1

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS
STANDARD REFERENCE MATERIAL 1010a
(ANSI and ISO TEST CHART No. 2)

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

**Interface of Preliminary Test Approach
and Empirical Bayes Approach
to Shrinkage Estimation**

by

ABDUNNABI M. ALI, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University
Ottawa, Canada
September, 1990

© Copyright
1990, Abdunnabi M. Ali

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60452-2

The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis,
**Interface of Preliminary Test Approach
and Empirical Bayes Approach
to Shrinkage Estimation**

submitted by
Abdunnabi M. Ali, Ph.D.
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Chairperson, Department of Mathematics and Statistics

Thesis Supervisor

External Examiner

Carleton University
September, 1990

This thesis is dedicated to my Mother and Father

Abstract

In this thesis, we deal with the estimation problem for specific models under uncertain prior information for both finite and large sample sizes. In particular, we consider the following problems:

- (i) Estimation of the mean vector of a multivariate normal distribution under symmetry and as a continuation we consider the estimation of means and treatment effects in one-way classification model.
- (ii) Preliminary test and empirical Bayes approach to shrinkage estimation of regression parameters under general restriction and we apply these results to estimate the population mean.
- (iii) Asymptotic theory for simultaneous estimation of Bernoulli means and the simple correlation coefficients.

In each problem, four basic estimators, namely, the unrestricted estimator (UE), restricted estimator (RE), preliminary test estimator (PTE) and shrinkage estimator (SE) are considered. Two more estimators, namely, the modified PTE (MPTE) and positive rule shrinkage estimator (PRSE) are added to the list with comments on the property. Moreover, we prove that the preliminary test approach to shrinkage estimation results in exactly the same expressions as the empirical Bayes approach under certain prior information.

We present the analysis of the estimators based on two decision principles, namely, frequentist decision principles and the Bayes risk principle. For a special

covariance matrix, we provide a max-min rule for the choice of level of the significance for the preliminary test estimator. The results show that the shrinkage estimator dominates the unrestricted estimator uniformly, except for the estimation of the population mean. It is shown that the empirical Bayes estimator has smaller Bayes risk than the unrestricted estimator when the dimension of the restrictive subspace exceeds three. Furthermore, neither the preliminary test estimator nor the shrinkage estimator dominate one another, except for large dimensions and a range of significance levels. The specific conditions which SE dominates PTE uniformly are given and the range for which the PTE dominates the SE is provided theoretically and numerically. Both the SE and the PTE perform well relative to the unrestricted and restricted estimators.

Application and comparison of the proposed estimators in the Binomial case are made using Efron and Morris baseball data. The comparison shows that the proposed estimators are quite comparable to the other estimators.

Acknowledgements

I am very grateful to a number of people who contributed, in one way or other, to my thesis. Firstly, my deep gratitude and appreciation to my thesis supervisor Dr. A. K. Md. Ehsanes Saleh, who first introduced me to the subject of the preliminary test and shrinkage estimation, and who, since the beginning, gave generously his time and effort towards the creation and correction of this work. His enthusiasm and intuition have been a source of inspiration to me. Indeed, this work would not have been completed without his great help and support throughout the course of this thesis. It is a great pleasure to thank my supervisory committee members, Dr. J. N. K. Rao and Dr. S. E. Mills for their constructive comments and encouragement. Thanks also to my teachers, whose contribution to my background in Statistics I will remember and appreciate for the rest of my life. My special thanks go to Dr. M. J. Moore for his support and encouragement.

Secondly, I would like to thank the Libyan people represented by Al-Fateh University in Tripoli, Libya for the financial support in offering me the scholarship for the period of my M.Sc. and Ph.D. studies.

Finally, I would like to express my deep appreciation to my family, especially to my wife for their patience and support during my study.

Table of Contents

Acceptance Sheet	ii
Abstract	iv
Acknowledgements	vi
Table of Contents	vii
List of Tables	x
List of Figures	xiv
List of Symbols	xvi
 Chapter 1: Introduction	 1
 Chapter 2: Estimation of the Mean Vector of a Multivariate Normal Distribution under Symmetry	 10
2.1 Introduction	10
2.2 The Proposed Estimators	12
2.3 Biases and Risks	14
2.4 Risk Analysis	21
2.5 Efficiency Analysis	24
2.6 Estimation of Means and Treatment Effects in a One-Way ANOVA Model	 41
2.6.1 The PT and Shrinkage Estimators	42
2.6.2 Biases and Risks	43
2.6.3 Analysis of the Estimators	46
2.6.4 Estimation of the Treatment Effects	58

Chapter 3: Preliminary Test and Empirical Bayes Approach to	
Shrinkage Estimation of Regression Parameters	60
3.1 Introduction	60
3.2 Preliminary Test Approach	62
3.3 The Bayesian Approach	63
3.3.1 Bayes Estimator	64
3.3.2 Complete and Sufficient Statistics	66
3.3.3 The Empirical Bayes Subspace Estimator	68
3.3.4 The Empirical Bayes Risk	68
3.3.5 Minimax Estimation	73
3.4 Frequentist Biases and Risks	75
3.5 Frequentist Risk Analysis	77
Chapter 4: Preliminary Test and Empirical Bayes Estimators	
of Finite Population Mean	81
4.1 Introduction	81
4.2 The Proposed Estimators	83
4.2.1 Preliminary Test Approach	83
4.2.2 The Bayesian Approach	85
4.3 The Prediction	87
4.3.1 The Proposed Estimators	87
4.3.2 Performance of the Prediction	88
4.4 Biases and MSE of Various Estimators	89
4.5 Performance of the Estimators	90
4.6 Bayes Risk and Analysis of the Estimators	104

Chapter 5: Asymptotic Theory for Simultaneous	
Estimation of Binomial Means	107
5.1 Introduction	107
5.2 The Proposed Estimators	108
5.2.1 Preliminary Test Approach	108
5.2.2 The Bayes and Empirical Bayes Approach	110
5.3 Asymptotic Risks and Pitman Alternatives	111
5.4 Asymptotic Risks and Bias	113
5.5 Analysis of the Estimators	121
5.5.1 ADR Analysis	121
5.5.2 Efficiency Analysis	127
5.6 Baseball Data Analysis	128
Chapter 6: Preliminary Test and Shrinkage Estimation	
for Simple Correlation	137
6.1 Introduction	137
6.2 The Proposed Estimators	138
6.3 Asymptotic Results	140
6.3.1 Pitman Alternatives	140
6.3.2 Asymptotic Theory	142
6.3.3 Asymptotic Biases and Risks	144
Chapter 7: Further Research	146
Appendix A	148
Appendix B: Computer Program	149
Bibliography	153

List of Tables

Table	Description	Page
2.1	Values of the maximum and minimum efficiencies and recommended significance levels for various values of ρ , $p=3$ and $n=6$	33
2.2	Values of the maximum and minimum efficiencies and recommended significance levels for various values of ρ , $p=3$ and $n=9$	34
2.3	Values of the maximum and minimum efficiencies and recommended significance levels for various values of ρ , $p=4$ and $n=10$	35
2.4	Values of the maximum and minimum efficiencies and recommended significance levels for various values of ρ , $p=4$ and $n=15$	36
2.5	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=6$	37
2.6	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=6$	38

2.7	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=15$	39
2.8	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=15$	40
2.9	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=4$	54
2.10	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=5$	55
2.11	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=6$	56
2.12	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=7$	57

4.1	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_{α}^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=5$ and $q=3$	98
4.2	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_{α}^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=6$ and $q=4$	99
4.3	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_{α}^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=7$ and $q=5$	100
4.4	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_{α}^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=5$ and $q=3$	101
4.5	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_{α}^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=6$ and $q=4$	102

4.6	Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=7$ and $q=5$	103
5.1	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for p -values	133
5.2	Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for p -values	134
5.3	The true values (θ_i^T), the unrestricted estimator, Efron and Morris estimators, Morris estimators, Albert empirical Bayes estimators and the proposed estimators ($\hat{\theta}_i^{PT}$) and ($\hat{\theta}_i^S$)	135
5.4	Computed values of the losses of the UMLE, Efron and Morris estimator, Morris estimator, Albert estimator and the proposed estimators ($\hat{\theta}_i^{PT}$) and ($\hat{\theta}_i^S$)	136

List of Figures

Figure	Description	Page
2.1	Relative efficiency of PTE to UMLE for $p=5$, $n=15$ and $\rho = -0.10$	28
2.2	Relative efficiency of PTE to UMLE for $p=4$, $n=9$ and $\rho = 0.30$	28
2.3	Dominance picture for the UMLE, RMLE, PTE and SE for $p=4$, $n=9$ and $\rho = -0.10$	29
2.4	Dominance picture for the UMLE, RMLE, PTE and SE for $p=4$, $n=9$ and $\rho = -0.10$	30
2.5	Dominance picture for the UMLE, RMLE, PTE and SE for $p=5$, $n=15$ and $\rho = 0.10$	31
2.6	Dominance picture for the UMLE, RMLE, PTE and SE for $p=5$, $n=15$ and $\rho = 0.10$	32
2.7	Performance picture for the UMLE, RMLE, PTE and SE for $p=4$ and $n=15$	52
2.8	Performance picture for the UMLE, RMLE, PTE and SE for $p=5$ and $n=20$	53
4.1	Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=6$, $q=4$ and $\alpha = 0.10, 0.15$	94

4.2	Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=6$, $q=4$ and $\alpha = 0.20, 0.50$	95
4.3	Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=7$, $q=5$ and $\alpha = 0.10, 0.15$	96
4.4	Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=7$, $q=5$ and $\alpha = 0.20, 0.50$	97
5.1	Dominance picture for UMLE, RMLE, PTE and SE for $p=5$	130
5.2	Dominance picture for UMLE, RMLE, PTE and SE for $p=5$	131
5.3	Dominance picture for UMLE, RMLE, PTE and SE for $p=20$	132

List of Symbols

Notation	Description
θ, μ, β	Parameters (scalar)
θ, μ, β	Parameters (column vector)
$N_p(\theta, \Sigma)$	Multivariate normal distribution with mean θ and covariance matrix Σ
$(1, \theta)$	Bernoulli distribution with parameter θ
I_p	Identity matrix
$1'_p$	Unit vector
$\sim$	Means distributed as
$\succ$	Means dominates
$\xrightarrow{a.s.}$	Means convergence almost surely
$\xrightarrow{P}$	Means convergence in probability
$\xrightarrow{D}$	Means convergence in distribution
$\stackrel{D}{=}$	Means equal in distribution
Δ	Non-centrality parameter
$\tilde{\theta}_n$	Unrestricted estimator
$\hat{\theta}_n$	Restricted estimator
$\hat{\theta}_n^{PT}$	Preliminary test estimator
$\hat{\theta}_n^S$	Shrinkage estimator
$\hat{\theta}_n^{S+}$	Positive rule shrinkage estimator
$\hat{\theta}_n^{PT+}$	Modified preliminary test estimator
$\hat{\theta}_n^B$	Bayes estimator
$\hat{\theta}_n^{EB}$	Empirical Bayes estimator
$R(\hat{\theta}_n, W)$	Frequentist risk
$R(\xi, \hat{\theta}_n)$	Bayes risk with respect to a prior distribution ξ

$R.E.(\hat{\theta}_n : \tilde{\theta}_n)$	The relative efficiency of $\hat{\theta}_n$ to $\tilde{\theta}_n$
$ch_{max}(A)$	The largest characteristic root of A
$ch_{min}(A)$	The smallest characteristic root of A
$Diag(\lambda_i)$	Diagonal of λ_i
UE	Unrestricted Estimator
RE	Restricted Estimator
PTE	Preliminary Test Estimator
SE	Shrinkage Estimator
UMLE	Unrestricted Maximum Likelihood Estimator
ULSE	Unrestricted Least Squares Estimator
RMLE	Restricted Maximum Likelihood Estimator
RLSE	Restricted Least Squares Estimator
MPTE	Modified Preliminary Test Estimator
PRSE	Positive Rule Shrinkage Estimator
BE	Bayes Estimator
EBE	Empirical Bayes Estimator
AR	Asymptotic Risk
ADF	Asymptotic Distribution Function
ADR	Asymptotic Distributional Risk
Var	Variance
MSE	Mean Square Error
R.E.	Relative Efficiency
RSL	Relative Savings Loss

Chapter 1

Introduction

The problem begins with an unknown state of nature or a parameter of interest $\theta = (\theta_1, \dots, \theta_p)' \in \Omega$. To this end a statistical experiment $\mathcal{E}$ is planned and then performed. This generates the sample $\mathbf{x}_n = (x_1, x_2, \dots, x_n)'$. Further information about θ is supposed to be obtained by careful analysis of the data $(\mathcal{E}, \mathbf{x}_n)$ in the light of all our prior information about θ and in the context of the particular inference problem related to θ . For example some relationship among the coordinates of θ , namely, $\mathbf{H}\theta = \mathbf{h}$ may be discovered and this information is incorporated in the particular inference on θ , where $\mathbf{H}$ is a given $q \times p$ matrix of rank q and $\mathbf{h}$ is a given $q \times 1$ vector. In general, matrices are represented by boldface upper case letters throughout this thesis, e.g., $\mathbf{A}$, Σ and boldface lower letters represent column vectors, e.g., $\mathbf{x}$, δ .

The unknown quantity θ which affects the decision process is commonly called the *state of nature*. In making decisions it is clearly important to consider what the possible states of nature are. The symbol Ω will be used to denote the set of all possible states of nature. Typically, when experiments are performed to obtain information about θ , the experiments are designed so that the observations are distributed according to some probability distribution which has θ as unknown parameter.

In such situations θ will be called the *parameter* and Ω the *parameter space*. Decisions are more commonly called *actions* in the literature. Particular actions will be denoted by $\mathbf{a} = (a_1, a_2, \dots, a_p)'$, while the set of all possible actions under consideration will be denoted $\mathcal{A}$.

When a statistical investigation is performed to obtain information about θ , the outcome (a random variable) will be denoted by X . Often X will be a vector, its components being independent observations from a common distribution. A particular realization of X will be denoted by $\mathbf{x}_n$. The set of possible outcomes is the *sample space*, and will be denoted $\mathcal{X}$. Usually $\mathcal{X}$ will be a subset of R^n , n -dimensional Euclidean space. The probability distribution of X will, of course, depend upon the unknown state of nature θ . Let $P_\theta(A)$ denote the probability of the event A ($A \subset \mathcal{X}$), when θ is the true state of nature. For going through the rituals of traditional sample space analysis of data we must begin with the invocation of the trinity $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. If the model is such that we can represent the family $\mathcal{P}$ as

$$\mathcal{P} = \{P_\theta, \theta \in \Omega\},$$

then we go about the business of analyzing the data according to standard methodology without having to worry about the relationship $\mathbf{H}\theta = \mathbf{h}$.

Unfortunately, instances of statistical models with $\mathcal{P}$ indexed by θ alone are very rare. Typically, we have to work with families that are indexed as

$$\mathcal{P} = \{P_\theta, \theta \in \omega\},$$

while it is suspected that $\theta \in \omega$, $\omega \subset \Omega$ [where ω specifies the relationships among the coordinates of θ .] If the inference problem related only to θ and if the information gained on $\theta \in \omega$ is of relevance to the problem and at the same time "uncertain", then we classify these kinds of problems as **statistical problems with uncertain prior information**. If the full model $\{P_\theta, \theta \in \Omega\}$ holds, then the statistical (*unrestricted*) solution holds truly to the problem. However, in the case of sub-model $\{P_\theta, \theta \in \omega \subset \Omega\}$, the (*restricted*) solution holds for the sub-model. However, if there is departure from the sub-model to full model, the restricted solution often may become *inconsistent* and *inefficient* for the full

model. Thus we need to strike at a balance between the two extreme solutions, namely, the *unrestricted* and the *restricted* ones.

The question arises: How to incorporate the *uncertain prior information* of the parameter into the inference of this parameter? In other words, how to combine the *unrestricted* and the *restricted* solutions optimally to minimize certain kinds of loss? During the last four and a half decades a large amount of effort and ingenuity has gone into the search for reasonable answers to this particular question.

Bancroft (1944) came up with the idea of preliminary testing followed by estimation of parameters. The problem arose basically in an ANOVA problem where the estimation of variances arose. The other problem considered at that time was the problem of testing equality of two means when the variances were suspected to be equal. Initial attempts of Bancroft were to find out the effect of preliminary test on the ultimate inference problem. He concentrated mainly on the estimation problem and calculated the bias and the mean square errors of the estimators that were developed by the preliminary test approach. Mosteller (1948) looked at the problem from a Bayesian point of view and almost gave the idea of empirical Bayes estimation before Robbins (1955). Kitagawa (1963) followed the idea of the preliminary test on successive occasions and obtained the distribution and moments of the preliminary test estimator in the two sample problem. Bozivich, Bancroft and Hartley (1956) studied the properties of the size of the test as well as power of a test after preliminary testing. Bennett (1952, 1956), Huntsberger (1955), Asano and Sato (1962) followed the application of the preliminary test approach in various directions. However, the theoretical developments were very skinny. Cohen (1965) proved that the preliminary test estimator was inadmissible. The implication being that the traditional sampling theory preliminary test estimator could be beaten. Bancroft argued that inadmissibility and unbiasedness have parallel notions which could be ignored. Both Bancroft and Cohen did not suggest a superior alternative.

Stein (1955a) discovered the surprising phenomenon that the sample mean vector was inadmissible for the estimation of the mean vector of a p -dimensional normal distribution for $p \geq 3$. James and Stein (1961) proposed an estimator

which dominate the sample mean uniformly for all values of θ . That put the preliminary test estimator in the background. Though research continued but the effect of such activity was almost nil. However, Han and Bancroft (1968) produced a simple paper on pooling means using the preliminary test approach and Bayes approach and gave rules to determine the size of the preliminary test for the equality of two means. Using this paper together with the paper of Mosteller (1948), Saleh (1973) began the study of the properties of the preliminary test estimator from scratch again. Subsequently, Saleh and Sen (1978) produced the first paper with a nonparametric approach to the preliminary test estimator. They continued to develop the properties of the preliminary test estimator applying the robust procedures from 1978 to date. The papers of Saleh and Sen (1985a, 1985b, 1985c, 1986a, 1986b) and Sen and Saleh (1985, 1987) mark the beginning of the shrinkage estimator using the preliminary test approach. It is professor J. N. K. Rao who pointed out the similarity between the preliminary test approach and empirical Bayes approach to shrinkage estimation of regression parameters in a parallelism problem. This led Ghosh, Saleh and Sen (1989) to verify the conjecture of Rao to establish that the estimator of a subset of regression parameters in a subhypotheses problem is the same as the preliminary test approach to shrinkage estimation of location type parameters .

In a Ph.D of Ahmed (1987), the preliminary test approach is documented in finite sample situation using normality assumption in the distribution. The book by Judge and Bock (1978) documents the properties of the preliminary test estimator and Stein-rule estimators separately. The thesis of Ahmed (1987) enlarges the scope of simultaneous properties of the preliminary test and shrinkage estimators.

In this thesis we consider the interface of the *preliminary test approach* and *empirical Bayes approach* to several standard problems not dealt with any where to understand the impact of such methodology.

For some nice accounts of the parametric theory of the preliminary test estimator (PTE) and shrinkage estimator (SE) in *the finite sample case*, we refer to Judge and Bock (1978), Anderson (1984), Arnold (1981), Berger (1980) and Ahmed (1987) among others. Shrinkage estimators have been proposed as alter-

natives to the unrestricted estimator (UE) in many cases (e.g., Judge and Bock (1978), Stein (1981) among others) and many authors have deduced them from an empirical Bayes approach see, e.g., Lindley and Smith (1972), Efron and Morris (1972a, 1972b), Berger and Chen (1987), Ghosh, Saleh and Sen (1989), Zellner (1971), Berger (1985), Lindley (1971a, 1971b) among others. In the literature normal Bayes and empirical Bayes method, for the finite population mean, have appeared due to Ericson (1969), Ghosh and Meeden (1986) and Lahiri and Peddada (1988). Some attempts have been made by Ghosh and Lahiri (1976) to relax the normality assumption. For general linear models, *the asymptotic theory* of the PTE has been treated in a nonparametric setup in Saleh and Sen (1978, 1983, 1984a, 1984b, 1985a, 1985b, 1985c, 1985d) and Sen and Saleh (1979, 1985, 1987). Also, *the asymptotic theory* of shrinkage estimation has been developed in a nonparametric setup by the same authors. Finite sample studies, on the estimation of binomial means, have been carried out by Albert (1984), Albert and Gupta (1981) among others.

Before going to different chapters we begin with the ideas behind the Preliminary test (PT) approach to shrinkage estimation. The preliminary test (PT) approach may be formulated in a decision theory setup as follows:

- (1) Experiment is reflected by the trinity $(\mathcal{X}, \mathcal{A}, P_{\theta})$, where $\theta \in \Omega$.
- (2) The constraint $\theta \in \omega$ is an uncertain prior information.
- (3) A non-empty set of actions denoted by $a \in \mathcal{A}$ is available to the statistician.
- (4) A key element of decision theory is the *loss function*. If a particular action a_1 is taken and θ_1 turns out to be the true state of nature, then a loss $L(\theta_1, a_1)$ will be incurred. Thus we will assume a *loss function* $L(\theta, a)$ is defined for all $(\theta, a) \in \Omega \times \mathcal{A}$.

Thus we have a statistical decision problem $(\Omega, \mathcal{A}, L)$ together with $(\mathcal{X}, \mathcal{A}, P_{\theta} | \theta \in \omega \text{ suspicious})$.

After the data $\mathbf{x}_n$ is obtained, the statistician chooses a decision rule $\delta(\mathbf{x}_n) \in \mathcal{A}$, $\delta(\mathbf{x}_n)$ is a function from $\mathcal{X}$ into $\mathcal{A}$. We are interested in the choice of $\delta(\mathbf{x}_n)$ which minimizes the *risk*, where the *risk* is the expected value of the loss function, i.e.,

$$R(\theta, \delta) = E[L(\theta, \delta(X))] = \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x}_n)) dF^X(\mathbf{x}_n|\theta).$$

We now formulate the *estimation* of θ when it is suspected that $\mathbf{H}\theta = \mathbf{h}$ may hold where θ is a p -dimensional vector and a location type parameter. Let $\mathbf{x}_n$ be a n -dimensional vector of observations. Furthermore, let $\mathcal{L}_n$ be an optimal test statistic for testing the general hypotheses

$$H_0 : \mathbf{H}\theta = \mathbf{h} \quad \text{against} \quad H_A : \mathbf{H}\theta \neq \mathbf{h}.$$

We may define the following estimators of θ .

- (i) *Unrestricted estimator* (UE). Let $\tilde{\theta}_n$ denote the estimator of θ using the full model.
- (ii) *Restricted estimator* (RE). Let $\hat{\theta}_n$ denote the restricted estimator of θ under the restriction $\mathbf{H}\theta = \mathbf{h}$. Then the restricted estimator (RE) performs better than the unrestricted estimator (UE) when $\mathbf{H}\theta = \mathbf{h}$ holds. However, as $\mathbf{H}\theta$ differs from $\mathbf{h}$, the restricted estimator (RE) may be considerably biased, inefficient and inconsistent while the performance of the unrestricted estimator (UE) remains steady over such departure. For this reason, it is desirable to develop estimator which is a compromise between the unrestricted estimator (UE) and the restricted estimator (RE) under uncertain prior information $\mathbf{H}\theta = \mathbf{h}$. We do this by using the preliminary test (PT) approach.
- (iii) *Preliminary test estimator* (PTE). Let $\mathcal{L}_{n,\alpha}$ be the α -level of the critical value of $\mathcal{L}_n$, then the PTE of θ is defined by

$$\hat{\theta}_n^{PT} = \hat{\theta}_n + [1 - I(\mathcal{L}_n \leq \mathcal{L}_{n,\alpha})] (\tilde{\theta}_n - \hat{\theta}_n),$$

where $I(A)$ is the indicator function of the set A . This type of estimator was proposed by Bancroft (1944). In this case the estimator $\hat{\theta}_n$ or $\tilde{\theta}_n$

is chosen according as H_0 is accepted or rejected. The performance of preliminary test estimator (PTE) depends on the size of the test α ($0 < \alpha < 1$). Though its performance is reasonable compared with the unrestricted estimator (UE) and the restricted estimator (RE), the discontinuous nature of the indicator function losses its total grip on the performance. Hence, we go to another choice of a function of $\mathcal{L}_n$.

- (iv) *Shrinkage estimator* (SE). A simple form of a shrinkage estimator of θ may be defined as

$$\hat{\theta}_n^S = \hat{\theta}_n + [1 - c\mathcal{L}_n^{-1}] (\tilde{\theta}_n - \hat{\theta}_n),$$

where c is a positive constant generally called the shrinkage constant and depends on the dimensional of θ and the matrix H . In this case, the SE is based on the classical Stein-rule estimator. Note that the shrinkage estimator (SE) is a smooth version of the preliminary test estimator (PTE). The two compromised estimators are similar looking where the test statistic $\mathcal{L}_n$ adjusts the estimators to departure from H_0 . When $\mathcal{L}_n$ is large, the preliminary test estimator, the shrinkage estimator and the unrestricted estimators are essentially equivalent, while for small values of $\mathcal{L}_n$ there is a difference.

Thus one has to study critically the performance characteristics of the four estimators for small values of $\mathcal{L}_n$ using the following *decision principles*:

- (i) *Frequentist Decision Principles*. Let $L(\delta, Q) = (\delta - \theta)'Q(\delta - \theta)$ a quadratic loss function with a positive definite matrix Q . Then the risk function of a decision rule δ is defined by

$$R(\delta, Q) = E_\theta[L(\delta, Q)] = E_\theta[(\delta - \theta)'Q(\delta - \theta)].$$

Sometimes we consider the Mahalanobis distance as loss and the corresponding risk. Let R_1 and R_2 be the risks of the estimators $\tilde{\theta}_n$ and $\hat{\theta}_n$ respectively, then the risk efficiency of $\hat{\theta}_n$ relative to $\tilde{\theta}_n$ is defined by

$$R.E.(\hat{\theta}_n : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, Q)}{R(\hat{\theta}_n, Q)} = \frac{R_1}{R_2}.$$

- (ii) *The Bayes Risk Principle.* The *Bayes risk* of a decision rule δ , with respect to a prior distribution π on Ω , is defined as

$$R(\pi, \delta) = E^\pi[R(\theta, \delta)].$$

A decision rule δ_1 is preferred to a rule δ_2 if

$$R(\pi, \delta_1) < R(\pi, \delta_2).$$

A decision rule which minimizes $R(\pi, \delta)$ with respect to δ is optimal; it is called a Bayes rule, and the risk called the Bayes risk for the prior π . For more details see Berger (1985).

In Chapter 2, we have presented the analysis of the Efron–Morris–Lindley estimator together with its generalization in the multivariate normal case, by considering the data $\mathbf{x}_i$; $i = 1, 2, \dots, n$ from a multivariate normal distribution $N_p(\mu, \Sigma)$, where Σ is unknown. We are interested in estimating the vector μ , when it is suspected that $\mu_1 = \mu_2 = \dots = \mu_p$. The performance of the four basic estimators are studied using the intraclass correlation matrix, which will be defined in this chapter. Furthermore, the ANOVA model is considered and the properties of the estimators are discussed. Chapter 3 contains the estimation of the regression parameters under suspected constraint $\mathbf{H}\theta = \mathbf{h}$. Here we bring in the interface of the preliminary test approach to shrinkage estimator of the regression parameters and empirical Bayes estimator to the same parameters. The performance of the estimators are studied using the two decision principles, namely, the frequentist decision principles and the Bayes risk principle. In Chapter 4, we make some applications of the results of regression estimators to estimate the mean in a finite population with auxiliary variables. The results show that the Stein phenomenon appears to be lost in such a case and parallels that of the preliminary test estimator of mean. The properties of the estimators will be considered using the frequentist risk analysis and the Bayes risk analysis in terms of the relative saving loss (RSL) introduced by Efron and Morris (1973). In Chapter 5, we analyze once again the baseball data of Efron and Morris (1975). We model the problem as p sets of

Bernoulli trials of size n_i ($i = 1, 2, \dots, p$) each and try to estimate the Bernoulli parameters. Thus, we proceed to the asymptotic setup to study the properties of the four basic estimators. In this regard, we also bring the Albert (1984) method of empirical Bayes estimator as well as Morris (1983) empirical Bayes estimator for the sake of comparison using Efron and Morris (1975) baseball data. In Chapter 6, we apply the results of Chapter 5 to the correlation estimation from several bivariate normal populations. Chapter 7 contains further research problems.

Chapter 2

Estimation of the Mean Vector of a Multivariate Normal Distribution under Symmetry

2.1 Introduction

In this chapter we consider the estimation of the mean vector of a p -variate normal distribution with unknown covariance matrix when it is a priori suspected that the component means are equal but unknown. Four basic estimators are considered, namely, the unrestricted maximum likelihood estimator (UMLE), restricted MLE (RMLE), preliminary test estimator (PTE) and shrinkage estimator (SE). The comparative properties of the estimators are studied with respect to a specific quadratic loss function using special covariance structure. Two more estimators, namely, the modified PTE (MPTE) and positive rule shrinkage estimator (PRSE) are added to the list with comments on the property.

For a special covariance matrix, namely, the intraclass correlation structure, we provide a max-min rule for the choice of level of the significance α for the preliminary test estimator (PTE). Tables for optimum selection of α are presented

for chosen sample sizes, p -values and ρ -values for the use of the preliminary test estimators. Further, tables are presented giving the maximum relative risk efficiencies of the restricted, preliminary test and shrinkage estimator relative to the unrestricted maximum likelihood estimator along with risk efficiency at the intersection of the preliminary test and the shrinkage estimator together with the intersecting point. For all values of ρ , the shrinkage estimator (SE) dominates UMLE uniformly, however, neither PTE nor SE dominates each other for all α , the level of significance of the PT though both fare well compared to UMLE and RMLE.

Suppose $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be a random sample from a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ and $\boldsymbol{\Sigma}$ is unknown, as special case, we assume that the covariance matrix $\boldsymbol{\Sigma}$ is of the intraclass correlation structure, i.e.,

$$\boldsymbol{\Sigma} = \sigma^2 [(1 - \rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p'], \quad \frac{-1}{(p-1)} < \rho < 1. \quad (2.1)$$

Let the null-hypothesis be

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_p = \mu \text{ (unknown)}. \quad (2.2)$$

We are primarily concerned with the estimation of the mean vector $\boldsymbol{\mu}$, when it is suspected but not sure that $\mu_1 = \mu_2 = \dots = \mu_p$, i.e., with uncertain prior information about $\boldsymbol{\mu}$. Furthermore, the objective of this study is to focus on the PTE and SE and to compare them with the UMLE and RMLE.

Da Silva and Han (1984) presented the analysis of the preliminary test estimator of one component of $\boldsymbol{\mu}$ only. In this work, we consider more general setting to estimate the vector $\boldsymbol{\mu}$ and we compare the first four estimators simultaneously. Also, the shrinkage estimator in (2.10) is a generalization of the estimator of Lindley (1962) in the discussion on a paper by C. M. Stein.

The proposed estimators are all presented in Section 2.2. The expressions for bias and the risk are outlined in Section 2.3. The risk analysis for the various estimators of $\boldsymbol{\mu}$ are presented in Section 2.4, and the the efficiency analysis results are considered in Section 2.5. In Section 2.6, we study the problem of treatment effects in a one-way ANOVA model.

2.2 The Proposed Estimators

The estimators are defined for general Σ first then specialized to the intraclass correlation structure. Let us define:

$$\tilde{\mu}_n = (\tilde{\mu}_1, \dots, \tilde{\mu}_p)', \quad \tilde{\mu}_i = \bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}, \quad i = 1, \dots, p, \quad (2.3)$$

$$\hat{\mu}_n = (\hat{\mu}_1, \dots, \hat{\mu}_p)' = \hat{\mu}_n \mathbf{1}_p, \quad \mathbf{1}_p' = (1, 1, \dots, 1), \quad (2.4)$$

where

$$\hat{\mu}_n = \bar{x} = \frac{1}{p} \sum_{i=1}^p \bar{x}_i = \frac{1}{p} \mathbf{1}_p' \tilde{\mu}_n.$$

Then, the unrestricted maximum likelihood estimator (UMLE) of μ is given by $\tilde{\mu}_n$ in (2.3), while under H_0 in (2.2), the restricted MLE (RMLE) is given by $\hat{\mu}_n$ in (2.4). We shall show that when H_0 holds, $\hat{\mu}_n$ has a smaller risk (under quadratic loss) than $\tilde{\mu}_n$, but when H_0 does not hold, $\tilde{\mu}_n$ may perform better than $\hat{\mu}_n \mathbf{1}_p$. As a result, when the prior information on H_0 is rather uncertain it may be desirable to have either a preliminary test estimator (PTE) denoted by $\hat{\mu}_n^{PT} = (\hat{\mu}_1^{PT}, \dots, \hat{\mu}_p^{PT})'$ or a shrinkage estimator (SE) denoted by $\hat{\mu}_n^S = (\hat{\mu}_1^S, \dots, \hat{\mu}_p^S)'$ of μ . In the case of a preliminary test the estimator $\hat{\mu}_n$ or $\tilde{\mu}_n$ is chosen according as H_0 is accepted or rejected. The SE is based on the classical Stein-rule. Unlike the case of $\hat{\mu}_n$, both the PTE and SE behave robustly against departures from the null hypothesis in (2.2) and have bounded risk. From these four basic estimators two more improved estimators may be defined; namely, Stein's positive-rule shrinkage estimator (PRSE), denoted by $\hat{\mu}_n^{S+} = (\hat{\mu}_1^{S+}, \dots, \hat{\mu}_p^{S+})'$ and the modified PTE (MPTE) denoted by $\hat{\mu}_n^{PT+} = (\hat{\mu}_1^{PT+}, \dots, \hat{\mu}_p^{PT+})'$.

For the preliminary test (PT) on H_0 in (2.2), we consider the Hotelling's T_n^2 test statistic with $(n-1)$ degrees of freedom (DF) for a $(p-1)$ -dimensional distribution.

$$T_n^2 = n \tilde{\mu}_n' \mathbf{C}' \mathbf{S}^{-1} \mathbf{C} \tilde{\mu}_n, \quad (2.5)$$

where

$$(n-1)\mathbf{S} = \mathbf{C} \left[\sum_{i=1}^n (\mathbf{x}_i - \tilde{\mu}_n)(\mathbf{x}_i - \tilde{\mu}_n)' \right] \mathbf{C}', \quad (2.6)$$

and

$$C = \begin{bmatrix} 1 - \frac{1}{p} & -\frac{1}{p} & -\frac{1}{p} & \dots & -\frac{1}{p} \\ -\frac{1}{p} & 1 - \frac{1}{p} & -\frac{1}{p} & \dots & -\frac{1}{p} \\ -\frac{1}{p} & -\frac{1}{p} & 1 - \frac{1}{p} & \dots & -\frac{1}{p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{p} & -\frac{1}{p} & -\frac{1}{p} & \dots & 1 - \frac{1}{p} \end{bmatrix}, \quad (2.7)$$

is a $p \times p$ idempotent and symmetric matrix of rank $(p - 1)$, satisfying $C1_p = 0$. Also, under H_0 in (2.2) we have $C\mu = 0$. Note that C can be written as

$$I_p - \frac{1}{p}1_p1_p'.$$

Under H_0 in (2.2), the critical value of the Hotelling's T_n^2 is equal to

$$T_{0,\alpha}^2 = \frac{(n-1)(p-1)}{(n-p+1)} F_{p-1, n-p+1}(\alpha), \quad (2.8)$$

where $F_{p-1, n-p+1}(\alpha)$ is the central F-distribution with $(p - 1)$ and $(n - p + 1)$ DF for a given level of significance α , $(0 < \alpha < 1)$. Then, the PTE of μ is defined by

$$\hat{\mu}_n^{PT} = \hat{\mu}_n1_p + [1 - I(T_n^2 \leq T_{0,\alpha}^2)] (\tilde{\mu}_n - \hat{\mu}_n1_p), \quad (2.9)$$

where $I(A)$ is the indicator function for the set A .

To introduce the SE, we propose

$$\hat{\mu}_n^S = \hat{\mu}_n1_p + [1 - k(n-1)T_n^{-2}] (\tilde{\mu}_n - \hat{\mu}_n1_p), \quad (2.10)$$

where k is the shrinkage factor is limited to the interval

$$0 < k < 2 \frac{(p-3)}{(n-p+3)}.$$

Also, we need to assume that $p > 3$. Note that whereas in (2.9) we have a dichotomous indicator function, in (2.10), $k(n-1)T_n^{-2}$ may be regarded as a smoother version. Therefore, the SE is a smoother version of the PTE, and both are based on the "pre-test" principle. Also, (2.9) is a convex combination of $\tilde{\mu}_n$ and $\hat{\mu}_n$, while (2.10) need not be especially because $k(n-1)T_n^{-2}$ may be greater than 1. In the spirit of Sclove, Morris and Radhakrishnan (1972), we may introduce the following the positive-rule SE (PRSE) :

$$\hat{\mu}_n^{S+} = \hat{\mu}_n1_p + [1 - k(n-1)T_n^{-2}]^+ (\tilde{\mu}_n - \hat{\mu}_n1_p), \quad (2.11)$$

where $y^+ = \max(0, y)$. Note that the PTE in (2.9) and the PRSE in (2.11) may not agree even if we let $(n-1)^{-1}T_{0,\alpha}^2 = k$. The justification of PTE in (2.9) lies in the preliminary test of H_0 , while in (2.10) overshrinking is a problem. This lead to the proposal of the PRSE in (2.11) which control the overshrinking problem in (2.10).

Finally, we propose MPTE of μ as follows:

$$\hat{\mu}_n^{PT+} = \hat{\mu}_n \mathbf{1}_p + [1 - k(n-1)T_n^{-2}] I(T_n^2 \geq T_{0,\alpha}^2)(\tilde{\mu}_n - \hat{\mu}_n \mathbf{1}_p), \quad (2.12)$$

where the restriction $T_{0,\alpha}^2 < (n-1)k$ is required. This estimator uniformly dominates $\hat{\mu}_n^{PT}$ but now we have the restriction $p \geq 4$. For $p < 4$ we are left with the PTE to improve over $\hat{\mu}_n$ or $\tilde{\mu}_n$. It is well-known that the positive-rule shrinkage estimator (PRSE) dominates the usual shrinkage estimator SE, while the MPTE dominates the usual PTE.

2.3 Biases and Risks

Let δ_n be an estimator of θ , $\mathbf{W}$ be a positive semi-definite (p.s.d) matrix, and consider the quadratic loss function

$$L(\delta_n, \theta) = n(\delta_n - \theta)' \mathbf{W}(\delta_n - \theta) = n \text{Tr} [\mathbf{W}(\delta_n - \theta)(\delta_n - \theta)'], \quad (2.13)$$

where $\text{Tr}(\mathbf{A}) = \text{trace of } \mathbf{A}$. Then the risk of δ_n is given by

$$R(\delta_n, \theta) = EL(\delta_n, \theta) = \text{Tr}(\mathbf{W}MSE), \quad (2.14)$$

where the mean square error (MSE) of δ_n is defined by

$$MSE = nE(\delta_n - \theta)(\delta_n - \theta)'. \quad (2.15)$$

In this section we present expressions for the bias, the mean square error (MSE) and the risk under loss function in (2.13) of the first four estimators in (2.3), (2.4), (2.9) and (2.10). Theorem 2.1 and 2.2 provide expressions for the bias and the MSE expressions respectively.

Theorem 2.1 *Biases of the UMLE, RMLE, PTE and SE are given by (2.16)–(2.19) respectively:*

$$B_1 = E(\tilde{\mu}_n - \mu) = 0, \quad (2.16)$$

$$B_2 = E(\hat{\mu}_n \mathbf{1}_p - \mu) = -\delta, \quad (2.17)$$

$$B_3 = E(\hat{\mu}_n^{PT} - \mu) = -\delta H_{p+1,m}(F^*; \Delta), \quad (2.18)$$

$$B_4 = E(\hat{\mu}_n^S - \mu) = -km\delta E(\chi_{p+1}^{-2}(\Delta)), \quad (2.19)$$

where

$$\Delta = n\delta'(\mathbf{C}\Sigma\mathbf{C}')^{-1}\delta, \quad \delta = \mathbf{C}\mu, \quad (2.20)$$

$\mathbf{C}$ is defined by (2.7) and $m = n - p + 1$, we use the symbol $H_{q_1, q_2}(\cdot; \Delta)$ to denote the cumulative distribution of a non-central F -distribution with (q_1, q_2) DF and non-centrality parameter Δ . Also, F^* is defined by

$$F^* = \frac{p-1}{p+1} F_{p-1,m}(\alpha),$$

where $F_{p-1,m}(\alpha)$ is the upper α probability point of a central F -distribution with $(p-1, m)$ DF.

Proof. The proof of the above theorem follows, by observing that $\tilde{\mu}_n - \hat{\mu}_n \mathbf{1}_p = \mathbf{C}\tilde{\mu}_n$ and making the following transformation

$$\mathbf{y} = \sqrt{n}\mathbf{C}\tilde{\mu}_n, \quad \mathbf{z} = \mathbf{D}\mathbf{y}, \quad (2.21)$$

where $\mathbf{D}$ is any symmetric and non-singular matrix such that $\mathbf{D}(\mathbf{C}\Sigma\mathbf{C}')\mathbf{D}' = \mathbf{I}_{p-1}$, then from Anderson (1984)

$$\frac{T_n^2}{n-1} = \frac{\chi_{p-1}^2(\Delta)}{\chi_m^2} = \frac{\mathbf{z}'\mathbf{z}}{\chi_m^2}, \quad (2.22)$$

where $\mathbf{z}'\mathbf{z}$ is distributed as a non-central chi-square distribution with $(p-1)$ DF and non-centrality parameter Δ and independent of the χ_m^2 variable and applying Theorem 1 from Appendix A.

Theorem 2.2 *The MSE of the UMLE, RMLE, PTE and SE under (2.15) are given by (2.23)–(2.26) respectively:*

$$\Gamma_1 = E\{n(\tilde{\mu}_n - \mu)(\tilde{\mu}_n - \mu)'\} = \Sigma, \quad (2.23)$$

$$\Gamma_2 = E\{n(\hat{\mu}_n 1_p - \mu)(\hat{\mu}_n 1_p - \mu)'\} = B + n\delta\delta', \quad (2.24)$$

$$\begin{aligned} \Gamma_3 &= E\{n(\hat{\mu}_n^{PT} - \mu)(\hat{\mu}_n^{PT} - \mu)'\} \\ &= \Sigma + (I_p - 2A)C\Sigma C' H_{p+1,m}(F^{**}; \Delta) \\ &\quad + \{2AH_{p+1,m}(F^{**}; \Delta) + (I_p - 2A)H_{p+3,m}(F^{**}; \Delta)\}n\delta\delta', \end{aligned} \quad (2.25)$$

$$\begin{aligned} \Gamma_4 &= E\{n(\hat{\mu}_n^S - \mu)(\hat{\mu}_n^S - \mu)'\} \\ &= \Sigma - km \left[2AE(\chi_{p+1}^{-2}(\Delta)) - k(m+2)E(\chi_{p+1}^{-4}(\Delta)) \right] C\Sigma C' \\ &\quad + km[4A + k(m+2)]n\delta\delta'E(\chi_{p+3}^{-4}(\Delta)), \end{aligned} \quad (2.26)$$

where

$$\begin{aligned} A &= (\Sigma C')(C\Sigma C')^{-1} \\ B &= \frac{1}{p^2} 1_p 1_p' \Sigma 1_p 1_p' \\ F^{**} &= \frac{p-1}{p+3} F_{p-1,m}(\alpha), \end{aligned}$$

and Δ and F^* are defined as in Theorem 2.1.

