}} V \} \]

\[
= a^2b^2k^{-2}n^{-1}(n - p - 2)^{-1}E\{ \rho_c(aW) - 1 \} E\{ V (A + \Sigma)^{-1} V \}. \tag{D–36}
\]

This along with Equation D–13 shows that

\[
L_{n,1} = a^2b^2(k - 1)(n - p - 2)^{-1}E\{ \rho_c(aW) - 1 \} \{ n(k - 1) + 1 \} B + tr(B)I_p \Sigma. \tag{D–37}
\]
Similarly,
\[ L_{n,2} = -abk^{-1}(1 - 1/n)\Sigma(A + \Sigma)^{-\frac{1}{2}}E\left\{(Z_nZ_n^T(\sum_{i=2}^n Z_iZ_i^T)^{-1}\{\rho_c - 1\})\right\} \times (A + \Sigma)^{-\frac{1}{2}}V \]  
\[ \times (A + \Sigma)^{-\frac{1}{2}}V \]. \hspace{1cm} (D–38)

Using Equation D–26, we obtain the following expression
\[ L_{n,2} = -abk^{-1}(1 - 1/n)\Sigma(A + \Sigma)^{-\frac{1}{2}}E\left\{(M_{2,1} - M_{1,1})(A + \Sigma)^{-\frac{1}{2}}V\right\} \]
\[ = -ab(k - 1)E\{\rho_c(aW) - 1\}B\Sigma. \hspace{1cm} (D–39) \]

Hence, combining Equations D–32, D–37 and D–39, we obtain
\[ E\{(\theta_i - \hat{\theta}_i^{EB})(\hat{\theta}_i^{EB} - \hat{\theta}_i^{LEB})^T\} \]
\[ = ab(k - 1)E\{\rho_c(aW) - 1\}\left\{ab(n - p - 2)^{-1}\{n(k - 1) + 1\} - 1\right\}B\Sigma \]
\[ + ab^2(k - 1)(n - p - 2)^{-1}E\{\rho_c(aW) - 1\}tr(B)\Sigma. \hspace{1cm} (D–40) \]

Let \( a_n = n - p - 2 \) and \( b_n = n^{-1}(k - 1)^{-1} \). Then, combining Equations D–17, 5–9, D–28 and D–40, and collecting the terms of \( \Sigma, B\Sigma \) and \( tr(B)\Sigma \) separately, Theorem 5.3.1 follows.

D.3 Proof of Theorem 5.4.1

Proof. Starting from Equation 5–17, we write
\[ R_1(\theta_i, \hat{\theta}_i^{EB}) = \Sigma + a^2b^2k^{-2}E\theta\{VS^{-1}(\bar{X}_i - \bar{X}_.)\bar{X}_i - \bar{X}_.)^T S^{-1}V\} \]
\[ + abk^{-1}E\theta\{(\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X}_.)^T S^{-1}V\} \]
\[ + abk^{-1}E\theta\{VS^{-1}(\bar{X}_i - \bar{X}_.)(\theta_i - \bar{X}_i)^T\}. \hspace{1cm} (D–41) \]

Consider now the following expectation
\[ E\theta\{(\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X}_.)^T S^{-1}V\} = n(k - 1)kE\theta\{(\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X}_.)^T S^{-1}\} \Sigma. \hspace{1cm} (D–42) \]
Write $\bar{X}^{(-i)} = (\bar{X}_1^T, \ldots, \bar{X}_{i-1}^T, \bar{X}_{i+1}^T, \ldots, \bar{X}_n^T)^T$ and using the multivariate version of Stein’s identity, given in Lemma 4.2.2, the expectation in the right hand site of Equation D–42 can be written as

$$E_\theta \{ \theta_i - \bar{X}_i \} (\bar{X}_i - \bar{X})^T S^{-1} = E_\theta \left[ E_\theta \{ \theta_i - \bar{X}_i \} (\bar{X}_i - \bar{X})^T S^{-1} | \bar{X}^{(-i)} \} \right]$$

$$= -\Sigma E_\theta \left[ \partial \{ S^{-1} (\bar{X}_i - \bar{X}) \} / \partial \bar{X}_i \right].$$  \(\text{(D–43)}\)

The calculation of the derivatives can be achieved by using the product rule as follows

$$\partial \{ S^{-1} (\bar{X}_i - \bar{X}) \} / \partial \bar{X}_i = \{ \partial S^{-1} / \partial \bar{X}_i \} (\bar{X}_i - \bar{X}) + S^{-1} \{ \partial (\bar{X}_i - \bar{X}) / \partial \bar{X}_i \}. \quad (\text{D–44})$$

