Topics in Minimax Shrinkage Estimation

by

Stavros Zinonos

A dissertation submitted to the
School of Graduate Studies
Rutgers, The State University of New Jersey
In partial fulfillment of the requirements
For the degree of
Doctor of Philosophy
Graduate Program in Statistics and Biostatistics
Written under the direction of
William E. Strawderman
And Approved by

New Brunswick, New Jersey
October, 2018
ABSTRACT OF THE DISSERTATION

Topics in Minimax Shrinkage Estimation

By Stavros Zinonos

Dissertation Director:

William E. Strawderman

The dissertation considers three different topics which pertain to minimax shrinkage estimation:

1) Minimax estimation of a mean vector with variable selection for classes of spherically symmetric distributions: The results of Zhou and Hwang [31] and Maruyama [22] are extended from the normal case with known scale, to scale mixtures of normals and more generally to spherically symmetric distributions with a residual vector. Slight extensions to the class of estimators to which the results pertain are also given.

2) Minimax shrinkage estimators of a location vector under concave loss: In particular it is shown for a wide class of concave loss functions, James-Stein and Baranchik-type estimators which dominate the “usual” estimator for quadratic loss also dominate for these concave losses. The distributions studied include multivariate normal distributions with covariance equal to a known multiple of the identity, normal distributions with an unknown scale times the identity, and general scale mixtures of multivariate normal distributions with an unknown scale.

3) Combining unbiased and possibly biased correlated estimators of a mean vector under general quadratic loss: The general approach is to use a shrinkage-type estimator which shrinks an unbiased estimator toward a biased estimator. Conditions under which the
combined estimator dominates the original unbiased estimator are given. Models studied include normal models with a known covariance structure, scale mixtures of normals, and more generally elliptically symmetric models with a known covariance structure. Elliptically symmetric models with a covariance structure known up to a multiple are also considered.
Acknowledgements

I would like to thank Professor William E. Strawderman for his many valuable remarks and insights which improved the quality of the thesis. His patience and support was much appreciated. I would also like to thank John B. Kostis, M.D., Director of the Cardiovascular Institute of New Jersey, for his support during the writing and editing of my thesis.
# Contents

ABSTRACT OF THE DISSERTATION ................................................................. ii

ACKNOWLEDGEMENTS ...................................................................................... iv

List of Figures ................................................................................................. vii

1  Introduction                                                                 1
   1.1  Outline .................................................................................................. 1
   1.2  Preliminaries ........................................................................................ 3
      1.2.1  Spherically Symmetric Distributions ............................................ 3
      1.2.2  Minimaxity for Spherically Symmetric Distributions .................... 7
      1.2.3  Normal Theory ............................................................................. 11

2  Minimax Shrinkage Estimation with Variable Selection .............................. 15
   2.1  Introduction: ....................................................................................... 15
   2.2  Results for the Normal Case ............................................................... 17
      2.2.1  Normal Theory ............................................................................. 17
      2.2.2  Simulation Results ...................................................................... 34
   2.3  Berger Class of Spherically Symmetric Distributions ......................... 39
   2.4  Normal Distribution with Residual Vector ......................................... 47
   2.5  Spherically Symmetric Distributions with Residual ......................... 52
   2.6  Summary and Conclusion .................................................................... 57

3  Minimaxity of Shrinkage Estimators Under Concave Loss ......................... 58
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Simulated Risk of Zhou and Hwang Estimator when p=6</td>
<td>35</td>
</tr>
<tr>
<td>2.2</td>
<td>Simulated Risk of Maruyama Estimator when p=6</td>
<td>36</td>
</tr>
<tr>
<td>2.3</td>
<td>Risk Comparisons in $V_1$ and $V_3$</td>
<td>36</td>
</tr>
<tr>
<td>2.4</td>
<td>Probability of Variables Being Selected</td>
<td>37</td>
</tr>
<tr>
<td>2.5</td>
<td>Zhou and Hwang Estimator for p=12</td>
<td>38</td>
</tr>
<tr>
<td>2.6</td>
<td>Zhou and Hwang Estimator for p=17</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>Risk of James-Stein Estimator (3.77) for Mixture Distribution (3.85) with Shape Parameter, a, and Rate Parameter, s, Under Quadratic Loss, $l(d, \theta) = |d - \theta|^2$</td>
<td>85</td>
</tr>
<tr>
<td>3.2</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Various $b$ for Mixture Parameters (3.85) $a = 3$, $s = 1$ and Loss, $l(t) = 1 - e^{-bt}$</td>
<td>86</td>
</tr>
<tr>
<td>3.3</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Various $b$ for Mixture Parameters (3.85) $a = 5$, $s = 2$ and Loss, $l(t) = 1 - e^{-bt}$</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Various Mixture Parameter (3.85) $a$, $s=3$ and Loss, $l(t) = 1 - e^{-2t}$</td>
<td>87</td>
</tr>
<tr>
<td>3.5</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Various Mixture Parameter (3.85) $a$, $s=1$ and Loss, $l(t) = 1 - e^{-5t}$</td>
<td>88</td>
</tr>
<tr>
<td>3.6</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Varying $b$, for Mixture Parameters (3.85) $a = 10$, $s = 1$ and Loss, $l(t) = 1 - e^{-bt}$</td>
<td>89</td>
</tr>
<tr>
<td>3.7</td>
<td>Simulated Risk of James-Stein Estimator (3.77) for Varying $b$ for Mixture Parameters (3.85) $a = 15$, $s = 1$ and Loss, $l(t) = 1 - e^{-bt}$</td>
<td>89</td>
</tr>
</tbody>
</table>
4.1 Simulated risk of $\delta_1(X,Y)$ (4.91) with $\sigma^2 = \tau^2 = 1$ for varying $\varrho$ \hspace{1cm} 138

4.2 Simulated risk for $\delta_1(X,Y)$ (4.91) for $\sigma^2 = 1$ and varying $\tau^2$ and $\varrho$ \hspace{1cm} 139

4.3 Simulated risk of $\delta_1(X,Y)$ (4.91) and $\delta_{1}^{+}(X,Y)$ (4.93) for $\sigma^2 = 1, \tau^2 = .5$ and varying $\varrho$ \hspace{1cm} 139

4.4 Simulated risk of $\delta_1(X,Y)$ (4.91) and risk of $\delta^{\text{combo}}(X,Y)$ (4.90) for $\sigma^2 = \tau^2 = 1$ and varying $\varrho$ \hspace{1cm} 141

4.5 Simulated risk of $\delta_1(X,Y)$ (4.91) vs combo estimator (4.90) for $\sigma^2 = 1, \varrho = -.5$ and varying $\tau^2$ \hspace{1cm} 142

4.6 Simulated risk of $\delta_1(X,Y)$ (4.91) and $\delta_{1}^{+}(X,Y)$ (4.93) versus risk of combo estimator (4.90) for $\sigma^2 = 1, \tau^2 = .5$ and $\varrho = -.5$ \hspace{1cm} 143
Chapter 1

Introduction

1.1 Outline

Certain aspects of minimax estimation of location for classes of spherically symmetric and elliptically symmetric distributions are discussed. In particular, three different topics will be presented. The first topic pertains to minimax estimators of location under quadratic loss, $\|d - \theta\|^2$, whose positive-part adaptively estimate certain subsets of the location vector as 0 while shrinking the remaining coordinates for classes of spherically symmetric distributions. The second topic pertains to minimax shrinkage estimators of location for classes of spherically symmetric distributions under concave loss, where conditions for domination of the “usual” minimax estimator $\delta(X) = X$ are the same as those for quadratic loss. The third topic discusses combination estimators that shrink an unbiased estimator toward a possibly biased correlated estimator where conditions on the shrinkage-type estimator are given to ensure domination over the original unbiased estimator for general quadratic loss, $l(d, \theta) = (d - \theta)'Q(d - \theta)$, $Q > 0$.

Chapter 1 discusses preliminary results for minimax estimation of a location vector, $\theta$, for spherically symmetric distributions under quadratic loss. Special attention is paid to normal distributions with covariance equal to a known multiple of the identity. Stein’s Lemma [25] is developed for the case that $X \sim N_p(\theta, \sigma^2 I_p)$, and is used to give conditions for minimaxity for Baranchik-type estimators. Generalizations of this technique are used
throughout in Chapters 2-4 which contain the major contributions of the thesis.

Chapter 2 studies minimax estimation of a mean vector with variable selection. A main contribution of the chapter is to extend the class of distributions to which the results of Zhou and Hwang [31], and Maruyama [22] apply from the normal case with known scale to scale mixtures of normal distributions, and more generally to spherically symmetric distributions with a residual vector while slightly extending the class of estimators for which the results apply.

Chapter 3 studies minimax shrinkage estimators of a location vector under concave loss. In particular it is shown for a wide class of concave loss functions, the James-Stein and Baranchik-type estimators which dominate the “usual” estimator for quadratic loss also dominate for these concave losses. The distributions studied include multivariate normal distributions with covariance matrix equal to a known multiple of the identity, normal distributions with covariance matrix equal to an unknown scale time the identity, and general scale mixtures of multivariate normal distributions with unknown scale.