Proof. Equations (2.23) and (2.24) can be verified easily. We shall prove the expressions (2.25) and (2.26). We define the following transformation

$$\mathbf{x} = \sqrt{n}(\tilde{\mu}_n - \mu), \quad (2.27)$$

and $\mathbf{y}$ is defined by (2.21). Then the joint density function of $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N_{2p} \left(\begin{bmatrix} 0 \\ \sqrt{n}\delta \end{bmatrix}, \begin{bmatrix} \Sigma & \Sigma C' \\ C\Sigma & C\Sigma C' \end{bmatrix} \right), \quad (2.28)$$

and

$$\sqrt{n}(\hat{\mu}_n^{PT} - \mu) = \mathbf{x} - \mathbf{y}I(T_n^2 \leq T_{0,\alpha}^2). \quad (2.29)$$

Therefore,

$$\begin{aligned}
 \Gamma_3 &= E \left\{ n(\hat{\mu}_n^{PT} - \mu)(\hat{\mu}_n^{PT} - \mu)' \right\} \\
 &= E(\mathbf{x}\mathbf{x}') - 2E \left[\mathbf{x}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] + E \left[\mathbf{y}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] \\
 &= \Sigma - 2E \left[\mathbf{x}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] + E \left[\mathbf{y}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right]. \quad (2.30)
 \end{aligned}$$

Now applying the transformation in (2.21) and equation (2.22), the third term of the expression in (2.30) can be evaluated as

$$E \left[\mathbf{y}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] = \mathbf{D}^{-1} E \left[\mathbf{z}\mathbf{z}' I \left(\frac{\mathbf{z}'\mathbf{z}}{\chi_m^2} \leq \frac{T_{0,\alpha}^2}{n-1} \right) \right] (\mathbf{D}')^{-1}. \quad (2.31)$$

Note that $\mathbf{z} \sim N(\sqrt{n}\mathbf{D}\delta, \mathbf{I}_{p-1})$. Hence, applying Theorem 2 from Appendix A, the expression in (2.31) reduces to

$$E \left[\mathbf{y}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] = \mathbf{C}\Sigma\mathbf{C}' H_{p+1,m}(F^*; \Delta) + n\delta\delta' H_{p+3,m}(F^{**}; \Delta). \quad (2.32)$$

Note that, it easy to show $\mathbf{C}\Sigma\mathbf{C}' = (\mathbf{D}'\mathbf{D})^{-1}$. The second term of expression in (2.30) can be evaluated as

$$\begin{aligned}
 E[\mathbf{x}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2)] &= E \left[E\{\mathbf{x}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) | \mathbf{y}\} \right] \\
 &= E \left[E(\mathbf{x} | \mathbf{y}) \mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right]. \quad (2.33)
 \end{aligned}$$

Using equation (2.28), one gets

$$\begin{aligned}
 E(\mathbf{x} | \mathbf{y}) &= \mu_{\mathbf{x}} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{y} - \mu_{\mathbf{y}}) \\
 &= \mathbf{A}(\mathbf{y} - \sqrt{n}\delta), \quad \mathbf{A} = (\Sigma\mathbf{C}')(\mathbf{C}\Sigma\mathbf{C}')^{-1}. \quad (2.34)
 \end{aligned}$$

Hence

$$\begin{aligned}
 E[\mathbf{x}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2)] &= E \left[\mathbf{A}(\mathbf{y} - \sqrt{n}\delta) \mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] \\
 &= \mathbf{A} E \left[\mathbf{y}\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2) \right] - \sqrt{n}\mathbf{A}\delta E[\mathbf{y}' I(T_n^2 \leq T_{0,\alpha}^2)] \\
 &= \mathbf{A} \{ \mathbf{C}\Sigma\mathbf{C}' H_{p+1,m}(F^*; \Delta) + n\delta\delta' H_{p+3,m}(F^{**}; \Delta) \\
 &\quad - n\delta\delta' H_{p+1,m}(F^*; \Delta) \}. \quad (2.35)
 \end{aligned}$$

The result in (2.35) follows by using Theorem 1 and 2 of Appendix A. Substituting (2.32) and (2.35) in (2.30) the result of equation (2.25) follows.

Next, we prove equation (2.26). Again, using the same transformation in (2.21) and (2.27), we have

$$\sqrt{n}(\hat{\mu}_n^S - \mu) = \mathbf{x} - \mathbf{y}k(n-1)T_n^{-2}. \quad (2.36)$$

Hence

$$\begin{aligned} \Gamma_4 &= E\{n(\hat{\mu}_n^S - \mu)(\hat{\mu}_n^S - \mu)'\} \\ &= E(\mathbf{x}\mathbf{x}') - 2E\{\mathbf{x}\mathbf{y}'k(n-1)T_n^{-2}\} + E\{\mathbf{y}\mathbf{y}'k^2(n-1)^2T_n^{-4}\} \\ &= \Sigma - 2kE\{\mathbf{x}\mathbf{y}'(n-1)T_n^{-2}\} + k^2E\{\mathbf{y}\mathbf{y}'(n-1)^2T_n^{-4}\}. \end{aligned} \quad (2.37)$$

Now, we evaluate the third term of (2.37). Therefore, by equation (2.21), (2.22) and the fact $E(\chi_m^4) = m(m+2)$, one gets

$$\begin{aligned} E[\mathbf{y}\mathbf{y}'(n-1)^2T_n^{-4}] &= E[\mathbf{y}\mathbf{y}'\chi_m^4(\chi_{p-1}^2)^{-2}] \\ &= m(m+2)\mathbf{D}^{-1}E\{\mathbf{z}\mathbf{z}'(\mathbf{z}'\mathbf{z})^{-2}\}(\mathbf{D}')^{-1} \\ &= m(m+2)\{\mathbf{C}\Sigma\mathbf{C}'E(\chi_{p+1}^{-4}(\Delta)) + n\delta\delta'E(\chi_{p+3}^{-4}(\Delta))\}. \end{aligned} \quad (2.38)$$

The result of (2.38) follows by applying Theorem 2 of Appendix A. Finally, we evaluate the second term of (2.37), i.e.,

$$\begin{aligned} E[\mathbf{x}\mathbf{y}'(n-1)T_n^{-2}] &= E[E\{\mathbf{x}\mathbf{y}'(n-1)T_n^{-2}|\mathbf{y}\}] \\ &= E[E(\mathbf{x}|\mathbf{y})\mathbf{y}'(n-1)T_n^{-2}] \\ &= E[\mathbf{A}(\mathbf{y} - \sqrt{n}\delta)\mathbf{y}'(n-1)T_n^{-2}] \\ &= \mathbf{A}\{E[\mathbf{y}\mathbf{y}'(n-1)T_n^{-2}] - \sqrt{n}\delta E[\mathbf{y}'(n-1)T_n^{-2}]\}. \end{aligned} \quad (2.39)$$

The result in (2.39) follows from equation (2.34). Now, similar to equation (2.38) and the fact $E(\chi_m^2) = m$, one gets

$$E[\mathbf{y}\mathbf{y}'(n-1)T_n^{-2}] = m\{\mathbf{C}\Sigma\mathbf{C}'E(\chi_{p+1}^{-2}(\Delta)) + n\delta\delta'E(\chi_{p+3}^{-2}(\Delta))\}. \quad (2.40)$$

Therefore, we only need to evaluate

$$\begin{aligned} E[\mathbf{y}'(n-1)T_n^{-2}] &= mE[\mathbf{z}'(\mathbf{z}'\mathbf{z})^{-2}](\mathbf{D}')^{-1} \\ &= m\sqrt{n}\delta'E(\chi_{p+1}^{-2}(\Delta)). \end{aligned} \quad (2.41)$$

The result of (2.41) follows by applying Theorem 1 of Appendix A. Substituting (2.41) and (2.40) in (2.39), we have

$$E[\mathbf{xy}'(n-1)\mathbf{T}_n^{-2}] = \text{Am} \left\{ \mathbf{C}\Sigma\mathbf{C}'E(\chi_{p+1}^{-2}(\Delta)) + n\delta\delta'E(\chi_{p+3}^{-2}(\Delta)) - n\delta\delta'E(\chi_{p+1}^{-2}(\Delta)) \right\}. \quad (2.42)$$

The result of equation (2.26) follows from (2.37), (2.38) and (2.42) after some simplification using the identity

$$E(\chi_p^{-2}(\Delta)) - E(\chi_{p+2}^{-2}(\Delta)) = 2E(\chi_{p+2}^{-4}(\Delta)) \quad \forall p > 2.$$

For the special case, $\Sigma = \sigma^2[(1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p']$, the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}\Sigma\mathbf{C}'$ reduce to the following:

$$\begin{aligned} \mathbf{A} &= \mathbf{I}_p \\ \mathbf{B} &= \frac{\sigma^2}{p}[1 + (p-1)\rho]\mathbf{1}_p\mathbf{1}_p' \\ \mathbf{C}\Sigma\mathbf{C}' &= \sigma^2(1-\rho)\mathbf{C}. \end{aligned}$$

Note that in this case

$$\Delta = \frac{n\delta'\delta}{\sigma^2(1-\rho)}.$$

Therefore, the expressions in (2.23)–(2.26) become

$$\Gamma_1 = E\{n(\tilde{\mu}_n - \mu)(\tilde{\mu}_n - \mu)'\} = \sigma^2[(1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p'], \quad (2.43)$$

$$\Gamma_2 = E\{n(\hat{\mu}_n\mathbf{1}_p - \mu)(\hat{\mu}_n\mathbf{1}_p - \mu)'\} = \frac{\sigma^2}{p}[1 + (p-1)\rho]\mathbf{1}_p\mathbf{1}_p' + n\delta\delta', \quad (2.44)$$

$$\begin{aligned} \Gamma_3 &= E\{n(\hat{\mu}_n^{PT} - \mu)(\hat{\mu}_n^{PT} - \mu)'\} \\ &= \sigma^2[(1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p'] - \sigma^2(1-\rho)\mathbf{C}H_{p+1,m}(F^*; \Delta) \\ &\quad + n\delta\delta'\{2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)\} \end{aligned} \quad (2.45)$$

$$\begin{aligned} \Gamma_4 &= E\{n(\hat{\mu}_n^S - \mu)(\hat{\mu}_n^S - \mu)'\} \\ &= \sigma^2[(1-\rho)\mathbf{I}_p + \rho\mathbf{1}_p\mathbf{1}_p'] - km\sigma^2(1-\rho)\mathbf{C}[2E(\chi_{p+1}^{-2}(\Delta)) \\ &\quad - k(m+2)E(\chi_{p+1}^{-4}(\Delta))] \\ &\quad + km[4 + k(m+2)]n\delta\delta'E(\chi_{p+3}^{-4}(\Delta)), \end{aligned} \quad (2.46)$$

Consequently, the risk function under quadratic loss in (2.13) for the first four estimators in (2.3), (2.4), (2.9) and (2.10) under the intraclass correlation matrix in (2.1) are given by

$$R(\tilde{\mu}_n, \mathbf{W}) = \sigma^2 [(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p')] , \quad (2.47)$$

$$R(\hat{\mu}_n\mathbf{1}_p, \mathbf{W}) = \frac{\sigma^2}{p} [1 + (p - 1)\rho] Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p') + n\delta' \mathbf{W} \delta, \quad (2.48)$$

$$\begin{aligned} R(\hat{\mu}_n^{PT}, \mathbf{W}) = & \sigma^2 [(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p')] \\ & - \sigma^2 (1 - \rho) Tr(\mathbf{W}\mathbf{C}) H_{p+1,m}(F^*; \Delta) \\ & + n\delta' \mathbf{W} \delta \{2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)\}, \end{aligned} \quad (2.49)$$

$$\begin{aligned} R(\hat{\mu}_n^S, \mathbf{W}) = & \sigma^2 [(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p')] \\ & - km\sigma^2 (1 - \rho) Tr(\mathbf{W}\mathbf{C}) \{2E(\chi_{p+1}^{-2}(\Delta)) - k(m + 2)E(\chi_{p+1}^{-4}(\Delta))\} \\ & + km[4 + k(m + 2)]n\delta' \mathbf{W} \delta E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (2.50)$$

The optimum value of k is

$$k_{opt} = \frac{p - 3}{m + 2}.$$

Hence, using the identity

$$E(\chi_{p+2}^{-2}(\Delta)) - (p - 2)E(\chi_{p+2}^{-4}(\Delta)) = \Delta E(\chi_{p+4}^{-4}(\Delta)),$$

as well as $k = (p - 3)/(m + 2)$ and $\Delta = n\delta' \delta / \sigma^2 (1 - \rho)$, the expression in (2.50) reduces to

$$\begin{aligned} R(\hat{\mu}_n^S, \mathbf{W}) = & \sigma^2 [(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p')] \\ & - \frac{m(p - 3)}{m + 2} \sigma^2 (1 - \rho) Tr(\mathbf{W}\mathbf{C}) \{(p - 3)E(\chi_{p+1}^{-4}(\Delta)) \\ & + \left[1 - \frac{(p + 1)n\delta' \mathbf{W} \delta}{2\Delta\sigma^2(1 - \rho)Tr(\mathbf{W}\mathbf{C})}\right] 2\Delta E(\chi_{p+3}^{-4}(\Delta))\}. \end{aligned} \quad (2.51)$$

For some numerical work and graphs we consider the particular case $\mathbf{W} = \sigma^{-2}\mathbf{I}_p$. Then the expressions in (2.47)–(2.50) reduce to

$$R(\tilde{\mu}_n, \sigma^{-2}\mathbf{I}_p) = p, \quad (2.52)$$

$$R(\hat{\mu}_n\mathbf{1}_p, \sigma^{-2}\mathbf{I}_p) = p\rho + (1 - \rho)(1 + \Delta), \quad (2.53)$$

$$R(\hat{\mu}_n^{PT}, \sigma^{-2} \mathbf{I}_p) = p - (1 - \rho)(p - 1)H_{p+1,m}(F^*; \Delta) \\ + \Delta(1 - \rho)\{2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)\}, \quad (2.54)$$

$$R(\hat{\mu}_n^S, \sigma^{-2} \mathbf{I}_p) = p - (p - 3)^2 m(m + 2)^{-1} (1 - \rho) E(\chi_{p-1}^{-2}(\Delta)). \quad (2.55)$$

Note that when $\rho = 0$, the intraclass correlation matrix in (2.1) reduces to $\Sigma = \sigma^2 \mathbf{I}_p$. Hence, the MSE and the risk of the first four estimators can be obtained easily by substituting $\rho = 0$ in the (2.23)–(2.26) and (2.47)–(2.50) respectively.

2.4 Risk Analysis

In this section we assume that the covariance matrix is the intraclass correlation structure as in (2.1) and we proceed with an analysis of the risks of the first four estimators for μ and determine their dominance properties. Firstly, we note that the risk of $\tilde{\mu}_n$ is a constant line with height

$$\sigma^2 [(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W} \mathbf{1}_p \mathbf{1}_p')].$$

For example, if $\mathbf{W} = \sigma^{-2} \mathbf{I}_p$, then the risk is equal to p . While, the risk for $\hat{\mu}_n \mathbf{1}_p$ is a function of Δ . Furthermore,

$$R(\hat{\mu}_n \mathbf{1}_p, \mathbf{W}) \stackrel{\leq}{\geq} R(\tilde{\mu}_n, \mathbf{W}) \quad \text{according as} \quad n\delta' \mathbf{W} \delta \stackrel{\leq}{\geq} \sigma^2 (1 - \rho) Tr(\mathbf{W} \mathbf{C}), \quad (2.56)$$

If $\mathbf{W} = \sigma^{-2} \mathbf{I}_p$, then

$$R(\hat{\mu}_n \mathbf{1}_p, \mathbf{W}) \stackrel{\leq}{\geq} R(\tilde{\mu}_n, \mathbf{W}) \quad \text{according as} \quad \Delta \stackrel{\leq}{\geq} (p - 1).$$

Thus, the knowledge of Δ determines which estimator is to be preferred.

Secondly,

$$R(\hat{\mu}_n^{PT}, \mathbf{W}) \stackrel{\leq}{\geq} R(\tilde{\mu}_n, \mathbf{W}) \quad \text{according as} \\ n\delta' \mathbf{W} \delta \{2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)\} \\ \stackrel{\leq}{\geq} \sigma^2 (1 - \rho) Tr(\mathbf{W} \mathbf{C}) H_{p+1,m}(F^*; \Delta). \quad (2.57)$$

For $\mathbf{W} = \sigma^{-2}\mathbf{I}_p$, the above equation (2.57) reduces to

$$R(\hat{\mu}_n^{PT}, \mathbf{W}) \leq R(\tilde{\mu}_n, \mathbf{W}) \quad \text{according as} \\ \Delta\{2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)\} \geq (p-1)H_{p+1,m}(F^*; \Delta). \quad (2.58)$$

Thus, we notice $\hat{\mu}_n^{PT}$ dominates $\tilde{\mu}_n$ whenever

$$\Delta \leq (p-1)H_{p+1,m}(F^*; \Delta)[2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)]^{-1}.$$

Also, we note that under H_0 the risks of $\tilde{\mu}_n$, $\hat{\mu}_n \mathbf{1}_p$ and $\hat{\mu}_n^{PT}$ may be ordered according to the magnitude of the risks as follows:

$$\hat{\mu}_n \mathbf{1}_p \succ \hat{\mu}_n^{PT} \succ \tilde{\mu}_n,$$

since the risk $\hat{\mu}_n^{PT}$ under H_0 is given by

$$\sigma^2 \left[(1-\rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W} \mathbf{1}_p \mathbf{1}_p') \right] - \sigma^2 (1-\rho) Tr(\mathbf{W} \mathbf{C}) H_{p+1,m}(F^*; 0),$$

which is less than $\sigma^2 \left[(1-\rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W} \mathbf{1}_p \mathbf{1}_p') \right]$.

Finally, note that $\mathbf{C}$ is an idempotent matrix with rank $(p-1)$. Thus, by the Courant theorem [see e.g., Rao (1973)],

$$\frac{\delta' \mathbf{W} \delta}{\delta' \delta} \leq ch_{\max}(\mathbf{W}) \quad \text{for all } \delta,$$

where $ch_{\max}(\mathbf{W})$ stands for the largest characteristic root of $\mathbf{W}$. Consequently, we characterize a class $\mathcal{W}$ by

$$\mathcal{W} = \left\{ \mathbf{W}_{p,s,d} : Tr(\mathbf{W} \mathbf{C}) / ch_{\max}(\mathbf{W}) \geq \frac{p+1}{2} \right\}. \quad (2.59)$$

Then, we obtain from (2.47), (2.51) and (2.59) the result:

$$\frac{R(\hat{\mu}_n^S, \mathbf{W})}{R(\tilde{\mu}_n, \mathbf{W})} \leq 1 \quad \forall \delta \quad \text{and} \quad \mathbf{W} \in \mathcal{W}. \quad (2.60)$$

The proof of (2.60) follows by examining the following term in (2.51).

$$1 - \frac{(p+1)n\delta' \mathbf{W} \delta}{2\Delta\sigma^2(1-\rho)Tr(\mathbf{W} \mathbf{C})},$$

which must be greater than or equal to zero.

If in particular $\mathbf{W} = \sigma^{-2}\mathbf{I}_p$ or $\sigma^{-2}\mathbf{C}$, then

$$Tr(\mathbf{WC})/ch_{max}(\mathbf{W}) = p - 1 \geq \frac{p+1}{2}, \quad \text{for } p \geq 4.$$

We assume $\mathbf{W}$ belongs to $\mathcal{W}$ given by (2.59) and the remaining discussion follows. $R(\hat{\mu}_n^S, \mathbf{W})$ is monotone in Δ , the smallest value is

$$\sigma^2 \left[(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p') \right] - \frac{m(p-3)}{(m+2)(p-1)} \sigma^2 (1 - \rho)Tr(\mathbf{WC}) \text{ at } \Delta = 0,$$

and the largest value is $\sigma^2 \left[(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p') \right]$. Thus, $\hat{\mu}_n^S$ dominates $\tilde{\mu}_n$ and is an admissible estimator when compared with UMLE. Let Δ_α be the point at which $R(\hat{\mu}_n^{PT}, \mathbf{W})$ and $R(\hat{\mu}_n^S, \mathbf{W})$ intersect for each α (level of significance), then for $\Delta \in (0, \Delta_\alpha)$, $\hat{\mu}_n^{PT}$ dominates $\hat{\mu}_n^S$ for a range of α -values, while for $\Delta \in (\Delta_\alpha, \infty)$, $\hat{\mu}_n^S$ dominates $\hat{\mu}_n^{PT}$. In general, $\hat{\mu}_n^S$ always dominates $\hat{\mu}_n^{PT}$ for large values of α . Also, for large values of Δ , the UMLE dominates the RMLE, PTE and SE, both the PTE and SE share a common property that as $\Delta \rightarrow \infty$ their risks converge to the common limit $\sigma^2 \left[(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p') \right]$. But the risk of SE for all value of Δ always lies below

$$\sigma^2 \left[(1 - \rho)Tr(\mathbf{W}) + \rho Tr(\mathbf{W}\mathbf{1}_p\mathbf{1}_p') \right].$$

The following theorem presents the dominance picture of the first four estimators for μ :

Theorem 2.3 Under H_0 the dominance picture of the estimators is:

$$\tilde{\mu}_n\mathbf{1}_p \succ \hat{\mu}_n^{PT} \succ \hat{\mu}_n^S \succ \tilde{\mu}_n, \quad (2.61)$$

whenever the quantile F^* satisfies

$$F^* > H_{p+1,m}^{-1} \left\{ \frac{(p-3)m}{(p-1)(m+2)} \right\}. \quad (2.62)$$

Proof. Consider the risk-differences under H_0

(i)

$$R(\tilde{\mu}_n, \mathbf{W}) - R(\hat{\mu}_n^S, \mathbf{W}) = km\sigma^2(1 - \rho)Tr(\mathbf{WC}) > 0$$

(ii)

$$R(\hat{\mu}_n^S, \mathbf{W}) - R(\hat{\mu}_n^{PT}, \mathbf{W}) = \sigma^2(1 - \rho)Tr(\mathbf{WC}) \\ \times \left[H_{p+1,m}(F^*; 0) - \frac{m(p-3)}{(m+2)(p-1)} \right] > 0.$$

for a range of α -values satisfying (2.62).

(iii)

$$R(\hat{\mu}_n^{PT}, \mathbf{W}) - R(\hat{\mu}_n \mathbf{1}_p, \mathbf{W}) = \sigma^2(1 - \rho)Tr(\mathbf{WC}) \\ \times [1 - H_{p+1,m}(F^*; 0)] > 0 \quad \text{for all } \alpha.$$

2.5 Efficiency Analysis

In this section we assume that the covariance matrix is the intraclass correlation as in (2.1). Also, we let $\mathbf{W} = \sigma^{-2}\mathbf{I}_p$ and we consider the risk efficiency of the four estimators for μ in (2.3), (2.4), (2.9) and (2.10) and provide a max-rule for the optimum choice of the level of significance α for the preliminary test of the null hypothesis in (2.2). Tables for the relative efficiency (maximum and minimum) and Δ -value at which minimum efficiency of $\hat{\mu}_n^{PT}$ occurs relative to $\tilde{\mu}_n$ are provided. Firstly, the relative efficiency of $\hat{\mu}_n \mathbf{1}_p$ compared to $\tilde{\mu}_n$ (the usual estimator) is given by

$$E_1(\Delta, \rho) = R.E.(\hat{\mu}_n \mathbf{1}_p : \tilde{\mu}_n) = \frac{R(\tilde{\mu}_n, \sigma^{-2}\mathbf{I}_p)}{R(\hat{\mu}_n \mathbf{1}_p, \sigma^{-2}\mathbf{I}_p)} = \left[\rho + \frac{1}{p}(1 - \rho)(1 + \Delta) \right]^{-1} \quad (2.63)$$

For a given n and p , $E_1(\Delta, \rho)$ is a function of Δ and ρ . The function $E_1(\Delta, \rho)$ has its maximum at $\Delta = 0$ with value

$$\left[\rho + \frac{1}{p}(1 - \rho) \right]^{-1} \quad (> 1).$$

For ρ fixed $E_1(\Delta, \rho)$ is a decreasing function of Δ . For the maximum $E_1(0, \rho)$ is also decreasing function of ρ .

Secondly, the relative efficiency of $\hat{\mu}_n^{PT}$ compared to $\tilde{\mu}_n$ is given by

$$E_3(\alpha, \Delta, \rho) = R.E.(\hat{\mu}_n^{PT} : \tilde{\mu}_n) = \frac{R(\tilde{\mu}_n, \sigma^{-2}\mathbf{I}_p)}{R(\hat{\mu}_n^{PT}, \sigma^{-2}\mathbf{I}_p)} = \frac{1}{1 + g_3(\Delta)}, \quad (2.64)$$

where

$$g_3(\Delta) = \frac{1}{p}(1 - \rho) \{ \Delta [2H_{p+1,m}(F^*; \Delta) - H_{p+3,m}(F^{**}; \Delta)] - (p-1)H_{p+1,m}(F^*; \Delta) \}.$$

For a given n and p , E_3 is a function of α , Δ and ρ . The function, $E_3(\alpha, \Delta, \rho)$ for $\alpha \neq 0$, has its maximum at $\Delta=0$ which will be denoted by E^* with value

$$E^* = [1 - p^{-1}(p-1)(1-\rho)H_{p+1,m}(F^*; 0)]^{-1} (> 1).$$

For fixed ρ , E^* is a decreasing function of α , while the minimum efficiency E_0 is an increasing function of α . For fixed α , E_0 is an increasing function of ρ . Also, for fixed values of α and ρ , $E_3(\alpha, \Delta, \rho)$ decreases as Δ increases from 0, crossing the line $E_3(1, \Delta, \rho) = 1$, attains a minimum value at a point Δ_0 and then increases asymptotically to 1. For any fixed value of α , the maximum value of $E_3(\alpha, 0, \rho)$ is a decreasing function of ρ . For any fixed value of ρ , the graph of any two function $E_3(\alpha_1, \Delta, \rho)$ and $E_3(\alpha_2, \Delta, \rho)$ ($\alpha_1 \neq \alpha_2$) never intersect above the line $E_3(1, \Delta, \rho) = 1$, but they do intersect below that line. In Figures 2.1-2.2, there are graphs of the relative efficiency of the PTE relative to the UMLE for various choices of α and for $p=5, 4$, $n=15, 9$ and $\rho = -0.10, 0.30$. To help the experimenter choose an estimator with maximum relative efficiency, we adopt the following procedure. If Δ is known and that $0 \leq \Delta \leq p-1$, then $\hat{\mu}_n 1_p$ is chosen, since E_1 has the largest relative efficiency for all Δ in this range. However, if Δ and ρ are both unknown, there is no way of choosing a uniformly best estimator. Hence, we shall use the following criterion for selecting the significance level of the preliminary test. It consists of considering two cases, namely,

- (i) Suppose the experimenter does not know the size of α but knows $\rho = \rho_0$ and wants to accept an estimator which has relative efficiency not less than E_0 . Then among the set of estimators with $\alpha \in A$, where $A = \{ \alpha : E_3(\alpha, \Delta, \rho_0) \geq E_0 \text{ for all } \Delta \}$, the estimator is chosen to maximize $E_3(\alpha, \Delta, \rho_0)$ over all $\alpha \in A$ and Δ . That is to say, we solve for α from the equation

$$\max_{\alpha} \min_{\Delta} E_3(\alpha, \Delta, \rho_0) = E_0.$$

The solution α^* is the optimum choice for α . Tables 2.1-2.4 provide the values of the maximum relative efficiency E^* , minimum relative efficiency E_0 and the recommended corresponding size of the preliminary test α for $p=3,4$ and for various values of n and ρ . For example, if $n=6$ and $p=3$ and the experimenter knows that $\rho = 0.30$ and is willing to accept an estimator with relative efficiency no less than 0.6516, then from Table 2.1 the criterion suggests that he selects $\alpha^* = 0.25$ as the significance level, the selected estimator can have maximum relative efficiency 1.3043.

- (ii) Suppose the experimenter does not know the size of Δ and ρ , but wants an estimator which has relative efficiency not less than E_0 . Then among the set of estimators with $\alpha \in B$, where $B = \{\alpha : E_3(\alpha, \Delta, \rho) \geq E_0 \text{ for all } \Delta \text{ and } \rho\}$, the estimator is chosen to maximize $E_3(\alpha, \Delta, \rho)$ over all $\alpha \in B$ and all Δ, ρ . That is to say, we solve for α from the equation

$$\max_{\alpha} \min_{\Delta} E_3(\alpha, \Delta, \rho) = E_0 \text{ for all } \rho.$$

The solution α^* is the optimum level of significance α . Again, the Tables 2.1-2.4 may be used for finding α^* . For example, suppose the experimenter has $n=10$, $p=4$ and $\rho \in D$, where

$$D = \{\rho : \rho = -0.30, -0.05, 0, .10, .30, .50, .70\}$$

he does not know ρ but is willing to accept an estimator with relative efficiency no less than 0.70. From Table 2.3, according to the criterion he should choose $\alpha^* = 0.25$. Using the recommended α^* , the experimenter can achieve a relative efficiency as large as 1.3880.

Finally, to compare $\hat{\mu}_n^S$ to $\tilde{\mu}_n$, we observe that

$$E_4(\Delta, \rho) = R.E.(\hat{\mu}_n^S : \tilde{\mu}_n) = \frac{1}{1 - g_4(\Delta)} \geq 1, \quad (2.65)$$

where

$$g_4(\Delta) = p^{-1}(p-3)^2 m(m+2)^{-1}(1-\rho)E(\chi_{p-1}^{-2}(\Delta)).$$

E_4 is a function of Δ and ρ . For a fixed ρ , E_4 is a decreasing function in Δ with maximum value $[1 - p^{-1}(p - 3)m(m + 2)^{-1}(1 - \rho)]^{-1}$ at $\Delta=0$ and as $\Delta \rightarrow \infty$, the relative efficiency asymptotically equals 1. Thus, $\hat{\mu}_n^S$ uniformly dominates $\tilde{\mu}_n$. We have also noticed that $\hat{\mu}_n^S$ does not dominate $\hat{\mu}_n^{PT}$ under H_0 for a range of α . However, except for a range of Δ near the origin, $\hat{\mu}_n^S$ dominates $\hat{\mu}_n^{PT}$ and both estimators share a common property that as $\Delta \rightarrow \infty$ their $R.E \rightarrow 1$. But the R.E of $\hat{\mu}_n^S$ always lies above this asymptotic value for all value of ρ , while R.E of $\hat{\mu}_n^{PT}$ may drop for some intermediate values of Δ . Thus, from the point of robust-efficiency both $\hat{\mu}_n^{PT}$ and $\hat{\mu}_n^S$ may be advocated leaning more towards $\hat{\mu}_n^S$, since the size of Δ is generally unknown and small. Tables 2.5–2.8 provide the maximum relative efficiencies (at $\Delta=0$) for the RMLE (E_1), PTE (E_3) and SE (E_4) relative to UMLE and the intersecting relative efficiencies (E_{Δ_α}) of the PTE and SE and the Δ_α -values at which the intersection occurs. Also, Tables 2.5–2.8 show the range of the values of Δ for which $\hat{\mu}_n^{PT}$ dominates $\hat{\mu}_n^S$ for selected values of α and ρ . For example, using Table 2.5, with $n=6$ and $p=4$, $\alpha = 0.10$, $\rho = 0.10$ and $0 \leq \Delta \leq 1.1126$, it appears that the PTE dominates the SE. But outside the interval the SE dominates the PTE. If $\Delta = 0$, then $\hat{\mu}_n^{PT}$ has larger efficiency ($E_1 = 3.0769$) than $\hat{\mu}_n^{PT}$ ($E_3 = 2.2348$) and $\hat{\mu}_n^S$ ($E_4 = 1.1561$) and the efficiency at the intersection point $\Delta_{.10} = 1.1126$ is equal to 1.1044. It is interesting to note that for large values of α and n , SE dominates PTE uniformly, i.e., when $\Delta_\alpha = 0$ in Tables 2.5–2.8. To examine this we have plotted the dominance pictures of the UMLE, RMLE, PTE and SE for $p=4$, $n=9$ and $p=5$, $n=15$ for various values of ρ and α . All these pictures are displayed in Figure 2.3–2.6. The computations in Tables 2.5–2.8 have been carried out using equations (2.63), (2.64) and (2.65) with a FORTRAN program called “PROGRAM MULTIVARIATE NORMAL”, which will be given in the Appendix B. To compute (2.64), we approximate the non-central F-distribution using the central Beta distribution. Also, the Figures have been created by SPSS Graphics (SPSSG). Finally, it is clear from the Tables 2.5–2.8 that the SE uniformly dominates the UMLE for all values of ρ .

Figure 2.1: Relative efficiency of PTE to UMLE for $p=5$, $n=15$ and $\rho = -0.10$

Figure 2.2: Relative efficiency of PTE to UMLE for $p=4$, $n=9$ and $\rho = 0.30$

Figure 2.3: Dominance picture for the UMLE, RMLE, PTE and SE for $p=4$, $n=9$ and $\rho = -0.10$

Figure 2.4: Dominance picture for the UMLE, RMLE, PTE and SE for $p=4$, $n=9$ and $\rho = -0.10$

Figure 7 Dominance picture for the UMLE, RMLE, PTE and SE for $p=5$, $n=15$ and $\rho = .10$

Figure 2.6: Dominance picture for the UMLE, RMLE, PTE and SE for $p=5$, $n=15$ and $\rho = 0.10$

Table 2.4: Values of the maximum and minimum efficiencies and recommended significance levels for various values of ρ , $p=4$ and $n=15$

[illegible]

Table 2.5: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=6$

$\alpha \backslash \rho$		-.10	-.05	.00	.10	.30	.50	.70
.05	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	3.9571	3.4882	3.1187	2.5735	1.9067	1.5144	1.2560
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.05}}$	1.1267	1.1202	1.1139	1.1013	1.0771	1.0539	1.0316
	$\Delta_{.05}$	1.2004	1.2004	1.2004	1.2004	1.2004	1.2004	1.2004
.10	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	3.0800	2.8140	2.5902	2.2348	1.7536	1.4429	1.2258
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.10}}$	1.1306	1.1239	1.1173	1.1044	1.0793	1.0554	1.0325
	$\Delta_{.10}$	1.1126	1.1126	1.1126	1.1126	1.1126	1.1126	1.1126
.15	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	2.5503	2.3824	2.2353	1.9895	1.6309	1.3818	1.1987
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.15}}$	1.1344	1.1275	1.1207	1.1073	1.0815	1.0569	1.0334
	$\Delta_{.15}$	1.0302	1.0302	1.0302	1.0302	1.0302	1.0302	1.0302
.20	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	2.1951	2.0820	1.9800	1.8033	1.5301	1.3289	1.1744
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.20}}$	1.1383	1.1312	1.1241	1.1104	1.0838	1.0584	1.0343
	$\Delta_{.20}$	.9496	.9496	.9496	.9496	.9496	.9496	.9496
.25	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.9407	1.8611	1.7878	1.6573	1.4461	1.2826	1.1523
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.25}}$	1.1424	1.1350	1.1278	1.1135	1.0861	1.0600	1.0352
	$\Delta_{.25}$	.8680	.8680	.8680	.8680	.8680	.8680	.8680

Table 2.6: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=6$

$\alpha \backslash \rho$		-.10	-.05	.00	.10	.30	.50	.70
.30	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.7500	1.6923	1.6383	1.5400	1.3750	1.2419	1.1324
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.30}}$	1.1468	1.1392	1.1317	1.1170	1.0887	1.0618	1.0362
	$\Delta_{.30}$	.7830	.7830	.7830	.7830	.7830	.7830	.7830
.35	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.6021	1.5594	1.5190	1.4440	1.3143	1.2060	1.1142
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.35}}$	1.1517	1.1438	1.1360	1.1208	1.0915	1.0637	1.0373
	$\Delta_{.35}$	.6919	.6919	.6919	.6919	.6919	.6919	.6919
.40	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.4846	1.4526	1.4220	1.3644	1.2622	1.1742	1.0977
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.40}}$	1.1574	1.1492	1.1411	1.1252	1.0948	1.0659	1.0385
	$\Delta_{.40}$	.5914	.5914	.5914	.5914	.5914	.5914	.5914
.45	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.3895	1.3653	1.3420	1.2976	1.2171	1.1460	1.0828
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.45}}$	1.1643	1.1556	1.1471	1.1305	1.0986	1.0685	1.0400
	$\Delta_{.45}$	.4772	.4772	.4772	.4772	.4772	.4772	.4772
.50	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.3114	1.2931	1.2753	1.2411	1.1780	1.1210	1.0692
	E_4	1.1976	1.1869	1.1765	1.1561	1.1173	1.0811	1.0471
	$E_{\Delta_{.50}}$	1.1728	1.1637	1.1547	1.1371	1.1035	1.0718	1.0419
	$\Delta_{.50}$	.3433	.3433	.3433	.3433	.3433	.3433	.3433

Table 2.7: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=15$

$\alpha \backslash \rho$		-.10	-.05	.00	.10	.30	.50	.70
.05	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	3.4765	3.1247	2.8376	2.3971	1.8292	1.4788	1.2411
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.05}}$	1.2093	1.1979	1.1868	1.1650	1.1238	1.0854	1.0495
	$\Delta_{.05}$	.9577	.9577	.9577	.9577	.9577	.9577	.9577
.10	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	2.6616	2.4747	2.3123	2.0441	1.6591	1.3962	1.2052
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.10}}$	1.2202	1.2081	1.1962	1.1732	1.1297	1.0893	1.0518
	$\Delta_{.10}$	.8248	.8248	.8248	.8248	.8248	.8248	.8248
.15	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	2.2126	2.0971	1.9929	1.8129	1.5355	1.3318	1.1757
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.15}}$	1.2298	1.2171	1.2047	1.1805	1.1350	1.0928	1.0537
	$\Delta_{.15}$	.7138	.7138	.7138	.7138	.7138	.7138	.7138
.20	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.9243	1.8467	1.7751	1.6474	1.4402	1.2793	1.1507
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.20}}$	1.2395	1.2261	1.2131	1.1878	1.1402	1.0963	1.0556
	$\Delta_{.20}$	.6090	.6090	.6090	.6090	.6090	.6090	.6090
.25	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.7224	1.6676	1.6162	1.5224	1.3641	1.2355	1.1292
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.25}}$	1.2498	1.2358	1.2221	1.1955	1.1457	1.0999	1.0577
	$\Delta_{.25}$	.5032	.5032	.5032	.5032	.5032	.5032	.5032

Table 2.8: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for various values of ρ , $p=4$ and $n=15$

$\alpha \backslash \rho$		-.10	-.05	.00	.10	.30	.50	.70
.30	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.5730	1.5330	1.4951	1.4246	1.3017	1.1984	1.1103
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.30}}$	1.2615	1.2466	1.2322	1.2042	1.1519	1.1040	1.0599
	$\Delta_{.30}$	.3910	.3910	.3910	.3910	.3910	.3910	.3910
.35	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.4581	1.4284	1.3998	1.3460	1.2499	1.1666	1.0937
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.35}}$	1.2752	1.2594	1.2440	1.2144	1.1592	1.1087	1.0625
	$\Delta_{.35}$	.2675	.2675	.2675	.2675	.2675	.2675	.2675
.40	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.3673	1.3449	1.3231	1.2817	1.2062	1.1391	1.0791
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.40}}$	1.2920	1.2750	1.2586	1.2268	1.1680	1.1145	1.0657
	$\Delta_{.40}$	.1271	.1271	.1271	.1271	.1271	.1271	.1271
.45	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.2941	1.2771	1.2604	1.2284	1.1691	1.1152	1.0661
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.45}}$	1.2941	1.2771	1.2604	1.2284	1.1691	1.1152	1.0661
	$\Delta_{.45}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
.50	E_1	5.7143	4.7059	4.0000	3.0769	2.1053	1.6000	1.2903
	E_3	1.2343	1.2213	1.2085	1.1838	1.1374	1.0944	1.0546
	E_4	1.3084	1.2903	1.2727	1.2389	1.1765	1.1200	1.0687
	$E_{\Delta_{.50}}$	1.2343	1.2213	1.2085	1.1838	1.1374	1.0944	1.0546
	$\Delta_{.50}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000

2.6 Estimation of Means and Treatment Effects in a One-Way ANOVA Model

Let x_{ij} , $j = 1, 2, \dots, n_i$; $i = 1, 2, \dots, p$ be independent normally distributed with mean θ_i and common variance σ^2 . Stating this assumption somewhat differently, we can specify the underlying model by writing

$$x_{ij} = \theta_i + \epsilon_{ij}, \quad \text{for } j = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, p,$$

where ϵ_{ij} are values assumed by independent random variables having normal distribution with mean 0 and the common variance σ^2 , i.e., $\epsilon_{ij} \sim N(0, \sigma^2)$. The likelihood function for the parameters given the vector $\mathbf{x}_i = (x_{i1}, \dots, x_{in_i})$ is

$$L(\mathbf{x}_i | \theta) = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_{ij} - \theta_i)^2}, \quad i = 1, 2, \dots, p, \quad (2.66)$$

where $\theta = (\theta_1, \dots, \theta_p)'$. If we define $n = \sum_{i=1}^p n_i$ and the null-hypothesis H_0 by

$$H_0 : \theta = \theta_0 \mathbf{1}_p, \quad (\text{unknown}) \quad (2.67)$$

then the likelihood function for the parameters given the vector $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ under H_0 is

$$L(\mathbf{x} | \theta) = \prod_{i=1}^p \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_{ij} - \theta_0)^2}. \quad (2.68)$$

We are primarily interested in the estimation of θ when we may have uncertain prior information in (2.67).

Let us define

$$\tilde{\theta}_n = (\tilde{\theta}_1, \dots, \tilde{\theta}_p)', \quad \tilde{\theta}_i = \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, \dots, p; \quad (2.69)$$

$$\hat{\theta}_n = (\hat{\theta}_1, \dots, \hat{\theta}_p)' = \hat{\theta}_n \mathbf{1}_p; \quad (2.70)$$

where

$$\hat{\theta}_n = \bar{x} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} x_{ij}, \quad \mathbf{1}_p = (1, 1, \dots, 1)'$$

Then, for the global model (2.66), the unrestricted maximum likelihood estimator (UMLE) of θ is given by $\tilde{\theta}_n$, while under the model (2.68), the restricted MLE

(RMLE) is given by $\hat{\theta}_n$. When H_0 holds, $\hat{\theta}_n$ has a smaller risk (under quadratic loss) than $\tilde{\theta}_n$. On the other hand, when H_0 does not hold, $\tilde{\theta}_n$ may perform better than $\hat{\theta}_n$. As a result, when the prior information on H_0 is rather uncertain it may be desirable to have either a preliminary test estimator (PTE) denoted by $\hat{\theta}_n^{PT} = (\hat{\theta}_1^{PT}, \dots, \hat{\theta}_p^{PT})'$ or a shrinkage estimator (SE) denoted by $\hat{\theta}_n^S = (\hat{\theta}_1^S, \dots, \hat{\theta}_p^S)'$ of θ .

2.6.1 The PT and Shrinkage Estimators

For the preliminary test (PT) on H_0 in (2.67), we consider the F-test statistic

$$L_n = \frac{(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \mathbf{D} (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p) / p - 1}{S_e^2}, \quad (2.71)$$

where

$$S_e^2 = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \tilde{\theta}_i)^2 / (n - p), \quad \text{and} \quad \mathbf{D} = \text{Diag}(n_i).$$

Under H_0 , L_n follows the central F-distribution with $(p - 1), (n - p)$ degrees of freedom (DF). Thus, for a given *level of significance* α ($0 < \alpha < 1$), the PTE of θ is defined by

$$\hat{\theta}_n^{PT} = \hat{\theta}_n \mathbf{1}_p + [1 - I(L_n \leq F_{p-1, n-p}(\alpha))] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (2.72)$$

where $I(A)$ is the indicator function for the set A .

To introduce the SE we follow Arnold (1981) and propose

$$\hat{\theta}_n^S = \hat{\theta}_n \mathbf{1}_p + [1 - cL_n^{-1}] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (2.73)$$

where we need to assume that $p > 3$ and c is the shrinkage constant.

Note that whereas in (2.72) we have a dichotomous indicator function, in (2.73), cL_n^{-1} may be regarded as a smoother version. Thus, the SE might be considered from a "pre-test" point of view. Also, (2.72) is a convex combination of $\tilde{\theta}_n$ and $\hat{\theta}_n$, while (2.73) need not be because cL_n^{-1} may be greater than 1.

2.6.2 Biases and Risks

Let $\mathbf{x}_n = \tilde{\theta}_n - \theta$ and $\mathbf{y}_n = \tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p = \mathbf{H}\tilde{\theta}_n$ then the following holds:

$$\begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} \sim N_{2p} \left\{ \begin{bmatrix} \mathbf{0} \\ \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{D}^{-1} & \mathbf{D}^{-1}\mathbf{H}' \\ \mathbf{H}\mathbf{D}^{-1} & \mathbf{D}^{-1}\mathbf{H}' \end{bmatrix} \right\}, \quad (2.74)$$

where

$$\delta = \mathbf{H}\theta, \quad \mathbf{H} = \mathbf{I}_p - \frac{\mathbf{J}\mathbf{D}}{n} \quad \text{and} \quad \mathbf{J} = \mathbf{1}_p \mathbf{1}_p'. \quad (2.75)$$

Note that $\mathbf{H}$ is an idempotent matrix of rank $p - 1$, satisfying $\mathbf{H}\mathbf{1}_p = \mathbf{0}$. Furthermore, under H_0 we have $\mathbf{H}\theta = \mathbf{0}$.

Let δ_n be an estimator of θ , $\mathbf{W}$ be a positive semi-definite (p.s.d) matrix, and consider the quadratic loss function

$$L(\delta_n, \theta) = (\delta_n - \theta)' \mathbf{W} (\delta_n - \theta) = \text{Tr} [\mathbf{W} (\delta_n - \theta)(\delta_n - \theta)'], \quad (2.76)$$

where $\text{Tr}(\mathbf{A}) = \text{trace of } \mathbf{A}$. Then the risk of δ_n is given by

$$R(\delta_n, \theta) = EL(\delta_n, \theta) = \text{Tr}(\mathbf{W} \text{MSE}), \quad (2.77)$$

where the mean square error (MSE) of δ_n is defined by

$$\text{MSE} = E(\delta_n - \theta)(\delta_n - \theta)'. \quad (2.78)$$

In this section we present expressions for the bias, the mean square error (MSE) and the risk under loss function in (2.76) of the first four estimators in (2.69), (2.70), (2.72) and (2.73). Theorem 2.4 and 2.5 provide expressions for the bias and the MSE expressions respectively.