It is easy to see that

$$\partial (\bar{X}_i - \bar{X}) / \partial \bar{X}_i = (1 - 1/n) I_p. \quad (\text{D–45})$$

Also, the following equality holds true

$$\partial S^{-1} / \partial \bar{X}_{ij} = -S^{-1} \{ \partial S / \partial \bar{X}_{ij} \} S^{-1}, \quad (\text{D–46})$$

where $\bar{X}_{ij}$ is the $j$th element of vector $\bar{X}_i$. We write $S = \sum_{m=1}^n \bar{X}_m \bar{X}_m^T - n \bar{X} \bar{X}^T$, and using the product rule again we see that

$$\frac{\partial S}{\partial \bar{X}_{ij}} = \frac{\partial \bar{X}_i}{\partial \bar{X}_{ij}} \bar{X}_i^T + \bar{X}_i \frac{\partial \bar{X}_i^T}{\partial \bar{X}_{ij}} - \bar{X}_i \frac{\partial \bar{X}_i}{\partial \bar{X}_{ij}} \bar{X}_i^T - \bar{X}_i \frac{\partial \bar{X}_i^T}{\partial \bar{X}_{ij}}$$

$$= (\bar{X}_i - \bar{X}) \frac{\partial \bar{X}_i}{\partial \bar{X}_{ij}} + \frac{\partial \bar{X}_i}{\partial \bar{X}_{ij}} (\bar{X}_i - \bar{X}). \quad (\text{D–47})$$

From Equations D–46 and D–47 follows that,

$$\partial S^{-1} / \partial \bar{X}_{ij} = -S^{-1} (\bar{X}_i - \bar{X}) f_j^T - f_j (\bar{X}_i - \bar{X})^T S^{-1}, \quad (\text{D–48})$$
where $f_j$ is the $j$th column of matrix $S^{-1} = (f_1, \ldots, f_p)$. Now, using Equation D–48, we see that

$$
(\partial S^{-1}/\partial \bar{X}_{ij})(\bar{X}_i - \bar{X}_\cdot) = -S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T f_j \\
-\text{tr}\{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T f_j\}.
$$

(D–49)

It follows that

$$
(\partial S^{-1}/\partial \bar{X}_i)(\bar{X}_i - \bar{X}_\cdot) = -S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1} \\
-\text{tr}\{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}\}.
$$

(D–50)

The result in Equation D–50, along with Equations D–44 and D–45, shows that

$$
\partial \{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}\}/\partial \bar{X}_i
= -S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1} \\
-\text{tr}\{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}\} S^{-1} + (1 - 1/n)S^{-1}.
$$

(D–51)

Also, using similar reasoning as in Equations D–6–D–12 we can show that

$$
E_{\theta}\{V S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}V\}
= k^2 n(k - 1)\{n(k - 1) + 1\}\Sigma E_{\theta}\{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}\} \Sigma
+ k^2 n(k - 1)\text{tr}\left[E_{\theta}\{S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1}\} \Sigma\right] \Sigma.
$$

(D–52)


\[\square\]

**D.4 Proof of Theorem 5.5.1**

Proof. Starting from Equation 5–21, we write

$$
R_1(\theta_i, \hat{\theta}_{c,i}^{LEB}) = \Sigma + a^2 b^2 k^{-2} E_{\theta}\{V S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1} V\rho_c^2\}
+ abk^{-1} E_{\theta}\{ (\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X}_\cdot)^T S^{-1} V \rho_c\}
+ abk^{-1} E_{\theta}\{ V S^{-1}(\bar{X}_i - \bar{X}_\cdot)(\theta_i - \bar{X}_i)^T \rho_c\}.
$$

(D–53)
Using Stein’s identity, provided in Lemma 4.2.2, we obtain an expression for

\[ E_{\theta}\{(\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X})^T S^{-1} V \rho_c\} \]

\[ = n(k - 1)k E_{\theta}\{(\theta_i - \bar{X}_i)(\bar{X}_i - \bar{X})^T S^{-1} \rho_c\} \Sigma \]

\[ = -n(k - 1)k \Sigma E_{\theta}\left[\partial\{S^{-1}(\bar{X}_i - \bar{X})\rho_c\}/\partial \bar{X}_i\right] \Sigma. \quad (D-54) \]