Chapter 4 studies combining unbiased and possibly biased correlated estimators of a mean vector under general quadratic loss. The general approach is to use a shrinkage type estimator which shrinks an unbiased estimator toward a biased estimator. Conditions under which the combined estimator dominates the original unbiased estimator are given. Models studied include normal models with a known covariance structure, scale mixtures of normals, and more generally elliptically symmetric models with a known covariance structure. Elliptically symmetric models with a covariance structure known up to a multiple are also considered.
1.2 Preliminaries

1.2.1 Spherically Symmetric Distributions

The class of spherically symmetric distributions play an important role in statistics. The class of distributions include the multivariate normal distribution with covariance equal to a known multiple of the identity, the class of scale mixtures of multivariate normal distributions with covariance equal to a known multiple of the identity, and the uniform distribution on a sphere. The results of Chapters 2-4 apply to classes of spherically symmetric distributions.

The following discussion closely follows that in Fourdinier, Strawderman, and Wells [13]. If $P$ is a spherically symmetric distribution about $\theta$, then for any Borel set $C$ and any orthogonal transformation $H$

$$P(HC + \theta) = P(C + \theta).$$

(1.1)

A random vector $X \in \mathbb{R}^n$ is spherically symmetric about the point $\theta$ if $X - \theta$ is orthogonally invariant. The notation

$$X \sim ss(\theta)$$

(1.2)

is used to denote that $X$ has a spherically symmetric distribution about $\theta$. An equivalent condition to $X \sim ss(\theta)$ is $X = Z + \theta$ where $Z \sim ss(0)$.

Uniform distributions on a sphere provide an important class of spherically symmetric distributions. The uniform distribution on a sphere is the unique orthogonally invariant distribution over a sphere [13]. It can be defined naturally through the uniform measure on the sphere. Let

$$S_R = \{x \in \mathbb{R}^n||x|| = R\}$$

(1.3)

and

$$B_R = \{x \in \mathbb{R}^n||x|| \leq R\}$$

(1.4)
denote the sphere and ball of radius R respectively. The uniform measure on the sphere

\[ \sigma_R(\Omega) = \frac{n}{R} \lambda\{ru \in \mathbb{R}^n | 0 < r < R; u \in \Omega\} \quad (1.5) \]

for any Borel Set \( \Omega \) of \( S_R \), where \( \lambda \) is the Lebesgue measure in \( \mathbb{R}^n \). Due to the orthogonal invariance of \( \lambda \)

\[ \sigma_R(H\Omega) = \sigma_R(\Omega) \]

for any orthogonal matrix \( H \) so that \( \sigma_R \) is an orthogonal invariant measure on \( S_r \). Let \( \Omega \) be any Borel subset of \( S_R \). The uniform distribution on \( S_R, U_R \), has the representation

\[ U_R(\Omega) = \frac{\sigma_R(\Omega)}{\sigma_R(S_R)}. \quad (1.6) \]

Defining \( S_{R,\theta} = \{x|\|x - \theta\| = R\} \), the uniform distribution over \( S_{R,\theta} \) is defined by

\[ U_{R,\theta}(\Omega) = U_R(\Omega - \theta) \quad (1.7) \]

for \( \Omega \) a Borel subset of \( S_{R,\theta} \).

\( S_R \) can be expressed through the usual polar coordinate parameterization. Let \( V = (0, \pi)^{n-2} \times (0, 2\pi) \) for \((\theta_1, \theta_2, ..., \theta_{n-2})' \in V \), and \( \phi_R(\theta_1, \theta_2, ..., \theta_{n-1}) = (x_1, x_2, ..., x_n)' \). Then

\[ x_1 = R\sin(\theta_1)\sin(\theta_2) ... \sin(\theta_{n-2})\sin(\theta_{n-1}) \]
\[ x_2 = R\sin(\theta_1)\sin(\theta_2) ... \sin(\theta_{n-2})\cos(\theta_{n-1}) \]
\[ x_3 = R\sin(\theta_1)\sin(\theta_2) ... \cos(\theta_{n-2}) \]
\[ \vdots \]
\[ x_{n-1} = R\sin(\theta_1)\cos(\theta_2) \]
\[ x_n = R\cos(\theta_1) \]

represents the usual polar coordinate parameterization. For any Borel subset of \( S_R \), the change of variables formula implies

\[ \sigma_R(\Omega) = \frac{n}{R} \lambda\{(0, R) \times \phi_R^{-1}(\Omega)\} = (1.8) \]
\[
\frac{n}{R} \int_0^R \int_{\phi_R^{-1}(\Omega)} r^{n-1} \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 d\theta_2 \ldots d\theta_{n-1} dr \\
= \frac{n R^{n}}{n} \int_{\phi_R^{-1}(\Omega)} \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 \ldots d\theta_{n-1} dr = \int_{\phi_R^{-1}(\Omega)} \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 \ldots d\theta_{n-1} dr.
\] (1.9)

An immediate consequence of (1.9) is for the distribution \( U_R \), the angles \( \theta_i \) are mutually independent with density proportional to \( \sin^{n-i-1} \) for \( 1 \leq i \leq n-2 \), and uniform on \( (0, 2\pi) \) for \( i = n-1 \). Additionally setting \( \Omega = S_R \) in (1.9) gives the relation

\[
\sigma_R(S_R) = R^{n-1} \sigma_1(S_1).
\]

For any integrable function \( h(x) \)

\[
\int_{B_R} h(x) dx = \int_0^R \int_V h(\phi_r) r^{n-1} \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 d\theta_2 \ldots d\theta_{n-1} dr = \int_0^\infty \int_{S_r} h(X) d\sigma(X) dr.
\] (1.10)

Setting \( h(x)=1 \) in (1.10) gives

\[
\lambda(B_R) = \int_0^R \int_{S_r} d\sigma(x) dr = \int_0^R \int_V r^{n-1} \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 d\theta_2 \ldots d\theta_{n-1} dr = \int_0^R r^{n-1} dr \int_V \sin^{n-2}(\theta_1) \ldots \sin(\theta_{n-2}) d\theta_1 d\theta_2 \ldots d\theta_{n-1} = \int_0^R r^{n-1} dr \sigma_1(S_1).
\] (1.11)

Since \( \sigma_1(S_1) = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} \), expressions (1.9) and (1.11) give the following well known formulas for the volume and surface area of \( B_R \) and \( S_R \) respectively,

\[
\lambda(B_R) = \frac{n R^{n}}{n} \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}, \text{ and } \sigma_R(S_R) = R^{n-1} \frac{2\pi^{n-2}}{\Gamma(\frac{n}{2})}.
\]

Fourdrinier, Strawderman, Wells [13] give the following representation for spherically symmetric distributions in terms of the uniform distribution, \( U_{R\theta} \), and the radial distribution, \( R \).
Theorem 1.1. (Fourdrinier, Strawderman, Wells [13]). A distribution $P$ in $\mathbb{R}^n$ is spherically symmetric about $\theta \in \mathbb{R}^n$ if and only if there exists a distribution $p$ in $\mathbb{R}_+$ such that $P(A) = \int_{\mathbb{R}_+} U_{r,\theta}(A)dp(r)$ for any Borel Set $A$ of $\mathbb{R}^n$. Furthermore, if a random vector $X$ has such a distribution $P$, then the radius $\|X - \theta\|$ has distribution $p$.

Proof. Sufficiency is immediate as the distribution $U_{r,\theta}$ is spherically symmetric about $\theta$ for any $r \geq 0$. To show the necessary condition it is sufficient to consider the case that $P \sim ss(0)$. Let $X$ have distribution $P$, with $P \sim ss(0)$, and let $v(x) = \|x\|$ with distribution $p$. For any Borel set $A \in \mathbb{R}^n$, $B \in \mathbb{R}_+$, and any orthogonal transformation $H$

$$
\int_B P(H^{-1}(A)|v = r)dp(r) = P(H^{-1}(A) \cap v^{-1}(B)) = P(H^{-1}(A \cap H(v^{-1}(B)))) = P(A \cap v^{-1}(B)) \tag{1.12}
$$

where (1.12) follows from $v$ being an orthogonality invariant function so that the inverse image of $v^{-1}(B)$ is equivalent to the inverse image of $H(v^{-1}(B))$, and the orthogonal invariance of the measure $P$. Therefore for any Borel set $A$

$$
P(H^{-1}(A)|v = r) = P(A|v = r) \tag{1.13}
$$

implying the result since the only orthogonally invariant distribution on $S_r$ is the uniform distribution on $S_r$.

As a consequence of Theorem 1.1 the moments of spherically symmetric distributions exhibit representation in terms of the radial distribution, $R$, and the uniform distribution $U_{R,\theta}$. The result for the first and second moments is given in Theorem 1.2.