Theorem 2.4 *Bias of the UMLE, RMLE, PTE and SE are given by (2.79)–(2.82) respectively:*

$$B_1 = E(\tilde{\theta}_n - \theta) = \mathbf{0}, \quad (2.79)$$

$$B_2 = E(\hat{\theta}_n \mathbf{1}_p - \theta) = -\delta, \quad (2.80)$$

$$B_3 = E(\hat{\theta}_n^{PT} - \theta) = -\delta G_{p+1,m}(F^*; \Delta), \quad (2.81)$$

$$B_4 = E(\hat{\theta}_n^S - \theta) = -cm\delta E(\chi_{p+1}^{-2}(\Delta)), \quad (2.82)$$

where

$$\Delta = \delta' \mathbf{D} \delta / \sigma^2, \quad \delta = \mathbf{H} \theta, \quad (2.83)$$

$\mathbf{H}$ is defined by (2.75) and $m = n - p$, we use the symbol $G_{q_1, q_2}(\cdot; \Delta)$ to denote the cumulative distribution of a non-central F -distribution with (q_1, q_2) DF and non-centrality parameter Δ . Also, F^* is defined by

$$F^* = \frac{p-1}{p+1} F_{p-1, m}(\alpha),$$

where $F_{p-1, m}(\alpha)$ is the upper α probability point of a central F -distribution with $(p-1, m)$ DF.

Proof. The proof of the above theorem follows, by observing that $\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p = \mathbf{H} \tilde{\theta}_n$ and making the following transformation

$$\mathbf{z} = \mathbf{C} \mathbf{y}_n / \sigma, \quad (2.84)$$

where $\mathbf{C}$ is any symmetric and non-singular matrix such that $\mathbf{C}(\mathbf{H} \mathbf{D}^{-1} \mathbf{H}') \mathbf{C}' = \mathbf{I}$, then L_n can be written as

$$L_n = \frac{\chi_{p-1}^2(\Delta)/p-1}{\chi_m^2/n-p} = \frac{\mathbf{z}' \mathbf{z}/p-1}{\chi_m^2/n-p}, \quad (2.85)$$

where $\mathbf{z}' \mathbf{z}$ is distributed as a non-central chi-square distribution with $(p-1)$ DF and non-centrality parameter Δ and independent of the χ_m^2 variable and applying Theorem 1 from Appendix A.

Theorem 2.5 *The MSE of the UMLE, RMLE, PTE and SE under (2.78) are given by (2.86)–(2.89) respectively:*

$$\Gamma_1 = E\{(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)'\} = \sigma^2 \mathbf{D}^{-1}, \quad (2.86)$$

$$\Gamma_2 = E\{(\hat{\theta}_n \mathbf{1}_p - \theta)(\hat{\theta}_n \mathbf{1}_p - \theta)'\} = \frac{\sigma^2}{n} \mathbf{J} + \delta \delta', \quad (2.87)$$

$$\begin{aligned} \Gamma_3 = E\{(\hat{\theta}_n^{PT} - \theta)(\hat{\theta}_n^{PT} - \theta)'\} &= \sigma^2 \mathbf{D}^{-1} - \sigma^2 \mathbf{J}^* G_{p+1, m}(F^*; \Delta) \\ &\quad + \delta \delta' \{2G_{p+1, m}(F^*; \Delta) - G_{p+3, m}(F^{**}; \Delta)\} \end{aligned} \quad (2.88)$$

$$\begin{aligned} \Gamma_4 &= E\{(\hat{\theta}_n^S - \theta)(\hat{\theta}_n^S - \theta)'\} \\ &= \sigma^2 \mathbf{D}^{-1} - cm \sigma^2 \mathbf{J}^* [2E(\chi_{p+1}^{-2}(\Delta)) - c(m+2)E(\chi_{p+1}^{-4}(\Delta))] \\ &\quad + cm[4 + c(m+2)]\delta \delta' E(\chi_{p+3}^{-4}(\Delta)), \end{aligned} \quad (2.89)$$

where

$$F^{**} = \frac{p-1}{p+3} F_{p-1,m}(\alpha), \quad \mathbf{J}^* = \mathbf{D}^{-1} - \frac{\mathbf{J}}{n},$$

Δ and F^* are defined as in Theorem 2.4.

Proof. The proof of the above theorem is similar to Theorem 2.2 in Section 2.3.

Consequently, the risk function under quadratic loss in (2.76) for the estimators in (2.69), (2.70), (2.72) and (2.73) are given by

$$R(\tilde{\theta}_n, \mathbf{W}) = \sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1}), \quad (2.90)$$

$$R(\hat{\theta}_n \mathbf{1}_p, \mathbf{W}) = \frac{\sigma^2}{n} \text{Tr}(\mathbf{W}\mathbf{J}) + \delta' \mathbf{W} \delta, \quad (2.91)$$

$$R(\hat{\theta}_n^{PT}, \mathbf{W}) = \sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1}) - \sigma^2 \text{Tr}(\mathbf{W}\mathbf{J}^*) G_{p+1,m}(F^*; \Delta) \\ + \delta' \mathbf{W} \delta \{2G_{p+1,m}(F^*; \Delta) - G_{p+3,m}(F^{**}; \Delta)\}, \quad (2.92)$$

$$R(\hat{\theta}_n^S, \mathbf{W}) = \sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1}) \\ - cm\sigma^2 \text{Tr}(\mathbf{W}\mathbf{J}^*) \{2E(\chi_{p+1}^{-2}(\Delta)) - c(m+2)E(\chi_{p+1}^{-4}(\Delta))\} \\ + cm[4 + c(m+2)]\delta' \mathbf{W} \delta E(\chi_{p+3}^{-4}(\Delta)). \quad (2.93)$$

The optimum value of c is

$$c_{opt} = \frac{p-3}{m+2}.$$

Hence, using the identity

$$E(\chi_{p+2}^{-2}(\Delta)) - (p-2)E(\chi_{p+2}^{-4}(\Delta)) = \Delta E(\chi_{p+4}^{-4}(\Delta)),$$

as well as $c = (p-3)/(m+2)$, the expression in (2.93) reduces to

$$R(\hat{\theta}_n^S, \mathbf{W}) = \sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1}) - \frac{m(p-3)}{m+2} \sigma^2 \text{Tr}(\mathbf{W}\mathbf{J}^*) \{(p-3)E(\chi_{p+1}^{-4}(\Delta)) \\ + \left[1 - \frac{(p+1)\delta' \mathbf{W} \delta}{2\Delta \sigma^2 \text{Tr}(\mathbf{W}\mathbf{J}^*)}\right] 2\Delta E(\chi_{p+3}^{-4}(\Delta))\}. \quad (2.94)$$

For some numerical work and graphs we consider the particular case $\mathbf{W} = \sigma^{-2}\mathbf{D}$.

Then the expressions in (2.90)–(2.93) reduce to

$$R(\tilde{\theta}_n, \sigma^{-2}\mathbf{D}) = p, \quad (2.95)$$

$$R(\hat{\theta}_n \mathbf{1}_p, \sigma^{-2}\mathbf{D}) = 1 + \Delta, \quad (2.96)$$

$$R(\hat{\theta}_n^{PT}, \sigma^{-2}D) = p - (p-1)G_{p+1,m}(F^*; \Delta) + \Delta \{2G_{p+1,m}(F^*; \Delta) - G_{p+3,m}(F^{**}; \Delta)\}, \quad (2.97)$$

$$R(\hat{\theta}_n^S, \sigma^{-2}D) = p - (p-3)^2 m(m+2)^{-1} E(\chi_{p-1}^{-2}(\Delta)). \quad (2.98)$$

2.6.3 Analysis of the Estimators

We now proceed with an analysis of the risks of the estimators and determine their dominance properties. Firstly, we note that

$$\frac{R(\hat{\theta}_n \mathbf{1}_p, \mathbf{W})}{R(\tilde{\theta}_n, \mathbf{W})} \leq 1 \quad \text{according as} \quad \delta' \mathbf{W} \delta \leq \sigma^2 \text{Tr}(\mathbf{WJ}^*), \quad (2.99)$$

Secondly,

$$\begin{aligned} \frac{R(\hat{\theta}_n^{PT}, \mathbf{W})}{R(\tilde{\theta}_n, \mathbf{W})} &\leq 1 \quad \text{according as} \\ &\delta' \mathbf{W} \delta \{2G_{p+1,m}(F^*; \Delta) - G_{p+3,m}(F^{**}; \Delta)\} \\ &\leq \sigma^2 \text{Tr}(\mathbf{WJ}^*) G_{p+1,m}(F^*; \Delta), \end{aligned} \quad (2.100)$$

and

$$\begin{aligned} \frac{R(\hat{\theta}_n^{PT}, \mathbf{W})}{R(\hat{\theta}_n \mathbf{1}_p, \mathbf{W})} &\leq 1 \quad \text{according as} \\ &\delta' \mathbf{W} \delta \{1 - 2G_{p+1,m}(F^*; \Delta) + G_{p+3,m}(F^{**}; \Delta)\} \\ &\leq \sigma^2 \text{Tr}(\mathbf{WJ}^*) [1 - G_{p+1,m}(F^*; \Delta)]. \end{aligned} \quad (2.101)$$

Finally, note that $\mathbf{H}$ is an idempotent matrix with rank $(p-1)$. Thus, by the Courant theorem [Rao (1973)],

$$\frac{\delta' \mathbf{W} \delta}{\delta' \mathbf{D} \delta} \leq ch_{\max}(\mathbf{W} \mathbf{D}^{-1}), \quad \text{for all } \delta,$$

where $ch_{\max}(\mathbf{W})$ stands for the largest characteristic root of $\mathbf{W}$. For convenience, we characterize a class $\mathcal{W}$ by

$$\mathcal{W} = \left\{ \mathbf{W}_{p,s,d} : \text{Tr}(\mathbf{WJ}^*) / ch_{\max}(\mathbf{W} \mathbf{D}^{-1}) \geq \frac{p+1}{2} \right\}. \quad (2.102)$$

Then, we obtain from (2.90), (2.93) and (2.102) the result:

$$\frac{R(\hat{\theta}_n^S, \mathbf{W})}{R(\tilde{\theta}_n, \mathbf{W})} \leq 1 \quad \forall \delta \quad \text{and} \quad \mathbf{W} \in \mathcal{W}. \quad (2.103)$$

The proof of (2.103) follows by examining the following term in (2.94):

$$1 - \frac{(p+1)\delta' \mathbf{W} \delta}{2\Delta \sigma^2 \text{Tr}(\mathbf{WJ}^*)},$$

which must be greater than or equal to zero. Note that for the dominance of $\hat{\theta}_n^S$ over $\tilde{\theta}_n$ for all $p \geq 4$, we need the typical choice of $\mathbf{W}$ given by (2.102). For an arbitrary choice of $\mathbf{W}$, the inequality in (2.102) may not hold simultaneously for all $\mathbf{D}$. In such a case, to induce the dominance of $\hat{\theta}_n^S$ over $\tilde{\theta}_n$, one may introduce the modified Stein-estimator. The modified Stein-estimator is similar to the normal theory case treated by Berger et al (1977) and to the nonparametric case discussed by Sen and Saleh (1985). In the sequel we assume $\mathbf{W}$ belongs to $\mathcal{W}$ given by (2.102) and the remaining discussion follows.

In the above we may note that $\text{Tr}(\mathbf{WJ}^*) > 0$ and

$$\sigma^2 ch_{\min}(\mathbf{WD}^{-1}) \leq \frac{\delta' \mathbf{W} \delta}{\Delta} \leq \sigma^2 ch_{\max}(\mathbf{WD}^{-1}),$$

where $ch_{\min}(\mathbf{W})$ stands for the smallest characteristic root of $\mathbf{W}$. Hence, (2.99) implies that $\hat{\theta}_n 1_p$ has a smaller risk than that of the risk of $\tilde{\theta}_n$ if $\Delta < \frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\max}(\mathbf{WD}^{-1})}$, i.e., $\hat{\theta}_n 1_p$ performs better than $\tilde{\theta}_n$ in the interval $\left[0, \frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\max}(\mathbf{WD}^{-1})}\right)$. Alternatively, when Δ deviates from the null hypothesis beyond $\frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\min}(\mathbf{WD}^{-1})}$, the risk of $\hat{\theta}_n 1_p$ grows and becomes unbounded whereas the risk of $\tilde{\theta}_n$ remains bounded. Hence departure from the null hypothesis, i.e., $\Delta \in \left(\frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\min}(\mathbf{WD}^{-1})}, \infty\right)$ is fatal to $\hat{\theta}_n 1_p$ but is of very little concern to $\tilde{\theta}_n$. For $\mathbf{W} = \sigma^{-2}\mathbf{D}$, we obtain $\frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\max}(\mathbf{WD}^{-1})} = \frac{\text{Tr}(\mathbf{WJ}^*)}{ch_{\min}(\mathbf{WD}^{-1})} = p - 1$. These findings are confirmed in Figures 2.7 ($p=4$, $n=15$) and 2.8 ($p=5$, $n=20$).

Using (2.92), we may point out some important characteristics of the risk of $\hat{\theta}_n^{PT}$. For $\delta = 0$, i.e., under the null hypothesis, the risk of $\hat{\theta}_n^{PT}$ reduces to

$$\sigma^2 [\text{Tr}(\mathbf{WD}^{-1}) - \text{Tr}(\mathbf{WJ}^*) G_{p+1,m}(F^2; 0)]$$

which is smaller than the risk of $\tilde{\theta}_n$. Also, for large deviations of δ from 0 (null hypothesis), i.e., as $\Delta \rightarrow \infty$ the risk of $\hat{\theta}_n^{PT}$ approaches to the risk of $\tilde{\theta}_n$ from above. Further, as δ departs from 0, i.e., as Δ grows, the value of risk of $\hat{\theta}_n^{PT}$ increases to a maximum after crossing the risk of $\tilde{\theta}_n$, then decreases towards the risk of $\tilde{\theta}_n$. Furthermore, we find that the risk of the $\hat{\theta}_n^{PT}$ is smaller than the risk of $\tilde{\theta}_n$ when

$$\Delta < \frac{Tr(WJ^*)}{ch_{max}(WD^{-1})(2 - \nu_0)}, \quad \nu_0 = \frac{G_{p+3,m}(F^{**}; \Delta)}{G_{p+1,m}(F^*; \Delta)} \leq 1$$

and for $\Delta > \frac{Tr(WJ^*)}{ch_{min}(WD^{-1})(2 - \nu_0)}$, the opposite conclusion holds, but as soon as $\Delta \rightarrow \infty$, the risk of $\hat{\theta}_n^{PT}$ converges to that of $\tilde{\theta}_n$. For $W = \sigma^{-2}D$, we obtain $\frac{Tr(WJ^*)}{ch_{min}(WD^{-1})} = \frac{Tr(WJ^*)}{ch_{max}(WD^{-1})} = p - 1$. In this case, for $\Delta \in [0, (p - 1)(2 - \nu_0)^{-1}]$, the PTE $\hat{\theta}_n^{PT}$ has a smaller risk than that of $\tilde{\theta}_n$, whereas the opposite is true for $\Delta \in ((p - 1)(2 - \nu_0)^{-1}, \infty)$.

Now, we compare $\hat{\theta}_n^{PT}$ and $\hat{\theta}1_p$. First, under the null hypothesis, i.e., $\delta = 0$ the risk of $\hat{\theta}1_p$ is $\frac{\sigma^2}{n}Tr(WJ)$ and

$$R(\hat{\theta}_n^{PT}, W) - R(\hat{\theta}1_p, W) = \sigma^2 Tr(WJ^*) [1 - G_{p+1,m}(F^*; 0)] > 0.$$

Hence, we conclude that under the null hypothesis $\hat{\theta}1_p$ performs better than $\hat{\theta}_n^{PT}$. Furthermore, $\hat{\theta}1_p$ performs better than $\hat{\theta}_n^{PT}$ for

$$\Delta < \frac{Tr(WJ^*)[1 - G_{p+1,m}(F^*; \Delta)]}{ch_{max}(WD^{-1})\{2(1 - G_{p+1,m}(F^*; \Delta)) - (1 - G_{p+3,m}(F^{**}; \Delta))\}}$$

and for

$$\Delta > \frac{Tr(WJ^*)[1 - G_{p+1,m}(F^*; \Delta)]}{ch_{min}(WD^{-1})\{2(1 - G_{p+1,m}(F^*; \Delta)) - (1 - G_{p+3,m}(F^{**}; \Delta))\}}$$

the opposite conclusion holds. For $W = \sigma^{-2}D$, the right-hand quantities are equal

$$\frac{(p - 1)[1 - G_{p+1,m}(F^*; \Delta)]}{2[1 - G_{p+1,m}(F^*; \Delta)] - [1 - G_{p+3,m}(F^{**}; \Delta)]}.$$

Considering the discussion above we conclude that none of the three estimators $\tilde{\theta}_n$, $\hat{\theta}_n 1_p$, and $\hat{\theta}_n^{PT}$ is inadmissible with respect to the other two.

The picture is somewhat different with respect to the shrinkage estimator. First, note from (2.103), that the risk of $\hat{\theta}_n^S$ is uniformly smaller than the risk of

$\tilde{\theta}_n$, where the upper limit is attained when $\Delta \rightarrow \infty$. This shows an inadmissibility of $\tilde{\theta}_n$ relative to $\hat{\theta}_n^S$.

To compare $\hat{\theta}_n^S$ and $\hat{\theta}_n 1_p$ under the null-hypothesis in (2.67) ($\delta = 0$) we have

$$\begin{aligned} R(\hat{\theta}_n^S, W) &= R(\hat{\theta}_n 1_p, W) + \sigma^2 \text{Tr}(WJ^*) \left[1 - \frac{m(p-3)}{(m+2)(p-1)} \right] \\ &> R(\hat{\theta}_n 1_p, W). \end{aligned} \quad (2.104)$$

Thus, the risk of $\hat{\theta}_n^S$ is greater than that of $\hat{\theta}_n 1_p$ for $\delta = 0$. However, as δ moves away from 0, $\delta'W\delta$ increases while $E(\chi_{p+1}^{-4}(\Delta))$ and $E(\chi_{p+3}^{-4}(\Delta))$ decrease, so the opposite conclusion holds. In general, $\hat{\theta}_n^S$ does not dominate $\hat{\theta}_n 1_p$ for Δ in the interval $[0, \Delta_*)$, where

$$\Delta_* = \frac{\text{Tr}(WJ^*) \left[1 - m(m+2)^{-1}(p-3) \{ E(\chi_{p+1}^{-2}(\Delta)) + \Delta E(\chi_{p+3}^{-4}(\Delta)) \} \right]}{ch_{\max}(WD^{-1}) \left\{ 1 - m(m+2)^{-1}(p-3)(p+1)E(\chi_{p+3}^{-4}(\Delta)) \right\}},$$

however, it does so for $\Delta \in (\Delta^*, \infty)$ where

$$\Delta^* = \frac{\text{Tr}(WJ^*) \left[1 - m(m+2)^{-1}(p-3) \{ E(\chi_{p+1}^{-2}(\Delta)) + \Delta E(\chi_{p+3}^{-4}(\Delta)) \} \right]}{ch_{\min}(WD^{-1}) \left\{ 1 - m(m+2)^{-1}(p-3)(p+1)E(\chi_{p+3}^{-4}(\Delta)) \right\}}.$$

Hence neither $\hat{\theta}_n^S$ nor $\hat{\theta}_n 1_p$ dominates the other.

Now, we compare $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$. First, consider the case when $\delta = 0$. In this case $R(\hat{\theta}_n^S, W) = \sigma^2 [\text{Tr}(WD^{-1}) - \frac{m(p-3)}{(m+2)(p-1)} \text{Tr}(WJ^*)]$. Hence we have

$$R(\hat{\theta}_n^S, W) - R(\hat{\theta}_n^{PT}, W) = \sigma^2 \text{Tr}(WJ^*) \left\{ G_{p+1,m}(F^*; 0) - \frac{m(p-3)}{(m+2)(p-1)} \right\}, \quad p \geq 4. \quad (2.105)$$

Thus the PTE $\hat{\theta}_n^{PT}$ has a smaller risk than that of the risk of $\hat{\theta}_n^S$ at $\delta = 0$ whenever

$$G_{p+1,m}(F^*; 0) > \frac{m(p-3)}{(m+2)(p-1)}, \quad (2.106)$$

otherwise the opposite conclusion holds. Thus, $\hat{\theta}_n^S$ does not dominate $\hat{\theta}_n^{PT}$ when the null hypothesis holds provided that (2.106) holds. Under this situation we can order dominance of the estimators under the null-hypothesis in (2.67) as follows

$$\hat{\theta}_n 1_p \succ \hat{\theta}_n^{PT} \succ \hat{\theta}_n^S \succ \tilde{\theta}_n, \quad (2.107)$$

where the notation $\succ$ means dominates. However, the dominance picture changes to $\hat{\theta}_n 1_p \succ \hat{\theta}_n^S \succ \hat{\theta}_n^{PT} \succ \tilde{\theta}_n$, whenever $G_{p+1,m}(F^*; 0) < \frac{m(p-3)}{(m+2)(p-1)}$. This is consistent with the Figures 2.7 and 2.8 and Tables 2.9–2.12. Thus, (2.106) specifies a range of the values of α for given $p \geq 4$ for which the PTE $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$. The picture changes as δ moves away from the origin (i.e., under the null-hypothesis). Note that $\tilde{\theta}_n$ has constant risk equal to $\sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1})$ and the risk of $\hat{\theta}_n 1_p$ depends on δ ; its risk becomes unbounded as δ moves further away from 0. As for $\hat{\theta}_n^{PT}$, the risk increases from the initial value

$$\sigma^2 [\text{Tr}(\mathbf{W}\mathbf{D}^{-1}) - \text{Tr}(\mathbf{W}\mathbf{J}^*)G_{p+1,m}(F^*; 0)]$$

to a maximum after crossing the risk of $\tilde{\theta}_n$, then decreases to the value $\sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1})$ which is the risk of $\tilde{\theta}_n$. Similarly, the risk of $\hat{\theta}_n^S$ with the initial value

$$\sigma^2 \left[\text{Tr}(\mathbf{W}\mathbf{D}^{-1}) - \frac{m(p-3)}{(m+2)(p-1)} \text{Tr}(\mathbf{W}\mathbf{J}^*) \right],$$

increases monotonically towards $\sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1})$ as Δ moves away from 0. The risk of $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$ intersect at the point $\Delta = \Delta_\alpha$ for each α ($0 < \alpha < 1$), if the condition in (2.106) is satisfied; otherwise there is no intersection. If $\Delta \in [0, \Delta_\alpha)$, then the PTE dominates the SE while outside this interval the SE dominates the PTE. Thus under the condition in (2.106) neither the PTE $\hat{\theta}_n^{PT}$ nor the SE $\hat{\theta}_n^S$ dominates each other. The SE $\hat{\theta}_n^S$ dominates the PTE $\hat{\theta}_n^{PT}$ uniformly if the condition in (2.106) is not satisfied. However, $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ share a common property as $\Delta \rightarrow \infty$, namely, their risk's converge to $\sigma^2 \text{Tr}(\mathbf{W}\mathbf{D}^{-1})$ which is the risk of $\tilde{\theta}_n$. But the risk function of $\hat{\theta}_n^S$ is always below this limit, whereas the risk function of $\hat{\theta}_n^{PT}$ exceeds this limit for some intermediate values of Δ depending on α , the level of significance of the test of (2.67). Finally, from the point of robust-efficiency both $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ may be advocated, leaning more towards $\hat{\theta}_n^S$, since the size of Δ is generally unknown and unlikely to be small.

Next, we consider the relative efficiency analysis of the estimators, for the specific case $\mathbf{W} = \sigma^{-2}\mathbf{D}$. Using (2.95)–(2.98), the relative efficiency (R.E.) of the

RMLE $\hat{\theta}_n 1_p$ to the UMLE $\tilde{\theta}_n$ is defined to be

$$R.E.(\hat{\theta}_n 1_p : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \sigma^{-2} \mathbf{D})}{R(\hat{\theta}_n 1_p, \sigma^{-2} \mathbf{D})} = \frac{p}{1 + \Delta}. \quad (2.108)$$

Similarly, the relative efficiency of $\hat{\theta}_n^{PT}$ and the SE $\hat{\theta}_n^S$ to the UMLE $\tilde{\theta}_n$ are given by (2.109)–(2.110) respectively:

$$R.E.(\hat{\theta}_n^{PT} : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \sigma^{-2} \mathbf{D})}{R(\hat{\theta}_n^{PT}, \sigma^{-2} \mathbf{D})} = \frac{1}{1 + g_1(\alpha, \Delta)}, \quad (2.109)$$

where

$$g_1(\alpha, \Delta) = \frac{1}{p} \{ \Delta [2G_{p+1,m}(F^*; \Delta) - G_{p+3,m}(F^{**}; \Delta)] - (p-1)G_{p+1,m}(F^*; \Delta) \}.$$

$$R.E.(\hat{\theta}_n^S : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \sigma^{-2} \mathbf{D})}{R(\hat{\theta}_n^S, \sigma^{-2} \mathbf{D})} = \frac{1}{1 - g_2(\Delta)}, \quad (2.110)$$

where

$$g_2(\Delta) = p^{-1}(p-3)^2 m(m+2)^{-1} E(\chi_{p-1}^{-2}(\Delta)).$$

Tables 2.9–2.12 provide the maximum relative efficiencies (at $\Delta = 0$) for $\hat{\theta}_n 1_p$ (E_1), $\hat{\theta}_n^{PT}$ (E_3) and $\hat{\theta}_n^S$ (E_4) relative to $\tilde{\theta}_n$, the intersecting relative efficiencies denoted by E_{Δ_α} of $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$ (which is the common efficiency of the PTE and the SE) and the Δ_α -values at which the intersection occurs. For example, using Table 2.10, with $p=5$, $n=12$, $\alpha = 0.15$ and $\Delta \in [0, 1.9109]$, it appears that $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$, but outside the interval domination is reversed. While, if the hypothesis in (2.67) is correct, i.e., $\Delta = 0$, then $\hat{\theta}_n 1_p$ has larger efficiency ($E_1 = 5.0000$) than $\hat{\theta}_n^{PT}$ ($E_3 = 2.3591$) and $\hat{\theta}_n^S$ ($E_4 = 1.4516$) relative to $\tilde{\theta}_n$. Note that $\Delta_\alpha = 0$ in Tables 2.9–2.12 means that no intersection between $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ occurs, and the corresponding relative efficiencies E_{Δ_α} are equal to E_3 , i.e., the SE $\hat{\theta}_n^S$ in (2.73) dominates the PTE $\hat{\theta}_n^{PT}$ in (2.72) uniformly when α -values are large. Figures 2.7–2.8 exhibit plots of the relative efficiency (R.E.) curves for $\hat{\theta}_n 1_p$, $\hat{\theta}_n^{PT}$, and $\hat{\theta}_n^S$ relative to $\tilde{\theta}_n$ versus Δ for various values of α , p and n . Apparently, $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$ for a range of α ($0 < \alpha < 1$) and for small values of Δ . Also, $\hat{\theta}_n^S$ dominates $\hat{\theta}_n^{PT}$ uniformly for large values of α .

Figure 2.7: Performance picture for the UMLE, RMLE, PTE and SE for $p=4$ and $n=15$

Figure 2.8: Performance picture for the UMLE, RMLE, PTE and SE for $p=5$ and $n=20$

Table 2.9: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=4$

$\alpha \backslash n$		9	12	15	25	30	40	50
.05	E_1	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000
	E_3	3.0001	2.9041	2.8505	2.7712	2.7526	2.7299	2.7167
	E_4	1.2174	1.2500	1.2683	1.2958	1.3023	1.3103	1.3151
	$E_{\Delta_{.05}}$	1.0960	1.1149	1.1264	1.1452	1.1500	1.1559	1.1595
	$\Delta_{.05}$	2.3292	2.1530	2.0489	1.8887	1.8502	1.8031	1.7754
.15	E_1	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000
	E_3	2.1236	2.0439	2.0026	1.9448	1.9317	1.9161	1.9070
	E_4	1.2174	1.2500	1.2683	1.2958	1.3023	1.3103	1.3151
	$E_{\Delta_{.15}}$	1.1100	1.1345	1.1496	1.1741	1.1803	1.1881	1.1928
	$\Delta_{.15}$	1.8925	1.6661	1.5411	1.3578	1.3150	1.2632	1.2329
.25	E_1	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000
	E_3	1.7063	1.6508	1.6227	1.5842	1.5755	1.5653	1.5593
	E_4	1.2174	1.2500	1.2683	1.2958	1.3023	1.3103	1.3151
	$E_{\Delta_{.25}}$	1.1249	1.1554	1.1746	1.2062	1.2143	1.2246	1.2308
	$\Delta_{.25}$	1.5046	1.2431	1.1009	.8944	.8465	.7887	.7550
.40	E_1	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000
	E_3	1.3740	1.3424	1.3268	1.3055	1.3008	1.2952	1.2919
	E_4	1.2174	1.2500	1.2683	1.2958	1.3023	1.3103	1.3151
	$E_{\Delta_{.40}}$	1.1582	1.2051	1.2359	1.2893	1.3008	1.2952	1.2919
	$\Delta_{.40}$	.8273	.4923	.3116	.0516	.0000	.0000	.0000
.50	E_1	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000	4.0000
	E_3	1.2426	1.2214	1.2109	1.1968	1.1937	1.1900	1.1879
	E_4	1.2174	1.2500	1.2683	1.2958	1.3023	1.3103	1.3151
	$E_{\Delta_{.50}}$	1.2012	1.2214	1.2109	1.1968	1.1937	1.1900	1.1879
	$\Delta_{.50}$	.1928	.0000	.0000	.0000	.0000	.0000	.0000

Table 2.10: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=5$

$\alpha \backslash n$		9	12	15	25	30	40	50
.05	E_1	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	E_3	3.6647	3.5101	3.4252	3.3015	3.2727	3.2380	3.2178
	E_4	1.3636	1.4516	1.5000	1.5714	1.5882	1.6087	1.6207
	$E_{\Delta_{.05}}$	1.1607	1.2088	1.2388	1.2887	1.3015	1.3177	1.3275
	$\Delta_{.05}$	2.9877	2.6393	2.4343	2.1238	2.0505	1.9616	1.9097
.15	E_1	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	E_3	2.4885	2.3591	2.2944	2.2062	2.1866	2.1633	2.1500
	E_4	1.3636	1.4516	1.5000	1.5714	1.5882	1.6087	1.6207
	$E_{\Delta_{.15}}$	1.1836	1.2506	1.2940	1.3682	1.3876	1.4122	1.4272
	$\Delta_{.15}$	2.4056	1.9109	1.6452	1.2712	1.1872	1.0873	1.0299
.25	E_1	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	E_3	1.9380	1.8476	1.8037	1.7453	1.7325	1.7174	1.7087
	E_4	1.3636	1.4516	1.5000	1.5714	1.5882	1.6087	1.6207
	$E_{\Delta_{.25}}$	1.2122	1.3033	1.3648	1.4738	1.5029	1.5403	1.5633
	$\Delta_{.25}$	1.8210	1.2240	.9181	.5036	.4131	.3067	.2462
.40	E_1	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	E_3	1.5045	1.4524	1.4277	1.3952	1.3881	1.3798	1.3751
	E_4	1.3636	1.4516	1.5000	1.5714	1.5882	1.6087	1.6207
	$E_{\Delta_{.40}}$	1.2858	1.4510	1.4277	1.3952	1.3881	1.3798	1.3751
	$\Delta_{.40}$	.7522	.0039	.0000	.0000	.0000	.0000	.0000
.50	E_1	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000	5.0000
	E_3	1.3333	1.2977	1.2809	1.2591	1.2543	1.2488	1.2456
	E_4	1.3636	1.4516	1.5000	1.5714	1.5882	1.6087	1.6207
	$E_{\Delta_{.50}}$	1.3333	1.2977	1.2809	1.2591	1.2543	1.2488	1.2456
	$\Delta_{.50}$	.0000	.0000	.0000	.0000	.0000	.0000	.0000

Table 2.12: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for n -values and $p=7$

$\alpha \backslash n$		9	12	15	25	30	40	50
.05	E_1	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
	E_3	5.0121	4.6836	4.5113	4.2712	4.2171	4.1527	4.1155
	E_4	1.4000	1.6897	1.8421	2.0588	2.1084	2.1681	2.2028
	$E_{\Delta_{.05}}$	1.1745	1.3101	1.3947	1.5414	1.5806	1.6309	1.6618
	$\Delta_{.05}$	4.7600	3.8221	3.2945	2.5407	2.3726	2.1747	2.0622
.15	E_1	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
	E_3	3.2276	2.9408	2.8124	2.6492	2.6145	2.5741	2.5511
	E_4	1.4000	1.6897	1.8421	2.0588	2.1084	2.1681	2.2028
	$E_{\Delta_{.15}}$	1.1871	1.3703	1.5028	1.7562	1.8276	1.9211	1.9795
	$\Delta_{.15}$	4.2441	2.7745	2.0376	1.1164	.9308	.7211	.6059
.25	E_1	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
	E_3	2.4043	2.2010	2.1150	2.0090	1.9869	1.9611	1.9465
	E_4	1.4000	1.6897	1.8421	2.0588	2.1084	2.1681	2.2028
	$E_{\Delta_{.25}}$	1.2049	1.4571	1.6654	2.0090	1.9869	1.9611	1.9465
	$\Delta_{.25}$	3.6097	1.6943	.8369	.0000	.0000	.0000	.0000
.40	E_1	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
	E_3	1.7661	1.6460	1.5974	1.5386	1.5265	1.5124	1.5045
	E_4	1.4000	1.6897	1.8421	2.0588	2.1084	2.1681	2.2028
	$E_{\Delta_{.40}}$	1.2525	1.6460	1.5974	1.5386	1.5265	1.5124	1.5045
	$\Delta_{.40}$	2.3008	.0000	.0000	.0000	.0000	.0000	.0000
.50	E_1	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000	7.0000
	E_3	1.5155	1.4316	1.3982	1.3582	1.3500	1.3405	1.3351
	E_4	1.4000	1.6897	1.8421	2.0588	2.1084	2.1681	2.2028
	$E_{\Delta_{.50}}$	1.3197	1.4316	1.3982	1.3582	1.3500	1.3405	1.3351
	$\Delta_{.50}$	1.0330	.0000	.0000	.0000	.0000	.0000	.0000

2.6.4 Estimation of the Treatment Effects

In this section, we consider the one-way classification model

$$x_{ij} = \mu + \beta_i + \epsilon_{ij}; \quad \text{for } j = 1, \dots, n_i, \quad i = 1, \dots, p,$$

where $\epsilon_{ij} \sim N(0, \sigma^2)$ and $\sum_{i=1}^p n_i \beta_i = 0$. We are interested in the estimation of the vector $\beta = (\beta_1, \dots, \beta_p)'$ when it is suspected that $\beta_1 = \beta_2 = \dots = \beta_p = 0$. In this case, the unrestricted estimator (UE) of β is given by

$$\tilde{\beta}_n = (\tilde{\beta}_1, \dots, \tilde{\beta}_p)', \quad \tilde{\beta}_i = \bar{x}_i - \bar{x}, \quad i = 1, \dots, p, \quad (2.111)$$

where $\bar{x}_i$ and $\bar{x}$ are defined by (2.69) and (2.70) respectively. Note that $\tilde{\beta}_n$ can be written as

$$\tilde{\beta}_n = \tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p.$$

Furthermore, the test statistics for

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0,$$

is given by L_n as defined by (2.71). Then the PTE of β is defined by

$$\hat{\beta}_n^{PT} = [1 - I(L_n \leq F_{p-1, n-p}(\alpha))] \tilde{\beta}_n, \quad (2.112)$$

and the corresponding shrinkage estimators (SE) is defined by

$$\hat{\beta}_n^S = \left[1 - \frac{p-3}{m+2} L_n^{-1}\right] \tilde{\beta}_n. \quad (2.113)$$

Finally, the positive rule shrinkage estimators is defined by

$$\hat{\beta}_n^{S+} = \left[1 - \frac{p-3}{m+2} L_n^{-1}\right] \left(L_n \geq \frac{p-3}{m+2}\right) \tilde{\beta}_n. \quad (2.114)$$

The following expressions are for the bias and the mean square error. Since the derivation of these formula is very similar to the one in Section 2.6.2, we omit the details of the derivation. The bias expressions for the estimators in (2.111), (2.112) and (2.113) are given by

$$B_1 = E(\tilde{\beta}_n - \beta) = 0, \quad (2.115)$$

$$B_3 = E(\hat{\beta}_n^{PT} - \beta) = -\delta G_{p+1, m}(F^*; \Delta), \quad (2.116)$$

$$B_4 = E(\hat{\beta}_n^S - \beta) = -\frac{p-3}{m+2} m \delta E(\chi_{p+1}^{-2}(\Delta)), \quad (2.117)$$

where Δ , δ , m , F^* and $G_{q_1, q_2}(\cdot; \Delta)$ are defined as in Theorem 2.4.

The mean square error MSE of the estimators in (2.111), (2.112) and (2.113) are given by

$$\Gamma_1 = E\{(\tilde{\beta}_n - \beta)(\tilde{\beta}_n - \beta)'\} = \sigma^2 J^*, \quad (2.118)$$

$$\begin{aligned} \Gamma_3 = E\{(\hat{\beta}_n^{PT} - \beta)(\hat{\beta}_n^{PT} - \beta)'\} &= \sigma^2 J^* - \sigma^2 J^* G_{p+1, m}(F^*; \Delta) \\ &\quad + \delta \delta' \{2G_{p+1, m}(F^*; \Delta) - G_{p+3, m}(F^{**}; \Delta)\} \end{aligned} \quad (2.119)$$

$$\begin{aligned} \Gamma_4 = E\{(\hat{\beta}_n^S - \beta)(\hat{\beta}_n^S - \beta)'\} \\ = \sigma^2 J^* - \frac{p-3}{m+2} m \sigma^2 J^* [2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))] \\ + m(m+2)^{-1}(p+1)(p-3)\delta \delta' E(\chi_{p+3}^{-4}(\Delta)), \end{aligned} \quad (2.120)$$

where F^{**} and J^* are defined as in Theorem 2.5.

Consequently, the risk function under quadratic loss in (2.76) for the estimators in (2.111), (2.112) and (2.113) are given by

$$R(\tilde{\beta}_n, W) = \sigma^2 \text{Tr}(WJ^*), \quad (2.121)$$

$$\begin{aligned} R(\hat{\beta}_n^{PT}, W) &= \sigma^2 \text{Tr}(WJ^*) - \sigma^2 \text{Tr}(WJ^*) G_{p+1, m}(F^*; \Delta) \\ &\quad + \delta' W \delta \{2G_{p+1, m}(F^*; \Delta) - G_{p+3, m}(F^{**}; \Delta)\}, \end{aligned} \quad (2.122)$$

$$\begin{aligned} R(\hat{\beta}_n^S, W) &= \sigma^2 \text{Tr}(WJ^*) \\ &\quad - \frac{p-3}{m+2} m \sigma^2 \text{Tr}(WJ^*) \{2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ &\quad + m(m+2)^{-1}(p+1)(p-3)\delta' W \delta E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (2.123)$$

The performance of the above estimators are exactly the same as the performance of the estimators in Section 2.6.3, by observing the similarity of the risk expressions and the analysis that follows.

Chapter 3

Preliminary Test and Empirical Bayes Approach to Shrinkage Estimation of Regression Parameters

3.1 Introduction

The problem of estimating the regression coefficients in the usual regression model is considered when it is a priori suspected that the coefficients may be restricted to a subspace. We shall be primarily concerned with preliminary test approach to the estimation of the parameters. As such four estimators, namely, the unrestricted least squares estimator (ULSE), restricted least squares estimator (RLSE), preliminary test estimator (PTE) and shrinkage estimator (SE) are given. Further, we consider the Bayes and Empirical Bayes approach to the estimation under certain prior distribution and show that the empirical Bayes estimator (EBE) is exactly equal to the SE. The RLSE has the smallest risk when the hypothesis that the regression coefficients satisfy the restriction holds. However, the risk of RLSE is unbounded when the parameter moves away from the subspace of the restriction. The SE has the smallest risk in most cases except when the parameter is near the null hypothesis; in such case the PTE is better.

Consider the general regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (3.1)$$

where $\mathbf{e} = (e_1, e_2, \dots, e_n)'$ and $\mathbf{e}$ is distributed normally with mean zero and covariance matrix $\sigma^2 \mathbf{I}_n$, $\sigma^2 > 0$ is unknown, that is $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, the notation $\sim$ means distributed as and $\mathbf{I}_n$ denotes the identity matrix of order n . Further, let $\mathbf{y}$ be an $n \times 1$ observable vector of random variables, $\mathbf{X}$ be an $n \times p$ matrix ($n > p$) of known fixed numbers and $\boldsymbol{\beta}$ be a $p \times 1$ vector of unknown parameters. We are mainly interested in the estimation of the vector $\boldsymbol{\beta}$ when it is a priori suspected that $\boldsymbol{\beta}$ may be restricted to the subspace defined by

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{h}, \quad (3.2)$$

where $\mathbf{H}$ is a given $q \times p$ matrix of rank q and $\mathbf{h}$ is a given $q \times 1$ vector. This is a very general restriction on the parameters of the regression model which includes as a special case the subhypothesis and the parallelism problem among others. Some details of special cases are given in Example 6.3.2 page 184–185 of Graybill (1976). For more general discussion on the hypothesis $\mathbf{H}\boldsymbol{\beta} = \mathbf{h}$ readers are referred to Searle (1971).

In regression problems, one may use the unrestricted least squares estimator (ULSE) to estimate the unknown parameters, if there are no restrictions on the parameters. But, quite often, some prior knowledge induces the experimenter to suspect the existence of a relation among the component of the parameters. In this chapter, we will be concerned with the general restriction in (3.2). The restricted least square estimator (RLSE) provides a first alternative, if the above relation holds. But if the parameters of the true model do not satisfy these relations, the RLSE has undesirable properties. Moreover, most of the time, the equality of these relations can not be taken for granted. Thus, an interesting alternative to the ULSE and RLSE is the preliminary test estimator, i.e., choose the ULSE or RLSE according to the result of a test about the null-hypothesis $\mathbf{H}\boldsymbol{\beta} = \mathbf{h}$. The method was proposed originally by Bancroft (1944), and has been developed later in a series of articles by Bancroft and Bancroft and Han (see e.g., Bancroft and

Han (1980) for an excellent review). However, despite many of their attractive features, Sclove et al (1972) have shown that the preliminary test estimators are not minimax for the quadratic loss function and are dominated by some shrinkage estimators, (see also Judge and Bock (1978)). The Bayesian approach allows, as usual, to incorporate both supposed relations and the degree of uncertainty about these relations through the prior distribution on the parameters.

Our main objective of this study is to show that the shrinkage estimator under a general uncertain sub-hypothesis stated earlier is an empirical Bayes estimator (EBE) and to provide a finite sample theory with normality assumption for the estimation of regression coefficients β .

In Section 3.2, we provide the preliminary test approach to the estimation. The empirical Bayes estimators and the analysis of the Bayes risk are considered in Section 3.3. In Section 3.4 contains the bias and the risk of the four estimators. Section 3.5 provides the frequentist risk analysis of the first four estimators.

3.2 Preliminary Test Approach

The unrestricted least squares estimator (ULSE) of β for the model (3.1) is given by

$$\tilde{\beta} = (X'X)^{-1}X'y = C^{-1}X'y, \quad C = X'X, \quad (3.3)$$

and the restricted least squares estimator (RLSE) of β under the restriction in (3.2) is given by

$$\hat{\beta} = \tilde{\beta} - C^{-1}H'(HC^{-1}H')^{-1}(H\tilde{\beta} - h). \quad (3.4)$$

The unbiased estimator of σ^2 is given by

$$S_e^2 = (n - p)^{-1} \| y - X\tilde{\beta} \|^2,$$

where $\| \cdot \|$ denotes the Euclidean norm.

To obtain the PTE, we first consider a test of the null hypothesis

$$H_0 : H\beta = h \quad \text{against} \quad H_A : H\beta \neq h.$$

The likelihood ratio test statistic is

$$L_n = \frac{(\mathbf{H}\tilde{\boldsymbol{\beta}} - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\boldsymbol{\beta}} - \mathbf{h})}{qS_e^2}.$$

The test statistic L_n follows a central F -distribution with $(q, n - p)$ degrees of freedom (DF) under H_0 . Let F_α be the upper 100 α percentile of the central F distribution with $(q, n - p)$ DF, then the PTE is proposed as a convex combination of $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}$ given by

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}} + [1 - I(L_n \leq F_\alpha)](\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}), \quad (3.5)$$

where $I(A)$ is the indicator of the set A .