We now continue by calculating the matrix derivative that appears in the last line of Equation D–54. First, the differentiation product rule shows that

\[ \frac{\partial[S^{-1}(\bar{X}_i - \bar{X})\rho_c]}{\partial \bar{X}_i} = \frac{\partial[S^{-1}(\bar{X}_i - \bar{X})]}{\partial \bar{X}_i} \rho_c + S^{-1}(\bar{X}_i - \bar{X}) \frac{\partial \rho_c}{\partial \bar{X}_i^T}. \quad (D-55) \]

In order to calculate \( \partial \rho_c / \partial \bar{X}_i \), we write the function \( \rho_c \) as

\[ \rho_c = I[||k_1 S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})|| \leq c] + \frac{c I[||k_1 S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})|| > c]}{||k_1 S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||}. \quad (D-56) \]

It follows that

\[ \partial \rho_c / \partial \bar{X}_i = \rho_d I[||k_1 S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})|| > c], \quad (D-57) \]

where

\[ \rho_d = c \partial\{k_1^2((\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X}))\}^{-\frac{1}{2}} / \partial \bar{X}_i \]

\[ = -\frac{c}{2k_1} ||S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^{-3} \frac{\partial}{\partial \bar{X}_i}\left\{(\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X})\right\}. \quad (D-58) \]

Further, it can be shown that

\[ \frac{\partial}{\partial \bar{X}_{ij}} (\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X}) = 2(1 - 1/n) f_j^T (\bar{X}_i - \bar{X}) \]

\[ -2 f_j^T (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X}), \quad (D-59) \]

and now it is easy to see that

\[ \frac{\partial}{\partial \bar{X}_i} (\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X}) = 2(1 - 1/n) S^{-1}(\bar{X}_i - \bar{X}) \]

\[ -2 S^{-1}(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})^T S^{-1}(\bar{X}_i - \bar{X}). \quad (D-60) \]
From Equations D–57, D–58 and D–60 follows that

\[
\frac{\partial \rho_c}{\partial \bar{X}_i^T} = -\frac{c}{k_1} I||k_1 S^{-\frac{1}{2}} (\bar{X}_i - \bar{X})|| > c \frac{1-n^{-1} - ||S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^2}{||S^{-\frac{1}{2}}(\bar{X}_i - \bar{X})||^3} \\
\times (\bar{X}_i - \bar{X})^T S^{-1}.
\]  

(D–61)

Combining Equations D–55, D–61 and D–51 we obtain that

\[
\frac{\partial}{\partial \bar{X}_i} [S^{-1}(\bar{X}_i - \bar{X}_.) \rho_c]/\partial \bar{X}_i \\
= -\text{tr}\{S^{-1}(\bar{X}_i - \bar{X}_.)(\bar{X}_i - \bar{X})^T\} S^{-1} \rho_c + (1 - 1/n) S^{-1} \rho_c \\
- S^{-1}(\bar{X}_i - \bar{X}_.)(\bar{X}_i - \bar{X})^T S^{-1} \rho_c \\
\times \left\{ (1 - 1/n)||S^{-\frac{1}{2}}(\bar{X}_i - \bar{X}_.)||^{-2} I||k_1 S^{-\frac{1}{2}}(\bar{X}_i - \bar{X}_.)|| > c \right\}. 
\]  

(D–62)

Further, similar calculations as in Equations D–6-D–12 show that

\[
E_\theta\{V S^{-1}(\bar{X}_i - \bar{X}_.)(\bar{X}_i - \bar{X}_.)^T S^{-1} V \rho_c^2\} \\
= k^2 n(k-1)\{n(k-1) + 1\} \Sigma E_\theta\{S^{-1}(\bar{X}_i - \bar{X}_.)(\bar{X}_i - \bar{X})^T S^{-1} \rho_c^2\} \Sigma \\
+ k^2 n(k-1) \text{tr}\left[ E_\theta\{S^{-1}(\bar{X}_i - \bar{X}_.)(\bar{X}_i - \bar{X})^T S^{-1} \rho_c^2\} \Sigma \right] \Sigma. 
\]  

(D–63)

The result of the Theorem follows from Equations D–53, D–54, D–62 and D–63. □
REFERENCES


BIOGRAPHICAL SKETCH

Georgios Papageorgiou was born in Larnaca, Cyprus on January 29 of 1978. He earned a bachelor’s degree in Statistics from the Athens University of Economics and Business in Athens, Greece, in 2000. After returning to Cyprus and working for one year for a market research company, he decided to pursue graduate studies. He received a Master of Statistics degree from the department of Statistics at the University of Florida and a Ph.D. in Statistics in August of 2007.