Theorem 1.2. Let $X \in \mathbb{R}^n$ be a random vector with a spherically symmetric distribution about $\theta \in \mathbb{R}^n$. Then the mean of $X$ exists iff the mean of $R = \|X - \theta\|$ exist, in which case $E[X] = \theta$. The covariance matrix of $X$ exists iff $E[R^2]$ is finite, in which case

$$
\text{Cov}(X) = \frac{E[R^2]}{n} I_n.
$$
Proof. It is sufficient to consider the case $\theta = 0$, as $X = Z + \theta$ with $Z \sim SS(0)$. Since $X = \frac{X}{\|X\|} \|X\|$, Theorem 1.1 implies that $X$ has the stochastic representation $X = UR$, where $U$ is the uniform distribution over the sphere and $R$ is the distribution of $\|U\|$, where $R$ and $U$ are independent. Therefore


exists iff $E[R]$ exist in which case $E[X] = 0$ since $E[U] = E[-U] = 0$ by the orthogonal invariance of $U$. The covariance matrix exist iff $E[R^2] < \infty$ since

$$E[\|X\|^2] = E[R^2\|U\|^2] = E[R^2]E[\|U\|^2] = E[R^2]$$

as $\|U\|^2 = 1$.

For $i \neq j$

$$E[U_iU_j] = E[-U_iU_j] = 0$$

since $U_iU_j$ has the same distribution as $-U_iU_j$ by orthogonal invariance. Since each of the marginals have the same distribution by orthogonal invariance, $E[U_i^2] = E[U_j^2]$ so that

$$1 = E[\sum_{i=1}^n U_i^2] = \sum_{i=1}^n E[U_i^2] = nE[U_j^2]$$

implying $E[U_j^2] = \frac{1}{n}$ for all $j$. Combing above

$$Cov(X) = Cov(RU) = E(RUU'R) =$$

$$E[R^2]E[UU'] = \frac{E[R^2]}{n} I_n$$

establishing the result.

\[ \blacksquare \]

1.2.2 Minimaxity for Spherically Symmetric Distributions

Let $X \sim ss(\theta)$. The case of estimating the location vector $\theta$ under quadratic loss, $\|d - \theta\|^2$, has been studied extensively. One particular approach is to find an estimator such that at
its maximal risk is minimal among all estimators. From a mathematical point of view, an estimator $\delta^M$ is minimax if $\delta^M$ satisfies

$$\inf_{\delta} \sup_{\theta} R(\theta, \delta) = \sup_{\theta \in \Theta} R(\theta, \delta^M). \quad (1.14)$$

For the class of spherically symmetric distributions, one such estimator is $\delta^M(X) = X$. This subsection establishes this result by determining the Bayes risk for a particular sequence of Bayes estimators. As a consequence of (1.14) any estimator such that

$$R(\theta, \delta) \leq R(\theta, \delta^M) \quad (1.15)$$

for all $\theta \in \Theta$ is also minimax.

For this subsection let the distribution of $X|\Theta$ have density $f(x|\theta)$ with $\Theta \sim \Lambda(\theta)$ having density $\lambda(\theta)$. The joint density,

$$\pi(x, \theta) = f(X|\theta)\lambda(\theta) \quad (1.16)$$

and the marginal density of $X$

$$m(X) = \int f(x|\theta)d\Lambda(\theta). \quad (1.17)$$

The Bayes risk for the estimator $\delta$

$$r(\Lambda, \delta) = E[L(\theta, \delta)] \quad (1.18)$$

where the expectation is taken with respect to the joint density (1.16) and $L$ is a loss function. As a consequence of Fubini’s Theorem the Bayes risk satisfies

$$E[E[L(\theta, \delta)|\Theta]] = \int R(\theta, \delta)d\Lambda(\theta) = E[E[L(\theta, \delta)|X]].$$
Any estimator $\delta_B$ that minimizes $\int R(\theta, \delta)d\Lambda(\theta)$ is known as a Bayes estimator with respect to the distribution $\Lambda$.

Supposing for the moment that $\Lambda(\theta) = \infty$ so that $\Lambda$ is not a probability distribution over $\theta$. It still may be the case that the posterior risk, $\int L(\theta, d)f(\|X - \theta\|^2)d\Lambda(\theta)$ is finite for each $x$, so that there is still a function $\delta_\Lambda$ that will minimize the posterior risk. We define the generalized Bayes estimator with respect to the measure $\Lambda(\theta)$ to be $\delta_\Lambda$ if the posterior expected loss, $E[L(\theta, \delta)(X)|X]$, is minimized at $\delta_\Lambda$ for all $x$. For quadratic loss, $\|d - \theta\|^2$, the Bayes estimator has form $E[\theta|X]$ as it minimize $E_\theta[\|\Theta - f(X)\|^2|X]$ for all $X$.

Theorem 1.3 gives sufficient conditions for an estimator $\delta$ to be minimax utilizing a sequence of Bayes risks. It can be found in Lehmann and Casella [19, Chapter 5: Theorem 1.12].

**Theorem 1.3.** (Lehmann and Casella [19]) Suppose that $\{\Lambda_n\}$ is a sequence of prior distributions with Bayes risks $r_{\Lambda_n}$ satisfying

$$r_\Lambda \leq r = \lim_{n \to \infty} r_{\Lambda_n}$$

for every prior distribution $\Lambda$ where $r_\Lambda$ is Bayes risk under $\Lambda$, and that $\delta$ is an estimator for which

$$\sup_\theta R(\theta, \delta) = r$$

Then $\delta$ is minimax estimator.

**Proof.** Suppose that $\delta'$ is any other estimator. Then

$$\sup_\theta R(\theta, \delta') \geq \int R(\theta, \delta')d\Lambda_n \geq r_{\Lambda_n}$$

by definition of the Bayes risk. Since this holds for every $n$, $\sup_\theta R(\theta, \delta') \geq \lim_{n \to \infty} r_{\Lambda_n} = r = \sup_\theta R(\delta, \theta')$ and therefore the estimator is minimax.
Theorem 1.4 establishes the minimaxity of the estimator $\delta^M(X) = X$ utilizing a sequence of priors that satisfy the conditions of Theorem 1.3. The sequence of priors was used in Strawderman [27] motivating the development of the James-Stein estimator.

**Theorem 1.4.** For a spherically symmetric distribution with finite second moments the estimator $\delta^M(X) = X$ is a minimax estimator of location under quadratic loss. Furthermore if $\delta(X)$ is another estimator for the location parameter $\theta$, the estimator $\delta(X)$ is minimax under square error loss provided its’ risk is less than $p\sigma^2$ uniformly in $\theta$ where $\sigma^2 I_p = \text{Cov}(X)$.

**Proof.** Let the prior distribution be $\pi(\theta) = f^{(n)}(\theta)$, the n fold convolution of f with itself. This implies the representation $\theta = \sum_{i=1}^{n} U_i$ where each $U_i$ is distributed as $f(u)$. Let $U_0 = X - \theta$. $U_0$ is independent of the other $U_i$’s as its’ distribution does not depend on $\theta$. Furthermore it has the same distribution as the other $U_i$’s.

With the above representation the Bayes estimator, $\delta_B(X) = E[\theta|X]$, has representation

$$E[\theta|X] = E[\sum_{i=1}^{n} U_i|X - \theta] = E[\sum_{i=1}^{n} U_i|\sum_{i=0}^{n} U_i] =$$

$$nE[U_j|\sum_{i=0}^{n} U_i] = nE[\sum_{i=0}^{n} U_i|\sum_{i=0}^{n} U_i] = \frac{nX}{n+1},$$

so that the Bayes risk

$$E[||\delta_B - \theta||^2] = E[||\frac{nX}{n+1} - \theta||^2] =$$

$$E\left[\frac{n^2\|X\|^2}{(n+1)^2} + \|\theta\|^2 - 2\frac{n}{n+1}X\theta\right] =$$

$$\frac{n^2(n+1)p\sigma^2}{(n+1)^2} + np\sigma^2 - 2\frac{n^2p\sigma^2}{n+1} =$$

$$\frac{n}{n+1}p\sigma^2.$$

Since

$$\lim_{n\to\infty} \frac{n}{(n+1)p\sigma^2} = p\sigma^2 = R(\theta, X) \text{ for all } \theta$$

Theorem 1.3 implies the estimator $\delta(X) = X$ is a minimax estimator with a constant risk of $p\sigma^2$. Hence any estimator with uniformly smaller risk than $p\sigma^2$ will be minimax.
1.2.3 Normal Theory

The class of Normal distributions with covariance equal to a known multiple of the identity are important examples of spherically symmetric distributions and play a central role in statistics. In this subsection we study minimax estimators of location under quadratic loss, \( \|d - \theta\|^2 \), when \( X \sim N_p(\theta, \sigma^2 I_p) \) with \( \sigma^2 \) known. The form of estimators studied are

\[
\delta(X) = X + \sigma^2 g(X)
\]  

(1.19)

and sufficient conditions on the function \( g \) will be given utilizing Stein’s Lemma [25]. A particular estimator in the class given by (1.19) is the Bayes estimator when \( X|\Theta \sim N_p(\theta, \sigma^2 I_p) \).