For Stein-rule shrinkage estimator (SE) of $\boldsymbol{\beta}$, we propose the following using the statistic L_n as

$$\hat{\boldsymbol{\beta}}^S = \hat{\boldsymbol{\beta}} + [1 - dL_n^{-1}](\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}), \quad (3.6)$$

where

$$d = \frac{(q-2)(n-p)}{q(n-p+2)} \quad \text{and} \quad q \geq 3.$$

This is a generalization of the usual James-Stein estimator. Note (3.6) is not a convex combination. Both (3.5) and (3.6) involve the statistic L_n which adjusts the estimator for departure from H_0 . For large values of L_n both (3.5) and (3.6) yield the ULSE, while for small values of L_n they may perform differently. However, a convex combination may be defined as

$$\hat{\boldsymbol{\beta}}^{S+} = \hat{\boldsymbol{\beta}} + [1 - dL_n^{-1}]^+(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}). \quad (3.7)$$

where $a^+ = \max(a, 0)$. This is known as positive part of (3.6). Such positive part estimators prevent overshrinking in (3.6). However, for simplicity, in the remainder of the chapter, we shall consider $\hat{\boldsymbol{\beta}}^S$ instead of $\hat{\boldsymbol{\beta}}^{S+}$.

3.3 The Bayesian Approach

In this section we consider the development of shrinkage estimators shrinking towards $\hat{\boldsymbol{\beta}}$. We approach this problem by using the empirical Bayes technique.

Properties of this estimator are compared with restricted and unrestricted least squares estimators under quadratic loss function in terms of the Bayes risk. The comparison shows that the EB technique has the attractive property that like the Bayes estimator (BE), the EB estimator remain the same for every $\mathbf{Q}$ in the quadratic loss function given in (3.35). Also, the EB estimators has smaller risk than the ULSE under certain restriction induced by the prior distribution. Now, we show that the shrinkage estimator in (3.6) is an empirical Bayes estimator (EBE) of β in the following way.

3.3.1 Bayes Estimator

Assume the model $\mathbf{y}|\beta \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$, where $\mathbf{I}_n$, $\mathbf{X}$ and β are defined as in Section 3.1. It is assumed that $\mathbf{X}'\mathbf{X}$ is non-singular. We assume the prior $\beta \sim N(\nu, \tau^2\mathbf{V})$, where ν is a p -component column vector and $\mathbf{V}$ is a $p \times p$ p.d matrix. The following theorem provides the posterior distribution of β given $\mathbf{y} = \underline{\mathbf{y}}$, where $\underline{\mathbf{y}}$ be a specific values of the vector $\mathbf{y}$. For the proof of this theorem see Berger (1985) or Ghosh, Saleh and Sen (1989), but we will give the proof for the sake of completeness.

Theorem 3.1 *The posterior distribution of β given $\mathbf{y} = \underline{\mathbf{y}}$ is*

$$N_p \left(\nu + \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1} \right)^{-1} \mathbf{X}'(\underline{\mathbf{y}} - \mathbf{X}\nu), \sigma^2 \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1} \right)^{-1} \right) \quad (3.8)$$

Proof. First note that

$$\begin{bmatrix} \mathbf{y} \\ \beta \end{bmatrix} \sim N_{n+p} \left\{ \begin{bmatrix} \mathbf{X}\nu \\ \nu \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{I}_n + \tau^2\mathbf{X}\mathbf{V}\mathbf{X}' & \tau^2\mathbf{X}\mathbf{V} \\ \tau^2\mathbf{V}\mathbf{X}' & \tau^2\mathbf{V} \end{bmatrix} \right\}. \quad (3.9)$$

Hence using the conditional formula the distribution of $\beta|\mathbf{y} = \underline{\mathbf{y}}$ is

$$N_p(\nu + \tau^2\mathbf{V}\mathbf{X}'(\sigma^2\mathbf{I}_n + \tau^2\mathbf{X}\mathbf{V}\mathbf{X}')^{-1}(\underline{\mathbf{y}} - \mathbf{X}\nu), \tau^2\mathbf{V} - \tau^4\mathbf{V}\mathbf{X}'(\sigma^2\mathbf{I}_n + \tau^2\mathbf{X}\mathbf{V}\mathbf{X}')^{-1}\mathbf{X}\mathbf{V}). \quad (3.10)$$

By exercise 2.9 page 33 of Rao (1973), one gets the following

$$(\sigma^2\mathbf{I}_n + \tau^2\mathbf{X}\mathbf{V}\mathbf{X}')^{-1} = (\tau^2)^{-1} \left[\frac{\sigma^2}{\tau^2}\mathbf{I}_n + \mathbf{X}\mathbf{V}\mathbf{X}' \right]^{-1}$$

$$\begin{aligned}
&= (\tau^2)^{-1} \left[\frac{\tau^2}{\sigma^2} \mathbf{I}_n - \frac{\tau^4}{\sigma^4} \mathbf{X} \left(\frac{\tau^2}{\sigma^2} \mathbf{X}'\mathbf{X} + \mathbf{V}^{-1} \right)^{-1} \mathbf{X}' \right] \\
&= (\tau^2)^{-1} \frac{\tau^2}{\sigma^2} \left[\mathbf{I}_n - \mathbf{X} \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1} \mathbf{X}' \right]. \quad (3.11)
\end{aligned}$$

Hence from (3.11) we have,

$$\begin{aligned}
\tau^2 \mathbf{V} \mathbf{X}' (\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{X} \mathbf{V} \mathbf{X}')^{-1} &= \frac{\tau^2}{\sigma^2} \left[\mathbf{V} \mathbf{X}' - \mathbf{V} \mathbf{X}' \mathbf{X} \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1} \mathbf{X}' \right] \\
&= \frac{\tau^2}{\sigma^2} \left\{ \mathbf{V} \mathbf{X}' - \left[\mathbf{V} - \frac{\sigma^2}{\tau^2} \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1} \right] \mathbf{X}' \right\} \\
&= \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1} \mathbf{X}'. \quad (3.12)
\end{aligned}$$

Also, from (3.12),

$$\begin{aligned}
\tau^2 \mathbf{V} \mathbf{X}' (\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{X} \mathbf{V} \mathbf{X}')^{-1} \mathbf{X} \mathbf{V} &= \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1} \mathbf{X}' \mathbf{X} \mathbf{V} \\
&= \mathbf{V} - \frac{\sigma^2}{\tau^2} \left(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1} \right)^{-1}. \quad (3.13)
\end{aligned}$$

The conclusion of theorem follows from (3.10), (3.12) and (3.13).

Now under the quadratic loss function, the Bayes estimator of β is given by the posterior mean. We choose $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$ as given in Ghosh, Saleh and Sen (1989). With this choice, the prior variance-covariance matrix is proportional to the variance covariance matrix of the usual ULSE $\tilde{\beta}$. This choice leads to useful James-Stein type EB estimators of β as shown by Arnold [(1981), page 168]. First, we note that with this choice $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$, the Bayes estimator is given by

$$\hat{\beta}^B = E(\beta|\mathbf{y}) = \nu + (1 - B)(\tilde{\beta} - \nu), \quad (3.14)$$

where $B = \frac{\sigma^2}{\sigma^2 + \tau^2}$. The posterior distribution of β given $\mathbf{y} = \underline{y}$ is

$$N_p \left(\nu + (1 - B)(\tilde{\beta} - \nu), \sigma^2(1 - B)\mathbf{C}^{-1} \right).$$

In an empirical Bayes framework, the parameters ν , τ^2 and σ^2 are unknown, and to be estimated from the marginal distribution of $\mathbf{y}$. Also, we take $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$, then from (3.9) the marginal distribution of $\mathbf{y}$ is given by

$$\mathbf{y} \sim N_n(\mathbf{X}\nu, \Sigma), \quad (3.15)$$

where

$$\Sigma = \sigma^2 \mathbf{I}_n + \tau^2 \mathbf{P}_X \quad \text{and} \quad \mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \quad (3.16)$$

Note that $\mathbf{P}_X$ is the orthogonal projection on the subspace spanned by the columns of the matrix $\mathbf{X}$.

3.3.2 Complete and Sufficient Statistics

First, we define the following norms

$$\begin{aligned} \|\mathbf{z}\|_{**}^2 &= \mathbf{z}'\mathbf{C}\mathbf{z} \quad \text{for } \mathbf{z} \in R^q \\ \|\mathbf{v}\|_*^2 &= \mathbf{v}'[\mathbf{I}_n - (1-B)\mathbf{P}_X]\mathbf{v} \quad \text{for } \mathbf{v} \in R^n. \end{aligned} \quad (3.17)$$

The next theorem provides a complete sufficient statistic for $(\nu, \sigma^2 + \tau^2, \sigma^2)$ based on the marginal distribution of $\mathbf{y}$.

Theorem 3.2 *Based on the marginal distribution of $\mathbf{y}$ and the assumption $\mathbf{H}\nu = \mathbf{h}$, we have*

$$(\hat{\beta}, (\mathbf{H}\tilde{\beta} - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta} - \mathbf{h}), S_e^2)$$

is a complete sufficient statistic for $(\nu, \sigma^2 + \tau^2, \sigma^2)$, where $S_e^2 = \frac{\|\mathbf{y} - \mathbf{X}\tilde{\beta}\|^2}{n-p}$, $\|\cdot\|$ denotes the Euclidean norm.

Proof. First, note that from (3.16) and exercise 2.9, page 33 of Rao (1973)

$$(\sigma^2 \mathbf{I}_n + \tau^2 \mathbf{P}_X)^{-1} = \sigma^{-2} [\mathbf{I}_n - (1-B)\mathbf{P}_X]. \quad (3.18)$$

This justifies the introduction of $\|\cdot\|_*^2$. Hence from (3.15), the probability density function (pdf) is proportional to

$$\exp \left\{ -(2\sigma^2)^{-1} \|\mathbf{y} - \mathbf{X}\nu\|_*^2 \right\}. \quad (3.19)$$

By the Pythagorean theorem, we have

$$\|\mathbf{y} - \mathbf{X}\nu\|_*^2 = \|\mathbf{y} - \mathbf{X}\tilde{\beta}\|_*^2 + \|\mathbf{X}\tilde{\beta} - \mathbf{X}\nu\|_*^2. \quad (3.20)$$

Since $\mathbf{X}\tilde{\beta} = \mathbf{P}_X \mathbf{y}$ and $\mathbf{P}_X \mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ on the subspace spanned by the columns of the matrix $\mathbf{X}$. Using $\mathbf{P}_X \mathbf{X} = \mathbf{X}$, $\mathbf{y} - \mathbf{X}\tilde{\beta} = (\mathbf{I}_n - \mathbf{P}_X)\mathbf{y}$ and (3.17), we have the following

$$\begin{aligned}\|\mathbf{X}\tilde{\beta} - \mathbf{X}\nu\|_*^2 &= (\mathbf{X}\tilde{\beta} - \mathbf{X}\nu)'[\mathbf{I}_n - (1 - B)\mathbf{P}_X](\mathbf{X}\tilde{\beta} - \mathbf{X}\nu) \\ &= (\tilde{\beta} - \nu)' \mathbf{X}'[\mathbf{I}_n - (1 - B)\mathbf{P}_X]\mathbf{X}(\tilde{\beta} - \nu) \\ &= B(\tilde{\beta} - \nu)' \mathbf{C}(\tilde{\beta} - \nu) \\ &= B \|\tilde{\beta} - \nu\|_{**}^2,\end{aligned}\tag{3.21}$$

and

$$\begin{aligned}\|\mathbf{y} - \mathbf{X}\tilde{\beta}\|_*^2 &= (\mathbf{y} - \mathbf{X}\tilde{\beta})'[\mathbf{I}_n - (1 - B)\mathbf{P}_X](\mathbf{y} - \mathbf{X}\tilde{\beta}) \\ &= \mathbf{y}'(\mathbf{I}_n - \mathbf{P}_X)'[\mathbf{I}_n - (1 - B)\mathbf{P}_X](\mathbf{I}_n - \mathbf{P}_X)\mathbf{y} \\ &= \mathbf{y}'(\mathbf{I}_n - \mathbf{P}_X)'(\mathbf{I}_n - \mathbf{P}_X)\mathbf{y} \\ &= \|\mathbf{y} - \mathbf{X}\tilde{\beta}\|^2.\end{aligned}\tag{3.22}$$

Now, $\hat{\beta} = \mathbf{P}_V \tilde{\beta}$, where $\mathbf{P}_V = \mathbf{I}_n - \mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}$. Here we assume that $\mathbf{h} = \mathbf{0}$ for the simplicity and it is clear that $\mathbf{P}_V$ is idempotent matrix and the ** orthogonal projection on the subspace R^q . Hence by the Pythagorean theorem, we have

$$\|\tilde{\beta} - \nu\|_{**}^2 = \|\tilde{\beta} - \hat{\beta}\|_{**}^2 + \|\hat{\beta} - \nu\|_{**}^2.\tag{3.23}$$

Note that

$$\|\tilde{\beta} - \hat{\beta}\|_{**}^2 = (\mathbf{H}\tilde{\beta} - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta} - \mathbf{h}).\tag{3.24}$$

Therefore, from (3.19)—(3.23), it follows that

$$\sigma^{-2} \|\mathbf{y} - \mathbf{X}\nu\|_*^2 = \sigma^{-2}(n - p)S_e^2 + (\sigma^2 + \tau^2)^{-1} \{ \|\tilde{\beta} - \hat{\beta}\|_{**}^2 + \|\hat{\beta} - \nu\|_{**}^2 \}.\tag{3.25}$$

Now, using the sufficiency and completeness result for the regular exponential family the theorem follows.

3.3.3 The Empirical Bayes Subspace Estimator

First, we observe that using (3.15)

$$E(\tilde{\beta}) = E(\mathbf{C}^{-1}\mathbf{X}'\mathbf{y}) = \nu,\tag{3.26}$$

and

$$\begin{aligned}
 \text{Var}(\tilde{\beta}) &= \text{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{Var}(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 &= (\sigma^2 + \tau^2)\mathbf{C}^{-1}.
 \end{aligned} \tag{3.27}$$

Now, we need to impose the following condition $\mathbf{H}\boldsymbol{\nu} = \mathbf{h}$. Under this condition, it follows that $E(\hat{\beta}) = \boldsymbol{\nu}$, and

$$(\mathbf{H}\tilde{\beta} - \mathbf{h}) \sim N\left(\mathbf{0}, (\sigma^2 + \tau^2)\mathbf{H}\mathbf{C}^{-1}\mathbf{H}'\right) \tag{3.28}$$

Therefore, it follows from (3.28),

$$\mathcal{L} = (\mathbf{H}\tilde{\beta} - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta} - \mathbf{h}) \sim (\sigma^2 + \tau^2)\chi_q^2, \tag{3.29}$$

accordingly, from Theorem 3.2 and (3.29) the UMVUE of $(\tau^2 + \sigma^2)^{-1}$ is $(q-2)\mathcal{L}^{-1}$. Also, from the above and Theorem 3.2 $\hat{\beta}$ is UMVUE of $\boldsymbol{\nu}$. Finally, the best scale invariant estimator of σ^2 is $\frac{(n-p)}{(n-p+2)}S_e^2$. Plugging the above estimator of $\boldsymbol{\nu}$, $(\sigma^2 + \tau^2)^{-1}$ and σ^2 in the Bayes estimator in (3.14), it follows that an empirical Bayes estimators of β denoted by $\hat{\beta}^{EB}$ is given by

$$\hat{\beta}^{EB} = \hat{\beta} + \left[1 - \frac{(n-p)(q-2)}{(n-p+2)}S_e^2\mathcal{L}^{-1}\right](\tilde{\beta} - \hat{\beta}). \tag{3.30}$$

Note that $L_n = \mathcal{L}/qS_e^2$. Hence the empirical Bayes estimator in (3.30) is equivalent to the shrinkage estimator in (3.6).

3.3.4 The Empirical Bayes Risk

In order to develop the empirical Bayes risk of the empirical Bayes estimators in (3.30), we need to state the following lemmas.

Lemma 3.1 (*Lemma 1. Ghosh, Saleh and Sen (1989)*)

If Z_1/Z_2 and Z_2 are independently distributed with $E|Z_1/Z_2| < \infty$, $E|Z_2| < \infty$ and $E|Z_2^{-1}| < \infty$, then

$$\begin{aligned}
 E(Z_1/Z_2) &= E(Z_1)/E(Z_2) \\
 E(Z_1/Z_2^2) &= E(Z_1)(E(Z_2))^{-1}E(Z_2^{-1})
 \end{aligned} \tag{3.31}$$

For the proof of this lemma see Ghosh, Saleh and Sen (1989).

Lemma 3.2 *For the marginal distribution of y ,*

$$\hat{\beta}, H\tilde{\beta} - h \quad \text{and} \quad y - X\tilde{\beta}$$

are mutually independent.

Proof. If we assume that $h = 0$ for simplicity, then one gets

$$\begin{aligned}\hat{\beta} &= P_V \tilde{\beta} = [C^{-1}X' - C^{-1}H'(HC^{-1}H')^{-1}HC^{-1}X'] y \\ H\tilde{\beta} &= HC^{-1}X'y.\end{aligned}$$

Since

$$[C^{-1}X' - C^{-1}H'(HC^{-1}H')^{-1}HC^{-1}X'] \Sigma X C^{-1}H' = 0, \quad (3.32)$$

by the Theorem 3.13a, page 51 of Arnold (1981), one gets $\hat{\beta}$ and $H\tilde{\beta} - h$ are independent. Next, we show that $H\tilde{\beta}$ and $y - X\tilde{\beta}$ are independent. Note that

$$\begin{aligned}H\tilde{\beta} &= HC^{-1}X'y \\ y - X\tilde{\beta} &= (I_n - P_X)y,\end{aligned}$$

by the same Theorem 3.13a, of Arnold (1981) and $X'P_X = X'$, we have

$$HC^{-1}X'\Sigma(I_n - P_X) = 0 \quad (3.33)$$

Finally, we show that $\hat{\beta}$ and $y - X\tilde{\beta}$ are independent, we observe that

$$[C^{-1}X' - C^{-1}H'(HC^{-1}H')^{-1}HC^{-1}X'] \Sigma(I_n - P_X) = 0 \quad (3.34)$$

Hence, the lemma follows from (3.32) – (3.34).

The next theorem provides the Bayes risk of the empirical Bayes estimator under the following quadratic loss function

$$L(\beta, a) = (a - \beta)'Q(a - \beta)/\sigma^2, \quad (3.35)$$

for every known positive definite matrix Q . Also, we assume that the prior (denoted by ξ) for β is given by

$$\xi \sim N(\nu, \tau^2 C^{-1}), \quad \text{where} \quad H\nu = h. \quad (3.36)$$

Theorem 3.3 Under the loss function in (3.95) and the prior in (3.96), the Bayes risk of $\hat{\beta}^{EB}$ is given by

$$\begin{aligned} R(\xi, \hat{\beta}^{EB}) &= \sigma^{-2} E \left[(\hat{\beta}^{EB} - \beta)' \mathbf{Q} (\hat{\beta}^{EB} - \beta) \right] \\ &= (1 - B) \text{Tr}(\mathbf{Q} \mathbf{C}^{-1}) + B \text{Tr} \left\{ \mathbf{Q} \mathbf{C}^{-1} \left[\mathbf{I}_n - \mathbf{H}'(\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1} \right] \right\} \\ &\quad + B \left(1 - \frac{(n-p)(q-2)}{q(n-p+2)} \right) \text{Tr}(\mathbf{W} \mathbf{H} \mathbf{C}^{-1} \mathbf{H}'), \end{aligned} \quad (3.37)$$

where $\mathbf{W} = (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{H}' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1}$ and $\mathbf{C}$ is defined as in (3.9).

Proof. Using a standard argument it follows that

$$R(\xi, \hat{\beta}^{EB}) = R(\xi, \hat{\beta}^B) + \sigma^{-2} E \left[(\hat{\beta}^{EB} - \hat{\beta}^B)' \mathbf{Q} (\hat{\beta}^{EB} - \hat{\beta}^B) \right]. \quad (3.38)$$

Note that $\hat{\beta}^B = E(\beta | \mathbf{y})$. Therefore,

$$\begin{aligned} R(\xi, \hat{\beta}^B) &= \sigma^{-2} E \left\{ E[(\hat{\beta}^B - \beta)' \mathbf{Q} (\hat{\beta}^B - \beta) | \mathbf{y}] \right\} \\ &= \sigma^{-2} E \{ \text{Tr} \mathbf{Q} \text{Var}(\beta | \mathbf{y}) \} \\ &= \sigma^{-2} E \text{Tr} \mathbf{Q} \left[\sigma^2 \tau^2 (\sigma^2 + \tau^2)^{-1} \right] \mathbf{C}^{-1} \\ &= (1 - B) \text{Tr}(\mathbf{Q} \mathbf{C}^{-1}). \end{aligned} \quad (3.39)$$

Now, one may use Theorem 3.1 with $\mathbf{V} = (\mathbf{X} \mathbf{X}')^{-1}$ and the equation (3.14) and (3.30) to get the identity

$$\hat{\beta}^{EB} - \hat{\beta}^B = B(\hat{\beta} - \nu) + \left(B - \frac{(n-p)(q-2)S_e^2}{(n-p+2)\mathcal{L}} \right) \mathbf{C}^{-1} \mathbf{H}' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} (\mathbf{H} \tilde{\beta} - \mathbf{h}), \quad (3.40)$$

where, $\mathcal{L}$ is defined as in (3.29). Since $S_e^2 = \frac{\|\mathbf{y} - \mathbf{X} \tilde{\beta}\|^2}{n-p}$, from Lemma 3.2, one gets the independence of β with $(S_e^2, \mathbf{H} \tilde{\beta} - \mathbf{h})$. Now, from (3.40), one gets

$$\begin{aligned} &E \left[(\hat{\beta}^{EB} - \hat{\beta}^B)' \mathbf{Q} (\hat{\beta}^{EB} - \hat{\beta}^B) \right] \\ &= B^2 E \left[(\hat{\beta} - \nu)' \mathbf{Q} (\hat{\beta} - \nu) \right] \\ &\quad + E \left[\left(B - \frac{(p-2)(q-2)S_e^2}{(n-p+2)\mathcal{L}} \right)^2 (\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) \right], \end{aligned} \quad (3.41)$$

where $\mathbf{W}$ is defined after (3.37). Next, $E[(\hat{\beta} - \nu)' \mathbf{Q}(\hat{\beta} - \nu)] = \text{Tr}(\mathbf{Q} \text{Var}(\hat{\beta}))$.

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var} [\tilde{\beta} - \mathbf{C}^{-1} \mathbf{H}'(\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} (\mathbf{H} \tilde{\beta} - \mathbf{h})] \\ &= \text{Var} [\{\mathbf{I}_n - \mathbf{C}^{-1} \mathbf{H}'(\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H}\} \tilde{\beta}] \\ &= (\sigma^2 + \tau^2) \mathbf{C}^{-1} [\mathbf{I}_n - \mathbf{H}'(\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1}]. \end{aligned}$$

Therefore,

$$E[(\hat{\beta} - \nu)' \mathbf{Q}(\hat{\beta} - \nu)] = \sigma^2 B^{-1} \text{Tr} [\mathbf{Q} \mathbf{C}^{-1} \{\mathbf{I}_n - \mathbf{H}'(\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1}\}]. \quad (3.42)$$

Also, using the independence of $\mathbf{H} \tilde{\beta} - \mathbf{h}$ and S_e^2 as well as $E(S_e^2) = \sigma^2$ and $E(S_e^4) = \frac{(n-p+2)}{(n-p)} \sigma^4$, we have

$$\begin{aligned} E \left[\left(B - \frac{(n-p)(q-2)S_e^2}{(n-p+2)\mathcal{L}} \right)^2 (\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) \right] \\ = B^2 E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h})] \\ - 2 \frac{B(n-p)(q-2)\sigma^2}{(n-p+2)} E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}] \\ + \frac{(q-2)^2(n-p)\sigma^4}{(n-p+2)} E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}^2]. \quad (3.43) \end{aligned}$$

Next, using (3.28), one gets

$$E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h})] = (\sigma^2 + \tau^2) \text{Tr}(\mathbf{W} \mathbf{H} \mathbf{C}^{-1} \mathbf{H}'). \quad (3.44)$$

Now, from Theorem 3.2, $\mathcal{L}$ is a function of the complete sufficient statistic. Also, it is easy to show that $(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}$ is ancillary. Hence from Basu's theorem, it follows that $(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}$ is distributed independently of $\mathcal{L}$. Now using Lemma 3.1, and (3.29) it follows after some simplifications that

$$E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}] = q^{-1} \text{Tr}(\mathbf{W} \mathbf{H} \mathbf{C}^{-1} \mathbf{H}') \quad (3.45)$$

$$E [(\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) / \mathcal{L}^2] = \frac{\text{Tr}(\mathbf{W} \mathbf{H} \mathbf{C}^{-1} \mathbf{H}')}{q(q-2)(\sigma^2 + \tau^2)}. \quad (3.46)$$

Now from (3.43)–(3.46), it follows that

$$\sigma^{-2} E \left[\left(B - \frac{(n-p)(q-2)S_e^2}{(n-p+2)\mathcal{L}} \right)^2 (\mathbf{H} \tilde{\beta} - \mathbf{h})' \mathbf{W} (\mathbf{H} \tilde{\beta} - \mathbf{h}) \right]$$

$$\begin{aligned}
&= BTr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}') - 2\frac{B(n-p)(q-2)}{q(n-p+2)}Tr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}') \\
&\quad + \frac{B(q-2)(n-p)}{q(n-p+2)}Tr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}') \quad (3.47)
\end{aligned}$$

The theorem follows from (3.38), (3.39), (3.41), (3.42) and (3.47).

The first term on the right hand side of (3.47) is the Bayes risk of the Bayes estimator under the chosen prior, while the sum of the second and third terms contribute to the risk of the empirical Bayes estimator due to its departure from the Bayes estimator.

Now, we compare the Bayes risk of $\hat{\beta}^{EB}$ with $\tilde{\beta}$ and $\hat{\beta}$. Under the loss function in (3.35) and the prior in (3.36), the respective Bayes risks of $\tilde{\beta}$ and $\hat{\beta}$ are given by

$$\begin{aligned}
R(\xi, \tilde{\beta}) &= (1-B)Tr(\mathbf{Q}\mathbf{C}^{-1}) + BTr\left\{\mathbf{Q}\mathbf{C}^{-1}[\mathbf{I}_n - \mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}]\right\} \\
&\quad + BTr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}'), \quad (3.48)
\end{aligned}$$

and

$$\begin{aligned}
R(\xi, \hat{\beta}) &= (1-B)Tr(\mathbf{Q}\mathbf{C}^{-1}) + BTr\left\{\mathbf{Q}\mathbf{C}^{-1}[\mathbf{I}_n - \mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}]\right\} \\
&\quad + (1-B)^2B^{-1}Tr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}'), \quad (3.49)
\end{aligned}$$

where $\mathbf{W}$ is defined as Theorem 3.3 and $\mathbf{C}$ as in (3.3). Hence from (3.37) and (3.48), we have

$$R(\xi, \tilde{\beta}) - R(\xi, \hat{\beta}^{EB}) = B\frac{(n-p)(q-2)}{(n-p+2)}Tr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}') > 0$$

so that $\hat{\beta}^{EB}$ has smaller Bayes risk than $\tilde{\beta}$. However from (3.37) and (3.49)

$$\begin{aligned}
R(\xi, \hat{\beta}) - R(\xi, \hat{\beta}^{EB}) &= \left\{(1-B)^2B^{-1} - B\left(1 - \frac{(n-p)(q-2)}{(n-p+2)q}\right)\right\} \times \\
&\quad Tr(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}') > 0 \quad (3.50)
\end{aligned}$$

$$\text{if and only if } (1-B)^2/B^2 > 1 - \frac{(n-p)(q-2)}{(n-p+2)q}.$$

It is interesting to notice that when the prior variance τ^2 is much smaller than the sampling variance σ^2 , i.e., B is close to 1, then $\hat{\beta}$ is expected to perform better than the empirical Bayes estimator under the prior in (3.36). But of course, if

$B \leq \frac{1}{2}$ if and only if $\sigma^2 \leq \tau^2$, i.e., when the prior variability is at least as large as the sampling variability, the empirical Bayes estimator has smaller risk than $\tilde{\beta}$. On the other hand, if the prior is believed to be true, the ideal estimator of β is the Bayes estimator $\hat{\beta}^B$. However, the empirical Bayes estimator is typically more robust than the Bayes estimator.

3.3.5 Minimax Estimation

The empirical Bayes estimator in (3.30) can be generalized to the class of the empirical Bayes estimator of the form

$$\hat{\beta}^k = \hat{\beta} + \left[1 - \frac{k(n-p)}{(n-p+2)} S_e^2 \mathcal{L}^{-1} \right] (\tilde{\beta} - \hat{\beta}), \quad (3.51)$$

where k is a constant. Note that for $k = q - 2$, $\hat{\beta}^{q-2} \equiv \hat{\beta}^{EB}$. Also, the Bayes risk in Theorem 3.3, under the quadratic loss function in (3.35) and the prior in (3.36) can be written as

$$\begin{aligned} R(\xi, \hat{\beta}^k) = & (1 - B) \text{Tr}(\mathbf{Q}\mathbf{C}^{-1}) + B \text{Tr} \left\{ \mathbf{Q}\mathbf{C}^{-1} [\mathbf{I}_n - \mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}] \right\} \\ & + B \left(1 - 2 \frac{k(n-p)}{q(n-p+2)} + \frac{k(q-2)(n-p)}{q(q-2)(n-p+2)} \right) \text{Tr}(\mathbf{W}\mathbf{H}\mathbf{C}^{-1}\mathbf{H}'), \end{aligned} \quad (3.52)$$

where $\mathbf{W}$ is given as in Theorem 3.3 and $\mathbf{C}$ as in (3.3). Hence the risk in (3.52) is minimized at $k = q - 2$. The constant $(q - 2)$ appear rather naturally in the estimator (3.30), but this constant $k = q - 2$ is optimal in the sense that minimizes the Bayes risk for every matrix $\mathbf{Q}$. Also, the empirical Bayes estimator in (3.30) can be written in the form

$$\hat{\beta}^G = \tilde{\beta} - \frac{\phi(F)}{F} (\tilde{\beta} - \hat{\beta}) \quad (3.53)$$

where

$$F = \mathcal{L} / \left(\frac{n-p}{n-p+2} S_e^2 \right), \quad \phi(F) = q - 2, \quad \text{for all } F.$$

We shall investigate conditions on ϕ under which the frequentist risk of $\hat{\beta}^G$ is less than the frequentist risk of $\tilde{\beta}$ under the loss function in (3.35). The next theorem provides this conditions and we will omit the proof of this theorem, since it is similar to the proof of Ghosh et al (1988).

Theorem 3.4 *Under the quadratic loss function in (3.95). If*

- (i) $2ch_{\max}(\mathbf{WHC}^{-1}\mathbf{H}') < \text{Tr}(\mathbf{WHC}^{-1}\mathbf{H}')$
- (ii) $0 < \phi(F) < 2 \left\{ \frac{\text{Tr}(\mathbf{WHC}^{-1}\mathbf{H}')}{ch_{\max}(\mathbf{WHC}^{-1}\mathbf{H}')} - 2 \right\}$
- (iii) $\phi(F) \uparrow F$.

Then

$$E \left[(\hat{\beta}^G - \beta)' \mathbf{Q} (\hat{\beta}^G - \beta) \right] < E \left[(\tilde{\beta} - \beta)' \mathbf{Q} (\tilde{\beta} - \beta) \right] \quad \text{for all } \beta \text{ and } \sigma^2.$$

where $ch_{\max}(A)$ = largest characteristic root of A .

For the empirical Bayes estimator if the condition (i). of Theorem 3.4 holds and

$$0 < q - 2 < 2 \left\{ \frac{\text{Tr}(\mathbf{WHC}^{-1}\mathbf{H}')}{ch_{\max}(\mathbf{WHC}^{-1}\mathbf{H}')} - 2 \right\},$$

then the empirical Bayes estimator dominates $\tilde{\beta}$ in a frequentist sense. In summary, the proposed EB estimator has several attractive features from both Bayesian and the frequentist standpoint. From a frequentist point of view, as discussed in the above, such an estimator is quite often minimax. From a Bayesian side, such an estimator dominates the ULSE for a wide range of priors. Finally, the EB estimator enjoys robustness against certain misspecified priors which a Bayes estimator under a normal prior, does not. In practice, it seems to be more appropriate to use the part positive empirical Bayes estimator.

3.4 Frequentist Biases and Risks

It is easy to note that the biases of the first four estimators are

$$E(\tilde{\beta} - \beta) = 0, \quad (3.54)$$

$$E(\hat{\beta} - \beta) = -\mathbf{C}^{-1}\mathbf{H}'(\mathbf{HC}^{-1}\mathbf{H}')^{-1}\delta; \quad \delta = \mathbf{H}\beta - \mathbf{h}, \quad (3.55)$$

$$E(\hat{\beta}^{Pr} - \beta) = -\mathbf{C}^{-1}\mathbf{H}'(\mathbf{HC}^{-1}\mathbf{H}')^{-1}\delta G_{q+2, n-p}(I_{\alpha}^*; \Delta), \quad (3.56)$$

$$E(\hat{\beta}^S - \beta) = -d_q \mathbf{C}^{-1}\mathbf{H}'(\mathbf{HC}^{-1}\mathbf{H}')^{-1}\delta E(\chi_{q+2}^{-2}(\Delta)), \quad (3.57)$$

where $\chi_{q+2}^2(\Delta)$ denotes a non-central chi-square variable with $q+2$ DF and non-centrality parameter $\Delta = \delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta/\sigma^2$, $l_{\alpha}^* = \frac{q}{q+2}F_{\alpha}$ and $G_{m,n}(\cdot; \Delta)$ is cumulative F-distribution function with (m, n) DF and non-centrality parameter Δ .

For the risks of the estimators, we consider quadratic loss function in (3.35), where $\mathbf{a}$ can be any one of the four estimators and $\mathbf{Q}$ is a known positive definite matrix. Thus we have the risk of $\mathbf{a}$ as

$$R(\mathbf{a}, \mathbf{Q}) = \sigma^{-2} E(\mathbf{a} - \beta)' \mathbf{Q} (\mathbf{a} - \beta) \quad (3.58)$$

Direct computation and following Judge and Bock (1978) lead to the following results

$$R(\tilde{\beta}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{Q}), \quad (3.59)$$

$$R(\hat{\beta}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{Q}) - \text{Tr}(\mathbf{A}_1) + \sigma^{-2} \eta_1' \mathbf{A}_1 \eta_1, \quad (3.60)$$

$$R(\hat{\beta}^{PT}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{Q}) - \text{Tr}(\mathbf{A}_1) G_{q+2, n-p}(l_{\alpha}^*; \Delta) + \sigma^{-2} \eta_1' \mathbf{A}_1 \eta_1 \{2G_{q+2, n-p}(l_{\alpha}^*; \Delta) - G_{q+4, n-p}(l_{\alpha}^{**}; \Delta)\}, \quad (3.61)$$

$$R(\hat{\beta}^S, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{Q}) - dq \text{Tr}(\mathbf{A}_1) \left\{ (q-2) E(\chi_{q+2}^{-4}(\Delta)) + \left[1 - \frac{(q+2)\eta_1' \mathbf{A}_1 \eta_1}{2\sigma^2 \Delta \text{Tr}(\mathbf{A}_1)} \right] 2\Delta E(\chi_{q+4}^{-4}(\Delta)) \right\}, \quad (3.62)$$

where $l_{\alpha}^{**} = \frac{q}{q+4}F_{\alpha}$ and

$$\begin{aligned} \eta_1' \mathbf{A}_1 \eta_1 &= \delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1} \mathbf{H}\mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1} \delta \\ \text{Tr}(\mathbf{A}_1) &= \text{Tr} \{ \mathbf{H}\mathbf{C}^{-1} \mathbf{Q} \mathbf{C}^{-1} \mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1} \}. \end{aligned} \quad (3.63)$$

To obtain the risk of $\hat{\beta}^S$, we have

$$\begin{aligned} R(\hat{\beta}^S, \mathbf{Q}) &= \sigma^{-2} E(\hat{\beta}^S - \beta)' \mathbf{Q} (\hat{\beta}^S - \beta) \\ &= \text{Tr}(\mathbf{C}^{-1} \mathbf{Q}) - 2d\sigma^{-2} E[L_n^{-1}(\tilde{\beta} - \beta)' \mathbf{Q} (\tilde{\beta} - \hat{\beta})] \\ &\quad + d^2\sigma^{-2} [L_n^{-2}(\tilde{\beta} - \hat{\beta})' \mathbf{Q} (\tilde{\beta} - \hat{\beta})]. \end{aligned}$$

There exists an orthogonal matrix $\mathbf{V}$ (see Judge and Bock (1978)) such that

$$\mathbf{V}\mathbf{C}^{-\frac{1}{2}}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-\frac{1}{2}}\mathbf{V}' = \begin{bmatrix} \mathbf{I}_q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{A} = \mathbf{V}\mathbf{C}^{-\frac{1}{2}}\mathbf{Q}\mathbf{C}^{-\frac{1}{2}}\mathbf{V}' = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_3 \\ \mathbf{A}_3' & \mathbf{A}_2 \end{bmatrix}. \quad (3.64)$$

The matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are of order q and $p - q$ respectively. Now we define the random variable

$$\omega = \mathbf{V}\mathbf{C}^{-\frac{1}{2}}\tilde{\beta} - \mathbf{V}\mathbf{C}^{-\frac{1}{2}}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{h}.$$

So $\omega \sim N(\eta, \sigma^2 \mathbf{I}_p)$, where

$$\eta = \mathbf{V}\mathbf{C}^{-\frac{1}{2}}\beta - \mathbf{V}\mathbf{C}^{-\frac{1}{2}}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{h}.$$

Partitioning the vectors $\omega = (\omega_1', \omega_2')'$ and $\eta = (\eta_1', \eta_2')'$ where ω_1' and ω_2' are independent sub-vector of order q and $p - q$ respectively. We obtain

$$\tilde{\beta} - \beta = \mathbf{C}^{-\frac{1}{2}}\mathbf{V}'(\omega - \eta),$$

and

$$L_n = \frac{\omega_1' \omega_1}{q S_e^2}, \quad \Delta = \frac{\eta_1' \eta_1}{\sigma^2}.$$

Thus applying the results from Judge and Bock (1978) Appendix B, as well as $E(S_e^2) = \sigma^2$ and $E(S_e^4) = \frac{(n-p+2)}{(n-p)}\sigma^4$ (see exercise 3 page 175 of Arnold (1981)), we have

$$\begin{aligned} R(\hat{\beta}^S, \mathbf{Q}) &= \text{Tr}(\mathbf{C}^{-1}\mathbf{Q}) \\ &\quad - 2d\sigma^{-2}E\left\{\frac{qS_e^2}{\omega_1' \omega_1} (\omega_1' \mathbf{A}_1 \omega_1 - \eta_1' \mathbf{A}_1 \omega_1 + \omega_2' \mathbf{A}_3' \omega_1 - \eta_2' \mathbf{A}_3' \omega_1)\right\} \\ &\quad + d^2\sigma^{-2}E\left\{\frac{q^2 S_e^4}{(\omega_1' \omega_1)} (\omega_1' \mathbf{A}_1 \omega_1)\right\} \\ &= \text{Tr}(\mathbf{C}^{-1}\mathbf{Q}) - dq\text{Tr}(\mathbf{A}_1) \{2E(\chi_{q+2}^{-2}(\Delta)) - (q-2)E(\chi_{q+2}^{-4}(\Delta))\} \\ &\quad + \sigma^{-2}\eta_1' \mathbf{A}_1 \eta_1 dq \{(q-2)E(\chi_{q+4}^{-4}(\Delta)) \\ &\quad + 2E(\chi_{q+2}^{-2}(\Delta)) - 2E(\chi_{q+4}^{-2}(\Delta))\} \end{aligned} \quad (3.65)$$

Now, we use the following identities to obtain the final form of the risk $\hat{\beta}^S$.

$$\begin{aligned} E(\chi_{p+2}^{-2}(\Delta)) - E(\chi_{p+2}^{-2}(\Delta)) &= 2E(\chi_{p+2}^{-4}(\Delta)) \quad \forall p > 2 \\ E(\chi_{p+2}^{-2}(\Delta)) - (p-2)E(\chi_{p+2}^{-4}(\Delta)) &= \Delta E(\chi_{p+4}^{-4}(\Delta)) \end{aligned} \quad (3.66)$$

Then, (3.65) reduces to the form in (3.62).

3.5 Frequentist Risk Analysis

In this section we provide the risk analysis of the estimators with the general loss function and give conditions on the $\mathbf{Q}$ matrix so that the shrinkage estimator dominates the usual estimator. Firstly, we note that

$$R(\hat{\beta}, \mathbf{Q})/R(\tilde{\beta}, \mathbf{Q}) \stackrel{\leq}{\geq} 1 \quad \text{according as} \quad \sigma^{-2}(\eta'_1 \mathbf{A}_1 \eta_1) \stackrel{\leq}{\geq} \text{Tr}(\mathbf{A}_1). \quad (3.67)$$

Secondly,

$$\begin{aligned} R(\hat{\beta}^{PT}, \mathbf{Q})/R(\tilde{\beta}, \mathbf{Q}) &\stackrel{\leq}{\geq} 1 \quad \text{according as} \\ \sigma^{-2}(\eta'_1 \mathbf{A}_1 \eta_1) \{2G_{q+2, n-p}(I^*; \Delta) - G_{q+4, n-p}(I^{**}; \Delta)\} &\stackrel{\leq}{\geq} \\ G_{q+2, n-p}(I^*; \Delta) \text{Tr}(\mathbf{A}_1), & \end{aligned} \quad (3.68)$$

and

$$\begin{aligned} R(\hat{\beta}^{PT}, \mathbf{Q})/R(\hat{\beta}, \mathbf{Q}) &\stackrel{\leq}{\geq} 1 \quad \text{according as} \\ \text{Tr}(\mathbf{A}_1) [1 - G_{q+2, n-p}(I^*; \Delta)] &\stackrel{\leq}{\geq} \\ \sigma^{-2}(\eta'_1 \mathbf{A}_1 \eta_1) [1 - \{2G_{q+2, n-p}(I^*; \Delta) - G_{q+4, n-p}(I^{**}; \Delta)\}] & \end{aligned} \quad (3.69)$$

Finally, we get from Courant theorem [see e.g., Rao (1973)]

$$\frac{\eta'_1 \mathbf{A}_1 \eta_1}{\eta'_1 \eta_1} = \frac{\delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}\mathbf{Q}\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta}{\delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta} \leq \lambda_1 \quad \text{for all } \delta \quad (3.70)$$

where λ_1 denotes the largest characteristic root of

$$\mathbf{B}^* = (\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}\mathbf{Q}\mathbf{C}^{-1}\mathbf{H}',$$

and from (3.63), we have $\text{Tr}(\mathbf{A}_1) = \text{Tr}(\mathbf{B}^*)$. Define now the class of positive semi-definite matrices $\mathcal{W}$ as

$$\mathcal{W} = \left\{ \mathbf{Q} : \frac{\text{Tr}(\mathbf{A}_1)}{\lambda_1} \geq \frac{q+2}{2} \right\}. \quad (3.71)$$

Then, we obtain from (3.59), (3.62) and (3.71) the result:

$$\frac{R(\hat{\beta}^S, \mathbf{Q})}{R(\tilde{\beta}, \mathbf{Q})} \leq 1 \quad \forall \eta_1 \quad \text{and} \quad \mathbf{Q} \in \mathcal{W}. \quad (3.72)$$

The proof of (3.72) follows by examining the following term in (3.62).

$$1 - \frac{(q+2)\eta'_1 A_1 \eta_1}{2\sigma^2 \Delta \text{Tr}(A_1)},$$

which must be greater than or equal to zero.

In particular if $Q = C$, we have $\lambda_1 = 1$ and rank of A_1 is q . Thus, Q belongs to $\mathcal{W}$. For arbitrary choice of Q , the inequality (3.71) may not hold simultaneously for all Q . In such a case, to include the dominance of $\hat{\beta}^S$ over $\tilde{\beta}$, one may introduce the modified Stein-rule estimator $\hat{\beta}^{MS}$ given by

$$\hat{\beta}^{MS} = \hat{\beta} + [I_p - dk_n L_n^{-1} Q^{-1} C] (\tilde{\beta} - \hat{\beta}), \quad (3.73)$$

where d is the shrinkage constant and k_n denotes the smallest characteristic root of $Q^{-1}C$. The modified Stein-rule estimator is similar to the normal theory case treated by Berger et al (1977) and it is not pursued here. We are content to assume (3.71) in our discussion in the sequel.

In (3.67) $\text{Tr}(A_1) > 0$, and hence (3.67) implies that only for η_1 close to 0, $\hat{\beta}$ performs better than $\tilde{\beta}$, i.e., $\hat{\beta}$ performs better than $\tilde{\beta}$ under H_0 . However, when η_1 deviates from 0, the risk of $\hat{\beta}$ grows and becomes unbounded while the risk of $\tilde{\beta}$ remains bounded. Hence departure from restriction of β on the planes $H\beta = h$ is fatal to $\hat{\beta}$, but is of least concern for $\tilde{\beta}$. Similarly, note that

$$G_{q+2,n-p}(l^*; \Delta) < G_{q,n-p}(l^*; \Delta) \leq G_{q,n-p}(l^*; 0) = 1 - \alpha, \quad (3.74)$$

for every $q \geq 3$, $0 < \alpha < 1$ and $\Delta > 0$, and that the left-hand-side converges to 0 as $\Delta \rightarrow \infty$ (i.e., η_1 moves away from 0). Also

$$G_{q+2,n-p}(l^*; \Delta) > G_{q+4,n-p}(l^{**}; \Delta), \quad \forall \alpha \in (0, 1), \quad q \geq 3 \quad \text{and} \quad \Delta > 0. \quad (3.75)$$

As a result, from (3.68), we conclude that the risk of the PTE is smaller than that of the risk of the ULSE in the neighborhood of η_1 . As η_1 moves away from 0, the opposite conclusion holds. Thus, the three estimators ULSE, RLSE and PTE is inadmissible with respect to the other two. The picture is somewhat different with respect to the shrinkage estimator $\hat{\beta}^S$.

Chapter 4

Preliminary Test and Empirical Bayes Estimators of Finite Population Mean

4.1 Introduction

In this chapter we study the problem of estimating a finite population mean when information on some relevant auxiliary variables is available but it is suspected that the auxiliary variables may have subspace restrictive contribution. We follow the developments of Chapter 3 and discuss four estimators, namely, the unrestricted estimator (UE), restricted estimator (RE), preliminary test estimator (PTE) and finally, shrinkage estimator (SE) of the finite population mean. Further, we show that the proposed SE is the empirical Bayes estimator (EBE) under a certain prior distribution. The comparisons show that neither the PTE nor the EBE dominates each other while they perform well relative to the UE and RE under the frequentist risk. Moreover, it is shown that the proposed EBE has smaller Bayes risk than that of the UE when the dimension of restrictive subspace exceeds three.

of/de

2

First, we note that the risk of $\hat{\beta}^S$ is uniformly smaller than the risk of $\tilde{\beta}$ under (3.71) and attains the upper limit $R(\tilde{\beta}, \mathbf{Q})$ as $\Delta \rightarrow \infty$. This shows inadmissibility of $\tilde{\beta}$ relative to $\hat{\beta}^S$. To compare $\hat{\beta}$ and $\hat{\beta}^S$, we may write

$$R(\hat{\beta}^S, \mathbf{Q}) = R(\hat{\beta}, \mathbf{Q}) + \text{Tr}(\mathbf{A}_1) - \sigma^{-2} \eta_1' \mathbf{A}_1 \eta_1 - dq \text{Tr}(\mathbf{A}_1) \times \\ \left\{ (q-2)E(\chi_{q+2}^{-4}(\Delta)) + \left[1 - \frac{(q+2)\eta_1' \mathbf{A}_1 \eta_1}{2\sigma^2 \Delta \text{Tr}(\mathbf{A}_1)} \right] 2\Delta E(\chi_{q+4}^{-4}(\Delta)) \right\}.$$

Under H_0 , this turns out to be

$$R(\hat{\beta}^S, \mathbf{Q}) = R(\hat{\beta}, \mathbf{Q}) + \frac{2(n-p+q)}{q(n-p+2)} \text{Tr}(\mathbf{A}_1) > R(\hat{\beta}, \mathbf{Q}). \quad (3.76)$$

Hence, the risk of $\hat{\beta}^S$ under H_0 is greater than that of the risk of $\hat{\beta}$. However, as η_1 moves away from 0, $\eta_1' \mathbf{A}_1 \eta_1$ increases while $E(\chi_{q+2}^{-4}(\Delta))$ and $E(\chi_{q+4}^{-4}(\Delta))$ decrease as do $G_{q+2, n-p}(l_\alpha^*; \Delta)$ and $G_{q+4, n-p}(l_\alpha^{**}; \Delta)$, so that reverse inequality holds in (3.76). Thus $\hat{\beta}^S$ does not dominate $\hat{\beta}$ near $\Delta = 0$. However, outside a small interval around $\eta_1 = 0$, $\hat{\beta}^S$ dominates $\hat{\beta}$. Hence, we recommend use of $\hat{\beta}^S$ over $\hat{\beta}$.

Now, we compare $\hat{\beta}^S$ and $\hat{\beta}^{PT}$. First, consider the case when H_0 holds i.e., $\eta_1 = 0$. In this case

$$R(\hat{\beta}^S, \mathbf{Q}) = R(\hat{\beta}^{PT}, \mathbf{Q}) + \text{Tr}(\mathbf{A}_1) \{G_{q+2, n-p}(l_\alpha^*; 0) - d\} \\ \geq R(\hat{\beta}^{PT}, \mathbf{Q}), \quad (3.77)$$

for all α such that l_α^* satisfies the inequality

$$l_\alpha^* > G_{q+2, n-p}^{-1}(d). \quad (3.78)$$

The risk of $\hat{\beta}^S$ is smaller than that of the risk of $\hat{\beta}^{PT}$ when the critical value l_α^* satisfies the opposite inequality in (3.78). Thus, $\hat{\beta}^S$ does not always dominate $\hat{\beta}^{PT}$ when the null-hypothesis H_0 holds. Thus, we may order the dominance picture under (3.78) as follows:

$$\hat{\beta} \succ \hat{\beta}^{PT} \succ \hat{\beta}^S \succ \tilde{\beta}. \quad (3.79)$$

The picture changes as η_1 moves away from 0. First, note that $\tilde{\beta}$ has the constant risk $\text{Tr}(\mathbf{C}^{-1}\mathbf{Q})$ while the risk of $\hat{\beta}$ depends on η_1 . Consequently, the risk of $\hat{\beta}$

becomes unbounded as η_1 moves away from 0. Also, as $\Delta \rightarrow \infty$ the risk of $\hat{\beta}^{PT}$ approaches $Tr(C^{-1}Q)$ which is the risk of $\tilde{\beta}$. Also, as η_1 moves away from 0, the risk of $\hat{\beta}^{PT}$ increases to a maximum after crossing the risk of $\tilde{\beta}$, then decreases to the minimum $Tr(C^{-1}Q)$. Hence, the risk of $\hat{\beta}^{PT}$ never blows up. Further, (3.78) specifies a range of α -values for which $\hat{\beta}^{PT}$ dominates $\hat{\beta}^S$ for $q \geq 3$. Thus, neither $\hat{\beta}^{PT}$ nor $\hat{\beta}^S$ dominates the other except that they share common property that as $\Delta \rightarrow \infty$ the risk of both becomes $Tr(C^{-1}Q)$. However, the risk of $\hat{\beta}^S$ is below the risk of $\tilde{\beta}$ while the risk of $\hat{\beta}^{PT}$ exceeds the risk of $\tilde{\beta}$ at some intermediate values of Δ depends on α .

Finally, we note that the application of $\hat{\beta}^S$ is constrained by the requirement $q \geq 3$ while $\hat{\beta}^{PT}$ does not but the use of $\hat{\beta}^{PT}$ is limited by the size of α which is given by (3.78) and the extent of departure of Δ from 0. Thus, from the point of robust efficiency both PTE and SE may be advocated leaning towards $\hat{\beta}^S$ since the size of Δ is generally unknown.

Chapter 4

Preliminary Test and Empirical Bayes Estimators of Finite Population Mean

4.1 Introduction

In this chapter we study the problem of estimating a finite population mean when information on some relevant auxiliary variables is available but it is suspected that the auxiliary variables may have subspace restrictive contribution. We follow the developments of Chapter 3 and discuss four estimators, namely, the unrestricted estimator (UE), restricted estimator (RE), preliminary test estimator (PTE) and finally, shrinkage estimator (SE) of the finite population mean. Further, we show that the proposed SE is the empirical Bayes estimator (EBE) under a certain prior distribution. The comparisons show that neither the PTE nor the EBE dominates each other while they perform well relative to the UE and RE under the frequentist risk. Moreover, it is shown that the proposed EBE has smaller Bayes risk than that of the UE when the dimension of restrictive subspace exceeds three.

Consider a finite population $\mathcal{U}$ with units labeled as $1, 2, \dots, N$. Let y be a characteristic of interest and y_i be the value of Y for the i^{th} unit in the population ($i = 1, 2, \dots, N$). The purpose of this chapter is to propose some estimators of the finite population mean

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i.$$

Quite often one has information about some auxiliary variables apart from the main characteristic Y . One can then use these auxiliary variables to improve on the estimation of the mean μ . Let $Y = (y_1, y_2, \dots, y_N)'$ and let X be an $N \times p$ matrix of auxiliary information such that $E(Y) = X\beta$, where $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ be weights of the auxiliary information given by X on Y . The X is naturally assumed to be known, while Y is completely unknown unless we sample the population $\mathcal{U}$. In addition, we assume that Y is a random sample of size N from a N -dimensional normal distribution with mean $X\beta$ and covariance matrix $\sigma^2 I_N$. A subset s of $\mathcal{U}$ is a sample from a finite population. Let S denote all possible samples from $\mathcal{U}$ of fixed size $< N$, then $P(s) \geq 0$ for all $s \in S$ and $\sum_{s \in S} P(s) = 1$, where $P(s)$ is the probability of selecting the sample s from S . Suppose $s = \{1, 2, \dots, n\}$ and $y_n = (y_1, y_2, \dots, y_n)'$ be the observed values on the characteristic Y from $\mathcal{U}$.

We wish to estimate μ under a quadratic loss when it suspected but not sure that $H\beta = h$ in the superpopulation, where H is a given $q \times p$ matrix of rank q and h is a given $q \times 1$ column vector. The estimation of μ is primarily based on a frequentist approach, namely, the preliminary test approach using the test-statistic for testing the null-hypothesis

$$H_0 : H\beta = h \quad \text{against} \quad H_A : H\beta \neq h.$$

In this study, we also consider the empirical Bayes approach with the superpopulation model under the general restriction, $H\beta = h$.

Preliminary test (PT) estimators together with the Bayes and empirical Bayes estimators are given in Section 4.2. The prediction problem is considered in Section 4.3. In Section 4.4, we present the exact mean square errors of the estimators based on PT approach and in Section 4.5, we compare them. Section 4.6, contains

the risk analysis of the Bayes and empirical Bayes estimators in terms of relative savings loss (RSL).

4.2 The Proposed Estimators

4.2.1 Preliminary Test Approach

Let $y_n = (y_1, y_2, \dots, y_n)'$ be a random sample from Y and X be an $N \times p$ matrix of auxiliary variables which we partition as X_1 and X_2 with dimension $n \times p$ and $(N - n) \times p$ respectively. Then, we may estimate the parameters β using X_1 matrix and predict the complementary vector $y_{N-n} = (y_{n+1}, \dots, y_N)'$ using the matrix X_2 . A convex combination of the mean of the components of y_n and the mean of the predicted y_{N-n} components gives the desired estimate of the population mean

$$\mu = (1 - f)\bar{y}_n + f\bar{y}_c,$$

where

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{y}_c = \frac{1}{N-n} \sum_{i=n+1}^N y_i \quad \text{and} \quad f = \frac{N-n}{N},$$

the finite population correction. Let $\tilde{\mu}_n$ be the unrestricted estimator (UE) while $\hat{\mu}_n$ be the restricted estimator (RE) of μ under $H_0 : H\beta = h$. $\hat{\mu}_n$ performs well when H_0 holds but a departure from H_0 plays havoc to $\hat{\mu}_n$ while $\tilde{\mu}_n$ performs well even under departure from H_0 . The expressions for $\tilde{\mu}_n$ and $\hat{\mu}_n$ are given by

$$\tilde{\mu}_n = (1 - f)\bar{y}_n + f\mathbf{a}'\tilde{\beta}_n, \quad (4.1)$$

$$\hat{\mu}_n = (1 - f)\bar{y}_n + f\mathbf{a}'\hat{\beta}_n, \quad (4.2)$$

where

$$f = \frac{N-n}{N}, \quad \mathbf{a}' = \frac{1}{N-n} \mathbf{1}'_{N-n} X_2, \quad \mathbf{1}'_{N-n} = (1, 1, \dots, 1),$$

and

$$\tilde{\beta}_n = (X_1' X_1)^{-1} X_1' y_n = C^{-1} X_1' y_n, \quad C = X_1' X_1,$$

$$\hat{\beta}_n = \tilde{\beta}_n - C^{-1} H' (H C^{-1} H')^{-1} (H \tilde{\beta}_n - h).$$

To propose a compromise estimator of μ based on $\tilde{\mu}_n$ and $\hat{\mu}_n$ we first test the null hypothesis

$$H_0 : \mathbf{H}\beta = \mathbf{h} \quad \text{against} \quad H_A : \mathbf{H}\beta \neq \mathbf{h}.$$

The likelihood ratio test-statistic for H_0 is given by

$$L_n = (\mathbf{H}\tilde{\beta}_n - \mathbf{h})'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\beta}_n - \mathbf{h})/qS_c^2,$$

where $S_c^2 = (n - p)^{-1} \parallel \mathbf{y}_n - \mathbf{X}_1 \tilde{\beta}_n \parallel^2$. The test statistic L_n follows a central F-distribution under H_0 with $(q, n - p)$ degrees of freedom (DF). Let F_α be the upper α percentile of the central F-distribution, the preliminary test estimator (PTE) is defined as

$$\hat{\mu}_n^{PT} = (1 - f)\bar{y}_n + f\mathbf{a}'\hat{\beta}_n^{PT}, \quad (4.3)$$

where

$$\hat{\beta}_n^{PT} = \hat{\beta}_n + [1 - I(L_n \leq F_\alpha)](\tilde{\beta}_n - \hat{\beta}_n),$$

where $I(A)$ is the indicator of the set A . It is clear that the $\hat{\mu}_n^{PT}$ is a convex combination of $\tilde{\mu}_n$ and $\hat{\mu}_n$. Due to the discontinuity of $I(L_n \leq F_\alpha)$, the performance of $\hat{\mu}_n^{PT}$ may not be so steady as $\tilde{\mu}_n$, hence we may consider smoother combination of $\tilde{\mu}_n$ and $\hat{\mu}_n$ given by the shrinkage estimator (SE)

$$\hat{\mu}_n^S = (1 - f)\bar{y}_n + f\mathbf{a}'\hat{\beta}_n^S, \quad (4.4)$$

with

$$\hat{\beta}_n^S = \hat{\beta}_n + \left[1 - \frac{(q - 2)(n - p)}{q(n - p + 2)}L_n^{-1}\right](\tilde{\beta}_n - \hat{\beta}_n),$$

or, we may consider the estimator given by

$$\hat{\mu}_n^{S+} = (1 - f)\bar{y}_n + f\mathbf{a}'\hat{\beta}_n^{S+}, \quad (4.5)$$

with

$$\hat{\beta}_n^{S+} = \hat{\beta}_n + \left[1 - \frac{(q - 2)(n - p)}{q(n - p + 2)}L_n^{-1}\right]I\left(L_n \geq \frac{(q - 2)(n - p)}{q(n - p + 2)}\right)(\tilde{\beta}_n - \hat{\beta}_n).$$

Note that $\hat{\mu}_n^{S+}$ is a convex combination of $\tilde{\beta}_n$ and $\hat{\beta}_n$, while $\hat{\mu}_n^S$ is not. Thus, $\hat{\mu}_n^{S+}$ controls over-shrinking of $\hat{\mu}_n^S$. Further, we observe that as $L_n \rightarrow \infty$, $\hat{\mu}_n^{S+} \rightarrow \tilde{\mu}_n$, while as $L_n \rightarrow \frac{(q-2)(n-p)}{q(n-p+2)}$, $\hat{\mu}_n^{S+} \rightarrow \hat{\mu}_n$.

4.2.2 The Bayesian Approach

Now, we show that (4.4) is an empirical Bayes estimator (EBE) of μ in the following way. Consider the model

$$Y|\beta \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_N),$$

that is $Y|\beta$ distributed normally with mean $\mathbf{X}\beta$ and covariance matrix $\sigma^2 \mathbf{I}_N$, $\sigma^2 > 0$ is unknown and $\mathbf{I}_N$ denotes the identity matrix of order N , this is known as the superpopulation. Let $\mathbf{X}$ be an $N \times p$ matrix of known fixed numbers which we partition as $\mathbf{X}_1$ and $\mathbf{X}_2$ with dimension $n \times p$ and $(N - p) \times p$ respectively. Note that if $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})$ then $\mathbf{X}_1$ can be written as $\mathbf{X}_1 = (\mathbf{x}_1 : \dots : \mathbf{x}_n)'$, $\mathbf{X}_2 = (\mathbf{x}_{n+1} : \dots : \mathbf{x}_N)'$ and $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]'$. Also, let β be a $p \times 1$ vector of unknown parameters. Further, we assume that the components of Y are independent and the prior $\beta \sim N(\nu, \tau^2 \mathbf{V})$, where ν is unknown a $p \times 1$ vector and $\mathbf{V}$ is a $p \times p$ matrix. This is the prior on the superpopulation parameter β . Under the above model and the square error loss function, namely

$$L(b, \mu) = (b - \mu)^2, \quad (4.6)$$

the Bayes estimator of μ is given by

$$\begin{aligned} \hat{\mu}_n^B = E(\mu|s, \mathbf{y}_n) &= N^{-1} \left[\sum_{i=1}^n y_i + \sum_{i=n+1}^N E(y_i|s, \mathbf{y}_n) \right] \\ &= N^{-1} \left[n\bar{y}_n + \sum_{i=n+1}^N E\{E(y_i|\beta, s, \mathbf{y}_n)|s, \mathbf{y}_n\} \right] \\ &= N^{-1} \left[n\bar{y}_n + \sum_{i=n+1}^N \mathbf{x}_i E(\beta|s, \mathbf{y}_n) \right] \\ &= (1-f)\bar{y}_n + f\mathbf{a}' E(\beta|s, \mathbf{y}_n) \\ &= (1-f)\bar{y}_n + f\mathbf{a}' \hat{\beta}_n^B, \end{aligned} \quad (4.7)$$

where $\mathbf{a}'$ is defined as in (4.2). Note that $\mathbf{a}'$ can be written as $\mathbf{a}' = (a_1, \dots, a_p)$, $a_j = \frac{1}{N-n} \sum_{i=n+1}^N x_{ij}$, $j = 1, \dots, p$. Also, as in Chapter 3,

$$\hat{\beta}_n^B = \nu + (\mathbf{X}_1' \mathbf{X}_1 + \frac{\sigma^2}{\tau^2} \mathbf{V}^{-1})^{-1} \mathbf{X}_1' (\mathbf{y}_n - \mathbf{X}_1 \nu). \quad (4.8)$$

For the proof of this see Chapter 3. In this chapter, we choose $V = (X_1' X_1)^{-1}$ as given in Ghosh et al (1989). Also, this choice of the prior was used by Zellner (1986) in econometric modeling, [see also Arnold (1981)]. We assume the prior (denoted by ξ) for β is given by

$$\xi \sim N(\nu, \tau^2 C^{-1}), \quad \text{where } H\nu = h. \quad (4.9)$$

Hence, under the model ξ , the expression in (4.8) simplifies to

$$\hat{\beta}_n^B = \nu + (1 - B)(\tilde{\beta}_n - \nu), \quad (4.10)$$

where $B = \frac{\sigma^2}{\sigma^2 + \tau^2}$ and $\tilde{\beta}_n$ is defined by (4.1). Therefore, the expression in (4.7) simplifies to

$$\hat{\mu}_n^B = (1 - f)\bar{y}_n + f a' [\nu + (1 - B)(\tilde{\beta}_n - \nu)]. \quad (4.11)$$

Note that when τ^2 is large compared to σ^2 , i.e., the case when we have little knowledge about the prior, the Bayes estimator $\hat{\mu}_n^B$ tends to the $\tilde{\mu}_n$. On the other hand, when the prior information outweighs the information contained in the data, i.e., τ^2 is small compared to σ^2 , $\hat{\mu}_n^B$ tends to $\hat{\mu}_n$.

For the empirical Bayes estimator, the unknown parameters ν , τ^2 , and σ^2 are needed to be estimated from the marginal distribution of y_n . One may consider the following class of estimators of μ , namely,

$$\hat{\mu}_n^K = (1 - f)\bar{y}_n + f a' \hat{\beta}_n^K, \quad (4.12)$$

where

$$\hat{\beta}_n^k = \hat{\beta}_n + \left[1 - \frac{k(n-p)}{q(n-p+2)} L_n^{-1} \right] (\tilde{\beta}_n - \hat{\beta}_n),$$

and k is a non-negative constant. Note that for $k = 0$ (i.e., $B = 0$), one gets the unrestricted estimators $\tilde{\mu}_n$ as in (4.1). Now we propose the following empirical Bayes estimators (EBE) of μ

$$\hat{\mu}_n^{EB} = (1 - f)\bar{y}_n + f a' \hat{\beta}_n^{EB}, \quad (4.13)$$

where

$$\hat{\beta}_n^{EB} = \hat{\beta}_n + \left[1 - \frac{(n-p)(q-2)}{q(n-p+2)} L_n^{-1} \right] (\tilde{\beta}_n - \hat{\beta}_n).$$

Note that the development of the Bayes and empirical Bayes estimators of β under these models can be found in Chapter 3. Also, the the constant $k = q - 2$ in (4.13) appear naturally, but this constant is optimal in the sense the Bayes risk of the $\hat{\mu}_n^K$ is minimized at $k = q - 2$. This will be proved in Section 4.6.

The positive part version of $\hat{\mu}_n^{EB}$ is given by

$$\hat{\mu}_n^{EB+} = (1 - f)\bar{y}_n + f\mathbf{a}' \left\{ \hat{\beta}_n + \left[1 - \frac{(n-p)(q-2)}{q(n-p+2)} L_n^{-1} \right]^+ (\tilde{\beta}_n - \hat{\beta}_n) \right\}, \quad (4.14)$$

where $a^+ = \max(a, 0)$. Such positive part estimators prevent overshrinking as does the positive part of James-Stein estimator in the canonical situation.

4.3 The Prediction

The word "prediction" is not applicable in real sense. Simply it means estimation but we shall continue to use prediction in the sense of estimation. In this section we provide the analysis of the estimation of the predictor

$$E(\mathbf{y}_{N-n}) = \mathbf{X}_2\beta,$$

when it is suspected that $H_0 : \mathbf{H}\beta = \mathbf{h}$ may hold. We may estimate the predictor $\mathbf{y}_{N-n}$ by $\mathbf{X}_2\beta^*$, where β^* is an appropriate estimator of β under the model $\mathbf{y}_n = \mathbf{X}_1\beta + \mathbf{e}_n$.

4.3.1 The Proposed Estimators

Under the models in Section 4.2, we may consider the following estimators of $\mathbf{y}_{N-n}$:

- (i) The unrestricted least squares predictor (ULSP) is given by

$$\tilde{\mathbf{y}}_{N-n} = \mathbf{X}_2\tilde{\beta}_n, \quad (4.15)$$

where $\tilde{\beta}_n$ is defined by (4.1).

(ii) The restricted least squares predictor (RLSP) is given by

$$\hat{y}_{N-n} = X_2 \hat{\beta}_n, \quad (4.16)$$

where $\hat{\beta}_n$ as in (4.2).

(iii) The preliminary test predictor (PTP) is given by

$$\hat{y}_{N-n}^{PT} = X_2 \hat{\beta}_n^{PT}, \quad (4.17)$$

where $\hat{\beta}_n^{PT}$ is defined by (4.3).

(iv) Finally, the shrinkage predictor (SP) is given by

$$\hat{y}_{N-n}^S = X_2 \hat{\beta}_n^S, \quad (4.18)$$

where $\hat{\beta}_n^S$ is defined by (4.4) and $q > 3$.

4.3.2 Performance of the Prediction

First, using the results in Section 3.5 Chapter 3, we note that the biases of the four estimators in (4.15), (4.16), (4.17) and (4.18) are given by

$$E(\tilde{y}_{N-n} - y_{N-n}) = 0, \quad (4.19)$$

$$E(\hat{y}_{N-n} - y_{N-n}) = -X_2 C^{-1} H' (H C^{-1} H')^{-1} \delta; \quad \delta = H\beta - h, \quad (4.20)$$

$$E(\hat{y}_{N-n}^{PT} - y_{N-n}) = -X_2 C^{-1} H' (H C^{-1} H')^{-1} \delta G_{q+2, n-p}(l_\alpha^*; \Delta), \quad (4.21)$$

$$E(\hat{y}_{N-n}^S - y_{N-n}) = -d q X_2 C^{-1} H' (H C^{-1} H')^{-1} \delta E(\chi_{q+2}^{-2}(\Delta)), \quad (4.22)$$

where

$$d = \frac{(q-2)(n-p)}{q(n-p+2)},$$

$\chi_{q+2}^{-2}(\Delta)$ is the inverted non-central chi-square variable with $q+2$ DF and non-centrality parameter

$$\Delta = \delta' (H C^{-1} H')^{-1} \delta / \sigma^2, \quad l_\alpha^* = \frac{q}{q+2} F_\alpha,$$

and $G_{m,n}(\cdot; \Delta)$ is cumulative F-distribution function with (m, n) DF and non-centrality parameter Δ .

Also, under the risk function in (3.58) and the results of Chapter 3, the risk of the four estimators in (4.15), (4.16), (4.17) and (4.18) are given by

$$R(\hat{y}_{N-n}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2), \quad (4.23)$$

$$R(\hat{y}_{N-n}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2) - \text{Tr}(\mathbf{A}_1) + \sigma^{-2} \eta_1' \mathbf{A}_1 \eta_1, \quad (4.24)$$

$$R(\hat{y}_{N-n}^{PT}, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2) - \text{Tr}(\mathbf{A}_1) G_{q+2, n-p}(l_\alpha^*; \Delta) + \sigma^{-2} \eta_1' \mathbf{A}_1 \eta_1 \{2G_{q+2, n-p}(l_\alpha^*; \Delta) - G_{q+4, n-p}(l_\alpha^{**}; \Delta)\}, \quad (4.25)$$

$$R(\hat{y}_{N-n}^S, \mathbf{Q}) = \text{Tr}(\mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2) - dq \text{Tr}(\mathbf{A}_1) \left\{ (q-2) E(\chi_{q+2}^{-4}(\Delta)) + \left[1 - \frac{(q+2) \eta_1' \mathbf{A}_1 \eta_1}{2\sigma^2 \Delta \text{Tr}(\mathbf{A}_1)} \right] 2\Delta E(\chi_{q+4}^{-4}(\Delta)) \right\}, \quad (4.26)$$

where $l_\alpha^{**} = \frac{q}{q+4} F_\alpha$ and

$$\eta_1' \mathbf{A}_1 \eta_1 = \delta' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2 \mathbf{C}^{-1} \mathbf{H}' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \delta$$

$$\text{Tr}(\mathbf{A}_1) = \text{Tr} \{ \mathbf{H} \mathbf{C}^{-1} \mathbf{X}_2' \mathbf{Q} \mathbf{X}_2 \mathbf{C}^{-1} \mathbf{H}' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \}.$$

The performance of the predictor estimators are similar to the performance of the estimators in Section 3.5. Furthermore, the shrinkage estimator dominates the usual estimators under the same matrix $\mathbf{W}$ in Chapter 3. This property does not hold for the estimation of the population mean which will be discussed in the next section.

4.4 Biases and MSE of Various Estimators

The bias and the mean square error (MSE) of an estimator e of μ are given by $\text{Bias}(e) = \sum_{s \in S} P(s) E(e - \mu)$ and $\text{MSE}(e) = \sum_{s \in S} P(s) E(e - \mu)^2$ respectively, where E denotes the expectation taken over the superpopulation model. Note that in the above expressions $P(s)$ does not depend on the estimator, hence we only need to determine $E(e - \mu)$ and $E(e - \mu)^2$ to compare the estimators. The $E(e - \mu)$ of the estimators in (4.1), (4.2), (4.3) and (4.4) are given by the expressions (4.27)–(4.30) respectively:

$$B_1 = E(\tilde{\mu}_n - \mu) = 0, \quad (4.27)$$

$$B_2 = E(\hat{\mu}_n - \mu) = -f\mathbf{a}'\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta; \quad \delta = \mathbf{H}\beta - \mathbf{h}, \quad (4.28)$$

$$B_3 = E(\hat{\mu}_n^{PT} - \mu) = -f\mathbf{a}'\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta G_{q+2, n-p}(l_\alpha^*; \Delta), \quad (4.29)$$

$$B_4 = E(\hat{\mu}_n^S - \mu) = -fdq\mathbf{a}'\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta E(\chi_{q+2}^{-2}(\Delta)), \quad (4.30)$$

where $\chi_{q+2}^{-2}(\Delta)$ is the inverted non-central chi-square variable with $q+2$ DF and non-centrality parameter

$$\Delta = \delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta/\sigma^2, \quad l_\alpha^* = \frac{q}{q+2}F_\alpha,$$

and $G_{m,n}(\cdot; \Delta)$ is cumulative F-distribution function with (m, n) DF and non-centrality parameter Δ .

The corresponding $E(e - \mu)^2$ of the four estimators in (4.1), (4.2), (4.3) and (4.4) may be shown to be

$$M_1 = E(\tilde{\mu}_n - \mu)^2 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}, \quad (4.31)$$

$$M_2 = E(\hat{\mu}_n - \mu)^2 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\{\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} - \mathcal{B}\} + f^2\mathcal{A}, \quad (4.32)$$

$$M_3 = E(\hat{\mu}_n^{PT} - \mu)^2 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\{\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} - \mathcal{B}G_{q+2, n-p}(l_\alpha^*; \Delta)\} \\ + f^2\mathcal{A}\{2G_{q+2, n-p}(l_\alpha^*; \Delta) - G_{q+4, n-p}(l_\alpha^{**}; \Delta)\}, \quad (4.33)$$

$$M_4 = E(\hat{\mu}_n^S - \mu)^2 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} \\ - f^2\sigma^2dq\mathcal{B}[2E\chi_{q+2}^{-2}(\Delta) - (q-2)E(\chi_{q+2}^{-4}(\Delta))] \\ + f^2dq(q+2)\mathcal{A}E(\chi_{q+4}^{-4}(\Delta)), \quad (4.34)$$

where

$$l_\alpha^{**} = \frac{q}{q+4}F_\alpha$$

and

$$\mathcal{A} = \mathbf{a}'\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\delta\delta'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}\mathbf{a}, \quad (4.35)$$

$$\mathcal{B} = \mathbf{a}'\mathbf{C}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{C}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{C}^{-1}\mathbf{a}. \quad (4.36)$$

4.5 Performance of the Estimators

In this section we proceed with an analysis of the frequentist risk of the four estimators in (4.1), (4.2), (4.3) and (4.4) and determine their dominance properties.

First, we consider the MSE of the four estimators under H_0 , i.e., $\delta = 0$. The expressions in (4.31)–(4.34) reduce to the following

$$M_1 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}, \quad (4.37)$$

$$M_2 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2 [\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} - \mathcal{B}], \quad (4.38)$$

$$M_3 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2 [\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} - \mathcal{B}G_{q+2,n-p}(l_\alpha^*; 0)], \quad (4.39)$$

$$M_4 = f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2 \left[\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} - \frac{(q-2)(n-p)}{q(n-p+2)}\mathcal{B} \right]. \quad (4.40)$$

Hence from (4.37)–(4.40), we note that the estimators may be ordered according to their MSE-values as

$$\hat{\mu}_n \succ \hat{\mu}_n^{PT} \succ \hat{\mu}_n^S \succ \tilde{\mu}_n, \quad (4.41)$$

where the notation $\succ$ means dominates, and $\hat{\mu}_n^{PT} \succ \hat{\mu}_n^S$ for a range of α -values satisfying

$$l_\alpha^* > G_{q+2,n-p}^{-1} \left\{ \frac{(q-2)(n-p)}{q(n-p+2)} \right\}. \quad (4.42)$$

Next, we consider the non-null situation, we note that from (4.34) $\hat{\mu}_n^S \succ \tilde{\mu}_n$ if

$$\frac{\mathcal{A}}{\mathcal{B}} \leq \frac{E(\chi_{q+2}^{-2}(\Delta)) + \Delta E(\chi_{q+4}^{-4}(\Delta))}{(q+2)E(\chi_{q+4}^{-4}(\Delta))} \sigma^2. \quad (4.43)$$

For a special case $\mathbf{H} = \mathbf{I}$, $\mathbf{C} = \mathbf{I}$, $\mathbf{a}'\mathbf{a} = 1$ and $\sigma^2 = 1$, the equation (4.43) reduces to

$$\Delta \leq \frac{E(\chi_{q+2}^{-2}(\Delta)) + \Delta E(\chi_{q+4}^{-4}(\Delta))}{(q+2)E(\chi_{q+4}^{-4}(\Delta))}. \quad (4.44)$$

Thus the knowledge of Δ determines which estimator is to be preferred. Further, we note that $\hat{\mu}_n^{PT}$ and $\hat{\mu}_n^S$ have the common MSE $f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}$ as $\Delta \rightarrow \infty$. The two estimators behave differently near the null hypothesis and the MSE of $\hat{\mu}_n^{PT}$ under H_0 is given by (4.39). On the other hand, as Δ diverts from 0 the values of MSE increases and crossing the line

$$f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a} \text{ at}$$

$$\frac{\mathcal{A}}{\mathcal{B}} = \frac{\sigma^2 G_{q+2,n-p}(l_\alpha^*; \Delta)}{2G_{q+2,n-p}(l_\alpha^*; \Delta) - G_{q+4,n-p}(l_\alpha^{**}; \Delta)}, \quad (4.45)$$

with maximum value depending on α ($0 < \alpha < 1$) then decreases to the value of the MSE of $\tilde{\mu}_n$. Hence we note that $\hat{\mu}_n^{PT} \succ \tilde{\mu}_n$ if

$$\frac{A}{B} \leq \frac{\sigma^2 G_{q+2,n-p}(l_\alpha^*; \Delta)}{2G_{q+2,n-p}(l_\alpha^*; \Delta) - G_{q+4,n-p}(l_\alpha^{**}; \Delta)}.$$

Similarly, the MSE of $\hat{\mu}_n^S$ has small values given by (4.40), then increases and crossing the line with value $f(1-f)\frac{\sigma^2}{n} + f^2\sigma^2\mathbf{a}'\mathbf{C}^{-1}\mathbf{a}$ at

$$\frac{A}{B} = \frac{E(\chi_{q+2}^{-2}(\Delta)) + \Delta E(\chi_{q+4}^{-4}(\Delta))}{(q+2)E(\chi_{q+4}^{-4}(\Delta))}\sigma^2.$$

Hence, $\hat{\mu}_n^S$ dominates $\tilde{\mu}_n$ whenever Δ satisfying (4.43).

Hence, except for a small interval near the null hypothesis neither $\hat{\mu}_n$ nor $\hat{\mu}_n^{PT}$ nor $\hat{\mu}_n^S$ dominates each other, while $\tilde{\mu}_n$ dominates all of them outside a certain interval. Tables 4.1–4.6 show the range of values of Δ for which $\hat{\mu}_n^{PT}$ and $\hat{\mu}_n^S$ dominate $\tilde{\mu}_n$ for selected values of N, n, p, q and α (level of significance), for the situation where

$$\mathbf{H} = \mathbf{I}, \quad \mathbf{C} = \mathbf{I}, \quad \mathbf{a}'\mathbf{a} = 1 \quad \text{and} \quad \sigma^2 = 1.$$

Note that in Tables 4.1–4.6, E_{Δ_1} denotes the relative efficiency at Δ_α^1 , which is the intersection point between the PTE and SE. Also, Δ_α^2 denotes the intersection point between the PTE and UE with common relative efficiency equal to 1 and Δ_S denotes the intersection point between the SE and UE with common relative efficiency equal to 1. The relative efficiency of the estimator $\hat{\mu}_n$ relative to $\tilde{\mu}_n$ is defined to be

$$R.E.(\hat{\mu}_n : \tilde{\mu}_n) = \frac{MSE(\tilde{\mu}_n)}{MSE(\hat{\mu}_n)}.$$

For example using Table 4.1, with $N = 50$, $p = 5$, $q = 3$, $n = 15$ and $\alpha = 0.10$ it appears that $E_{\Delta_{1,10}} = 1.1299$, $\Delta_{1,10}^1 = 0.6817$, $\Delta_{1,10}^2 = 0.8167$ and $\Delta_S = 1.4155$, i.e., if $\Delta \in [0, .6817]$ then the PTE dominates the SE but outside the interval the SE dominates the PTE, if $\Delta \in [0, .8167]$ then PTE dominates UE. Also, if $\Delta \in [0, 1.4115]$ then the SE dominates the UE, but outside the interval the UE dominates the SE. In summary, we note that the SE has a wide range compare to the PTE and in all cases the range does not exceeds 1.5. Therefore, except for this small range the usual estimator $\tilde{\mu}_n$ dominates $\hat{\mu}_n^{PT}$ and $\hat{\mu}_n^S$. Note that in

Tables 4.1–4.6, for large values of α or n , we have $\Delta_\alpha^1 = 0$, i.e., the SE dominates the PTE and $E_{\Delta_\alpha^1}$ is equal to the maximum relative efficiency of the PTE relative to the UE. To examine this we have plotted the performance pictures of the UE, RE, PTE and SE for selected values of N , n , p , q and α , for the situation where

$$\mathbf{H} = \mathbf{I}, \quad \mathbf{C} = \mathbf{I}, \quad \mathbf{a}'\mathbf{a} = 1 \quad \text{and} \quad \sigma^2 = 1.$$

All these pictures are displayed in Figures 4.1–4.4. The computations in Tables 4.1–4.6 have been carried out using equations (4.31)–(4.34). To compute (4.33), we approximate the non-central F-distribution using the central Beta distribution. Also, the Figures have been created by SPSS Graphics (SPSSG).

Figure 4.1: Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=6$, $q=4$ and $\alpha = 0.10, 0.15$.

Figure 4.2: Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=6$, $q=4$ and $\alpha = 0.20, 0.50$.

Figure 4.3: Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=7$, $q=5$ and $\alpha = 0.10, 0.15$.

Figure 4.4: Performance picture for the UE, RE, PTE and SE of the finite population mean μ , for $N=40$, $n=10$, $p=7$, $q=5$ and $\alpha = 0.20, 0.50$.