The Bayes estimator has the form

\[
E[\Theta|X] = E[X + \Theta - X|X] = X - \frac{\int (x-\theta)e^{-\frac{\|x-\theta\|^2}{2\sigma^2}}d\Lambda(\theta)}{\int e^{-\frac{\|x-\theta\|^2}{2\sigma^2}}d\Lambda(\theta)} = X - \sigma^2 \nabla m(X) 
\]

The form of Stein’s Lemma [27] utilized is given by Theorem 1.5. Let \( g(x) \) be a weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \). The divergence of the function \( g(X) \), \( \text{div}_x(g(x)) \), is defined to be

\[
\text{div}_x(g(x)) = \sum_{i=1}^{p} \frac{\partial}{\partial x_i} g_i(X)
\]  

(1.20)

where \( \frac{\partial}{\partial x_i} \) is the weak partial derivative of the function \( g_i \) with respect to the coordinate \( x_i \).

Theorem 1.5. (Fourdinier and Strawderman [12]) Let \( X \sim N_p(\theta, \sigma^2 I_p) \). Assuming that \( g(x) \) is a weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \)

\[
E_{\theta,\sigma^2}[(X-\theta)'g(X)] = E_{\theta,\sigma^2}[\sigma^2 \text{div}_x(g(X))].
\]

Proof. Let the \( i \)'th component of \( g(X) \) be \( g_i(X) \). Since

\[
\int_{\mathbb{R}^p} (x_i - \theta_i) g_i(x)e^{-\frac{\sum_{i=1}^{p} (x_i - \theta_i)^2}{2\sigma^2}}dx = \int_{\mathbb{R}^p} g_i(X)\sigma^2 \frac{\partial}{\partial x_i} e^{-\frac{\sum_{i=1}^{p} (x_i - \theta_i)^2}{2\sigma^2}}dx
\]

by weak differentiability of \( g(X) \)
\[ \int_{\mathbb{R}^p} g_i(X) \sigma^2 \frac{\partial}{\partial x_i} e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}} \, dx = \int_{\mathbb{R}^p} \sigma^2 \left( \frac{\partial}{\partial x_i} g_i(X) \right) e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}}. \]

Therefore

\[ E_{\theta, \sigma^2}[(X - \theta)' g(X)] = \]

\[ \int_{\mathbb{R}^p} \sum_{i=1}^p (x_i - \theta_i) g_i(x) (2\pi \sigma^2)^{-\frac{p}{2}} e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}} \, dx = \]

\[ \sum_{i=1}^p \int_{\mathbb{R}^p} (x_i - \theta_i) g_i(x) (2\pi \sigma^2)^{-\frac{p}{2}} e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}} \, dx = \]

\[ \sum_{i=1}^p \int_{\mathbb{R}^p} \sigma^2 \left( \frac{\partial}{\partial x_i} g_i(x) \right) (2\pi \sigma^2)^{-\frac{p}{2}} e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}} \, dx = \]

\[ \int_{\mathbb{R}^p} \sigma^2 \text{div}_x (g(x)) (2\pi \sigma^2)^{-\frac{p}{2}} e^{-\frac{\sum_{i=1}^p (x_i - \theta_i)^2}{2\sigma^2}} \, dx = \]

\[ E[\sigma^2 \text{div}_x (g(X))]. \]

\[ \Box \]

Using Theorem 1.5 the risk under quadratic loss of the estimator (1.19) with weakly differentiable function \( g \), satisfies

\[ E[\|X + \sigma^2 g(X)\|^2] = \]

\[ E[\|X - \theta\|^2 + \|\sigma^2 g(X)\|^2 + 2\sigma^2(X - \theta)' g(X)] = \]

\[ E[\|X - \theta\|^2] + \sigma^4 E[\|g(X)\|^2] + \sigma^2(2(X - \theta)' g(X))] = \]

\[ p\sigma^2 + \sigma^4 E[\|g(X)\|^2] + \sigma^2 E[2\text{div}_x (g(X))] = \]

\[ p\sigma^2 + \sigma^4 E[\|g(X)\|^2 + 2\text{div}_x (g(X))]. \quad (1.21) \]

Using expression (1.21) Theorem 1.6 gives sufficient conditions on the function \( g \) so that estimators of the form (1.19) are minimax.

**Theorem 1.6.** (Fourdinier and Strawderman [12]) Let \( X \sim N_p(\theta, \sigma^2 I_p) \) with \( \sigma^2 \) known. Let \( \delta(X) = X + \sigma^2 g(X) \) be an estimator for the location parameter \( \theta \) with \( E_\theta[\|g(X)\|^2] \leq \infty \). A sufficient condition for the estimator \( \delta(X) \) to be minimax under quadratic loss is
\[ \|g(X)\|^2 + 2\text{div}_x(g(X)) \leq 0 \text{ for } x \text{ a.e.} \]

**Proof.** The condition \( E_\theta\|g(X)\|^2 \) is a sufficient for the risk of the estimator to be finite. Using expression (1.21) the difference in risk between the estimator \( \delta^M(X) = X \) and the estimator \( \delta(X) \), \( \triangle_{\theta,\sigma^2}(X,\delta) \), satisfies

\[ \triangle_{\theta,\sigma^2}(X,\delta) = -E_\theta[\sigma^4(\|g(X)\|^2 + 2\text{div}_x(g(X))]. \]

Since \( \sigma^4 \) is positive and \( \|g(x)\|^2 + 2\text{div}_x(g(X)) \) is non-positive for \( x \) a.e., it follows that

\[ R(\theta,X) - R(\theta,\delta) = -E_\theta[\sigma^4(\|g(X)\|^2 - 2\text{div}_x(g(X)))] \geq 0 \text{ for all } \theta \]

so that

\[ R(\theta,X) \geq R(\theta,\delta) \]

uniformly in \( \theta \). Since the estimator, \( \delta^M(X) = X \) is minimax by Theorem 1.4, the estimator \( \delta(X) \) is minimax as well.

The following Corollary found in Lehmann and Casella [19, Chapter 5: Theorem 5.5], originally due to Baranchik [3], gives condition for the minimaxity of Baranchik-type estimators of the form

\[ \delta(X) = X - \sigma^2 r(\|X\|^2)\frac{X}{\|X\|^2}. \quad (1.22) \]

This class includes the James-Stein estimators and the positive-part versions of James-Stein estimators. Additionally Stawderman [28] gives classes of admissible Bayes estimators for \( p \geq 5 \) which belong to the class of estimators given in (1.22).

**Corollary 1.1.** *(Leehman and Casella [19] )* Let \( X \sim N_p(\theta,\sigma^2 I_p) \) with \( \sigma^2 \) known and \( p \geq 3 \). The estimator

\[ \delta(X) = X + \sigma^2 g(X) = X - \sigma^2 \frac{r(\|X\|^2)}{\|X\|^2}X \]

is minimax under quadratic loss once \( r(\cdot) \) satisfies the conditions
i) $r(X)$ is non decreasing and differentiable a.e.,

ii) $0 < r(X) \leq 2(p - 2)$.

Proof. Let $g(X) = \frac{-r(||X||^2)}{||X||^2}$.

$$||g(X)||^2 = \frac{r^2(||X||)}{||X||^2}$$

and

$$2div_X(g(X)) = -2(p - 2) \frac{r(||X||^2)}{||X||^2} - 4r'(||X||^2)$$

so that

$$||g(X)||^2 + 2div_X(g(X)) = \frac{r(||X||^2)}{||X||^2} [r(||X||^2) - 2(p - 2)] - 4r'(||X||^2) \leq \frac{r(||X||^2)}{||X||^2} [r(||X||^2) - 2(p - 2)]$$

since $r'(t) \geq 0$ a.e. by assumption. Since $0 \leq r(t) \leq 2(p - 2)$

$$\frac{r(||X||^2)}{||X||^2} [r(||X||^2) - 2(p - 2)] \leq \frac{r(||X||^2)}{||X||^2} [2(p - 2) - 2(p - 2)] \leq 0$$

so that Theorem 1.6 implies the result.
Chapter 2

Minimax Shrinkage Estimation with Variable Selection

2.1 Introduction:

We study minimax estimators of the mean vector for spherically symmetric distributions under square error loss which dominate the standard minimax estimator $\delta_0(X) = X$. We are particularly interested in minimax estimators whose positive parts adaptively estimate certain subset of the mean as 0, while shrinking the remaining coordinates. The results may be viewed as extensions of Hwang and Zhou [31] and of Maruyama [22] from the Gaussian case to generally spherically symmetric distributions.

Specifically, let $X$ be a spherically symmetric distribution with density given by $f(\|x - \theta\|^2)$ where $\dim(X) = \dim(\theta) = p \geq 3$, and let the loss function for estimation of $\theta$ be given by

$$L(\theta, d) = \|d - \theta\|^2.$$  \hfill (2.1)

We study minimaxity of estimators of the form:

$$\delta(X) = (\delta_1(X), \delta_2(X), ..., \delta_p(X))'.$$  \hfill (2.2)
where
\[ \delta_i(X) - (1 - h_i(X))X_i \quad (2.3) \]

and
\[ h_i(X) = h_i(X_1^2, X_2^2, \ldots, X_p^2). \quad (2.4) \]

When such an estimator is minimax, and when \( f \) is unimodal, the positive-part estimator \( \delta^+(X) \) with
\[ \delta_i^+(X) = (1 - h_i(X))_+X_i \quad (2.5) \]
is also minimax (and in fact dominates \( \delta(X) \)) and additionally may allow adaptively selected subsets of the coordinates to be estimated by 0. Hence minimaxity and variable selection are simultaneously achieved.