Table 4.1: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=5$ and $q=3$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.0809	1.1100	1.1243	1.1324	1.1368	1.1385	1.1367
	$\Delta_{.05}^1$	.8353	.7676	.7355	.7170	.7051	.6967	.6905
	$\Delta_{.05}^2$	.9169	.8747	.8547	.8432	.8358	.8307	.8269
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.10	$E_{\Delta_{.10}^1}$	1.0939	1.1299	1.1476	1.1574	1.1628	1.1649	1.1627
	$\Delta_{.10}^1$	.7638	.6817	.6448	.6241	.6108	.6015	.5947
	$\Delta_{.10}^2$	.8649	.8167	.7959	.7844	.7772	.7722	.7686
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.15	$E_{\Delta_{.15}^1}$	1.1057	1.1478	1.1684	1.1800	1.1863	1.1888	1.1862
	$\Delta_{.15}^1$	.7020	.6103	.5699	.5473	.5328	.5227	.5153
	$\Delta_{.15}^2$	.8246	.7758	.7558	.7449	.7382	.7336	.7302
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.20	$E_{\Delta_{.20}^1}$	1.1175	1.1659	1.1900	1.2035	1.2111	1.2140	1.2112
	$\Delta_{.20}^1$	.6437	.5426	.4981	.4730	.4570	.4458	.4376
	$\Delta_{.20}^2$	.7912	.7437	.7249	.7148	.7086	.7043	.7013
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.25	$E_{\Delta_{.25}^1}$	1.1300	1.1861	1.2146	1.2308	1.2401	1.2439	1.2408
	$\Delta_{.25}^1$	.5848	.4723	.4222	.3938	.3755	.3628	.3533
	$\Delta_{.25}^2$	.7624	.7171	.6996	.6904	.6846	.6807	.6779
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.40	$E_{\Delta_{.40}^1}$	1.1834	1.2881	1.3507	1.3906	1.4157	1.4026	1.3822
	$\Delta_{.40}^1$	.3652	.1800	.0898	.0363	.0010	.0000	.0000
	$\Delta_{.40}^2$	.6934	.6566	.6429	.6358	.6314	.6285	.6264
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.50	$E_{\Delta_{.50}^1}$	1.2650	1.2958	1.2786	1.2683	1.2603	1.2520	1.2404
	$\Delta_{.50}^1$	.1008	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.6565	.6255	.6142	.6083	.6048	.6024	.6006

Table 4.2: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=6$ and $q=4$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.1096	1.1714	1.2040	1.2231	1.2343	1.2396	1.2377
	$\Delta_{.05}^1$	.8400	.7473	.7021	.6758	.6587	.6467	.6378
	$\Delta_{.05}^2$	.9476	.9054	.8844	.8722	.8643	.8588	.8547
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.10	$E_{\Delta_{.10}^1}$	1.1273	1.2094	1.2537	1.2799	1.2954	1.3029	1.3007
	$\Delta_{.10}^1$	.7764	.6539	.5971	.5647	.5439	.5294	.5187
	$\Delta_{.10}^2$	.9084	.8548	.8312	.8181	.8099	.8042	.8000
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.15	$E_{\Delta_{.15}^1}$	1.1455	1.2483	1.3056	1.3401	1.3610	1.3714	1.3691
	$\Delta_{.15}^1$	.7155	.5680	.5007	.4624	.4378	.4206	.4080
	$\Delta_{.15}^2$	.8752	.8171	.7932	.7803	.7723	.7668	.7628
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.20	$E_{\Delta_{.20}^1}$	1.1651	1.2932	1.3681	1.4148	1.4438	1.4588	1.4569
	$\Delta_{.20}^1$	.6533	.4791	.3990	.3532	.3236	.3029	.2876
	$\Delta_{.20}^2$	.8460	.7865	.7631	.7507	.7430	.7378	.7341
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.25	$E_{\Delta_{.25}^1}$	1.1878	1.3504	1.4530	1.5203	1.5640	1.5879	1.5877
	$\Delta_{.25}^1$	.5863	.3787	.2810	.2244	.1875	.1615	.1423
	$\Delta_{.25}^2$	.8196	.7604	.7379	.7262	.7189	.7140	.7105
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.40	$E_{\Delta_{.40}^1}$	1.3052	1.5823	1.5421	1.5188	1.5006	1.4820	1.4560
	$\Delta_{.40}^1$	.3008	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.40}^2$	.7517	.6982	.6793	.6695	.6636	.6597	.6568
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.50	$E_{\Delta_{.50}^1}$	1.4386	1.3696	1.3445	1.3302	1.3192	1.3083	1.2931
	$\Delta_{.50}^1$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.7125	.6648	.6483	.6399	.6349	.6315	.6291

Table 4.3: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=50$, $p=7$ and $q=5$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.1065	1.2004	1.2527	1.2846	1.3042	1.3146	1.3141
	$\Delta_{.05}^1$	.8637	.7468	.6886	.6545	.6323	.6166	.6051
	$\Delta_{.05}^2$	.9686	.9275	.9061	.8935	.8852	.8794	.8751
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.10	$E_{\Delta_{.10}^1}$	1.1214	1.2493	1.3253	1.3731	1.4033	1.4197	1.4202
	$\Delta_{.10}^1$	.8152	.6533	.5757	.5310	.5021	.4819	.4670
	$\Delta_{.10}^2$	.9411	.8840	.8582	.8439	.8348	.8285	.8240
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.15	$E_{\Delta_{.15}^1}$	1.1375	1.3037	1.4098	1.4798	1.5255	1.5516	1.5543
	$\Delta_{.15}^1$	.7653	.5613	.4642	.4082	.3719	.3466	.3278
	$\Delta_{.15}^2$	.9159	.8500	.8228	.8081	.7989	.7927	.7882
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.20	$E_{\Delta_{.20}^1}$	1.1557	1.3717	1.5250	1.6336	1.7089	1.7546	1.7633
	$\Delta_{.20}^1$	.7118	.4605	.3385	.2672	.2207	.1880	.1637
	$\Delta_{.20}^2$	.8923	.8214	.7940	.7795	.7705	.7644	.7601
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.25	$E_{\Delta_{.25}^1}$	1.1771	1.4676	1.7100	1.9051	2.0551	2.0814	2.0038
	$\Delta_{.25}^1$	.6522	.3401	.1815	.0865	.0235	.0000	.0000
	$\Delta_{.25}^2$	.8699	.7965	.7694	.7553	.7466	.7408	.7366
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.40	$E_{\Delta_{.40}^1}$	1.2910	1.6729	1.6200	1.5903	1.5678	1.5452	1.5142
	$\Delta_{.40}^1$	.3876	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.40}^2$	.8076	.7349	.7104	.6981	.6907	.6857	.6822
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.50	$E_{\Delta_{.50}^1}$	1.5325	1.4323	1.3989	1.3804	1.3666	1.3531	1.3350
	$\Delta_{.50}^1$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.7687	.7003	.6782	.6673	.6607	.6564	.6533

Table 4.4: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=5$ and $q=3$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.0816	1.1112	1.1262	1.1350	1.1406	1.1444	1.1468
	$\Delta_{.05}^1$	.8353	.7676	.7355	.7170	.7051	.6967	.6905
	$\Delta_{.05}^2$	.9169	.8747	.8547	.8432	.8358	.8307	.8269
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.10	$E_{\Delta_{.10}^1}$	1.0947	1.1314	1.1498	1.1606	1.1675	1.1720	1.1750
	$\Delta_{.10}^1$	.7638	.6817	.6448	.6241	.6108	.6015	.5947
	$\Delta_{.10}^2$	.8649	.8167	.7959	.7844	.7772	.7722	.7686
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.15	$E_{\Delta_{.15}^1}$	1.1067	1.1495	1.1710	1.1837	1.1918	1.1971	1.2006
	$\Delta_{.15}^1$	.7020	.6103	.5699	.5473	.5328	.5227	.5153
	$\Delta_{.15}^2$	.8246	.7758	.7558	.7449	.7382	.7336	.7302
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.20	$E_{\Delta_{.20}^1}$	1.1186	1.1679	1.1929	1.2078	1.2174	1.2237	1.2279
	$\Delta_{.20}^1$	.6407	.5426	.4981	.4730	.4570	.4458	.4376
	$\Delta_{.20}^2$	.7912	.7437	.7249	.7148	.7086	.7043	.7013
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.25	$E_{\Delta_{.25}^1}$	1.1313	1.1883	1.2180	1.2358	1.2474	1.2552	1.2604
	$\Delta_{.25}^1$	.5848	.4723	.4222	.3938	.3755	.3628	.3533
	$\Delta_{.25}^2$	.7624	.7171	.6996	.6904	.6846	.6807	.6779
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.40	$E_{\Delta_{.40}^1}$	1.1852	1.2919	1.3570	1.4001	1.4302	1.4238	1.4172
	$\Delta_{.40}^1$	.3652	.1800	.0898	.0363	.0010	.0000	.0000
	$\Delta_{.40}^2$	.6934	.6566	.6429	.6358	.6314	.6285	.6264
	Δ_S	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155	1.4155
.50	$E_{\Delta_{.50}^1}$	1.2677	1.2998	1.2822	1.2743	1.2683	1.2638	1.2599
	$\Delta_{.50}^1$	.1008	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.6565	.6255	.6142	.6083	.6048	.6024	.6006

Table 4.5: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=6$ and $q=4$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.1106	1.1734	1.2072	1.2278	1.2414	1.2507	1.2569
	$\Delta_{.05}^1$	.8400	.7473	.7021	.6758	.6587	.6467	.6378
	$\Delta_{.05}^2$	.9476	.9054	.8844	.8722	.8643	.8588	.8547
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.10	$E_{\Delta_{.10}^1}$	1.1285	1.2120	1.2579	1.2861	1.3048	1.3177	1.3265
	$\Delta_{.10}^1$	.7764	.6539	.5971	.5647	.5439	.5294	.5187
	$\Delta_{.10}^2$	.9084	.8548	.8312	.8181	.8099	.8042	.8000
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.15	$E_{\Delta_{.15}^1}$	1.1468	1.2515	1.3108	1.3481	1.3731	1.3905	1.4025
	$\Delta_{.15}^1$	.7155	.5680	.5007	.4624	.4378	.4206	.4080
	$\Delta_{.15}^2$	.8752	.8171	.7932	.7803	.7723	.7668	.7628
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.20	$E_{\Delta_{.20}^1}$	1.1667	1.2971	1.3747	1.4250	1.4596	1.4840	1.5012
	$\Delta_{.20}^1$	.6533	.4791	.3990	.3532	.3236	.3029	.2876
	$\Delta_{.20}^2$	.8460	.7865	.7631	.7507	.7430	.7378	.7341
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.25	$E_{\Delta_{.25}^1}$	1.1896	1.3552	1.4617	1.5342	1.5858	1.6232	1.6502
	$\Delta_{.25}^1$	.5863	.3787	.2810	.2244	.1875	.1615	.1423
	$\Delta_{.25}^2$	.8196	.7604	.7379	.7262	.7189	.7140	.7105
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.40	$E_{\Delta_{.40}^1}$	1.3085	1.5917	1.5531	1.5326	1.5192	1.5089	1.5001
	$\Delta_{.40}^1$	.3008	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.40}^2$	.7517	.6982	.6793	.6695	.6636	.6597	.6568
	Δ_S	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433	1.3433
.50	$E_{\Delta_{.50}^1}$	1.4438	1.3748	1.3506	1.3379	1.3296	1.3234	1.3181
	$\Delta_{.50}^1$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.7125	.6648	.6483	.6399	.6349	.6315	.6291

Table 4.6: Values of the intersecting efficiencies for PTE and SE for each α with the intersecting point Δ_α^1 -values, the intersecting point between the PTE and UE and the intersecting point between the SE and UE Δ_S -values for various values of the sample size n , $N=70$, $p=7$ and $q=5$

$\alpha \backslash n$		10	15	20	25	30	35	40
.05	$E_{\Delta_{.05}^1}$	1.1074	1.2028	1.2569	1.2910	1.3140	1.3300	1.3413
	$\Delta_{.05}^1$	.8637	.7468	.6886	.6545	.6323	.6166	.6051
	$\Delta_{.05}^2$	.9686	.9275	.9061	.8935	.8852	.8794	.8751
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.10	$E_{\Delta_{.10}^1}$	1.1225	1.2525	1.3310	1.3820	1.4172	1.4421	1.4598
	$\Delta_{.10}^1$	.8152	.6533	.5757	.5310	.5021	.4819	.4670
	$\Delta_{.10}^2$	.9411	.8840	.8582	.8439	.8348	.8285	.8240
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.15	$E_{\Delta_{.15}^1}$	1.1388	1.3077	1.4174	1.4922	1.5453	1.5839	1.6119
	$\Delta_{.15}^1$	.7653	.5613	.4642	.4082	.3719	.3466	.3278
	$\Delta_{.15}^2$	.9159	.8500	.8228	.8081	.7989	.7927	.7882
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.20	$E_{\Delta_{.20}^1}$	1.1571	1.3770	1.5356	1.6518	1.7390	1.8050	1.8546
	$\Delta_{.20}^1$	.7118	.4605	.3385	.2672	.2207	.1880	.1637
	$\Delta_{.20}^2$	.8923	.8214	.7940	.7795	.7705	.7644	.7601
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.25	$E_{\Delta_{.25}^1}$	1.1788	1.4746	1.7261	1.9356	2.1094	2.1683	2.1425
	$\Delta_{.25}^1$	.6522	.3401	.1815	.0865	.0235	.0000	.0000
	$\Delta_{.25}^2$	.8699	.7965	.7694	.7553	.7466	.7408	.7366
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.40	$E_{\Delta_{.40}^1}$	1.2941	1.6844	1.6332	1.6068	1.5898	1.5770	1.5661
	$\Delta_{.40}^1$	.3876	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.40}^2$	.8076	.7349	.7104	.6981	.6907	.6857	.6822
	Δ_S	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925	1.2925
.50	$E_{\Delta_{.50}^1}$	1.5392	1.4387	1.4062	1.3896	1.3790	1.3711	1.3644
	$\Delta_{.50}^1$	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	$\Delta_{.50}^2$	.7687	.7003	.6782	.6673	.6607	.6564	.6533

4.6 Bayes Risk and Analysis of the Estimators

In this section we compare the empirical Bayes estimators proposed in Section 4.2 with the unrestricted estimators (UE) $\tilde{\mu}_n$ in (4.1). The comparisons would be made in term of the relative savings loss (RSL) introduced by Efron and Morris (1973) and used by Ghosh and Lahiri (1987) and Lahiri and Peddada (1988). Let $\Gamma(\xi, b)$ denotes the Bayes risk of the estimator b under the loss in (4.6) and the prior ξ in (4.9). Thus, $\Gamma(\xi, b) = \sum_{s \in S} P(s) E(b - \mu)^2$. Then RSL of $\hat{\mu}_n^{EB}$ with respect to $\tilde{\mu}_n$ is given by

$$\begin{aligned} RSL(\xi; \hat{\mu}_n^{EB}, \tilde{\mu}_n) &= \frac{\Gamma(\xi, \hat{\mu}_n^{EB}) - \Gamma(\xi, \hat{\mu}_n^B)}{\Gamma(\xi, \tilde{\mu}_n) - \Gamma(\xi, \hat{\mu}_n^B)} \\ &= 1 - \frac{\Gamma(\xi, \tilde{\mu}_n) - \Gamma(\xi, \hat{\mu}_n^{EB})}{\Gamma(\xi, \tilde{\mu}_n) - \Gamma(\xi, \hat{\mu}_n^B)}. \end{aligned} \quad (4.46)$$

Note that $RSL(\xi; \hat{\mu}_n^{EB}, \tilde{\mu}_n) < 1$ if and only if $\Gamma(\xi, \hat{\mu}_n^{EB}) < \Gamma(\xi, \tilde{\mu}_n)$. The closer the value of $RSL(\xi; \hat{\mu}_n^{EB}, \tilde{\mu}_n)$ to zero, the more superior is the estimator $\hat{\mu}_n^{EB}$ with respect to $\tilde{\mu}_n$.

The following theorem provides the Bayes risk of $\hat{\mu}_n^B$, $\tilde{\mu}_n$ and $\hat{\mu}_n^K$ under the prior in (4.9) and loss function in (4.6).

Theorem 4.1 *Under the loss function in (4.6) and the prior in (4.9), for $q \geq 3$, the Bayes risk of $\hat{\mu}_n^B$, $\tilde{\mu}_n$ and $\hat{\mu}_n^K$ are given by*

$$(i) \quad \Gamma(\xi, \hat{\mu}_n^B) = f^2 \sigma^2 \left\{ (1 - B) \sum_{s \in S} P(s) \mathbf{a}' \mathbf{C}^{-1} \mathbf{a} + (N - n)^{-1} \right\}, \quad (4.47)$$

$$(ii) \quad \Gamma(\xi, \tilde{\mu}_n) = f^2 \sigma^2 \left\{ \sum_{s \in S} P(s) \mathbf{a}' \mathbf{C}^{-1} \mathbf{a} + (N - n)^{-1} \right\}, \quad (4.48)$$

$$\begin{aligned} (iii) \quad \Gamma(\xi, \hat{\mu}_n^K) &= \Gamma(\xi, \hat{\mu}_n^B) + f^2 \sigma^2 B \sum_{s \in S} P(s) \mathbf{a}' \mathbf{C}^{-1} \mathbf{a} \\ &\quad - f^2 \sigma^2 B \{1 - g(k, n, p)\} \sum_{s \in S} P(s) \mathbf{B}, \end{aligned} \quad (4.49)$$

where

$$g(k, n, p) = 1 - \frac{2k(n-p)}{q(n-p+2)} + \frac{k^2(n-p)}{q(q-2)(n-p+2)},$$

$\mathbf{C}$ is defined as in (4.1) and B is defined as in (4.10).

Proof. (i)

$$\Gamma(\xi, \hat{\mu}_n^B) = \sum_{s \in S} P(s) E(\hat{\mu}_n^B - \mu)^2. \quad (4.50)$$

Using the fact $\hat{\mu}_n^B - \mu = f(\mathbf{a}'\hat{\beta}_n^B - \bar{y}_c)$, $\bar{y}_c = \frac{1}{N-n} \sum_{i=n+1}^N y_i$,

$$\Gamma(\xi, \hat{\mu}_n^B) = \sum_{s \in S} P(s) f^2 E(\mathbf{a}'\hat{\beta}_n^B - \bar{y}_c)^2.$$

Under the model ξ in (4.9), we have

$$\begin{aligned} E(\bar{y}_c^2) &= \text{Var}(\bar{y}_c) + (E(\bar{y}_c))^2 \\ &= \text{Var}(E(\bar{y}_c|\beta)) + E(\text{Var}(\bar{y}_c|\beta)) + \{E[E(\bar{y}_c|\beta)]\}^2 \\ &= \text{Var}(\mathbf{a}'\beta) + E((N-n)^{-1}\sigma^2) + (\mathbf{a}'\nu)^2 \\ &= \tau^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a} + (N-n)^{-1}\sigma^2 + (\mathbf{a}'\nu)^2, \end{aligned} \quad (4.51)$$

$$\begin{aligned} E(\mathbf{a}'\hat{\beta}_n^B \bar{y}_c) &= E\{\mathbf{a}'[\nu + (1-B)(\tilde{\beta}_n - \nu)]\bar{y}_c\} \\ &= E(\mathbf{a}'\nu \bar{y}_c) + (1-B)E[\mathbf{a}'(\tilde{\beta}_n - \nu)\bar{y}_c] \\ &= (\mathbf{a}'\nu)^2 + (1-B)E[\mathbf{a}'\tilde{\beta}_n \bar{y}_c] - (1-B)(\mathbf{a}'\nu)^2. \end{aligned} \quad (4.52)$$

Now, we need to evaluate

$$\begin{aligned} E[\mathbf{a}'\tilde{\beta}_n \bar{y}_c] &= E\{E(\mathbf{a}'\tilde{\beta}_n)\bar{y}_c|\beta\} \\ &= E\{E(\mathbf{a}'\tilde{\beta}_n|\beta)E(\bar{y}_c|\beta)\} \\ &= E(\mathbf{a}'\beta)^2 = \tau^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a} + (\mathbf{a}'\nu)^2. \end{aligned} \quad (4.53)$$

Substituting (4.53) in (4.52), one gets

$$E(\mathbf{a}'\hat{\beta}_n^B \bar{y}_c) = (\mathbf{a}'\nu)^2 + (1-B)\tau^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a}. \quad (4.54)$$

Using the fact that $\mathbf{a}'\hat{\mu}_n^B \sim N(\mathbf{a}'\nu, \tau^2(1-B)\mathbf{a}'\mathbf{C}^{-1}\mathbf{a})$, one gets

$$E(\mathbf{a}'\hat{\mu}_n^B)^2 = (\mathbf{a}'\nu)^2 + (1-B)\tau^2 \mathbf{a}'\mathbf{C}^{-1}\mathbf{a}. \quad (4.55)$$

Hence using (4.50)–(4.55) and the fact $\sum_{s \in S} P(s) = 1$, part (i) of the theorem follows after some simplification.

(ii) Part (ii) of the theorem is proved by putting $B = 0$ in part (i).

(iii) Using a standard argument it follows that

$$\Gamma(\xi, \hat{\mu}_n^K) = \Gamma(\xi, \hat{\mu}_n^B) + \sum_{s \in S} P(s) E(\hat{\mu}_n^K - \hat{\mu}_n^B)^2. \quad (4.56)$$

Now, we observe that $\hat{\mu}_n^K - \hat{\mu}_n^B = f\mathbf{a}'(\hat{\beta}_n^K - \hat{\beta}_n^B)$. Hence, the result of part (iii) of the theorem follows using Theorem 3.3 of Chapter 3.

The next theorem provides the $RSL(\xi; \hat{\mu}_n^K, \tilde{\mu}_n)$ and the optimum choice for the shrinking factor k .

Theorem 4.2 *Under the prior ξ in (4.9) and loss function in (4.6), we have the following*

(i)

$$RSL(\xi; \hat{\mu}_n^K, \tilde{\mu}_n) = 1 - [1 - g(k, n, p)]\mathcal{D}, \quad (4.57)$$

(ii) *For fixed n, p $g(k, n, p)$ is minimized at $k = q - 2$, where $g(k, n, p)$ is defined as in Theorem 4.1 and*

$$\mathcal{D} = \frac{\sum_{s \in S} P(s) \mathbf{a}' \mathbf{C}^{-1} \mathbf{H}' (\mathbf{H} \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H} \mathbf{C}^{-1} \mathbf{a}}{\sum_{s \in S} P(s) \mathbf{a}' \mathbf{C}^{-1} \mathbf{a}}, \quad (4.58)$$

Proof. (i) The proof of the theorem follows immediately from Theorem 4.1 and equation (4.46).

(ii) The proof of part (ii) is straightforward.

Note that the proposed empirical Bayes estimator $\hat{\mu}_n^{EB}$ is the best in the general class of the form $\hat{\mu}_n^K$ given in (4.12) when $q \geq 3$. Substituting $k = q - 2$ in part (i) of Theorem 4.2, we have

$$\begin{aligned} RSL(\xi; \hat{\mu}_n^{EB}, \tilde{\mu}_n) &= 1 - \frac{(n-p)(q-2)}{q(n-p+2)} \mathcal{D}, \\ &< 1 \quad \forall \quad n > p, \quad \forall \quad q \geq 3, \quad 0 < \mathcal{D} < 1. \end{aligned} \quad (4.59)$$

Chapter 5

Asymptotic Theory for Simultaneous Estimation of Binomial Means

5.1 Introduction

In this chapter, we study the relevant asymptotic theory assuming unequal Bernoulli trials without transformation. Efron and Morris (1975) considered a James–Stein (1961) estimator to predict the batting averages of major league players after using an arc sine transformation of the batting average. Also, one may mention the Bayes and empirical Bayes estimators with a mixture of a product of Beta distributions due to Albert (1984) and Albert and Gupta (1981, 1983).

Let $x_{i1}, x_{i2}, \dots, x_{in_i}, i = 1, 2, \dots, p$ be a random sample of size n_i from the i^{th} set of Bernoulli trials with parameters $(1, \theta_i)$. The likelihood function for the parameters given the vector $\mathbf{x} = (x_{11}, \dots, x_{1n_1}, \dots, x_{p1}, \dots, x_{pn_p})$ is

$$L(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^p \prod_{j=1}^{n_i} \theta_i^{x_{ij}} (1 - \theta_i)^{1-x_{ij}}, \quad (5.1)$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$. If we define $n = n_1 + n_2 + \dots + n_p$, and the null-hypothesis H_0 by

$$H_0 : \theta_1 = \theta_2 = \dots = \theta_p = \theta_0, (\text{unknown}) \quad (5.2)$$

then the likelihood function under H_0 is

$$L(\mathbf{x}|\theta) = \theta^{\sum_{i=1}^p y_i} (1 - \theta)^{n - \sum_{i=1}^p y_i}, \quad (5.3)$$

where

$$y_i = \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, 2, \dots, p.$$

We are primarily interested in the estimation of θ when we may have uncertain prior information in (5.2). We shall find it convenient to formulate this problem in an asymptotic setup.

Along with preliminary notions, the proposed estimators are all presented in Section 5.2. The concept of asymptotic risk (AR) and its relation to asymptotic distributional risk (ADR) is outlined in Section 5.3. Asymptotic bias (AB), and ADR results for the various estimators of θ are presented in Section 5.4; and the related asymptotic dominance results and efficiencies are considered in Section 5.5. Data analysis of Efron and Morris (1975) baseball data is made in Section 5.6.

5.2 The Proposed Estimators

5.2.1 Preliminary Test Approach

Let us define

$$\tilde{\theta}_n = (\tilde{\theta}_1, \dots, \tilde{\theta}_p)', \quad \tilde{\theta}_i = \frac{y_i}{n_i}, \quad i = 1, 2, \dots, p; \quad (5.4)$$

$$\hat{\theta}_n = (\hat{\theta}_1, \dots, \hat{\theta}_p)' = \hat{\theta}_n \mathbf{1}_p; \quad (5.5)$$

where

$$\hat{\theta}_n = \frac{\sum_{i=1}^p y_i}{n}, \quad \mathbf{1}_p = (1, 1, \dots, 1)'. \quad (5.6)$$

Then, for the global model (5.1), the unrestricted maximum likelihood estimator (UMLE) of θ is given by $\tilde{\theta}_n$, while under the model (5.3), the restricted MLE (RMLE) is given by $\hat{\theta}_n$. When H_0 holds, $\hat{\theta}_n$ has a smaller risk (under quadratic loss) than $\tilde{\theta}_n$. On the other hand, when H_0 does not hold, $\tilde{\theta}_n$ may perform better than $\hat{\theta}_n$. As a result, when the prior information on H_0 is rather uncertain it

may be desirable to have either a preliminary test estimator (PTE) denoted by $\hat{\theta}_n^{PT} = (\hat{\theta}_1^{PT}, \dots, \hat{\theta}_p^{PT})'$ or a shrinkage estimator (SE) denoted by $\hat{\theta}_n^S = (\hat{\theta}_1^S, \dots, \hat{\theta}_p^S)'$ of θ . In the case of a preliminary test the estimator $\hat{\theta}_n$ or $\tilde{\theta}_n$ is chosen according as H_0 is accepted or rejected. The SE is based on the classical Stein-rule. Unlike the case of $\hat{\theta}_n$, both the PTE and SE behave robustly against departures from the null hypothesis in (5.2) [see e.g., Sen (1986)] and have bounded risk even when n is very large. From these four basic estimators two more improved estimators may be defined; namely, Stein's positive-rule shrinkage estimator (PRSE), denoted by $\hat{\theta}_n^{S+} = (\hat{\theta}_1^{S+}, \dots, \hat{\theta}_p^{S+})'$ and the modified PTE (MPTE) denoted by $\hat{\theta}_n^{PT+} = (\hat{\theta}_1^{PT+}, \dots, \hat{\theta}_p^{PT+})'$. The primary objective of this study is to focus on the PTE and SE and to compare them with the UMLE and RMLE.

For the preliminary test (PT) on H_0 in (5.2), we consider the chi-square test statistic

$$L_n = n(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \hat{\Sigma}_n^{-1} (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.7)$$

where

$$\hat{\Sigma}_n^{-1} = \frac{1}{\hat{\theta}_n(1 - \hat{\theta}_n)} \Lambda, \quad \Lambda = \text{Diag}(\lambda_i), \quad \lambda_{ni} = \frac{n_i}{n} \quad \text{and} \quad \lim_{n \rightarrow \infty} \lambda_{ni} = \lambda_i.$$

Under H_0 , for large n , L_n closely follows the central chi-square distribution with $(p-1)$ degrees of freedom (DF). Thus, for a given level of significance α ($0 < \alpha < 1$), the critical value of L_n may be approximated by $\chi_{p-1, \alpha}^2$, the upper $100\alpha\%$ point of the chi-square distribution with $(p-1)$ DF. Then, the PTE of θ is defined by

$$\hat{\theta}_n^{PT} = \hat{\theta}_n \mathbf{1}_p + [1 - I(L_n \leq \chi_{p-1, \alpha}^2)] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.8)$$

where $I(A)$ is the indicator function for the set A .

To introduce the SE we follow Arnold (1981) and propose

$$\hat{\theta}_n^S = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.9)$$

where we need to assume that $p > 3$.

Note that whereas in (5.8) we have a dichotomous indicator function, in (5.9), $(p-3)L_n^{-1}$ may be regarded as a smoother version. Thus, the SE might be considered from a "pre-test" point of view. Also, (5.8) is a convex combination of

$\tilde{\theta}_n$ and $\hat{\theta}_n$, while (5.9) need not be especially because $(p-3)L_n^{-1}$ may be greater than 1. In the spirit of Sclove et al. (1972), we may introduce the following the *positive-rule* SE (PRSE) :

$$\hat{\theta}_n^{S+} = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}]^+ (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.10)$$

where $y^+ = \max(0, y)$. Note that (5.8) and (5.10) may not agree even if we let $\chi_{p-1, \alpha}^2 = p-3$. The justification of (5.8) lies in the preliminary test of H_0 , whereas (5.10) is intended to control possible overshrinking in (5.9).

Finally, we propose the MPTE of θ as follows:

$$\hat{\theta}_n^{PT+} = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}] I(L_n \geq \chi_{p-1, \alpha}^2) (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.11)$$

where the restriction $\chi_{p-1, \alpha}^2 < (p-3)$ is required. This estimator uniformly dominates $\hat{\theta}_n^{PT}$ but now we have the restriction $p \geq 4$. For $p < 4$ we are left with the PTE to improve over $\hat{\theta}_n$ or $\tilde{\theta}_n$. It is well-known [see e.g., Sclove et al. (1972)] that the positive-rule estimator PRSE dominates the usual shrinkage estimator SE, while the MPTE dominates the usual PTE.

5.2.2 The Bayes and Empirical Bayes Approach

Now, we describe the empirical Bayes estimator for this problem due to Albert (1984). He has considered this problem by assuming the prior distribution of θ_i , $i = 1, 2, \dots, p$ to be independent and identically distributed as beta ($K\eta, K(1-\eta)$), where K is an unknown positive number and η an unknown number in the interval $(0, 1)$ and obtained the following Bayes estimator denoted by $\hat{\theta}_n^B$:

$$\hat{\theta}_n^B = \hat{\theta}_n \mathbf{1}_p + [1 - \varphi(K)] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.12)$$

where

$$\varphi(K) = \frac{K}{\bar{n} + K}, \quad \bar{n} = \frac{1}{p} \sum_{i=1}^p n_i,$$

which depends on the unknown parameter K . To obtain the EBE, Albert (1984) estimated this unknown function $\varphi(K)$ by its posterior mean, given by

$$\begin{aligned} \widehat{\varphi(K)} &= E[\varphi(K)|y] \\ &= \frac{\int_0^\infty \varphi(K) m(y|K) dK/K}{\int_0^\infty m(y|K) dK/K}, \end{aligned}$$

where $\mathbf{y} = (y_1, \dots, y_p)$, y_i is distributed as binomial distribution with parameters n_i and θ_i and $m(\mathbf{y}|K)$ is the marginal density of $\mathbf{y}$ which is proportional to

$$\int_0^1 L(\mathbf{y}|K, \eta) \phi^*(\eta) d\eta,$$

where $\phi^* = [\eta(1 - \eta)]^{-1}$,

$$L(\mathbf{y}|K, \eta) = \prod_{i=1}^p \binom{n_i}{y_i} (K\eta)^{y_i} (K(1 - \eta))^{n_i - y_i} / K^{(n_i)},$$

and $a^{(b)} = a(a + 1) \dots (a + b - 1)$ ($a^{(0)} = 1$). Then the empirical Bayes estimator (EBE) is defined by

$$\hat{\theta}_n^{EB} = \hat{\theta}_n \mathbf{1}_p + [1 - \varphi(\widehat{K})] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.13)$$

For the proof of these results see Albert (1984).

In the baseball data example which will be discussed in Section 5.6, the function $\varphi(K)$ is estimated by plotting the likelihood of marginal density $m(\mathbf{y}|K)$ against the values of φ ($0 \leq \varphi \leq 1$). The estimate of $\varphi(K)$ is equal to .90, i.e., $\varphi(\widehat{K}) = 0.90$. Note that the empirical Bayes approach is more difficult to apply than the preliminary test approach we discussed earlier. In the preliminary test approach we need only to compute L_n the test statistic to find $\hat{\theta}_n^S$.

5.3 Asymptotic Risks and Pitman Alternatives

Let δ_n be an estimator of θ , $\mathbf{W}$ be a positive semi-definite (p.s.d) matrix, and consider the *quadratic loss function*

$$L_n(\delta_n, \theta) = n(\delta_n - \theta)' \mathbf{W} (\delta_n - \theta) = n \text{Tr} [\mathbf{W} (\delta_n - \theta) (\delta_n - \theta)'], \quad (5.14)$$

where $\text{Tr}(\mathbf{A}) = \text{trace of } \mathbf{A}$. Then the risk of δ_n is given by

$$R_n(\delta_n, \theta) = EL_n(\delta_n, \theta) = \text{Tr}(\mathbf{W} \mathbf{V}_n), \quad (5.15)$$

where

$$\mathbf{V}_n = nE(\delta_n - \theta)(\delta_n - \theta)'$$

For the p -dimensional unit cube Ω , let $\omega(\subset \Omega)$ be the subspace for which θ satisfies the hypothesis in (5.2). Then, by virtue of the consistency of the L_n test, we note that for any fixed $\theta \notin \omega$, $L_n \xrightarrow{P} +\infty$, as $n \rightarrow \infty$ ($\xrightarrow{P}$ means convergence in probability). Thus, for any fixed $\theta \notin \omega$, as n increases $\hat{\theta}_n^{PT}$, $\hat{\theta}_n^S$, $\hat{\theta}_n^{S+}$ and $\hat{\theta}_n^{PT+}$ are equivalent in probability to the UMLE $\tilde{\theta}_n$, while $\hat{\theta}_n$ will have unbounded risk. The situation is different, however, when $\theta \in \omega$, i.e., θ belongs to a shrinking neighborhood of ω . For this reason, we consider a sequence $\{K_{(n)}\}$ of Pitman-alternatives:

$$K_{(n)}: \theta_{(n)} = \theta_0 \mathbf{1}_p + n^{-\frac{1}{2}} \xi; \quad \xi = (\xi_1, \dots, \xi_p)', \quad \theta_0 \mathbf{1}_p \in \omega, \quad (5.16)$$

where the ξ_i are fixed numbers. Note that for $\xi = 0$, $\theta_{(n)} = \theta_0 \mathbf{1}_p \in \omega$, so that (5.2) is a particular case of (5.16).

For the estimator δ_n , the asymptotic risk (AR) under $K_{(n)}$ is defined by

$$R(\delta; \xi) = \lim_{n \rightarrow \infty} R_n(\delta_n; \theta_{(n)}), \quad (5.17)$$

whenever the limit exists. We introduce the asymptotic distribution function (ADF) of δ_n by

$$G_{(\delta)}^*(\mathbf{x}) = \lim_{n \rightarrow \infty} P \{ n^{\frac{1}{2}}(\delta_n - \theta_{(n)}) \leq \mathbf{x} | K_{(n)} \}, \quad (5.18)$$

whenever the limit exists. We let

$$V_{(\delta)}^* = \int \dots \int \mathbf{x} \mathbf{x}' dG_{(\delta)}^*(\mathbf{x}). \quad (5.19)$$

Then, the asymptotic distributional risk (ADR) of δ_n is defined by

$$R^*(\delta, \xi) = \text{Tr}(\mathbf{W} V_{(\delta)}^*). \quad (5.20)$$

One may need extra regularity conditions to evaluate (5.17) rather than (5.20). This point has been explained in detail in various other contexts, e.g., in Sen (1984), Sen and Saleh (1985) and Saleh and Sen (1985), among others. Fortunately, in this specific context, we shall see that both (5.17) and (5.20) can be evaluated under the same set of regularity conditions, although it is a lot simpler to work with (5.20). As such, in Section 5.4 we shall study the ADR results for the various estimators.

5.4 Asymptotic Risks and Bias

First, we consider the case of *fixed alternatives*. Note that by (5.8),

$$\begin{aligned}
 n(\hat{\theta}_n^{PT} - \tilde{\theta}_n)' \mathbf{W}(\hat{\theta}_n^{PT} - \tilde{\theta}_n) &= n(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \mathbf{W}(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p) I(L_n < \chi_{p-1, \alpha}^2) \\
 &= \{L_n I(L_n < \chi_{p-1, \alpha}^2)\} \{n(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \mathbf{W}(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p) L_n^{-1}\} \\
 &\leq \{L_n I(L_n < \chi_{p-1, \alpha}^2)\} ch_{\max}(\mathbf{W} \mathbf{A}^{-1}) \\
 &\leq \{L_n I(L_n < \chi_{p-1, \alpha}^2)\} Tr(\mathbf{W} \mathbf{A}^{-1}) \\
 &\leq \{L_n I(L_n < \chi_{p-1, \alpha}^2)\} \left(\sum_{j=1}^p (|w_{jj}| \lambda_j^{-1}) \right), \tag{5.21}
 \end{aligned}$$

where $ch_{\max}(A)$ = largest characteristic root of A . Also, note that

$$E\{L_n I(L_n < \chi_{p-1, \alpha}^2) | \theta \notin \omega\} \leq \chi_{p-1, \alpha}^2 P\{L_n < \chi_{p-1, \alpha}^2 | \theta \notin \omega\}.$$

Therefore, by the consistency of the test based on L_n ,

$$E\{L_n I(L_n < \chi_{p-1, \alpha}^2) | \theta \notin \omega\} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \tag{5.22}$$

Thus, for fixed $\theta \notin \omega$, $\hat{\theta}_n^{PT}$ and $\tilde{\theta}_n$ are asymptotically risk-equivalent. A very similar treatment holds for $\hat{\theta}_n^S$, $\hat{\theta}_n^{S+}$ and $\hat{\theta}_n^{PT+}$. For $\hat{\theta}_n^S$ we note that on the set $\{L_n > 0\}$

$$\begin{aligned}
 n(\hat{\theta}_n^S - \tilde{\theta}_n)' \mathbf{W}(\hat{\theta}_n^S - \tilde{\theta}_n) &= (p-3)^2 L_n^{-2} \{n(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \mathbf{W}(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)\} \\
 &\leq (p-3)^2 L_n^{-1} ch_{\max}(\mathbf{W} \mathbf{A}^{-1}), \tag{5.23}
 \end{aligned}$$

while on the set $\{L_n = 0\}$, we have $\hat{\theta}_n^S = \tilde{\theta}_n = \hat{\theta}_n \mathbf{1}_p$. Thus, if we can show that

$$E\{L_n^{-1} I(L_n > 0) | \theta \notin \omega\} \rightarrow 0, \quad \text{as } n \rightarrow \infty, \tag{5.24}$$

then $\hat{\theta}_n^S$ and $\tilde{\theta}_n$ become asymptotically risk-equivalent for every $\theta \notin \omega$. Now L_n is non-negative and for every $\theta \notin \omega$, $L_n \xrightarrow{P} +\infty$, as $n \rightarrow \infty$; hence, to show (5.24) holds, it suffices to show that for every positive $\epsilon > 0$, we have

$$E\{L_n^{-1} I(0 < L_n \leq \epsilon | \theta \notin \omega\} \rightarrow 0, \quad \text{as } n \rightarrow +\infty. \tag{5.25}$$

To prove (5.25). First, we define

$$k_n = \max_{1 \leq j \leq p} n |\tilde{\theta}_j - \hat{\theta}|. \quad (5.26)$$

Then

$$L_n \geq \frac{4}{n} k_n^2, \quad \text{with probability 1.} \quad (5.27)$$

Now $k_n = 0$ yields $L_n = 0$ and hence, $\hat{\theta}_n^S = \tilde{\theta}_n = \hat{\theta}_n \mathbf{1}_p$. Thus, we exclude the case $k_n = 0$. In this case $k_n > 0$ and it can be less than 1, with probability 1; in any case $k_n > \frac{1}{n}$, with probability 1 (excluding the case $k_n = 0$). Thus, on the set $\{L_n > 0\}$, we may write

$$L_n^{-1} I(0 < L_n \leq \epsilon) = L_n^{-1} I(0 < L_n < \frac{4}{n}) + L_n^{-1} I(\frac{4}{n} \leq L_n \leq \epsilon). \quad (5.28)$$

Let

$$k_{0n} = \frac{1}{2} \sqrt{n\epsilon}. \quad (5.29)$$

Then, using (5.27) and (5.29) we obtain for every θ [under $K_{(n)}$ as well]

$$\begin{aligned} E_{\theta} \left\{ L_n^{-1} I(0 < L_n \leq \epsilon) \right\} &\leq \frac{n}{4} [E_{\theta} \{ k_n^{-2} I(\frac{1}{n} < k_n < 1) \} \\ &\quad + E_{\theta} \{ k_n^{-2} I(1 < k_n < k_{0n}) \}]. \end{aligned} \quad (5.30)$$

Replacing k_n by $[k_n]$, the largest integer contained in k_n , the second term in (5.30) (for $k_n > 1$), satisfies the inequality

$$\begin{aligned} E_{\theta} \{ k_n^{-2} I(1 < k_n < k_{0n}) \} &\leq E_{\theta} \{ [k_n]^{-2} I(1 \leq [k_n] \leq k_{0n}) \} \\ &= \sum_{k=1}^{k_{0n}} k^{-2} P_{\theta} \{ [k_n] = k \} \\ &\leq 2 \sum_{k=1}^{k_{0n}-1} \{ k^2 (k+1) \}^{-1} P_{\theta} \{ [k_n] \leq k \} \\ &\quad + k_{0n}^{-2} P_{\theta} \{ [k_n] = k_{0n} \}. \end{aligned} \quad (5.31)$$

Now, the degrees of freedom for L_n is $(p-1)$, hence there are $k^{(p-1)}$ configurations for which $k_n \leq k$. For each configuration the probability is $O(n^{-(p-1)/2})$. Thus,

$P_{\theta}\{k_n \leq k\} = O(k^{p-1}n^{-(p-1)/2})$ for every $k \leq k_{0n}$. Hence, by (5.30) and (5.31)

$$\begin{aligned}
 & \frac{n}{4} E_{\theta} \{k_n^{-2} I(1 \leq k_n \leq k_{0n})\} \\
 & \leq \frac{n}{2} \left\{ \sum_{k=1}^{k_{0n}-1} O(k^{(p-4)} \cdot n^{-(p-1)/2}) + O(k_{0n}^{p-3} \cdot n^{-(p-1)/2}) \right\} \\
 & = O(n^{\frac{-(p-1)}{2}+1}) \left\{ \sum_{k=1}^{k_{0n}-1} O(k^{(p-4)}) + O(k_{0n}^{(p-3)}) \right\} \\
 & = O\left(n^{-\frac{(p-1)}{2}+1} \cdot k_{0n}^{(p-3)}\right) \\
 & = O\left(\epsilon^{\frac{(p-3)}{2}} \cdot n^{-\frac{(p-1)}{2}+1} \cdot n^{\frac{(p-1)}{2}-1}\right) \\
 & = O(\epsilon^{\frac{(p-3)}{2}}), \quad p \geq 4.
 \end{aligned} \tag{5.32}$$

Since ϵ is arbitrarily small (5.32) may be made small. For the first term we set $[nk_n] = k_n^*$ so that on $\frac{1}{n} \leq k_n < 1$, we obtain $1 \leq k_n^* < n$; thus,

$$\begin{aligned}
 & \frac{n}{4} E_{\theta} \{k_n^{-2} I(\frac{1}{n} \leq k_n < 1)\} \\
 & \leq \frac{n^3}{4} E_{\theta} \{k_n^{*-2} I(1 \leq k_n^* \leq n-1)\} \\
 & = \frac{n^3}{4} \sum_{k^*=1}^{n-1} k^{*-2} P_{\theta}(k_n^* = k^*)
 \end{aligned} \tag{5.33}$$

By a similar argument utilized before, one may show that the last term is $O(n^{\frac{-(p-3)}{2}})$, $p \geq 4$. Hence, we have proved that $E_{\theta} L_n^{-1} I(0 < L_n < \epsilon) \rightarrow 0$, as $n \rightarrow \infty$.

Next, note that for any $\theta \notin \omega$, $\hat{\theta}_n 1_p - \theta \xrightarrow{a.s.} \eta (\neq 0)$, as $n \rightarrow \infty$. ($\xrightarrow{a.s.}$ means convergence almost surely), so that as $n \rightarrow \infty$

$$n(\hat{\theta}_n 1_p - \theta)' W(\hat{\theta}_n 1_p - \theta) \xrightarrow{P} +\infty. \tag{5.34}$$

Consequently, the asymptotic risk of $\hat{\theta}_n 1_p$, for any $\theta \notin \omega$ approaches $+\infty$. Also, note that

$$nE(\tilde{\theta}_n - \theta)(\tilde{\theta}_n - \theta)' = \text{Diag} \left\{ \left(\frac{\theta_1(1-\theta_1)}{\lambda_1}, \dots, \frac{\theta_p(1-\theta_p)}{\lambda_p} \right) \right\},$$

so that the asymptotic risk of UMLE is bounded for every $\theta \in \Omega$; hence, we arrive at the following theorem.

Theorem 5.1 For any fixed $\theta \notin \omega$, $\hat{\theta}_n 1_p$ has asymptotic risk $+\infty$, while $\hat{\theta}_n^S$, $\hat{\theta}_n^{S+}$, $\hat{\theta}_n^{PT}$, $\hat{\theta}_n^{PT+}$, and $\tilde{\theta}_n$ are asymptotically risk-equivalent, with bounded risk.

Next, we consider the case of Pitman alternatives in (5.16). As we pointed out earlier we shall mainly study ADR properties though similar AR properties hold. Note that $\hat{\theta}_n 1_p$ and $\tilde{\theta}_n$ both have non-negative elements bounded by 1, so finite moments of $\sqrt{n}(\tilde{\theta}_n - \theta)$ and $\sqrt{n}(\hat{\theta}_n 1_p - E\hat{\theta}_n 1_p)$ of all finite orders exist and have finite limit as $n \rightarrow \infty$. Also, under $\{K_{(n)}\}$ in (5.16),

$$E[\hat{\theta}_n 1_p | K_{(n)}] - \theta = O(n^{-\frac{1}{2}}),$$

so that in this case, $n\|\tilde{\theta}_n - \hat{\theta}_n 1_p\|^2 = O(1)$, in L_∞ norm. Thus, for (5.8)–(5.11), the convergence in distribution will ensure the convergence in second moment, and hence (5.20) will ensure (5.17). For the SE in (5.9), PRSE in (5.10) and MPTE in (5.11) we need a more refined treatment. Note that (5.23) applies as well for $\{K_{(n)}\}$, and hence, if we are able to show (5.24) holds under $\{K_{(n)}\}$ (as well as H_0), then again (5.20) would ensure (5.17). Thus for our purpose it suffices to consider the ADR in (5.20) through the asymptotic distribution in (5.18). We work out the ADR results in this section. For the study of the asymptotic distribution theory and the ADR results, we assume the usual regularity conditions hold. Further, we define $\theta_0 1_p$ as in (5.16) so that $\theta_0 1_p \in \omega$, then we have the following theorem.

Theorem 5.2 Under $\{K_{(n)}\}$ in (5.16) and assumed regularity conditions, the following hold

$$(i) \quad x_n = \sqrt{n}(\tilde{\theta}_n - \theta_0 1_p) \xrightarrow{D} N_p(\xi, \Sigma_0), \quad \text{as } n \rightarrow \infty, \quad (5.35)$$

where $\Sigma_0 = \theta_0(1 - \theta_0)\Lambda^{-1}$, [Λ defined by (5.7)];

$$(ii) \quad y_n = \sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n 1_p) \xrightarrow{D} N_p(J\xi, \Sigma_0 J'), \quad \text{as } n \rightarrow \infty, \quad (5.36)$$

where $J = I_p - 1_p 1_p' \Lambda$, [1_p defined by (5.6)];

$$(iii) \quad z_n = \sqrt{n}(\hat{\theta}_n 1_p - \theta_0 1_p) \xrightarrow{D} N_p(0, B), \quad \text{as } n \rightarrow \infty, \quad (5.37)$$

where $\lambda = (\lambda_1, \dots, \lambda_p)'$, $B = \theta_0(1 - \theta_0)1_p 1_p'$, and we assume $\lambda' \xi = 0$;

$$(iv) \begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} \xrightarrow{D} N_{2p} \left\{ \begin{bmatrix} \boldsymbol{\xi} \\ \mathbf{J}\boldsymbol{\xi} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 \mathbf{J}' \\ \mathbf{J}\boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_0 \mathbf{J}' \end{bmatrix} \right\}, \quad \text{as } n \rightarrow \infty; \quad (5.38)$$

$$\begin{bmatrix} \mathbf{z}_n \\ \mathbf{y}_n \end{bmatrix} \xrightarrow{D} N_{2p} \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{J}\boldsymbol{\xi} \end{bmatrix}, \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_0 \mathbf{J}' \end{bmatrix} \right\}, \quad \text{as } n \rightarrow \infty; \quad (5.39)$$

(v) L_n is distributed asymptotically as a noncentral chi-square with $(p-1)$ DF, and noncentrality parameter

$$\Delta = \boldsymbol{\delta}' \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\delta}, \quad \boldsymbol{\delta} = \mathbf{J}\boldsymbol{\xi}. \quad (5.40)$$

where $\xrightarrow{D}$ means convergence in distribution.

Proof. The first three parts of the theorem are proved by computing the moment-generating function (MGF) of the random variable $\mathbf{x}_n$, $\mathbf{y}_n$ and $\mathbf{z}_n$ and applying the continuity theorem [Theorem 14.3-3 of Bishop, Fienberg and Holland (1975)]. Note that the proof of these results is similar to the Theorem 14.3-5 of Bishop, Fienberg and Holland (1975). Part (iv) of the theorem follows by applying the previous parts of the theorem. Next we prove the last part of the theorem, thus we need to state the following lemmas.

Lemma 5.1 If $X_n \xrightarrow{D} X$ and g be a continuous function, then $g(X_n) \xrightarrow{D} g(X)$.

For a proof of this result, see Rao (1973), p. 124.

Lemma 5.2 If $\mathbf{u} = (u_1, u_2, \dots, u_p)$ has the multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $Y = \mathbf{u}' \mathbf{A} \mathbf{u}$ for some symmetric matrix $\mathbf{A}$, then $Y \sim \sum_{i=1}^p \psi_i Z_i^2$, where $Z_1^2, \dots, Z_p^2$ are independent non-central chi square variables with one degree of freedom, non-centrality parameter δ_i (say) and $\psi_1, \dots, \psi_p$ are the eigenvalues of $(\mathbf{A}^{\frac{1}{2}})' \boldsymbol{\Sigma} \mathbf{A}^{\frac{1}{2}}$.