Hwang and Zhou [31] and Maruyama [22] have studied classes of such procedures in the Gaussian case, basing the shrinkage on the \( \ell_p \)-norm of \( X \): \( 1 < p < 2 \) for Hwang and Zhou, and general \( p \) for Maruyama. We extend slightly the classes of minimax estimators studied in these papers and extend the results to certain spherically symmetric distributions studied by Berger [4]. We also extend the result to the general class of spherically symmetric distributions with a residual vector. In particular let
\[ \left( \begin{array}{c} X \\ U \end{array} \right) \sim f(\|x - \theta\|^2 + \|U\|^2) \quad (2.6) \]

where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = k \geq 1 \), and let
\[ L(\theta, d) = \frac{\|d - \theta\|^2}{E\|U\|^2}. \quad (2.7) \]

In this case the dominating estimators are of the form
\[ \delta_i(X, U) = (1 - \frac{U'h_i(X)}{k + 2})X. \quad (2.8) \]
The basic tool in the Gaussian setting is the Stein unbiased estimator of risk technique as in Zhou and Hwang [31] and Maruyama [22]. Generalizations of this technique to the spherically symmetric setting form the basis of the generalizations to general spherically symmetric distributions.

Section 2.1 considers generalizations of classes of minimax estimators in the Gaussian case. Particular attention is paid to the class of Psuedo-Bayes estimators. Section 2.2 provides numerical simulations of the risk for classes of estimators developed in Section 2.1 when the \( \dim(X)=6 \). Section 3 studies extensions to the class of spherically symmetric distributions introduced by Berger [4]. Section 4 extends the results of Section 2 to spherically symmetric normal distributions with a residual vector. Section 5 develops results for the general class of spherically symmetric distributions with a residual vector. Section 6 gives some concluding remarks.

2.2 Results for the Normal Case

2.2.1 Normal Theory

In this Section, \( X \sim N_p(\theta, \sigma^2 I_p) \) with loss given by (2.1). We initially focus on what are sometimes referred to as Pseudo-Bayes estimators, by which we mean (in this chapter) estimators of the form

\[
\delta(X) = X + \sigma^2 g(X)
\]

(2.9)

where

\[
g(X) = \frac{\nabla m(X)}{m(X)}.
\]

(2.10)

The function \( m(X) \) is referred to as a psuedo-marginal. If \( m(X) \) were a true marginal distribution corresponding to a generalized or proper prior \( \pi(\theta) \), then \( \delta \) would be a generalized or proper Bayes estimator. In this section in particular, \( g \) will also have \( i^{th} \) coordinate
of the form

\[ g_i(X) = -h_i(X_1^2, X_2^2, \ldots, X_p^2)X_i. \]  \hspace{1cm} (2.11)

Conditions in which thresholding each of the coordinates of \( \delta(X) \) will yield an estimator with smaller risk under square error loss are established by Theorem 2.1. It will be general enough so that it can apply to the class of minimax estimators that will be developed later on in the Section to generate a new class of estimators that are able to select between the full model and reduced models and still be minimax. To that end the following is a generalization of Theorem 4 in Zhou and Hwang [31].

**Theorem 2.1.** Suppose \( X \) is a random variable in \( \mathbb{R}^p \) with density \( f((x_1 - \theta_1)^2, \ldots, (x_p - \theta_p)^2) \) where \( f \) is symmetric, unimodal, and non-increasing in each of the coordinates separately for each fixed value of the other coordinates. Let \( \delta(X) = (\delta_1(X), \ldots, \delta_p(X))' \) be an estimator of \( \theta \) such that:

\[ \delta_i(X) = (1 - h_i(X))X_i \]

with \( h_i(X) \) symmetric in \( X \) for all \( i \). Let

\[ \delta^+(X) = ((1 - h_1(X))_+X_1, \ldots, (1 - h_p(X))_+X_p)' = (\delta^+_1, \ldots, \delta^+_p)' \]

then

\[ E_\theta(\theta_i - \delta^+_i)^2 \leq E_\theta(\theta_i - \delta_i)^2. \]

Furthermore if there exist an \( i \) such that \( P_\theta(h_i(X) > 1) > 0 \), then

\[ E_\theta(\theta_i - \delta^+_i)^2 < E_\theta(\theta_i - \delta_i)^2. \]

**Proof:** Since

\[ E_\theta(\theta_i - \delta^+_i)^2 - E_\theta(\theta_i - \delta_i)^2 = \]

\[ E_\theta[(\delta^+_i)^2 - \delta_i^2](I_{\{h_i(X) < 1\}}(X) + I_{\{h_i(X) \geq 1\}}(X))] \]

\[ -2\theta_i E_\theta[\delta^+_i - \delta_i](I_{\{h_i(X) < 1\}}(X) + I_{\{h_i(X) \geq 1\}}(X))], \]

it is sufficient to consider the case \( h_i(X) \geq 1 \), as \( h_i(X) < 1 \) implies \( \delta^+_i(X) = \delta_i(X) \) and
\[ E_{\theta}[(\delta_i^{+2} - \delta_i^2)(I_{\{h_i(X) < 1\}}(X))] - 2\theta_i E_{\theta}[(\delta_i^+ - \delta_i)(I_{\{h_i(X) < 1\}}(X))] = 0. \]

Conditioning the expectation of

\[ E_{\theta}[(\delta_i^{+2} - \delta_i^2)(I_{\{h_i(X) \geq 1\}}(X))] - 2\theta_i E_{\theta}[(\delta_i^+ - \delta_i)(I_{\{h_i(X) \geq 1\}}(X))] \tag{2.12} \]

on the event

\[ A := \{ X|h_i(X) = h \geq 1 \text{ and for } j \neq i X_j = x_j \} \]

expression (2.12) is expressible as

\[ E[E_{\theta}[(\delta_i^{+2} - \delta_i^2)|A]] - 2\theta_i E[E_{\theta}[(\delta_i^+ - \delta_i)|A]]. \tag{2.13} \]

The conditional expectation in (2.13)

\[ E_{\theta}[(\delta_i^{+2} - \delta_i^2)|A] - 2\theta_i E_{\theta}[(\delta_i^+ - \delta_i)|A] = 2\theta_i (1 - h) E_{\theta}[X_i|A] - E[\delta_i^2|A] \]

as \( \delta_i^+ = 0 \) when \( x \in A \). If

\[ \theta_i(E_{\theta}[X_i|A] \geq 0) \tag{2.14} \]

then

\[ E_{\theta}[(\delta_i^{+2} - \delta_i^2)|A] - 2\theta_i E_{\theta}[(\delta_i^+ - \delta_i)|A] \leq 0 \]

as \( (1 - h) \leq 0 \) if \( h \geq 1 \) so that \( 2\theta_i (1 - h) E_{\theta}[X_i|A] \) is non-positive, and \( \delta_i^2 \geq 0 \) implies \( -E_{\theta}[\delta_i^2|A] \leq 0 \). By the symmetry of \( h \), any solution of \( h_i(X) = h \) comes in pairs. Therefore if \( y_i \) is a solution to

\[ h(x_1, \cdots, y_i, \cdots, x_p) = h, \]

then \(-y_i\) is another solution, and

\[ \theta_i|y_i|f((x_1 - \theta_1), \cdots, (|y_i| - \theta_i), \cdots, (x_p - \theta_p)) - \]

\[ \theta_i|y_i|f((x_1 - \theta_1), \cdots, (-|y_i| - \theta_i), \cdots, (x_p - \theta_p)) \leq 0 \]
by the assumption that $f$ is non-increasing in each of the coordinates separately for each fixed value of the other coordinates. It follows that

$$\theta_i E_\theta[X_i|A] \leq 0.$$  

Since (2.14) is satisfied

$$E_\theta[(\delta^+_i - \delta^-_i)|A] - 2\theta_i E_\theta[(\delta^+_i - \delta^-_i)|A] \leq 0. \quad (2.15)$$

Upon taking the expectation of (2.15)

$$E_\theta(\theta_i - \delta^+_i)^2 - E_\theta(\theta_i - \delta^-_i)^2 = E_\theta(\delta^+_i - \delta^-_i)^2 - 2\theta_i E_\theta(\delta^+_i - \delta^-_i) \leq 0 \quad (2.16)$$

establishing $E_\theta(\theta_i - \delta^+_i)^2 \leq E_\theta(\theta_i - \delta^-_i)^2$. Further supposing that $P_\theta(h_i(X) > 1) > 0$, $E_\theta(\delta^+_i - \delta^-_i)^2 - 2\theta_i E_\theta(\delta^+_i - \delta^-_i)$ is strictly bounded above by $E_\theta(\delta^+_i - \delta^-_i)^2$ which is strictly negative implying $E_\theta(\theta_i - \delta^+_i)^2 < E_\theta(\theta_i - \delta^-_i)^2$.