For the proof of this lemma see Bishop et al. (1975), p. 474.

Lemma 5.3 If $\mathbf{C}^2 = \mathbf{C}$ and $\mathbf{C}' = \mathbf{C}$, then the eigenvalues of $\mathbf{C}$ are all either 0 or 1. The number of eigenvalues equal to 1 is equal to the trace of $\mathbf{C}$, where

$$\text{Tr}(\mathbf{C}) = \sum_i c_{ii}.$$

For the proof of this result see Bishop et al. (1975), p. 473.

Now we prove the last part of the theorem. Note that L_n can be written as

$$L_n = \mathbf{y}'_n \Sigma_0^{-1} \mathbf{y}_n. \quad (5.41)$$

Now, by part (ii) of this theorem, we have $\mathbf{y}_n \xrightarrow{D} \mathbf{y}$, where $\mathbf{y}$ has the multivariate normal distribution with mean δ and covariance matrix $\Sigma_0 - \mathbf{B}$, i.e., $N(\delta, \Sigma_0 - \mathbf{B})$. Note that $\Sigma_0 \mathbf{J}'$ can be written as $\Sigma_0 - \mathbf{B}$. Let $g(\mathbf{u}) = \mathbf{u}' \Sigma_0^{-1} \mathbf{u}$. Evidently, g is a continuous function of $\mathbf{u}$. By Lemma 5.1 and the expression (5.41), we have

$$L_n = \mathbf{y}'_n \Sigma_0^{-1} \mathbf{y}_n \xrightarrow{D} \mathbf{y}' \Sigma_0^{-1} \mathbf{y}. \quad (5.42)$$

Thus the asymptotic distribution of L_n under $\{K_{(n)}\}$ is the distribution of $\mathbf{y}' \Sigma_0^{-1} \mathbf{y}$, where $\mathbf{y} \sim N(\delta, \Sigma_0 - \mathbf{B})$. This reduces the problem to finding the distribution of a quadratic form of a multivariate normal random vector. If we apply Lemma 5.2, we see that

$$\mathbf{y}' \Sigma_0^{-1} \mathbf{y} \stackrel{D}{=} \sum_{i=1}^p \psi_i Z_i^2,$$

where ψ_i are the eigenvalues of

$$\mathbf{C} = \Sigma_0^{-\frac{1}{2}} (\Sigma_0 - \mathbf{B}) \Sigma_0^{-\frac{1}{2}} = \mathbf{I}_p - \Sigma_0^{-\frac{1}{2}} \mathbf{B} \Sigma_0^{-\frac{1}{2}}. \quad (5.43)$$

Note that $\Sigma_0^{-\frac{1}{2}} \mathbf{B} \Sigma_0^{-\frac{1}{2}}$ can be written in the form of one matrix with diagonal λ_i , $i = 1, 2, \dots, p$. Also, it can be shown that $\mathbf{C}$ is idempotent and symmetric matrix. Applying Lemma 5.3 to $\mathbf{C}$ defined as in (5.43) gives the following result:

$$\begin{aligned} \text{Tr}(\mathbf{C}) &= \text{Tr}(\mathbf{I}_p - \Sigma_0^{-\frac{1}{2}} \mathbf{B} \Sigma_0^{-\frac{1}{2}}) \\ &= p - 1, \quad \text{since} \quad \sum_{i=1}^p \lambda_i = 1. \end{aligned}$$

Hence

$$\mathbf{y}' \Sigma_0^{-1} \mathbf{y} \stackrel{D}{=} \sum_{i=1}^{p-1} \psi_i Z_i^2 = \chi_{p-1}^2(\Delta),$$

i.e., we have established that L_n has an asymptotic chi square distribution with $p-1$ degrees of freedom and the non-centrality parameter Δ .

We denote a multinormal distribution function with mean vector μ and dispersion matrix Σ by $\Phi(\mathbf{x}; \mu, \Sigma)$ and the corresponding density function by

$\phi_p(\mathbf{x}; \mu, \Sigma)$. Also, let $H_p(\mathbf{x}; \Delta)$ denote the noncentral chi-square distribution function with p DF and noncentrality parameter $\Delta(\geq 0)$ and $E(\chi_m^{-2r}(\Delta)) = \int_0^\infty x^{-r} dH_m(x; \Delta)$. Note that by (5.8) and Theorem 5.2 we have

$$\sqrt{n}(\hat{\theta}_n^{PT} - \theta_0 \mathbf{1}_p) = \mathbf{x}_n - \mathbf{y}_n I(L_n \leq \chi_{p-1, \alpha}^2). \quad (5.44)$$

By (5.38)–(5.44) and the general formulation of Saleh and Sen (1984a), we arrive at the following theorem:

Theorem 5.3 *Under $\{K_{(n)}\}$ in (5.16) and the assumed regularity conditions, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^{PT} - \theta_0 \mathbf{1}_p)$ is given by*

$$\begin{aligned} G_p^{PT}(\mathbf{x}) &= \Phi_p(\mathbf{x}; \mathbf{0}, \mathbf{B}) H_{p-1}(\chi_{p-1, \alpha}^2; \Delta) \\ &+ \int \dots \int_{M(\delta)} \Phi_p(\mathbf{x} - \boldsymbol{\xi} - \mathbf{z}; \mathbf{0}, \mathbf{B}) d\Phi_p(\mathbf{x}; \mathbf{0}, \Sigma_0 \mathbf{J}'), \end{aligned} \quad (5.45)$$

where

$$M(\delta) = \{\mathbf{z}; (\mathbf{z} + \mathbf{J}\boldsymbol{\xi})'(\mathbf{z} + \mathbf{J}\boldsymbol{\xi}) \geq \chi_{p-1, \alpha}^2\}, \quad (5.46)$$

and $\mathbf{J}, \mathbf{B}$ and Δ are all defined as in Theorem 5.2.

Note that the density function corresponding to $G_p^{PT}(\mathbf{x})$ is

$$\begin{aligned} g_p^{PT}(\mathbf{x}) &= \phi_p(\mathbf{x}; \mathbf{0}, \mathbf{B}) H_{p-1}(\chi_{p-1, \alpha}^2; \Delta) \\ &+ \int \dots \int_{M(\delta)} \phi_p(\mathbf{x} - \boldsymbol{\xi} - \mathbf{z}; \mathbf{0}, \mathbf{B}) d\phi_p(\mathbf{z}; \mathbf{0}, \Sigma_0 \mathbf{J}') \end{aligned} \quad (5.47)$$

Next, by virtue of (5.7), (5.9) and Theorem 5.2, we obtain the following theorem:

Theorem 5.4 *Under $\{K_{(n)}\}$ in (5.16) and the assumed regularity conditions,*

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^S - \theta_0 \mathbf{1}_p) &= \mathbf{x}_n - (p-3)\mathbf{y}_n \{\mathbf{y}_n' \Sigma_0^{-1} \mathbf{y}_n\}^{-1} \\ &\xrightarrow{D} (\mathbf{x} + \boldsymbol{\xi}) - (p-3)(\mathbf{y} + \mathbf{J}\boldsymbol{\xi}) \times \\ &\quad \{(\mathbf{y} + \mathbf{J}\boldsymbol{\xi})' \Sigma_0^{-1} (\mathbf{y} + \mathbf{J}\boldsymbol{\xi})\}^{-1} \end{aligned} \quad (5.48)$$

where

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N_{2p} \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_0 & \Sigma_0 \mathbf{J}' \\ \mathbf{J} \Sigma_0 & \Sigma_0 \mathbf{J}' \end{bmatrix} \right\}.$$

Very similar representations hold for the positive rule estimator in (5.10). Under $\{K_{(n)}\}$ in (5.16), and by virtue of the general equivalence result in (5.17) and (5.20) and by Theorems 5.2, 5.3 and 5.4, we arrive at following expressions for the asymptotic bias and risks. Since the derivation of these formula are very similar to the cases treated in Chapter 2, we omit these details of the derivation. The asymptotic bias expressions for the first four estimators in (5.4), (5.5), (5.8) and (5.9) are given by

$$B_1(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\tilde{\theta}_n - \theta_{(n)})\} = 0, \quad (5.49)$$

$$B_2(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})\} = -\delta, \quad (5.50)$$

$$B_3(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n^{PT} - \theta_{(n)})\} = -\delta H_{p+1}(\chi_{p-1, \alpha}^2; \Delta), \quad (5.51)$$

$$B_4(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n^S - \theta_{(n)})\} = -(p-3)\delta E(\chi_{p+1}^{-2}(\Delta)), \quad (5.52)$$

where Δ is defined by (5.40). Also, the asymptotic distributional covariance matrices of the first four estimators in (5.4), (5.5), (5.8) and (5.9) are given by

$$\Gamma_1(\delta) = \lim_{n \rightarrow \infty} E\{n(\tilde{\theta}_n - \theta_{(n)})(\tilde{\theta}_n - \theta_{(n)})'\} = \Sigma_0, \quad (5.53)$$

$$\Gamma_2(\delta) = \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})'\} = \mathbf{B} + \delta\delta', \quad (5.54)$$

$$\begin{aligned} \Gamma_3(\delta) &= \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n^{PT} - \theta_{(n)})(\hat{\theta}_n^{PT} - \theta_{(n)})'\} \\ &= \Sigma_0 - \Sigma_0 H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) \\ &\quad + \delta\delta' \{2H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) - H_{p+3}(\chi_{p-1, \alpha}^2; \Delta)\}, \end{aligned} \quad (5.55)$$

$$\begin{aligned} \Gamma_4(\delta) &= \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n^S - \theta_{(n)})(\hat{\theta}_n^S - \theta_{(n)})'\} \\ &= \Sigma_0 - (p-3)\Sigma_0 \{2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ &\quad + (p-3)(p+1)\delta\delta' E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (5.56)$$

Consequently, the asymptotic distributional risk function under quadratic loss in (5.14) for the first four estimators are given by

$$R(\tilde{\theta}_n, \mathbf{W}) = \text{Tr}(\mathbf{W}\Sigma_0), \quad (5.57)$$

$$R(\hat{\theta}_n \mathbf{1}_p, \mathbf{W}) = \text{Tr}(\mathbf{W}\mathbf{B}) + \delta'\mathbf{W}\delta, \quad (5.58)$$

$$R(\hat{\theta}_n^{PT}, \mathbf{W}) = \text{Tr}(\mathbf{W}\Sigma_0) - \text{Tr}(\mathbf{W}\Sigma_0)H_{p+1}(\chi_{p-1, \alpha}^2; \Delta)$$

$$+ \delta' W \delta \{2H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta)\}, \quad (5.59)$$

$$\begin{aligned} R(\hat{\theta}_n^S, W) = & Tr(W\Sigma_0) - (p-3)Tr(W\Sigma_0) \{2E(\chi_{p+1}^{-2}(\Delta)) \\ & - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ & + (p-3)(p+1)\delta' W \delta E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (5.60)$$

Hence, using the identity

$$E(\chi_{p+2}^{-2}(\Delta)) - (p-2)E(\chi_{p+2}^{-4}(\Delta)) = \Delta E(\chi_{p+4}^{-4}(\Delta)),$$

the expression in (5.60) reduces to

$$\begin{aligned} R(\hat{\theta}_n^S, W) = & Tr(W\Sigma_0) - (p-3)Tr(W\Sigma_0) \{(p-3)E(\chi_{p+1}^{-4}(\Delta)) \\ & + \left[1 - \frac{(p+1)\delta' W \delta}{2\Delta Tr(W\Sigma_0)}\right] 2\Delta E(\chi_{p+3}^{-4}(\Delta))\}. \end{aligned} \quad (5.61)$$

For some numerical work and graphs we consider the particular case $W = \Sigma_0^{-1}$.

Then the expressions in (5.57)–(5.60) reduce to

$$R(\tilde{\theta}_n, \Sigma_0^{-1}) = p, \quad (5.62)$$

$$R(\hat{\theta}_n 1_p, \Sigma_0^{-1}) = 1 + \Delta, \quad (5.63)$$

$$\begin{aligned} R(\hat{\theta}_n^{PT}, \Sigma_0^{-1}) = & p - p H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) \\ & + \Delta \{2H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta)\}, \end{aligned} \quad (5.64)$$

$$\begin{aligned} R(\hat{\theta}_n^S, \Sigma_0^{-1}) = & p - p(p-3) \{2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ & + (p-3)(p+1)\Delta E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (5.65)$$

5.5 Analysis of the Estimators

5.5.1 ADR Analysis

We now proceed with an analysis of the risks of the estimators and determine their asymptotic dominance properties. Firstly, we note that

$$\frac{R(\hat{\theta}_n 1_p, W)}{R(\tilde{\theta}_n, W)} \stackrel{\leq}{>} 1 \quad \text{according as} \quad \delta' W \delta \stackrel{\leq}{>} Tr(WA), \quad (5.66)$$

where $A = \Sigma_0 - B$. Secondly,

$$\begin{aligned} \frac{R(\hat{\theta}_n^{PT}, W)}{R(\tilde{\theta}_n, W)} &\leq 1 \quad \text{according as} \\ &\delta' W \delta \{2H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) - H_{p+3}(\chi_{p-1, \alpha}^2; \Delta)\} \\ &\leq Tr(W \Sigma_0) H_{p+1}(\chi_{p-1, \alpha}^2; \Delta), \end{aligned} \quad (5.67)$$

and

$$\begin{aligned} \frac{R(\hat{\theta}_n^{PT}, W)}{R(\hat{\theta}_n 1_p, W)} &\leq 1 \quad \text{according as} \\ &\delta' W \delta \{1 - 2H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) + H_{p+3}(\chi_{p-1, \alpha}^2; \Delta)\} \\ &\leq Tr(WA) - Tr(W \Sigma_0) H_{p+1}(\chi_{p-1, \alpha}^2; \Delta). \end{aligned} \quad (5.68)$$

Finally, note that J is an idempotent matrix with rank $(p-1)$, and the rank of Σ_0 is greater than $(p-1)$. Thus, by the Courant theorem [Rao (1973)],

$$\frac{\delta' W \delta}{\delta' \Sigma_0^{-1} \delta} \leq ch_{\max}(W \Sigma_0), \quad \text{for all } \delta,$$

where $ch_{\max}(W)$ stands for the largest characteristic root of W . For convenience, we characterize a class $\mathcal{W}$ by

$$\mathcal{W} = \left\{ W_{p.s.d} : Tr(W \Sigma_0) / ch_{\max}(W \Sigma_0) \geq \frac{p+1}{2} \right\}. \quad (5.69)$$

Then, we obtain from (5.57), (5.60) and (5.69) the result:

$$\frac{R(\hat{\theta}_n^S, W)}{R(\tilde{\theta}_n, W)} \leq 1 \quad \forall \delta \quad \text{and} \quad W \in \mathcal{W}. \quad (5.70)$$

The proof of (5.70) follows by examining the following term in (5.61).

$$1 - \frac{(p+1)\delta' W \delta}{2\Delta Tr(W \Sigma_0)},$$

which must be greater than or equal to zero. Note that for the dominance of $\hat{\theta}_n^S$ over $\tilde{\theta}_n$ for all $p \geq 4$, we need the typical choice of W given by (5.69). For an arbitrary choice of W , the inequality in (5.69) may not hold simultaneously for

all Σ_0 . In such a case, to induce the dominance of $\hat{\theta}_n^S$ over $\tilde{\theta}_n$, one may introduce the modified Stein-estimator, $\hat{\theta}_n^{MS}$, given by

$$\hat{\theta}_n^{MS} = \hat{\theta}_n \mathbf{1}_p + [\mathbf{I}_p - cd_n L_n^{-1} \mathbf{W}^{-1} \hat{\Sigma}_n^{-1}] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (5.71)$$

where c is the shrinkage constant and $d_n = ch_{\min}(\mathbf{W}^{-1} \hat{\Sigma}_n^{-1})$. The modified Stein-estimator is similar to the normal theory case treated by Berger et al (1977) and to the nonparametric case discussed by Sen and Saleh (1985). Sen (1986) provides a detailed discussion of a general form of the shrinkage estimator, with arbitrary $\mathbf{W}$, which is also quite applicable to categorical data. In the sequel we assume $\mathbf{W}$ belongs to $\mathcal{W}$ given by (5.69) and the remaining discussion follows.

In the above we may note that $Tr(\mathbf{W}\mathbf{A}) > 0$ and

$$ch_{\min}(\mathbf{W}\Sigma_0) \leq \frac{\delta' \mathbf{W} \delta}{\Delta} \leq ch_{\max}(\mathbf{W}\Sigma_0),$$

where $ch_{\min}(\mathbf{W})$ stands for the smallest characteristic root of $\mathbf{W}$. Hence, (5.66) implies that $\hat{\theta}_n \mathbf{1}_p$ has a smaller ADR than that of the ADR of $\tilde{\theta}_n$ if $\Delta < \frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\max}(\mathbf{W}\Sigma_0)}$, i.e., $\hat{\theta}_n \mathbf{1}_p$ performs better than $\tilde{\theta}_n$ in the interval $[0, \frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\max}(\mathbf{W}\Sigma_0)})$. Alternatively, when Δ deviates from the null hypothesis beyond $\frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\min}(\mathbf{W}\Sigma_0)}$, the ADR of $\hat{\theta}_n \mathbf{1}_p$ grows and becomes unbounded whereas the ADR of $\tilde{\theta}_n$ remains bounded. Hence departure from the null hypothesis, i.e., $\Delta \in (\frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\min}(\mathbf{W}\Sigma_0)}, \infty)$ is fatal to $\hat{\theta}_n \mathbf{1}_p$ but is of very little concern to $\tilde{\theta}_n$. For $\mathbf{W} = \Sigma_0^{-1}$, we obtain $\frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\max}(\mathbf{W}\Sigma_0)} = \frac{Tr(\mathbf{W}\mathbf{A})}{ch_{\min}(\mathbf{W}\Sigma_0)} = p - 1$. These findings are confirmed in Figures 5.1 ($p=5$) and 5.3 ($p=20$).

Similarly, we note that $H_m(x; \Delta)$ is a decreasing function of m (DF) and Δ . In fact for fixed x we have

$$\lim_{m \rightarrow \infty} H_m(x; \Delta) = \lim_{\Delta \rightarrow \infty} H_m(x; \Delta) = 0,$$

and in particular, we have

$$H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) < H_{p-1}(\chi_{p-1, \alpha}^2; \Delta) \leq H_{p-1}(\chi_{p-1, \alpha}^2; 0) = 1 - \alpha, \quad (5.72)$$

for every $p \geq 2$, $0 < \alpha < 1$ and $\Delta > 0$. Using these results in (5.59), we may point out some important characteristics of the ADR of $\hat{\theta}_n^{PT}$. For $\delta = 0$, i.e., under

the null hypothesis, the ADR of $\hat{\theta}_n^{PT}$ reduces to $Tr(W\Sigma_0) \{1 - H_{p+1}(\chi_{p-1,\alpha}^2; 0)\}$ which is smaller than the ADR of $\tilde{\theta}_n$. Also, for large deviations of δ from 0 (null hypothesis), i.e., as $\Delta \rightarrow \infty$ the ADR of $\hat{\theta}_n^{PT}$ approaches to the ADR of $\tilde{\theta}_n$ from above. Similar results hold when $p \rightarrow \infty$. Further, as δ departs from 0, i.e., as Δ grows, the value of ADR of $\hat{\theta}_n^{PT}$ increases to a maximum after crossing the ADR of $\tilde{\theta}_n$, then decreases towards the ADR of $\tilde{\theta}_n$. Furthermore, we find that the ADR of the $\hat{\theta}_n^{PT}$ is smaller than the ADR of $\tilde{\theta}_n$ when

$$\Delta < \frac{Tr(W\Sigma_0)}{ch_{max}(W\Sigma_0)(2 - \nu_0)}, \quad \nu_0 = \frac{H_{p+3}(\chi_{p-1,\alpha}^2; \Delta)}{H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)} \leq 1$$

and for $\Delta > \frac{Tr(W\Sigma_0)}{ch_{min}(W\Sigma_0)(2 - \nu_0)}$, the opposite conclusion holds, but as soon as $\Delta \rightarrow \infty$, the ADR of $\hat{\theta}_n^{PT}$ converges to that of $\tilde{\theta}_n$. For $W = \Sigma_0^{-1}$, we obtain $\frac{Tr(W\Sigma_0)}{ch_{min}(W\Sigma_0)} = \frac{Tr(W\Sigma_0)}{ch_{max}(W\Sigma_0)} = p$. In this case, for $\Delta \in [0, p(2 - \nu_0)^{-1}]$, the PTE $\hat{\theta}_n^{PT}$ has a smaller ADR than that of $\tilde{\theta}_n$, whereas the opposite is true for $\Delta \in (p(2 - \nu_0)^{-1}, \infty)$.

Now, we compare $\hat{\theta}_n^{PT}$ and $\hat{\theta}1_p$. First, under the null hypothesis, i.e., $\delta = 0$ the ADR of $\hat{\theta}1_p$ is $Tr(W\Sigma_0) - Tr(WA)$ and $R(\hat{\theta}_n^{PT}, W) - R(\hat{\theta}1_p, W) > 0$ whenever

$$H_{p+1}(\chi_{p-1,\alpha}^2; 0) < \frac{Tr(WA)}{Tr(W\Sigma_0)}. \quad (5.73)$$

Hence, we conclude that under the null hypothesis $\hat{\theta}1_p$ performs better than $\hat{\theta}_n^{PT}$ for a range of α for which (5.73) is satisfied. However, for a fixed δ as $\alpha \rightarrow 0$, the ADR of $\hat{\theta}_n^{PT}$ approach $\delta'W\delta$ while the ADR of $\hat{\theta}1_p$ remains unchanged. In such a case $\hat{\theta}1_p$ performs poorly than $\hat{\theta}_n^{PT}$. In the sequel we assume α to be bounded away from 0 satisfying (5.73) and the remaining discussion follows. Then, $\hat{\theta}1_p$ performs better than $\hat{\theta}_n^{PT}$ for

$$\Delta < \frac{Tr(WA) - Tr(W\Sigma_0)H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)}{ch_{max}(W\Sigma_0)\{2(1 - H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)) - (1 - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta))\}}$$

and for

$$\Delta > \frac{Tr(WA) - Tr(W\Sigma_0)H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)}{ch_{min}(W\Sigma_0)\{2(1 - H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)) - (1 - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta))\}}$$

the opposite conclusion holds. For $W = \Sigma_0^{-1}$, the right-hand quantities are equal to

$$\frac{p\{1 - H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)\} - 1}{2(1 - H_{p+1}(\chi_{p-1,\alpha}^2; \Delta)) - (1 - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta))}.$$

For large p , this expression approaches $p-1$. This is consistent with the Figures 5.1 and 5.3 and Tables 5.1 and 5.2. Considering the discussion above we conclude that none of the three estimators $\tilde{\theta}_n$, $\hat{\theta}_n 1_p$, and $\hat{\theta}_n^{PT}$ is asymptotically inadmissible with respect to the other two.

The picture is somewhat different with respect to the shrinkage estimator. First, note from (5.70), that the ADR of $\hat{\theta}_n^S$ is uniformly smaller than the ADR of $\tilde{\theta}_n$, where the upper limit is attained when $\Delta \rightarrow \infty$. This shows the asymptotic inadmissibility of $\tilde{\theta}_n$ (under $K_{(n)}$) relative to $\hat{\theta}_n^S$. To compare $\hat{\theta}_n^S$ and $\hat{\theta}_n 1_p$ under the null-hypothesis in (5.2) ($\delta = 0$) we have

$$\begin{aligned} R(\hat{\theta}_n^S, W) &= R(\hat{\theta}_n 1_p, W) + Tr(WA) - \frac{(p-3)}{(p-1)} Tr(W\Sigma_0) \\ &> R(\hat{\theta}_n 1_p, W). \end{aligned} \quad (5.74)$$

Thus, the asymptotic risk of $\hat{\theta}_n^S$ is greater than that of $\hat{\theta}_n 1_p$ for $\delta = 0$. However, as δ moves away from 0, $\delta'W\delta$ increases while $E(\chi_{p+1}^{-4}(\Delta))$ and $E(\chi_{p+3}^{-4}(\Delta))$ decrease, so the opposite conclusion holds. In general, $\hat{\theta}_n^S$ does not dominate $\hat{\theta}_n 1_p$ for Δ in the interval $[0, \Delta_*)$, where

$$\Delta_* = \frac{Tr(WA) - (p-3)Tr(W\Sigma_0) \{E(\chi_{p+1}^{-2}(\Delta)) + \Delta E(\chi_{p+3}^{-4}(\Delta))\}}{ch_{max}(W\Sigma_0) \{1 - (p-3)(p+1)E(\chi_{p+3}^{-4}(\Delta))\}},$$

however, it does so for $\Delta \in (\Delta^*, \infty)$ where

$$\Delta^* = \frac{Tr(WA) - (p-3)Tr(W\Sigma_0) \{E(\chi_{p+1}^{-2}(\Delta)) + \Delta E(\chi_{p+3}^{-4}(\Delta))\}}{ch_{min}(W\Sigma_0) \{1 - (p-3)(p+1)E(\chi_{p+3}^{-4}(\Delta))\}}.$$

Hence neither $\hat{\theta}_n^S$ nor $\hat{\theta}_n 1_p$ asymptotically dominates the other (under $K_{(n)}$).

Now, we compare $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$. First, consider the case when $\delta = 0$. In this case $R(\hat{\theta}_n^S, W) = 2(p-1)^{-1}Tr(W\Sigma_0)$. Hence we have

$$R(\hat{\theta}_n^S, W) - R(\hat{\theta}_n^{PT}, W) = Tr(W\Sigma_0) \left\{ H_{p+1}(\chi_{p-1, \alpha}^2; 0) - \frac{p-3}{p-1} \right\}, \quad p \geq 4. \quad (5.75)$$

Thus the PTE $\hat{\theta}_n^{PT}$ has a smaller ADR than that of the ADR of $\hat{\theta}_n^S$ at $\delta = 0$ whenever

$$H_{p+1}(\chi_{p-1, \alpha}^2; 0) > \frac{p-3}{p-1}, \quad (5.76)$$

otherwise the opposite conclusion holds. Thus, $\hat{\theta}_n^S$ does not dominate $\hat{\theta}_n^{PT}$ when the null hypothesis holds provided that $H_{p+1}(\chi_{p-1,\alpha}^2; 0) > (p-3)/(p-1)$. Under this situation we can order dominance of the estimators under the null-hypothesis in (5.2) as follows

$$\hat{\theta}_n 1_p \succ \hat{\theta}_n^{PT} \succ \hat{\theta}_n^S \succ \tilde{\theta}_n, \quad (5.77)$$

where the notation $\succ$ means dominates. However, the dominance picture changes to $\hat{\theta}_n 1_p \succ \hat{\theta}_n^S \succ \hat{\theta}_n^{PT} \succ \tilde{\theta}_n$, whenever $H_{p+1}(\chi_{p-1,\alpha}^2; 0) < \frac{p-3}{p-1}$. Figure 5.3 represents this situation. Further, we can specify the condition in (5.76) equivalently as

$$2h_{p+1}(\chi_{p-1,\alpha}^2; 0) \leq \frac{2}{p-1} + \alpha \quad (5.78)$$

where $h_{p+1}(\cdot; 0)$ is the p.d.f of a chi-square variable with $(p+1)$ DF and α is the level of significance of the test of the null-hypothesis in (5.2). Thus, (5.78) specifies a range of the values of α for given $p \geq 4$ for which the PTE $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$. The picture changes as δ moves away from the origin (i.e., under the null-hypothesis). Note that $\tilde{\theta}_n$ has constant risk equal to $Tr(W\Sigma_0)$ and the risk of $\hat{\theta}_n 1_p$ depends on δ ; its risk becomes unbounded as δ moves further away from 0. As for $\hat{\theta}_n^{PT}$, the risk increases from the initial value $Tr(W\Sigma_0)\{1 - H_{p+1}(\chi_{p-1,\alpha}^2; 0)\}$ to a maximum after crossing the risk of $\tilde{\theta}_n$, then decreases to the value $Tr(W\Sigma_0)$ which is the ADR of $\tilde{\theta}_n$. Similarly, the risk of $\hat{\theta}_n^S$ with the initial value $\frac{2}{p-1}Tr(W\Sigma_0)$ increases monotonically towards $Tr(W\Sigma_0)$ as Δ moves away from 0. The ADR of $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$ intersect at the point $\Delta = \Delta_\alpha$ for each α ($0 < \alpha < 1$), if the condition in (5.76) is satisfied, otherwise there is no intersection. If $\Delta \in [0, \Delta_\alpha)$, then the PTE dominates the SE while outside this interval the SE dominates the PTE. Thus under the condition in (5.76) neither the PTE $\hat{\theta}_n^{PT}$ nor the SE $\hat{\theta}_n^S$ dominates each other under $\{K_{(n)}\}$. The SE $\hat{\theta}_n^S$ dominates the PTE $\hat{\theta}_n^{PT}$ uniformly if the condition in (5.76) is not satisfied. However, $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ share a common property as $\Delta \rightarrow \infty$, namely, their ADR's converge to $Tr(W\Sigma_0)$ which is the risk of $\tilde{\theta}_n$. But the risk function of $\hat{\theta}_n^S$ is always below this limit, whereas the risk function of $\hat{\theta}_n^{PT}$ exceeds this limit for some intermediate values of Δ depending on α , the level of significance of the test of (5.2).

To improve over $\hat{\theta}_n^{PT}$ ($\hat{\theta}_n^S$) one may use $\hat{\theta}_n^{PT+}$ ($\hat{\theta}_n^{S+}$) as studied by Sclove, Morris and Radhakrishnan(1972) in the multivariate normal mean problem. This relative dominance of (5.10) over (5.9) remains intact under our asymptotic setup, hence, we may advocate the use of this modified SE. In terms of ordering, MPTE comes in between the usual PTE in (5.8) and SE, while in general, the relative ADR picture with PTE and SE remains applicable to this case as well. Finally, from the point of robust-efficiency both $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ may be advocated, leaning more towards $\hat{\theta}_n^S$, since the size of Δ is generally unknown and unlikely to be small.

5.5.2 Efficiency Analysis

Next, we consider the relative efficiency analysis of the estimators, for the specific case $W = \Sigma_0^{-1}$. Using (5.62)–(5.65), the relative efficiency (R.E.) of the RMLE $\hat{\theta}_n 1_p$ to the UMLE $\tilde{\theta}_n$ is defined to be

$$R.E.(\hat{\theta}_n 1_p : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \Sigma_0^{-1})}{R(\hat{\theta}_n 1_p, \Sigma_0^{-1})} = \frac{p}{1 + \Delta}. \quad (5.79)$$

Similarly, the relative efficiency of $\hat{\theta}_n^{PT}$ and the SE $\hat{\theta}_n^S$ to the UMLE $\tilde{\theta}_n$ are given by (5.80)–(5.81) respectively:

$$R.E.(\hat{\theta}_n^{PT} : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \Sigma_0^{-1})}{R(\hat{\theta}_n^{PT}, \Sigma_0^{-1})} = \frac{1}{1 + g_1(\alpha, \Delta)}, \quad (5.80)$$

where

$$g_1(\alpha, \Delta) = \frac{\Delta}{p} \{2H_{p+1}(\chi_{p-1, \alpha}^2; \Delta) - H_{p+3}(\chi_{p-1, \alpha}^2; \Delta)\} - H_{p+1}(\chi_{p-1, \alpha}^2; \Delta).$$

$$R.E.(\hat{\theta}_n^S : \tilde{\theta}_n) = \frac{R(\tilde{\theta}_n, \Sigma_0^{-1})}{R(\hat{\theta}_n^S, \Sigma_0^{-1})} = \frac{1}{1 + g_2(\Delta)}, \quad (5.81)$$

where

$$g_2(\Delta) = \frac{\Delta}{p} (p-3)(p+1)E(\chi_{p+3}^{-4}(\Delta)) - (p-3) \{2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))\}.$$

Tables 5.1–5.2 provide the maximum relative efficiencies (at $\Delta = 0$) for $\hat{\theta}_n 1_p$ (E_1), $\hat{\theta}_n^{PT}$ (E_3) and $\hat{\theta}_n^S$ (E_4) relative to $\tilde{\theta}_n$, the intersecting relative efficiencies

denoted by E_{Δ_α} of $\hat{\theta}_n^S$ and $\hat{\theta}_n^{PT}$ (which is the common efficiency of the PTE and the SE) and the Δ_α -values at which the intersection occurs. For example, using Table 5.1, with $p=6$, $\alpha = 0.10$, and $\Delta \in [0, 1.1514]$, it appears that $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$, but outside the interval domination is reversed. While, if the hypothesis in (5.2) is correct, i.e., $\Delta = 0$, then $\hat{\theta}_n 1_p$ has larger efficiency ($E_1 = 6.0000$) than $\hat{\theta}_n^{PT}$ ($E_3 = 4.2350$) and $\hat{\theta}_n^S$ ($E_4 = 2.5000$) relative to $\tilde{\theta}_n$. Note that $\Delta_\alpha = 0$ in Tables 5.1–5.2 means that no intersection between $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ occurs, and the corresponding relative efficiencies E_{Δ_α} are equal to E_3 , i.e., the SE $\hat{\theta}_n^S$ in (5.9) dominates the PTE $\hat{\theta}_n^{PT}$ in (5.8) when p -values or α -values are large. Figures 5.1–5.2 exhibit plots of the relative efficiency (R.E.) curves for $\hat{\theta}_n 1_p$, $\hat{\theta}_n^{PT}$, and $\hat{\theta}_n^S$ relative to $\tilde{\theta}_n$ versus Δ for various values of α and $p=5$. Apparently, $\hat{\theta}_n^{PT}$ dominates $\hat{\theta}_n^S$ for a range of α ($0 < \alpha < 1$) and for small values of Δ . Figure 5.3 exhibits plots of the relative efficiency (R.E.) curves for $\hat{\theta}_n 1_p$, $\hat{\theta}_n^{PT}$, and $\hat{\theta}_n^S$ relative to $\tilde{\theta}_n$ versus Δ for various values of α and $p=20$. Apparently, $\hat{\theta}_n^S$ dominates $\hat{\theta}_n^{PT}$ for large values of p .

5.6 Baseball Data Analysis

Now, we illustrate the above theory with an analysis of the famous baseball data of Efron and Morris (1975). In this data a James–Stein estimator is used to predict batting averages of 18 major league players in the remainder of the 1970 season. The number of hits y_i in the first 45 bats is observed for each player i , ($i = 1, 2, \dots, 18$). The problem is to estimate $\theta = (\theta_1, \dots, \theta_{18})'$, where θ_i denotes the final season batting average of player i . Efron and Morris (1975) used an arc sine transformation on each y_i to obtain approximate normality and then used a James–Stein estimator on the transformed counts, denoted by $\hat{\theta}_n^{EM}$. Also, Morris (1983) proposed an empirical Bayes estimator denoted by $\hat{\theta}_n^M$ to estimate θ in this problem. Albert (1984) proposed an empirical Bayes estimator given by (5.13) to estimate θ . Note that for this example L_n in (5.7) is equal to 19.045 and $\hat{\theta}_n = 0.2654$. We present in Table 5.3 the true batting averages (θ^T), the UMLE ($\tilde{\theta}_n$), $\hat{\theta}_n^{EM}$, $\hat{\theta}_n^M$, $\hat{\theta}_n^{EB}$, $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$. The true batting averages, obtained from

Morris (1983), have mean 0.267 and s.d. 0.037. In order to assess the performance of the various estimators, loss defined by $(\hat{\theta}_i - \theta_i^T)^2 / (0.037)^2$ for each estimator $\hat{\theta}_i$, $i = 1, \dots, 18$ along with the average loss (due to the estimators) have been tabulated in Table 5.4. The PTE $\hat{\theta}_n^{PT} = (.265, \dots, .265)'$ is the result of testing the null hypothesis $H_0 : \theta_1 = \theta_2 = \dots = \theta_p$ at 15% level of significance. In Table 5.4 we have denoted by L_{MLE} , L_{EM} , L_M , L_{EB} , L_{PT} , and L_S the losses due to the estimators, $\tilde{\theta}_n$ (UMLE), $\hat{\theta}_n^{EM}$, $\hat{\theta}_n^M$, $\hat{\theta}_n^{EB}$, $\hat{\theta}_n^{PT}$ and $\hat{\theta}_n^S$ respectively. A comparison of the average loss of various estimators show that our estimator $\hat{\theta}_n^S$ is quite comparable to that of $\hat{\theta}_n^{EM}$, $\hat{\theta}_n^M$ and $\hat{\theta}_n^{EB}$, (and $\hat{\theta}_n^{PT}$ is not far behind). If we apply Morris's ((1983), (3.8) page 32) approximation formula, $\hat{\theta}_i^M = .208\tilde{\theta}_i + 0.2105$ obtained by an empirical Bayes approach we get exactly $\hat{\theta}_i^S$ ($i = 1, \dots, 18$), which supports the validity of our asymptotic theory as well.

Figure 5.1: Dominance picture for UMLE, RMLE, PTE and SE for $p=5$

Figure 5.2: Dominance picture for UMLE, RMLE, PTE and SE for $p=5$

Figure 5.3: Dominance picture for UMLE, RMLE, PTE and SE for $p=20$

Table 5.1: Maximum relative efficiencies of the RMLE, PTE, and SE and the intersecting efficiencies for the PTE and SE for each α with corresponding Δ_α -values for p-values

$\alpha \backslash p$		4	6	8	10	12	14	16
.05	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	5.9971	7.3761	8.3363	9.0666	9.6514	10.1362	10.5481
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.05}}$	1.2617	1.8128	2.4102	3.0680	3.8010	4.6266	5.5660
	$\Delta_{.05}$	2.0988	1.8974	1.7020	1.5048	1.3048	1.1025	.8983
.10	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	3.5396	4.2350	4.7092	5.0648	5.3466	5.5783	5.7739
	E_4	1.5000	2.500	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.10}}$	1.2913	1.9996	2.9013	4.0761	5.3466	5.5783	5.7739
	$\Delta_{.10}$	1.6932	1.1514	.7081	.3048	.0000	.0000	.0000
.15	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	2.6425	3.1003	3.4085	3.6375	3.8177	3.9652	4.0891
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.15}}$	1.3199	2.2163	3.4085	3.6375	3.8177	3.9652	4.0891
	$\Delta_{.15}$	1.3593	.5482	.0000	.0000	.0000	.0000	.0000
.20	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	2.1684	2.5038	2.7274	2.8925	3.0218	3.1272	3.2155
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.20}}$	1.3511	2.4973	2.7274	2.8925	3.0218	3.1272	3.2155
	$\Delta_{.20}$	1.0462	.0044	.0000	.0000	.0000	.0000	.0000
.25	E_1	4.0000	6.0000	8.0000	10.0000	12.0000	14.0000	16.0000
	E_3	1.8729	2.1328	2.3049	2.4314	2.5300	2.6101	2.6771
	E_4	1.5000	2.5000	3.5000	4.5000	5.5000	6.5000	7.5000
	$E_{\Delta_{.25}}$	1.3873	2.1328	2.3049	2.4314	2.5300	2.6101	2.6771
	$\Delta_{.25}$	.7340	.0000	.0000	.0000	.0000	.0000	.0000

Table 5.3: The true values (θ_i^T), the unrestricted estimator, Efron and Morris estimators, Morris estimators, Albert empirical Bayes estimators and the proposed estimators ($\hat{\theta}_i^{PT}$) and ($\hat{\theta}_i^S$)

i	θ_i^T	$\tilde{\theta}_i$	$\hat{\theta}_i^{EM}$	$\hat{\theta}_i^M$	$\hat{\theta}_i^{EB}$	$\hat{\theta}_i^{PT}$	$\hat{\theta}_i^S$
1	0.346	0.400	0.290	0.308	0.279	0.265	0.294
2	0.300	0.378	0.286	0.301	0.277	0.265	0.289
3	0.279	0.356	0.281	0.295	0.274	0.265	0.245
4	0.223	0.333	0.277	0.288	0.272	0.265	0.280
5	0.276	0.311	0.273	0.281	0.270	0.265	0.275
6	0.273	0.311	0.273	0.281	0.270	0.265	0.275
7	0.266	0.289	0.268	0.274	0.268	0.265	0.270
8	0.211	0.267	0.264	0.267	0.265	0.265	0.266
9	0.271	0.244	0.259	0.260	0.263	0.265	0.261
10	0.232	0.244	0.259	0.260	0.263	0.265	0.261
11	0.266	0.222	0.254	0.252	0.261	0.265	0.256
12	0.258	0.222	0.254	0.252	0.261	0.265	0.256
13	0.306	0.222	0.254	0.252	0.261	0.265	0.256
14	0.267	0.222	0.254	0.252	0.261	0.265	0.256
15	0.228	0.222	0.254	0.252	0.261	0.265	0.256
16	0.288	0.200	0.249	0.244	0.259	0.265	0.252
17	0.318	0.178	0.233	0.236	0.257	0.265	0.247
18	0.200	0.156	0.208	0.227	0.254	0.265	0.242
<i>Mean</i>	0.267	0.265	0.261	0.266	0.265	0.265	0.263
<i>s.d.</i>	0.037	0.068	0.019	0.022	0.007	0.000	0.014

Table 5.4: Computed values of the losses of the UMLE, Efron and Morris estimator, Morris estimator, Albert estimator and the proposed estimators ($\hat{\theta}_i^{PT}$) and ($\hat{\theta}_i^S$)

i	L_{MLE}	L_{EM}	L_M	L_{EB}	L_{PT}	L_S
1	2.130	2.291	1.055	3.279	4.793	1.975
2	4.444	0.143	0.001	0.386	0.895	0.088
3	4.331	0.003	0.187	0.018	0.143	0.844
4	8.839	2.130	3.086	1.754	1.289	2.373
5	0.895	0.007	0.018	0.026	0.088	0.001
6	1.055	0.090	0.047	0.007	0.047	0.003
7	0.386	0.003	0.047	0.003	0.001	0.012
8	2.291	2.052	2.291	2.130	2.130	2.210
9	0.533	0.105	0.088	0.047	0.026	0.073
10	0.105	0.533	0.573	0.702	0.795	0.614
11	1.414	0.105	0.143	0.018	0.001	0.073
12	0.947	0.012	0.026	0.007	0.036	0.003
13	5.154	1.975	2.130	1.479	1.228	1.826
14	1.479	0.123	0.164	0.026	0.003	0.088
15	0.026	0.494	0.421	0.795	1.000	0.573
16	5.657	1.111	1.414	0.614	0.386	0.947
17	14.317	5.278	4.912	2.718	2.052	3.682
18	1.414	0.047	0.533	2.130	3.086	1.289
Mean	3.079	0.912	0.952	0.897	1.000	0.926
s.d.	3.565	1.346	1.315	1.045	1.270	1.051

Chapter 6

Preliminary Test and Shrinkage Estimation for Simple Correlation

6.1 Introduction

An important concept that will be of interest in studying regression theory and in studying relationships among random variables is correlation. In this chapter we discuss the preliminary test and shrinkage estimation for correlation coefficients.

Let $(x_{i1}, y_{i1}), \dots, (x_{in_i}, y_{in_i})$ be a random sample of size n_i from the i^{th} bivariate normal distribution $N(\mu_{i1}, \mu_{i2}, \sigma_{i1}^2, \sigma_{i2}^2, \rho_i)$; $i = 1, 2, \dots, p$. Then the maximum likelihood estimator (MLE) of $\rho = (\rho_1, \dots, \rho_p)'$ is given by

$$\mathbf{r} = (r_1, r_2, \dots, r_p)', \quad (6.1)$$

where

$$r_i = \frac{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)}{\sqrt{\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}}, \quad i = 1, \dots, p;$$

and

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad \bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.$$

We are interested in the estimation of ρ when it is suspected that we may have

uncertain prior information

$$\rho_1 = \rho_2 = \cdots = \rho_p = \rho, \quad \text{unknown.} \quad (6.2)$$

We shall find it convenient to formulate this problem in an asymptotic setup.

Along with preliminary notions, the proposed estimators are all presented in Section 6.2. The concept of asymptotic risk (AR) and its relation to asymptotic distributional risk (ADR) and Asymptotic bias (AB) are presented in Section 6.3.