To establish minimax estimators of the form

$$\delta_i(X) = (1 - h_i(X))X_i$$

the following Lemma will be used in Theorem 2.2 to bound the risk of $\delta(X)$. It is a straightforward application of the Correlation Inequality that can be found in Casella and Berger [7, Theorem 4.7.9].

**Lemma 2.1.** Let $f$ and $h$ be continuous monotonic functions defined over $[a, b] \subseteq \mathbb{R}$ into $\mathbb{R}$. For any finite collection $\{x_i\}_{i=1}^n \subseteq [a, b]$

i) If $f$ and $h$ are both monotonic increasing functions then

$$\left(\sum_{i=1}^n f(x_i)\right)\left(\sum_{i=1}^n h(x_i)\right) \leq \sum_{i=1}^p f(x_i)h(x_i) \quad (2.17)$$
ii) If $f$ is a monotonic increasing function and $h$ is a monotonic decreasing function then

$$n \sum_{i=1}^{n} f(x_i)h(x_i) \leq (\sum_{i=1}^{n} f(x_i))(\sum_{i=1}^{n} h(x_i))$$ (2.18)

For estimators of the form

$$\delta(X) = X + \sigma^2 g(X) = X + \sigma^2 \frac{\nabla m(X)}{m(X)}$$ (2.19)

where $g$ is given by (2.11), the following Theorem gives conditions on the function $g$ so that $\delta(X)$ is minimax under quadratic loss, $\|d - \theta\|^2$.

**Theorem 2.2.** Let $X$ be distributed as $N_p(\theta, \sigma^2 I_p)$ with the location vector $\theta$ unknown. Let $\delta_{ZH}(X) = X + \sigma^2 \frac{\nabla m(X)}{m(X)} = X + \sigma^2 g(X)$ with $g(X)$ weakly differentiable, $E_\theta\|g(X)\|^2 < \infty$, and $\sqrt{m(x)} = j(D)$ where $D = \sum_{i=1}^{p} h(x_i^2)$.

Assume $h(t)$ satisfies the following conditions:

i) The second derivative with respect to $t$ of $h$ exist for all $t \geq 0$.

ii) $\forall t \geq 0, h(t) \geq 0, h'(t) \geq 0, \text{ and } h''(t) \leq 0$.

iii) There exist an $A > 0$ such that $2h''(t)t + h'(t) \geq Ah'(t) \forall t \geq 0$.

iv) There exist a $B > 0$ such that $h(t) \geq Bh'(t)t \forall t \geq 0$.

Assume $j(t)$ satisfies the following conditions:

v) The second derivative of $j$ with respect to $t$ exist for all $t \geq 0$ a.e..

vi) $\forall t \geq 0, j'(t) \leq 0, \text{ and } j''(t) \geq 0$ a.e..

vii) There exist a $C > 0$ such that $-j'(t) \geq Ctj''(t) \forall t \geq 0$. a.e..

Then a sufficient condition for the estimator $\delta_{ZH}(X)$ of $\theta$ under square error loss to be minimax is $\frac{pABC}{2} \geq 1$. 

Proof: A sufficient condition under square error loss for the risk of the estimator $\delta_ZH(X)$ to exist is $E_\theta\|g(X)\|^2 < \infty$. Furthermore if $\triangle_x j(D) \leq 0$ a.e. then the estimator $\delta_ZH(X)$ will be minimax. To provide sufficient conditions in which $\triangle_x j(D) \leq 0$ note that the second partial derivative of $j(D)$ with respect to $x_i$ is

$$\frac{\partial^2}{\partial x_i^2} j(D) = 4j''(D)[2h''(x_i^2)x_i^2 + h'(x_i^2)].$$

so that the Laplacian with respect to $x$ is given by

$$\triangle_x j(D) = 4j''(D)[\sum_{i=1}^p (h'(x_i^2))^2 x_i^2] + 2j'(D)[\sum_{i=1}^p 2h''(x_i^2)x_i^2 + h'(x_i^2)]. \quad (2.20)$$

From (2.20) $\triangle_x j(D) \leq 0$ is logically equivalent to

$$2j'(D)[\sum_{i=1}^p [2h''(x_i^2)x_i^2 + h'(x_i^2)]] \leq -4j''(D)[\sum_{i=1}^p (h'(x_i^2))^2 x_i^2]. \quad (2.21)$$

Except for a set of measure 0,

$$-2j'(D)[\sum_{i=1}^p 2h''(x_i^2)x_i^2 + h'(x_i^2)] \geq \quad (2.22)$$

(by assumption vii)

$$j''(D)[2C(\sum_{i=1}^p h(x_i^2)][\sum_{i=1}^p 2h''(x_i^2)x_i^2 + h'(x_i^2)] \geq \quad (2.23)$$

(by assumption iii)

$$j''(D)[2CA(\sum_{i=1}^p h(x_i^2))(\sum_{i=1}^p h'(x_i^2)) \geq \quad (2.24)$$

(by Lemma 2.1 and assumption ii)

$$j''(D)[2CAp(\sum_{i=1}^p h(x_i^2)h'(x_i^2))] \geq \quad (2.25)$$

(since $\frac{pABC}{2} > 1$)

$$j''(D)[\frac{4}{B}(\sum_{i=1}^p h(x_i^2)h'(x_i^2))] \geq \quad (2.26)$$
(by assumption iv)

\[ 4j''(D)[\sum_{i=1}^{p} h'(x_i^2)x_i^2]^2. \]  

(2.27)

Since \(-2j'(D)[\sum_{i=1}^{p} 2h''(x_i^2)x_i^2 + h'(x_i^2)] \geq -2j'(D)[\sum_{i=1}^{p} 2h''(x_i^2)x_i^2 + h'(x_i^2)]\), \(\triangle_x j(D) \leq 0\) a.e. and thus the estimator \(\delta_Z H(X)\) is minimax.

\[ \blacksquare \]

**Remark 1:** Since \(h\) and \(h'\) are both positive functions, the correlation inequality is not needed to show that

\[ \sum_{i=1}^{p} h(x_i^2)h'(x_i^2) \leq D \sum_{i=1}^{p} h'(x_i^2) \]

however it seems that the bound then will not be strict enough, since for many typical choices for the functions \(h\) and \(j\), \(\frac{ABC}{2} \geq 1\) will not be satisfied.

**Remark 2:** If \(h(t) = t^{\frac{d}{2}}\) then condition i) and ii) implies that \(d \in [1, 2]\). The second derivative being negative is needed for the correlation inequality to go in the right direction in the proof for Theorem 2.2.

The following 2 examples illustrate how Theorems 2.1 and 2.2 can be used to create classes of minimax estimators that adaptability select certain coordinates to be 0. The first Example is that of Zhou and Hwang [31], and the second example shows that Theorem 2.2 extends slightly the class of minimax estimators.

**Example 2.1.** *(Zhou and Hwang [31])* Let \(X \sim N_p(\theta, I_p)\). The estimator of Zhou and Hwang in [31] is of the form:

\[ \delta(X) = (\delta_1(X), \delta_2(X), ..., \delta_p(X)) \]  

(2.28)
where
\[
\delta_i(X) = (1 - \frac{c}{\|X\|_2^2 - b_i X_i}) x_i. \tag{2.29}
\]

The estimator will be minimax \([22]\) provided \(p \geq 3\),
\[
0 < c \leq 2(p - 2) - 2b(p - 1),
\]
and
\[
0 < b \leq \frac{p - 2}{p - 1}. \tag{2.30}
\]

Furthermore Zhou and Hwang \([31]\) showed that the thresholded estimator with coordinates of the form
\[
\delta_i^+(X) = (1 - \frac{c}{\|X\|_2^2 - b_i X_i})_+ x_i \tag{2.31}
\]
will dominate \(\delta(X)\), and thus is a minimax estimator for \(\theta\) under quadratic loss.