6.2 The Proposed Estimators

Let us define

$$\theta = (\theta_1, \dots, \theta_p)', \quad \text{where} \quad \theta_i = \frac{1}{2} \log \frac{1 + \rho_i}{1 - \rho_i},$$

and the null hypothesis H_0 by

$$H_0 : \theta_1 = \theta_2 = \cdots = \theta_p = \theta_0, \quad \text{unknown.} \quad (6.3)$$

Then the unrestricted maximum likelihood estimator (UMLE) of θ is given by

$$\tilde{\theta}_n = (\tilde{\theta}_1, \dots, \tilde{\theta}_p)', \quad \tilde{\theta}_i = \frac{1}{2} \log \frac{1 + r_i}{1 - r_i}, \quad i = 1, 2, \dots, p; \quad (6.4)$$

while under the model (6.3), the restricted MLE (RMLE) is given by

$$\hat{\theta}_n = (\hat{\theta}_n, \dots, \hat{\theta}_n)' = \hat{\theta}_n \mathbf{1}_p; \quad (6.5)$$

where

$$\hat{\theta}_n = \sum_{i=1}^p (n_i - 3) \tilde{\theta}_i / n, \quad n = \sum_{i=1}^p (n_i - 3), \quad \mathbf{1}_p = (1, 1, \dots, 1)'$$

If the bias in mean $\tilde{\theta}_n$ can be neglected, the test for the hypothesis in (6.2), i.e., $\rho_1 = \dots = \rho_p$ is equivalent to the test of equality of the mean values of $\tilde{\theta}_n$, i.e., the hypothesis in (6.3). Furthermore, $\tilde{\theta}_i$ is distributed as normal with mean θ_i and variance $(n_i - 3)^{-1}$. i.e., $\tilde{\theta}_i \sim N(\theta_i, \frac{1}{n_i - 3})$. For a proof of these results, see Rao (1973) pages no. 433-434..

When H_0 holds, $\hat{\theta}_n$ has a smaller risk (under quadratic loss) than $\tilde{\theta}_n$. On the other hand, when H_0 does not hold, $\tilde{\theta}_n$ may perform better than $\hat{\theta}_n$. As a result, when the prior information on H_0 is rather uncertain it may be desirable to have either a preliminary test estimator (PTE) denoted by $\hat{\theta}_n^{PT} = (\hat{\theta}_1^{PT}, \dots, \hat{\theta}_p^{PT})'$ or a shrinkage estimator (SE) denoted by $\hat{\theta}_n^S = (\hat{\theta}_1^S, \dots, \hat{\theta}_p^S)'$ of θ . In the case of a preliminary test the estimator $\hat{\theta}_n$ or $\tilde{\theta}_n$ is chosen according as H_0 is accepted or rejected. The SE is based on the classical Stein-rule. Unlike the case of $\hat{\theta}_n$, both the PTE and SE behave robustly against departures from the null hypothesis in (6.3) and have bounded risk even when n is very large. From these four basic estimators two more improved estimators may be defined; namely, Stein's positive-rule shrinkage estimator (PRSE), denoted by $\hat{\theta}_n^{S+} = (\hat{\theta}_1^{S+}, \dots, \hat{\theta}_p^{S+})'$ and the modified PTE (MPTE) denoted by $\hat{\theta}_n^{PT+} = (\hat{\theta}_1^{PT+}, \dots, \hat{\theta}_p^{PT+})'$. The primary objective of this study is to focus on the PTE and SE and to compare them with the UMLE and RMLE.

For the preliminary test (PT) on H_0 in (6.3), we consider the chi-square test statistic

$$\begin{aligned} L_n &= \sum_{i=1}^p (\tilde{\theta}_i - \hat{\theta}_n)^2 (n_i - 3) \\ &= n(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p)' \Lambda (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \end{aligned} \quad (6.6)$$

where

$$\Lambda = \text{Diag}(\lambda_i), \quad \lambda_{n_i} = \frac{n_i - 3}{n} \quad \text{and} \quad \lim_{n \rightarrow \infty} \lambda_{n_i} = \lambda_i.$$

Under H_0 , for large n , L_n closely follows the central chi-square distribution with $(p-1)$ degrees of freedom (DF). Thus, for a given level of significance α ($0 < \alpha < 1$), the critical value of L_n may be approximated by $\chi_{p-1, \alpha}^2$, the upper $100\alpha\%$ point of the chi-square distribution with $(p-1)$ DF. Then, the PTE of θ is defined by

$$\hat{\theta}_n^{PT} = \hat{\theta}_n \mathbf{1}_p + [1 - I(L_n \leq \chi_{p-1, \alpha}^2)] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (6.7)$$

where $I(A)$ is the indicator function for the set A .

To introduce the SE we propose

$$\hat{\theta}_n^S = \hat{\theta}_n \mathbf{1}_p + [1 - (p-3)L_n^{-1}] (\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p), \quad (6.8)$$

where we need to assume that $p > 3$.

Note that whereas in (6.7) we have a dichotomous indicator function, in (6.8), $(p-3)L_n^{-1}$ may be regarded as a smoother version. Thus, the SE might again be considered from a "pre-test" point of view. Also, (6.7) is a convex combination of $\tilde{\theta}_n$ and $\hat{\theta}_n$, while (6.8) need not be especially because $(p-3)L_n^{-1}$ may be greater than 1. In the spirit of Sclove et al. (1972), we may introduce the following the *positive-rule* SE (PRSE) :

$$\hat{\theta}_n^{S+} = \hat{\theta}_n 1_p + [1 - (p-3)L_n^{-1}]^+ (\tilde{\theta}_n - \hat{\theta}_n 1_p), \quad (6.9)$$

where $y^+ = \max(0, y)$. Note that (6.7) and (6.9) may not agree even if we let $\chi_{p-1, \alpha}^2 = p-3$. The justification of (6.7) lies in the preliminary test of H_0 , whereas (6.9) is intended to control possible overshrinking in (6.8).

Finally, we propose the MPTE of θ as follows:

$$\hat{\theta}_n^{PT+} = \hat{\theta}_n 1_p + [1 - (p-3)L_n^{-1}] I(L_n \geq \chi_{p-1, \alpha}^2) (\tilde{\theta}_n - \hat{\theta}_n 1_p), \quad (6.10)$$

where the restriction $\chi_{p-1, \alpha}^2 < (p-3)$ is required. This estimator uniformly dominates $\hat{\theta}_n^{PT}$ but now we have the restriction $p \geq 4$. For $p < 4$ we are left with the PTE to improve over $\hat{\theta}_n$ or $\tilde{\theta}_n$. It is well-known that the positive-rule estimator PRSE dominates the usual shrinkage estimator SE, while the MPTE dominates the usual PTE.

6.3 Asymptotic Results

6.3.1 Pitman Alternatives

Here, we shall consider the asymptotic risks for the UMLE, RMLE, PTE and SE of θ . Let δ_n be an estimator of θ , W be a positive semi-definite (p.s.d) matrix, and consider the *quadratic loss function*

$$L_n(\delta_n, \theta) = n(\delta_n - \theta)'W(\delta_n - \theta) = n\text{Tr}[W(\delta_n - \theta)(\delta_n - \theta)'], \quad (6.11)$$

where $\text{Tr}(A)$ =trace of A . Then the risk of δ_n is given by

$$R_n(\delta_n, \theta) = EL_n(\delta_n, \theta) = \text{Tr}(WV_n), \quad (6.12)$$

where

$$V_n = nE(\delta_n - \theta)(\delta_n - \theta)'$$

For the p -dimensional space Ω , let $\omega(\subset \Omega)$ be the subspace for which θ satisfies the hypothesis in (6.3). Then, by virtue of the consistency of the L_n test, we note that for any fixed $\theta \notin \omega$, $L_n \xrightarrow{P} +\infty$, as $n \rightarrow \infty$ ($\xrightarrow{P}$ means convergence in probability). Thus, for any fixed $\theta \notin \omega$, as n increases $\hat{\theta}_n^{PT}$, $\hat{\theta}_n^S$, $\hat{\theta}_n^{S+}$ and $\hat{\theta}_n^{PT+}$ are equivalent in probability to the UMLE $\tilde{\theta}_n$, while $\hat{\theta}_n$ will have unbounded risk. The situation is different, however, when $\theta \in \omega$, i.e., θ belongs to a shrinking neighborhood of ω . For this reason, we consider a sequence $\{K_{(n)}\}$ of Pitman-alternatives:

$$K_{(n)}: \theta_{(n)} = \theta_0 \mathbf{1}_p + n^{-\frac{1}{2}} \xi; \quad \xi = (\xi_1, \dots, \xi_p)', \quad \theta_0 \mathbf{1}_p \in \omega, \quad (6.13)$$

where the ξ_i are fixed numbers. Note that for $\xi = 0$, $\theta_{(n)} = \theta_0 \mathbf{1}_p \in \omega$, so that (6.3) is a particular case of (6.13).

For the estimator δ_n , the asymptotic risk (AR) under $K_{(n)}$ is defined by

$$R(\delta; \xi) = \lim_{n \rightarrow \infty} R_n(\delta_n; \theta_{(n)}), \quad (6.14)$$

whenever the limit exists. We introduce the asymptotic distribution function (ADF) of δ_n by

$$G_{(\delta)}^*(\mathbf{x}) = \lim_{n \rightarrow \infty} P \{ n^{\frac{1}{2}}(\delta_n - \theta_{(n)}) \leq \mathbf{x} | K_{(n)} \}, \quad (6.15)$$

whenever the limit exists. We let

$$V_{(\delta)}^* = \int \dots \int \mathbf{x} \mathbf{x}' dG_{(\delta)}^*(\mathbf{x}). \quad (6.16)$$

Then, the asymptotic distributional risk (ADR) of δ_n is defined by

$$R^*(\delta, \xi) = \text{Tr}(\mathbf{W} V_{(\delta)}^*). \quad (6.17)$$

One may need extra regularity conditions to evaluate (6.14) rather than (6.17). This point has been explained in detail in various other contexts, e.g., in Sen (1984), Sen and Saleh (1985) and Saleh and Sen (1985), among others. Fortunately, in this specific context, we shall see that both (6.14) and (6.17) can be evaluated under the same set of regularity conditions, although it is a lot simpler to work with (6.17). Now, we shall study the ADR results for the various estimators.

6.3.2 Asymptotic Theory

Next theorem provides the asymptotic relation between the estimators. The derivation of this theorem is very similar to one in Chapter 5.

Theorem 6.1 *For any fixed $\theta \notin \omega$, $\hat{\theta}_n \mathbf{1}_p$ has asymptotic risk $+\infty$, while $\hat{\theta}_n^S$, $\hat{\theta}_n^{S+}$, $\hat{\theta}_n^{PT}$, $\hat{\theta}_n^{PT+}$, and $\tilde{\theta}_n$ are asymptotically risk-equivalent, with bounded risk.*

Next, we consider case of Pitman alternatives in (6.13). As we pointed out earlier we shall mainly study ADR properties though similar AR properties hold. Note that $\hat{\theta}_n \mathbf{1}_p$ and $\tilde{\theta}_n$ both have non-negative elements bounded by 1, so finite moments of $\sqrt{n}(\tilde{\theta}_n - \theta)$ and $\sqrt{n}(\hat{\theta}_n \mathbf{1}_p - E\hat{\theta}_n \mathbf{1}_p)$ of all finite orders exist and have finite limit as $n \rightarrow \infty$. Also, under $\{K_{(n)}\}$ in (6.13),

$$E[\hat{\theta}_n \mathbf{1}_p | K_{(n)}] - \theta = O(n^{-\frac{1}{2}}),$$

so that in this case, $n\|\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p\|^2 = O(1)$, in L_∞ norm. Thus, for (6.7)–(6.10), the convergence in distribution will ensure the convergence in second moment, and hence (6.17) will ensure (6.14). For the SE in (6.8), PRSE in (6.9) and MPTE in (6.10) we need a more refined treatment. Thus for our purpose it suffices to consider the ADR in (6.17) through the asymptotic distribution in (6.15). We work out the ADR results in this section. For the study of the asymptotic distribution theory and the ADR results, we assume the usual regularity conditions hold. Further, we define $\theta_0 \mathbf{1}_p$ as in (6.13) so that $\theta_0 \mathbf{1}_p \in \omega$, then we have the following theorem.

Theorem 6.2 *Under $\{K_{(n)}\}$ in (6.13) and assumed regularity conditions, the following hold*

$$(i) \quad \mathbf{x}_n = \sqrt{n}(\tilde{\theta}_n - \theta_0 \mathbf{1}_p) \xrightarrow{D} N_p(\xi, \Lambda^{-1}), \quad \text{as } n \rightarrow \infty, \quad (6.18)$$

where Λ defined by (6.6);

$$(ii) \quad \mathbf{y}_n = \sqrt{n}(\tilde{\theta}_n - \hat{\theta}_n \mathbf{1}_p) \xrightarrow{D} N_p(\mathbf{J}\xi, \Lambda^{-1}\mathbf{J}'), \quad \text{as } n \rightarrow \infty, \quad (6.19)$$

where $\mathbf{J} = \mathbf{I}_p - \mathbf{1}_p \mathbf{1}_p' \Lambda$, $[\mathbf{1}_p]$ defined by (6.5);

$$(iii) \quad \mathbf{z}_n = \sqrt{n}(\hat{\theta}_n \mathbf{1}_p - \theta_0 \mathbf{1}_p) \xrightarrow{D} N_p(\mathbf{0}, \mathbf{B}), \quad \text{as } n \rightarrow \infty, \quad (6.20)$$

where $\lambda = (\lambda_1, \dots, \lambda_p)'$, $\mathbf{B} = \mathbf{1}_p \mathbf{1}_p'$, and we assume $\lambda' \xi = 0$;

$$(iv) \quad \begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} \xrightarrow{D} N_{2p} \left\{ \begin{bmatrix} \xi \\ \mathbf{J}\xi \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1} \mathbf{J}' \\ \mathbf{J}\Lambda^{-1} & \Lambda^{-1} \mathbf{J}' \end{bmatrix} \right\}, \quad \text{as } n \rightarrow \infty; \quad (6.21)$$

$$\begin{bmatrix} \mathbf{z}_n \\ \mathbf{y}_n \end{bmatrix} \xrightarrow{D} N_{2p} \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{J}\xi \end{bmatrix}, \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \Lambda^{-1} \mathbf{J}' \end{bmatrix} \right\}, \quad \text{as } n \rightarrow \infty; \quad (6.22)$$

(v) L_n is distributed asymptotically as a noncentral chi-square with $(p-1)$ DF, and noncentrality parameter

$$\Delta = \delta' \Lambda \delta, \quad \delta = \mathbf{J}\xi. \quad (6.23)$$

where $\xrightarrow{D}$ means convergence in distribution.

We denote a multinormal distribution function with mean vector μ and dispersion matrix Σ by $\Phi(\mathbf{x}; \mu, \Sigma)$ and the corresponding density function by $\phi_p(\mathbf{x}; \mu, \Sigma)$. Also, let $H_p(\mathbf{x}; \Delta)$ denote the noncentral chi-square distribution function with p DF and noncentrality parameter $\Delta (\geq 0)$ and $E(\chi_m^{-2r}(\Delta)) = \int_0^\infty x^{-r} dH_m(x; \Delta)$. Note that by (6.7) and Theorem 6.2 we have

$$\sqrt{n}(\hat{\theta}_n^{PT} - \theta_0 \mathbf{1}_p) = \mathbf{x}_n - \mathbf{y}_n I(L_n \leq \chi_{p-1, \alpha}^2). \quad (6.24)$$

By (6.21)–(6.24) and the general formulation of Saleh and Sen (1984a), we arrive at the following theorem:

Theorem 6.3 Under $\{K_{(n)}\}$ in (6.13) and the assumed regularity conditions, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n^{PT} - \theta_0 \mathbf{1}_p)$ is given by

$$\begin{aligned} G_p^{PT}(\mathbf{x}) &= \Phi_p(\mathbf{x}; \mathbf{0}, \mathbf{B}) H_{p-1}(\chi_{p-1, \alpha}^2; \Delta) \\ &+ \int \dots \int_{M(\delta)} \Phi_p(\mathbf{x} - \xi - \mathbf{z}; \mathbf{0}, \mathbf{B}) d\Phi_p(\mathbf{x}; \mathbf{0}, \Lambda^{-1} \mathbf{J}'), \end{aligned} \quad (6.25)$$

where

$$M(\delta) = \{\mathbf{z}; (\mathbf{z} + \mathbf{J}\xi)'(\mathbf{z} + \mathbf{J}\xi) \geq \chi_{p-1, \alpha}^2\}, \quad (6.26)$$

and $\mathbf{J}, \mathbf{B}$ and Δ are all defined as in Theorem 6.2.

Note that the density function corresponding to $G_p^{PT}(\mathbf{x})$ is

$$g_p^{PT}(\mathbf{x}) = \phi_p(\mathbf{x}; \mathbf{0}, \mathbf{B}) H_{p-1}(\chi_{p-1, \alpha}^2; \Delta) + \int \dots \int_{M(\delta)} \phi_p(\mathbf{x} - \xi - \mathbf{z}; \mathbf{0}, \mathbf{B}) d\phi_p(\mathbf{z}; \mathbf{0}, \Lambda^{-1} \mathbf{J}') \quad (6.27)$$

Next, by virtue of (6.6), (6.8) and Theorem 6.2, we obtain the following theorem:

Theorem 6.4 Under $\{K_{(n)}\}$ in (6.13) and the assumed regularity conditions,

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^S - \theta_0 \mathbf{1}_p) &= \mathbf{x}_n - (p-3)\mathbf{y}_n \{\mathbf{y}_n' \Lambda \mathbf{y}_n\}^{-1} \\ &\xrightarrow{D} (\mathbf{x} + \xi) - (p-3)(\mathbf{y} + \mathbf{J}\xi) \times \\ &\quad \{(\mathbf{y} + \mathbf{J}\xi)' \Lambda (\mathbf{y} + \mathbf{J}\xi)\}^{-1} \end{aligned} \quad (6.28)$$

where

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim N_{2p} \left\{ \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Lambda^{-1} & \Lambda^{-1} \mathbf{J}' \\ \mathbf{J} \Lambda^{-1} & \Lambda^{-1} \mathbf{J}' \end{bmatrix} \right\}.$$

Very similar representations hold for the positive rule estimator in (6.9).

6.3.3 Asymptotic Biases and Risks

Under $\{K_{(n)}\}$ in (6.13), and by virtue of the general equivalence result in (6.14) and (6.17) and by Theorems 6.2, 6.3 and 6.4, we arrive at the following expressions for the asymptotic bias and risks. Since the derivation of these formula are very similar to the cases treated in Chapter 5, we omit these details of the derivation. The asymptotic bias expressions for the first four estimators in (6.4), (6.5), (6.7) and (6.8) are given by

$$B_1(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\tilde{\theta}_n - \theta_{(n)})\} = 0, \quad (6.29)$$

$$B_2(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})\} = -\delta, \quad (6.30)$$

$$B_3(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n^{PT} - \theta_{(n)})\} = -\delta H_{p+1}(\chi_{p-1, \alpha}^2; \Delta), \quad (6.31)$$

$$B_4(\delta) = \lim_{n \rightarrow \infty} E\{\sqrt{n}(\hat{\theta}_n^S - \theta_{(n)})\} = -(p-3)\delta E(\chi_{p+1}^{-2}(\Delta)), \quad (6.32)$$

where Δ is defined by (6.23). Also, the asymptotic distributional covariance matrices of the first four estimators in (6.4), (6.5), (6.7) and (6.8) are given by

$$\Gamma_1(\delta) = \lim_{n \rightarrow \infty} E\{n(\tilde{\theta}_n - \theta_{(n)})(\tilde{\theta}_n - \theta_{(n)})'\} = \Lambda^{-1}, \quad (6.33)$$

$$\Gamma_2(\delta) = \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})(\hat{\theta}_n \mathbf{1}_p - \theta_{(n)})'\} = \mathbf{B} + \delta\delta', \quad (6.34)$$

$$\begin{aligned} \Gamma_3(\delta) &= \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n^{PT} - \theta_{(n)})(\hat{\theta}_n^{PT} - \theta_{(n)})'\} \\ &= \Lambda^{-1} - \Lambda^{-1}H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) \\ &\quad + \delta\delta' \{2H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta)\}, \end{aligned} \quad (6.35)$$

$$\begin{aligned} \Gamma_4(\delta) &= \lim_{n \rightarrow \infty} E\{n(\hat{\theta}_n^S - \theta_{(n)})(\hat{\theta}_n^S - \theta_{(n)})'\} \\ &= \Lambda^{-1} - (p-3)\Lambda^{-1} \{2E(\chi_{p+1}^{-2}(\Delta)) - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ &\quad + (p-3)(p+1)\delta\delta'E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (6.36)$$

Consequently, the asymptotic distributional risk function under quadratic loss in (6.11) for the first four estimators are given by

$$R(\tilde{\theta}_n, \mathbf{W}) = \text{Tr}(\mathbf{W}\Lambda^{-1}), \quad (6.37)$$

$$R(\hat{\theta}_n \mathbf{1}_p, \mathbf{W}) = \text{Tr}(\mathbf{W}\mathbf{B}) + \delta'\mathbf{W}\delta, \quad (6.38)$$

$$\begin{aligned} R(\hat{\theta}_n^{PT}, \mathbf{W}) &= \text{Tr}(\mathbf{W}\Lambda^{-1}) - \text{Tr}(\mathbf{W}\Lambda^{-1})H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) \\ &\quad + \delta'\mathbf{W}\delta \{2H_{p+1}(\chi_{p-1,\alpha}^2; \Delta) - H_{p+3}(\chi_{p-1,\alpha}^2; \Delta)\}, \end{aligned} \quad (6.39)$$

$$\begin{aligned} R(\hat{\theta}_n^S, \mathbf{W}) &= \text{Tr}(\mathbf{W}\Lambda^{-1}) - (p-3)\text{Tr}(\mathbf{W}\Lambda^{-1}) \{2E(\chi_{p+1}^{-2}(\Delta)) \\ &\quad - (p-3)E(\chi_{p+1}^{-4}(\Delta))\} \\ &\quad + (p-3)(p+1)\delta'\mathbf{W}\delta E(\chi_{p+3}^{-4}(\Delta)). \end{aligned} \quad (6.40)$$

The performance of the above estimators are exactly the same as the performance of the estimators in Section 5.5, by observing the similarity of the risk expressions and the analysis that follows.

Chapter 7

Further Research

In chapter 2, we have studied the estimation of the mean vector under suspected symmetry when the covariance matrix is the intraclass correlation matrix. The Hotelling's T^2 -test used may not be the optimum test for the test of symmetry. There is need to investigate the properties of the estimator using the appropriate test statistic, i.e., likelihood ratio test and see whether the results are comparable.

In Chapter 3, we have studied the problem of estimating the regression coefficient vector under the suspected null hypothesis $H\beta = h$. The next problem would be to study the stochastic hypothesis $H\beta = h + \eta$ where $\eta \sim N(0, \sigma^2 \Omega)$ and Ω is known matrix. Under this model what would be the form of the shrinkage estimator? Will it retain the same properties as given in Chapter 3. Another problem would be to study multivariate extension of the univariate regression problem. Finally, one could study the effect of various null hypothesis on the PTE as well as SE. Do they retain the same properties or there any loss gain of the risk as one varies the null hypothesis.

Through out the thesis, we have studied the properties using quadratic risks. It will be a reasonable to provide good estimators of the risk for practitioners of the statistical techniques we have provided.

Another burning problem is the study of confidence sets of the PTE and SE. This problem will result in a lot of analytical tool which may later be used to other related fields such as Pitman- Nearness of estimators.

Empirical Bayes approach has been studied here relative to preliminary test

approach. There are various methods of obtaining shrinkage estimators. How do they compare with each other? It will be interesting to study comparative easiness for each methods.

In the thesis, we have considered only linear estimators assuming normality. One may used some other method to remove the normality conditions. This will provide robust statistical procedure for the various estimators.

It will be interesting to study the problem of Binomial means by using some predicting method using regression method and compare the two estimators.

While the work of the thesis looks at relationship of preliminary test approach and the empirical Bayes approach to shrinkage estimation, one may proceed with the preliminary test approach to obtain reasonable estimators in the absence of prior knowledge on the distribution of parameters of interest. One problem is the simultaneous estimation of several location parameters when the distribution of each component is of the form $\frac{1}{\sigma} f_0 \left(\frac{x - \mu_\alpha}{\sigma} \right)$, $\alpha = 1, 2, \dots, s$; where f_0 is known. The problem may be formulated in the following way based on order statistic:

Let $x_{1:N_\alpha}^{(\alpha)} \leq \dots \leq x_{N_\alpha:N_\alpha}^{(\alpha)}$ be an ordered sample of size N_α from the α^{th} continuous distribution $\frac{1}{\sigma} f_0 \left(\frac{x - \mu_\alpha}{\sigma} \right)$ where μ_α and σ are the location and scale parameters respectively. Let $x_{n_{\alpha 1}:N_\alpha}^{(\alpha)} \leq \dots \leq x_{n_{\alpha k}:N_\alpha}^{(\alpha)}$ be a subset of the order statistics of the α^{th} sample ($\alpha = 1, 2, \dots, s$). We consider the estimation of the vector $\mu = (\mu_1, \dots, \mu_s)'$ when it is suspected that $\mu = 1_p \mu_0$, (μ_0 fixed). Then, following the work in the thesis one may consider the UE, RE, PTE and SE using best linear unbiased estimators (BLUE). In this case, the BLUE of $\mu = (\mu_1, \dots, \mu_s)'$ based on the subset of the order statistics will be reasonable together with the asymptotic theory. Here one will have to invoke the Pitman alternative to obtain the asymptotic distributional properties of the estimators and their quadratic risks. The detailed work on this problem will need suitable test statistic for testing H_0 and study of the properties of the estimators. The work on this thesis will be used to work the details of the problem mentioned above, thus enlarge the scope of application for various distributions used in practice.

Appendix A

Theorem 1 If the $(p \times 1)$ vector $\mathbf{z}$ is distributed normally with mean $\boldsymbol{\theta}$ vector and covariance matrix $\mathbf{I}_p$, then

$$E[\phi(\mathbf{z}'\mathbf{z})] = \theta E[\phi(\chi_{(p+2, \lambda)}^2)],$$

where

$$\lambda = \boldsymbol{\theta}'\boldsymbol{\theta}/2.$$

Theorem 2 If the $(p \times 1)$ vector $\mathbf{z}$ is distributed normally with mean vector $\boldsymbol{\theta}$ and covariance matrix $\mathbf{I}_p$, then

$$E[\phi(\mathbf{z}'\mathbf{z})\mathbf{z}\mathbf{z}'] = E[\phi(\chi_{(p+2, \lambda)}^2)] \mathbf{I}_p + E[\phi(\chi_{(p+4, \lambda)}^2)] \boldsymbol{\theta}\boldsymbol{\theta}',$$

where λ as in Theorem 1.

Theorem 3 If $\mathbf{z}$ is a $(p \times 1)$ normally distributed random vector with mean $\boldsymbol{\theta}$ vector and covariance $\mathbf{I}_p$ and $\mathbf{A}$ is any positive definite symmetric matrix, then

$$E[\phi(\mathbf{z}'\mathbf{z})\mathbf{z}'\mathbf{A}\mathbf{z}] = E[\phi(\chi_{(p+2, \lambda)}^2)] \text{Tr}(\mathbf{A}) + E[\phi(\chi_{(p+4, \lambda)}^2)] \boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta},$$

where λ as in Theorem 1. For the proof of these results, see Judge and Bock (1978) Appendix B.

Appendix B: Computer Program

In this appendix, we present the computer program called "PROGRAM MULTIVARIATE NORMAL" which generates Tables 2.5-2.8 in Chapter 2. This program written in FORTRAN using some functions from NAG FORTRAN Library, namely, beta distribution (G01BDF) and inverse F-distribution (G01CBF).

```

C      Constants
      DOUBLE PRECISION RHOMAT(1:7), ALPHAMAT(1:10)
      INTEGER NMAT(1:4)
C      Variables
      INTEGER N,M,P,I,J,A,IFAIL,LIM,Q,K,Q1
      DOUBLE PRECISION TSTAR, ALPHA, BETA,F1(0:50), XSQ,F2(0:50)
      DOUBLE PRECISION RELMAX(7,10), RE3MAX(7,10),X1,X2,TDSTAR,
:      RE4MAX(7,10), DELTA0(7,10), EQUALVALUE(7,10)
C      Functions and functions from
C      NAG Fortran Library.
      DOUBLE PRECISION RE1, RE3, RE4, G01CBF, G01BDF
C      Common blocks 'BLANK', 'FBLOCK', defined
      COMMON XSQ, P, N,M
      COMMON /FBLOCK/ F1,/F2BLOCK/ F2, LIM
C      Input data for n, RHO and ALPHA
      DATA (NMAT(I),I=1,4)/6,9,10,15/
:      (RHOMAT(I),I=1,7)/-0.10,-0.05,0,0.1,0.3,0.5,0.7/
:      (ALPHAMAT(I),I=1,10)/0.05,0.1,0.15,0.2,0.25,0.3,0.35
:      ,0.40,.45,0.50/
C      Begin
      LIM=25
      DO 1050 P=4,5
        Q=P-1
        DO 1040 I=1,4
          N=NMAT(I)
          M=N-Q
          DO 1020 A=1,10
            IFAIL=0
            ALPHA=ALPHAMAT(A)
            BETA=1-ALPHA
            TSTAR=(DBLE(Q)/DBLE(Q+2))*G01CBF(BETA,Q,M,IFAIL)
            TDSTAR=(DBLE(Q)/DBLE(Q+4))*G01CBF(BETA,Q,M,IFAIL)
            X1=TSTAR/((DBLE(M)/DBLE(Q+2))+TSTAR)
            X2=TDSTAR/((DBLE(M)/DBLE(Q+4))+TDSTAR)
            DO 1000 K=0,LIM+2
              Q1=Q+2+2*K
              F1(K)=G01BDF(X1,DBLE(Q1)/2,DBLE(M)/2,IFAIL)
              F2(K)=G01BDF(X2,DBLE(Q1+2)/2,DBLE(M)/2,IFAIL)
1000          CONTINUE
            DO 1010 J=1,7
              XSQ=RHOMAT(J)
              RELMAX(J,A)=RE1(DBLE(0))
              RE3MAX(J,A)=RE3(DBLE(0))
              RE4MAX(J,A)=RE4(DBLE(0))
              CALL EQUALIZE(DELTA0(J,A))
              EQUALVALUE(J,A)=RE3(DELTA0(J,A))
1010          CONTINUE
1020          CONTINUE
C      Output
      PRINT'(/T6A,I2,A,I2)', 'P=',P, 'N=',N
      PRINT'(T6,A,7F9.2)', '$\alpha \backslash$ \rho$ & ',
:      (RHOMAT(J),J=1,7)
      DO 1030 A=1,10
        ALPHA=ALPHAMAT(A)
        PRINT'(T2,F4.2,A,7F9.4)', ALPHA, '& $E_1$ & ',
:      (RE1MAX(J,A), J=1,7)
        PRINT'(T6,A,7F9.4)', '& $E_3$ & ', (RE3MAX(J,A),J=1,7)
        PRINT'(T6,A,7F9.4)', '& $E_4$ & ', (RE4MAX(J,A),J=1,7)
        PRINT'(T6,A,F3.2,A,7F9.4)', '& $E_{\Delta}$ & ', ALPHA,
:      '$ & ', (EQUALVALUE(J,A),J=1,7)
        PRINT'(T6,A,F3.2,A,7F9.4)', '& $\Delta_{\alpha}$ & ',
:      (DELTA0(J,A),J=1,7)
1030          CONTINUE
1040          CONTINUE
1050          CONTINUE

```

```

C
C      Function RE1
      DOUBLE PRECISION FUNCTION RE1(DELTA)
C      Declarations
      DOUBLE PRECISION DELTA, XSQ
      INTEGER P,N
      COMMON XSQ, P, N
C      Begin
      RE1=P/((1-XSQ)*(1+2*DELTA)+P*XSQ)
      END

C
C      Function RE3
      DOUBLE PRECISION FUNCTION RE3(DELTA)
C      Declarations
      DOUBLE PRECISION DELTA, G, H2, H4, XSQ
      INTEGER P, N
      COMMON XSQ, P, N
C      Begin
      CALL H(DELTA, H2, H4)
      G=((1-XSQ)*(P-1)*H2)-(2*(1-XSQ)*DELTA*(2*H2-H4))
      RE3=P/(P-G)
      END

C
C      Subroutine H
      SUBROUTINE H(DELTA,H2,H4)
C      Declarations
      INTEGER I, LIM
      DOUBLE PRECISION CO(0:100), F1(0:50),H2,H4,DELTA,F2(0:50)
      COMMON /F1BLOCK/ F1,/F2BLOCK/ F2, LIM
C      Begin
      CO(0)=EXP(-1*DELTA/2)
      DO 1000 I=1,LIM
          CO(I)=CO(I-1)*DELTA/(2*I)
1000  CONTINUE
      H2=0
      H4=0
      DO 1010 I=0, LIM
          H2=H2+F1(I)*CO(I)
          H4=H4+F2(I)*CO(I)
1010  CONTINUE
      END

C
C      Function RE4
      DOUBLE PRECISION FUNCTION RE4(DELTA)
C      Declarations
      DOUBLE PRECISION DELTA, G, E, XSQ,
      :      ROH,L1
      INTEGER P, N, M
      COMMON XSQ, P, N,M
C      Begin
      ROH=1-XSQ
      L1=(DBLE(M)/DBLE(M+2))*DBLE(P-3)*ROH*DBLE(P-3)
      G=L1*E(DELTA,P-1)
      RE4=P/(P-G)
      END

C
C      Function E
      DOUBLE PRECISION FUNCTION E(DELTA, P)
C      Declarations
      DOUBLE PRECISION DELTA, SUM, FACTOR
      INTEGER P, R, LIM
C      Begin
      IF (DELTA .EQ. 0) THEN
          LIM=1
      ELSE

```

```
ENDIF
SUM=1/DBLE(P-2)
FACTOR=DBLE(1)
DO 1000 R=1, LIM
    FACTOR=FACTOR*DELTA/(2*R)
    SUM=SUM+(FACTOR/(P-2+(2*R)))
1000 CONTINUE
E=EXP(-1*DELTA/2)*SUM
END

C
C Subroutine to find out where RE3 and RE4 are equal, at EQUALAT
SUBROUTINE EQUALIZE(EQUALAT)
C Declarations
DOUBLE PRECISION EQUALAT, DIF, RE3, RE4, STEP, D, TEMP
INTEGER I
PARAMETER (D=0.0000001)
C Begin
STEP=DBLE(1)
C TEMP is the estimate for where RE3 and RE4 are equal
TEMP=DBLE(1)
C First find unit interval which contains the (first) equal point,
DO 1000 I=1,50
    DIF=RE4(TEMP)-RE3(TEMP)
    IF (DIF .GE. 0)GOTO 1010
    TEMP=TEMP+1
1000 CONTINUE
1010 CONTINUE
C and now cut the step size in half each time,
DO 1020 I=1, INT(LOG(D)/LOG(.5))
    IF (DIF.EQ.0)GOTO 1030
C else.
    STEP=STEP/2
    TEMP=TEMP-(ABS(DIF)/DIF)*STEP
    DIF=RE4(TEMP)-RE3(TEMP)
1020 CONTINUE
1030 CONTINUE
EQUALAT=TEMP
END
```

Bibliography

- [1] Ahmed, S. E. (1987). *Estimation Strategies under Uncertain Prior Information*. Ph.D Thesis, Carleton University.
- [2] Albert, J. H. (1984). *Empirical Bayes estimation of a set of binomial probabilities*. J. Statist. Comput. Simul. 20, 129–144.
- [3] Albert, J. H. and Gupta, A. K. (1981). *Bayesian methods for binomial data with applications to a nonresponse problem*. Technical report, Department of Mathematics and Statistics, Bowling Green State University.
- [4] Albert, J. H. and Gupta, A. K. (1983). *Estimation in contingency tables using prior information*. J. Roy. Statist. Soc. (Ser. B) 45, 60–69.
- [5] Anderson, T. W. (1984). *An Introduction to Multivariate Analysis*. John Wiley & Sons, New York.
- [6] Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. John Wiley & Sons, New York.
- [7] Asano, C. and Sato, S. (1962). *A bivariate analogue of pooling of data*. Bulletin of Mathematical Statistics 10 (3–4), 39–59.
- [8] Bancroft, T. A. (1944). *On biases in estimation due to the use of preliminary tests of significance*. Ann. Math. Statist. 15, 190–204.

- [9] Bancroft, T. A. and Han, C. P. (1980). *Inference based on conditionally specified ANOVA models incorporating preliminary testing*. Handbook of Statistics, 6, P. R. Krishnaiah (Ed.). North Holland, New York, 407-441.
- [10] Bennett, B. M. (1952). *Estimation of means on the basis of preliminary tests of significance*. Ann. Inst. Statist. Math. 4, 31-43.
- [11] Bennett, B. M. (1956). *On the use of preliminary test in certain statistical procedures*. Ann. Inst. Statist. Math. 8, 45-57.
- [12] Berger, J. O. (1980). *A robust generalized Bayes estimation and confidence region for multivariate normal mean*. Ann. Statist. 8, 716-761.
- [13] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- [14] Berger, J. and Chen, S. Y. (1987). *Minimaxity of empirical Bayes estimators derived from subjective hyperpriors*. In Advances in Multivariate Statistical analysis, A. K. Gupta. (Ed.). 1-12, D. Reidel Publishing Company.
- [15] Bishop, Y., Fienberg, S., and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Massachusetts.
- [16] Bozivich, H., Bancroft, T. A. and Hartley, H. O. (1956). *Power of analysis of variance test procedures for certain incompletely specified models*. Ann. Math. Statist. 27, 1017-1043.
- [17] Cohen, A. (1965). *Estimates of the linear combination of parameters in the mean vector of a multivariate distribution*. Ann. Math. Statist. 36, 78-87.
- [18] Da Silva, A. G. and Han, C-P. (1984). *Pooling means in a multivariate normal population*. ESTADISTICA- Journal of the Inter-American Statistical Institute 36, nos. 126-127, 63-75.
- [19] Efron, B. and Morris, C. (1972a). *Limiting the risk of Bayes and empirical Bayes estimators. —Part II: The empirical Bayes case*. J. Amer. Statist. Assoc. 67, 130-139.

- [20] Efron, B. and Morris, C. N. (1972b). *Empirical Bayes on vector observations—an extension of Stein's method*. Biometrika **59**, 335–347.
- [21] Efron, B. and Morris, C. (1973). *Stein's estimation rule and its competitors—an empirical Bayes approach*. J. Amer. Statist. Assoc. **68**, 117–130.
- [22] Efron, B. and Morris, C. N. (1975). *Data analysis using Stein's estimators and its generalizations*. J. Amer. Statist. Assoc. **70**, 311–319.
- [23] Ericson, W. A. (1969). *Subjective Bayesian models in sampling finite populations (with discussions)*. J. Roy. Statist. Soc. (Ser. B) **31**, 195–233.
- [24] Ghosh, M. and Meeden, G. (1986). *Empirical Bayes estimation in finite population sampling*. J. Amer. Statist. Assoc. **81**, 1058–1062.
- [25] Ghosh, M. and Lahiri (1987). *Robust empirical Bayes estimation of means from stratified samples*. J. Amer. Statist. Assoc. **82**, 1153–1162.
- [26] Ghosh, M., Saleh, A. K. Md. E. and Sen, P. K. (1989). *Empirical Bayes subset estimation in regression models*. Statistics & Decisions **7**, 15–35.
- [27] Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Wadsworth Publishing Company, Belmont, California.
- [28] Han, C.-P. and Bancroft, T. (1968). *On pooling means when variance is unknown*. J. Amer. Statist. Assoc. **63**, 1333–1342.
- [29] Huntsberger, D. V. (1955). *A generalization of a preliminary test procedure for pooling data*. Ann. Math. Statist. **26**, 734–743.
- [30] James, W. and Stein, C. (1961). *Estimation with quadratic loss*. Proceeding of the Fourth Berkeley symposium on Mathematical Statistics and probability (Ulcla Press, Berkeley, California), 361–379.
- [31] Judge, G. G. and Bock, M. E. (1978). *Statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics*. North-Holland: Amsterdam.

- [32] Kitagawa, T. (1963). *Estimation after preliminary test of significance*. University of California Publication in Statistics. (Ser. B), 147-186.
- [33] Lahiri, P. and Peddada, S. D. (1989). *Bayes and empirical Bayes estimator of finite population mean*. Unpublished report.
- [34] Lindley, D. V. (1962). *Contribution to discussion on paper by C. M. Stein*. J. Roy. Statist. Soc. (Ser. B) 24, 285-287.
- [35] Lindley, D. V. (1971a). *Bayesian statistics*. A review. SIAM, Philadelphia.
- [36] Lindley, D. V. (1971b). *The estimation of many parameters*. In Foundations of Statistical Inference, V. P. Godambe and D. A. Sprott (Eds.). Holt, Rinehart, and Winston, Toronto.
- [37] Lindley, D. and Smith, A. F. M. (1972). *Bayes estimate for the linear model (with discussion)*. J. Roy. Statist. Soc. (Ser. B) 34, 1-41.
- [38] Morris, C. N. (1983). *Parametric empirical Bayes confidence intervals*. Scientific Inference, Data Analysis, and Robustness. G. E. P. Box, T. Leonard and Chien-Fu Wu (Eds.). Academic Press, 25-50.
- [39] Mosteller, F. (1948). *On pooling data*. J. Amer. Statist. Assoc. 43, 231-242.
- [40] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. Wiley, New York.
- [41] Robbins, H. (1955). *An empirical Bayes approach to statistics*. In Proc. Third Berkeley Symp. Math. Statist. probab. 1, 157-164. University of California Press, Berkeley.
- [42] Saleh, A. K. Md. (1973). *Pretest estimator of means of multivariate normal distribution*. (Ser. 94) Carleton Mathematical.
- [43] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1978). *Non-parametric estimation of location parameter after a preliminary test on regression*. Ann. Statist. 6, 154-168.

- [44] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1983). *Nonparametric tests of location parameter after a preliminary test on regression in the multivariate case*. Commun. Statist.-Theory and Methods 12 (16), 1855-1872.
- [45] Saleh, A.K.Md. Ehsanes and Sen, P. K. (1984a). *Least squares and rank order preliminary test estimation in general multivariate linear models*. Proc. of ISI, Golden Jubilee Conference on Statistics: Application and New Directions. (Dec. 16-Dec. 19, 1981), 237-253.
- [46] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1984b). *Non-parametric preliminary test inference*. Handbook of Statistics. 4, : P. R. Krishnaiah and P. K. Sen (Eds.). North-Holland Publishers, 275-279.
- [47] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985a). *On Shrinkage M-estimators of location parameters*. Commun. Statist.-Theory and Methods 24, 2313-2329.
- [48] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985b). *Preliminary test predicted in general multivariate linear models*. Proceedings of the Pacific Area Statistical Conference, M. Matusita (Ed.). North-Holland Publishers. Held December 15-19, 1982, 619-638.
- [49] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985c). *Shrinkage least squares estimation in a general multivariate linear model*. The Proceeding of the 5th Pannonian Symposium on Math. Statistics. J. Mogyorodi, I. and Vince, W. Wertz (Eds.). 307-325.
- [50] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1985d). *Nonparametric shrinkage estimation in a parallelism problem*. Sankhyā (Ser. A) 47, 156-165.
- [51] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1986a). *On shrinkage least square estimation in a parallelism problem*. Commun. Statist. 15 (5), 1451-1466.
- [52] Saleh, A. K. Md. Ehsanes and Sen, P. K. (1986b). *On shrinkage R-estimation in a multiple regression model*. Commun. Statist. 15 (7), 2229-2244.

- [53] Sclove, S. L., Morris, C. and Radhakrishnan, R. (1972). *Non-optimality of preliminary test estimators of the mean of a multivariate normal distribution*. Ann. Math. Statist. **43**, 1481-1490.
- [54] Searle, S. R. (1971). *Linear Model*. Wiley, New York.
- [55] Sen, P. K. and Saleh, A. K. Md. Ehsanes (1979). *Nonparametric estimation of location parametric after a preliminary test on regression in multivariate case*. J. Multivariate Anal. (Ser. 2) **9**, 332-341.
- [56] Sen, P. K. (1984). *A James-Stein detour of U-statistics*. Commun. Statist.-Theory and Methods **A13**, 2725-2747.
- [57] Sen, P. K. and Saleh, A. K. Md. Ehsanes (1985). *On some shrinkage estimators of multivariate location*. Ann. Statist. **13**, 272-281.
- [58] Sen, P. K. (1986). *On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators*. Sankhyā, Indian Journal of Statistics. (Ser. A) **48**, 354-372.
- [59] Sen, P. K. and Saleh, A. K. Md. Ehsanes (1987). *On preliminary test and shrinkage M-estimation in linear models*. Ann. Statist. (Ser. 4) **15**, 1580-1592.
- [60] Stein, C. (1955a). *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. In Proc, Third Berkeley Symp. Math. Statist. Probab. **1**, 197-206. University of California Press, Berkeley.
- [61] Stein, C. (1955b). *A necessary and sufficient condition for admissibility*. Ann. Math. Statist. **26**, 518-522.
- [62] Stein, C. (1981). *Estimation of the mean of a multivariate distribution*. Ann. Statist. **9**, 1135-1151.
- [63] Zellner, A. (1986). *On assessing prior distributions and Bayesian regression analysis with g-prior distribution*. In Bayesian Inference and Decision Techniques, P. K. Goel and A. Zellner (Eds.).

END

1|2|0|4|9|1

FIN