This result follows from Theorem 2.2 by setting \(j(t) = t^{-\frac{a}{2}}\) with \(a > 0\), and \(h(t) = t^{\frac{a}{2}}\) with \(d \in (1, 2]\) in Theorem 2.2. With these choices the \(i\)th coordinate of the estimator \(\delta_{ZH}\), denoted by \(\delta_{ZH,i}\), is
\[
\delta_{ZH,i} = x_i + \left( -\frac{a x_i^d}{\sum_{i=1}^p |x_i|^d} \right) x_i.
\]

To check the finiteness of the risk it is sufficient to show that \(E_\theta[||g(X)||^2] < \infty\). Since
\[
E_\theta[||g(X)||^2] = E_\theta[\frac{a x_i^d}{\sum_{i=1}^p (x_i^d)^{\frac{d}{2}}} \sum_{i=1}^p (x_i^d)^{\frac{d}{2}} x_i^2] \leq
\]
\[
K_1 E_\theta[\frac{\sum_{i=1}^p (x_i^d)^{\frac{d}{2}}}{(\sum_{i=1}^p (x_i^d)^{\frac{d}{2}})^2}] \tag{2.32}
\]
for some positive constant \(K_1\), and since the terms in the sum of (2.32)
\[
\frac{x_i^d}{(\sum_{i=1}^p |x_i|^d)^{\frac{d}{2}}} = \frac{|x_i|^{2d-2}}{(\sum_{i=1}^p |x_i|^d)^{\frac{d}{2}}} \leq
\]
\[
\frac{(\sum_{i=1}^p |x_i|^d)^{\frac{d}{2}}}{(\sum_{i=1}^p |x_i|^d)^{\frac{d}{2}}} \leq \frac{1}{(\sum_{i=1}^p x_i^d)^{\frac{d}{2}}} \leq (1)^{\frac{2d-2}{d}} \frac{K_2}{(\sum_{i=1}^p x_i^d)^{\frac{d}{2}}} \tag{2.33}
\]
by Jensen’s inequality as \(d \in (1, 2]\), it follows that
\[
\frac{1}{K_2} \sum_{i=1}^{P} (|x_i|^{d}) \cdot \frac{2}{d} \leq (\sum_{i=1}^{P} |x_i|^{d}) \cdot \frac{2}{d}.
\]

Combining the bounds given in (2.32) and (2.33)

\[E_{\theta}[\|g(X)\|^2] \leq pK_1K_2E[\frac{1}{\sum_{i=1}^{P} x_i^2}] \leq K_3E[\frac{1}{\sum_{i=1}^{P} x_i^2}]\]

which is finite as long as \(p \geq 3\).

To check when conditions i)-iv) are satisfied in Theorem 2.2, conditions i-ii) are satisfied once \(d \in (1, 2]\), as \(h(t)\) and \(h'(t)\) are both non negative \(\forall t \geq 0\), while \(h''(t) \leq 0 \forall t \geq 0\). To find the constant \(A\) of condition iii), the expression

\[2h''(t) + h'(t) = 2 \left(\frac{d}{2}\right) \left(\frac{d}{2} - 1\right) t^{(\frac{d}{2}) - 1} + \frac{d}{2} t^{(\frac{d}{2}) - 1} = \]

\[\left(\frac{d}{2}\right) \left(\frac{d}{2} - 1\right) t^{d - 1} \geq A_{\frac{d}{2}} t^{(\frac{d}{2}) - 1} = Ah'(t).\]

is satisfied once \(A = [d - 1]\).

To find the constant \(B\) of condition iv), the expression

\[h(t) = t^{\left(\frac{d}{2}\right)} \geq B \left(\frac{d}{2}\right) t^{(\frac{d}{2}) - 1} = Bh'(t)t\]

when \(B = \frac{2}{a}\).

With \(j(t) = t^{(-\frac{d}{2})}\) and \(a > 0\), \(j'(t) \leq 0\), and \(j''(t) \geq 0\) which satisfy conditions v) and vi) of Theorem 2.2.

Setting \(C = \frac{4}{a+1}\) satisfies condition vii) in Theorem 2.2 as

\[-j'(t) \leq \frac{t}{j''(t)} = \frac{t}{(\frac{d}{2} + 1)} \geq Ct.\]

Assuming \(E[\|g(X)\|^2] \leq \infty\) is satisfied, Theorem 2.2 states that a sufficient condition for the estimator \(\delta_{ZH}(X)\) to be minimax under square error is

\[\frac{pABC}{2} = \frac{8p[\frac{d-1}{2d}]}{2(d+4)} \geq 1,\]

which is equivalent to
\[0 < a \leq 4p\left\lceil \frac{d-1}{d} \right\rceil - 4.\]

Setting \(c = \frac{ad}{2}\), which is the shrinkage constant of \(\delta_{ZH}\), the new bounds satisfy:

\[0 < a_2^d \leq 2p(d - 1) - 4d \iff 0 < c < 2p(d - 1) - 2d.\]

Setting \(d = 2 - b\) in \(\delta_{ZH}\), so that \(\delta_{ZH}\) has the same form as the estimator given in (2.43), the bounds for \(b\) satisfy

\[0 < c < 2(p - 2) - 2b(p - 1)\]

which is equivalent to the bounds given by (2.30).

Theorem 2.1 implies the positive-part version of the estimator \(\delta_{ZH}\) will dominate \(\delta_{ZH}\) since the density of \(X\) is symmetric, unimodal, and increasing in each coordinate separately for each fixed value of the others, and \(\delta_{ZH}\) is of the form

\[\delta_{ZH_i}(x) = (1 - h_i(X))x_i = (1 - \frac{c}{||X||_2^2 - |x_i|^2})x_i\]

with \(h_i(X)\) symmetric in \(X\) for all \(i\). Therefore the positive-part version of \(\delta_{ZH}\), \(\delta_{ZH_i}^+\), with \(i\)th coordinate given in (2.31) will be minimax.

Example 2.2. Let \(X \sim N_p(\theta, I_p)\). Then estimators whose \(i\)th coordinate is of the form

\[\delta_{S_i}^+(X) = (1 - \frac{2dk_1(x_i^2)_{d-2}}{(\sum_{i=1}^{p} |x_i|^d + c)^{\frac{d}{a}}})_+x_i\]

will be minimax once

\[p\left\lceil \frac{d-1}{d} \right\rceil \left[ \frac{1}{a + \frac{k_1}{a+1}} \right] > 1\]

\[\text{and}\]

\[p > 4da\]

for \(a > 1, c > 0, k_1 > 0, \text{ and } d \in (1, 2]\). As in Example 2.1, Theorem 2.2 is used to derive a minimax estimator where the result of Theorem 2.1 apply.
Let \( j(t) = e^{-k_1 \int_0^t (r+c)^a dr} \) for \( a > 1, c > 0, \) and \( k_1 > 0. \) Then

\[
 j'(t) = -k_1(t+c)^{-a}j(t) \leq 0
\]

for all \( t > 0, \) and

\[
 j''(t) = [k_1^2(t+c)^{-2a} + k_1a]j(t) > 0
\]

for \( t > 0. \) The ratio

\[
 \frac{-j'(t)}{j''(t)} = (t+c)\left[\frac{(t+c)^{a-1}}{k_1 + a(t+c)^{a-1}}\right]
\]

is a monotonic increasing function for \( t > 0, \) so that condition vii) in Theorem 2.2 is satisfied once \( C = \frac{1}{a + \frac{k_1}{ca-1}}. \)

Let \( h(t) = t^\frac{d}{2} \) with \( d \in (1,2] \) as in Example 2.1. The function \( h(t) \) satisfies condition i-iv) for \( A = d - 1 \) and \( B = \frac{d}{2}. \) Therefore the estimator \( \delta_S(X) \) whose \( i^{th} \) component is of the form

\[
 \delta_S(X)_i = (1 - \frac{2dk_1(x_i^2)^{\frac{d-2}{2}}}{(\sum_{i=1}^p |x_i|^d + c)^a})x_i
\]

will be minimax once

\[
 \frac{p^{ABC}}{2} = p^{(d-1)}\left[\frac{1}{a + \frac{k_1}{ca-1}}\right] > 1
\]

provided the risk of the estimator if finite.

A sufficient condition for the risk to be finite is \( E[\|g(X)\|^2] \leq \infty. \) Since

\[
 \frac{1}{\sum_{i=1}^p (x_i^2)^\frac{d}{2} + c} \leq \frac{p^d}{(\sum_{i=1}^p x_i^2)^\frac{d}{2}}
\]

it follows that

\[
 \left(\frac{1}{(\sum_{i=1}^p |x_i|^d + c)^a}\right)^{4a} \leq \left(\frac{p^d}{(\sum_{i=1}^p x_i^2)^\frac{d}{2}}\right)^{4a}.
\]

By the Cauchy-Schwarz inequality

\[
 E_\theta[\|g(X)\|^2] \leq \sqrt{E_\theta[16d^4k_1^4(\sum_{i=1}^p (x_i^2)^{d-1})] \sqrt{E_\theta[\frac{1}{\sum_{i=1}^p |x_i|^d + c}]} < \infty \]
once \( p > 4d \). This follows as \( \mathbb{E}_\theta \left[ 16d^4 k_1^4 \left( \sum_{i=1}^{p} (x_i^2)^{d-1} \right)^2 \right] \) is finite since all positive moments of a normal distribution exist, while using expression (2.38)

\[
\mathbb{E}_\theta \left[ \left( \frac{1}{\sum_{i=1}^{p} |x_i|^d + r} \right)^{4a} \right] \leq K \mathbb{E}_\theta \left[ \left( \sum_{i=1}^{p} x_i^2 \right)^{2da} \right]
\]

for some positive constant \( K \) which is finite if \( p > 4d \) since the expected value is the \( 2da \)th moment with respect to an inverse \( \chi^2_p \) distribution.

Since the distribution of \( X \) is a spherical multivariate normal distribution with \( \sigma = 1 \), and \( \delta_S \) is of the form

\[
\delta_S(X)_i = (1 - \frac{2dk_1(x_i^2)^{d-2}}{(\sum_{i=1}^{p} |x_i|^d + c)^{a}})x_i = (1 - h_i(X))x_i
\]

with \( h_i(X) \) symmetric in the \( i \)th coordinate for every \( i \), Theorem 2.1 is satisfied so that the positive-part version of \( \delta_S \), with \( i \)th coordinate given by

\[
\delta_S^+(X)_i = (1 - \frac{2dk_1(x_i^2)^{d-2}}{(\sum_{i=1}^{p} |x_i|^d + c)^{a}})_+ x_i
\]

will be minimax and dominate the estimator \( \delta_S(X) \) provided

\[
p \left( \frac{d-1}{d} \right) \left[ \frac{1}{a + \frac{k_1}{c^2+r}} \right] \geq 1
\]

(2.41)

and

\[
p > 4d
\]

(2.42)

were \( a > 1, c > 0, k_1 > 0, \) and \( d \in (1, 2] \).

Under the conditions given in Example 2.1, which is the basic result of Hwang and Zhou [31], the class of estimators with \( i \)th coordinate of the form

\[
\delta_i^+(X) = (1 - \frac{c}{\|X\|_{2-\delta} \|X_i\|_{2-\delta}})_+ x_i
\]

(2.43)

which is in the class of Psuedo-Bayes estimators, are simultaneously minimax and allow adaptively selected coordinates to be estimated by 0. More specifically the \( i \)th coordinate
is set to 0 once

\[ |X_i|^b \leq \frac{c}{\|X_i\|^2 - b}. \]  \hspace{1cm} (2.44)

Maruyama [22] extends the class of \( \ell_d \) norms that can be used to threshold estimators and allows all \( d > 0 \). Theorem 2.3 generalizes the results of Maruyama to a slightly larger class of estimators. The class of estimators considered in Theorem 2.3 are no longer considered Psuedo-Bayes, but are of the form \( \delta(X) = X + \sigma^2 g(X) \).

**Theorem 2.3.** Let \( X \sim N_p(\theta, \sigma^2 I_p) \) with the location vector \( \theta \) unknown. Let the \( i^{th} \) coordinate of the estimator \( \delta_M(X) = X + \sigma^2 g(X) \) be given by

\[ \delta_{M,i}(X) = x_i - \sigma^2 c v(D) w(x_i^2) x_i \]

with \( c > 0 \) and \( D = \sum_{i=1}^{p} h(x_i^2) \). Suppose further that \( g(X) \) is weakly differentiable, \( E_\theta \|g(X)\|^2 < \infty \), and the following assumption hold:

Assume \( v(t) \) satisfies the following conditions:

i) The first derivative with respect to \( t \) exist for all \( t > 0 \).

ii) \( \forall t > 0, v(t) > 0 \), and \( v'(t) \leq 0 \).

iii) There exist a constant \( F > 0 \) such that \( \frac{-4v'(t)}{v(t)} \leq F, \forall t > 0 \)

iv) There exist a constant \( E > 0 \) such that \( v(D) \sum_{i=1}^{p} w(x_i^2) x_i^2 \leq E \) for all \( x_i \).

Assume that \( w(t) \) satisfies the following conditions:

v) \( w(t) \) is differentiable with respect to \( t \) for all \( t > 0 \).

vi) \( w(t) > 0 \), and \( w'(t) < 0 \) for all \( t > 0 \).

vii) There exist a constant \( A < 0 \) such that \( Aw(t) \leq w'(t)t, \forall t \geq 0 \).

viii) For \( t > 0 \) \( w(t)t \) is monotonic increasing in \( t \).

Assume that \( h(t) \) satisfies the following conditions:
ix) $h(t)$ is differentiable in $t$ for all $t > 0$.

x) $h(t) > 0$ and $h'(t) \geq 0$ for all $t > 0$.

xi) There exist a constant $B > 0$ such that $h'(t)t \leq Bh(t)$, $\forall t \geq 0$.

Then the estimator $\delta_M(X)$ will be minimax under square error loss provided the inequality

$$0 \leq cE + FB \leq 4p(A + \frac{1}{2})$$

(2.45)

holds.

**Proof:** $E_\theta \|g(X)\|^2 < \infty$ is sufficient for the risk of $\delta_M(X)$ to be finite. Without loss of generality assume $\sigma^2 = 1$. The partial derivative with respect to $x_i$ of $g_i(X)$ is

$$\frac{\partial}{\partial x_i} - cv(D)w(x_i^2)x_i =$$

$$-c(2v'(D)h'(x_i^2)x_i^2 + 2v(D)w'(x_i^2)x_i^2 + v(D)w(x_i^2))$$

and

$$\|g(X)\|^2 = c^2v^2(D)\sum_{i=1}^p w^2(x_i^2)x_i^2.$$

Let

$$L_1 = \frac{v(D)\sum_{i=1}^p [w(x_i^2)(w(x_i^2)x_i^2)]}{\sum_{i=1}^p w(x_i^2)},$$

$$L_2 = -\frac{4v'(D)\sum_{i=1}^p h'(x_i^2)x_i^2w(x_i^2)}{v(D)\sum_{i=1}^p w(x_i^2)},$$

and

$$L_3 = 4\left(\frac{\sum_{i=1}^p [w'(x_i^2)x_i^2 + w(x_i^2)]}{\sum_{i=1}^p w(x_i^2)}\right).$$

Outside a set of measure 0, $\|g(X)\|^2 + 2div(g(X))$ is expressible as

$$cv(D)\sum_{i=1}^p w(x_i^2)[cL1 + L2 - L3].$$

(2.46)
By Stein’s inequality a sufficient condition for the estimator $\delta_M(x)$ to be minimax is expression (2.46) being non-positive. Since $cv(D) \sum_{i=1}^p w(x_i^2)$ is positive by assumption ii) and vi), expression (2.46) being non positive is equivalent to

$$cL_1 + L_2 \leq L_3.$$ \hspace{1cm} (2.47)

By assumption viii)

$$\frac{v(D) \sum_{i=1}^p w(x_i^2)x_i^2}{p} \leq \frac{E}{p}$$

so that $L_1$ satisfies

$$cL_1 \leq \frac{cE}{p} \hspace{1cm} (2.48)$$

by Lemma 2.1 since $w(t)$ is decreasing (assumption vi) and $w(t)t$ is increasing in $t$. $L_2$ satisfies

(by assumption ii) $L_2 \leq \frac{F \sum_{i=1}^p h'(x_i)w(x_i^2)}{\sum_{i=1}^p h(x_i^2)(\sum_{i=1}^p w(x_i^2))} \leq$

(by assumption xi) $\frac{FB \sum_{i=1}^p h(x_i^2)w(x_i^2)}{p(\sum_{i=1}^p h(x_i^2)(\sum_{i=1}^p w(x_i^2)) = FB \frac{p}{p}} \hspace{1cm} (2.49)$

where (2.49) follows from Lemma 2.1 Since $h$ is monotonic increasing (assumption x) and $w$ is monotonic decreasing (assumption vi).

Furthermore by assumption vii) $L_3$ satisfies

$$4(A + \frac{1}{2}) = 4 \frac{\sum_{i=1}^p [Aw(x_i^2)] + w(x_i^2)}{\sum_{i=1}^p w(x_i^2)} \leq L_3.$$

Therefore once $\frac{cE + FB}{p} \leq 4(A + \frac{1}{2})$

$$cL_1 + L_2 \leq \frac{cE + FB}{p} \leq 4(A + \frac{1}{2}) \leq L_3 \hspace{1cm} (2.50)$$

so that (2.46) non-positive and $\delta_M(X)$ is minimax.
Example 2.3. (Maruyama [22]) Let $X \sim N_p(\theta, I_p)$ and let $\delta_M(X)$ be an estimator of $\theta$ with $i^{th}$ coordinate of the form

$$
\delta_{M_i}(x) = (1 - \frac{c}{\sum_{i=1}^{p}(x_i^2)^{\frac{d}{2}}(\frac{2-a}{d})})x_i.
$$

The results of Maruyama [22] imply $\delta_M(X)$ is minimax once

$$
0 \leq c_{\max}\{1, p^{1-(\frac{2-a}{d})}\} \leq 2(p - 2 - a(p - 1)) \quad (2.51)
$$

for $p \geq 3$, and $0 \leq a \leq \frac{p-1}{p-2}$.

This result follows from Theorem 2.3 by letting: $h(t) = t^{\frac{d}{2}}$, $v(t) = t^{-\left(\frac{2-a}{d}\right)}$, and $w(t) = t^{-\left(\frac{d}{2}\right)}$. For the function $h(t)$ conditions ix-xi are satisfied once $d > 0$ with $B = \frac{d}{2}$.

Checking the conditions for $w$ where $w(t) = t^{-\left(\frac{d}{2}\right)}$, conditions vii and viii will be satisfied with $0 < a < 2$. The constant $A = -\frac{a}{2}$ will satisfy condition vii as

$$
Aw(t) = At^{-\left(\frac{d}{2}\right)} = -(\frac{a}{2})t^{-\left(\frac{d}{2}\right)} = w(t)t.
$$

Checking the conditions for the function $v$, conditions i and ii are satisfied for $0 < a < 2$. Condition iii is satisfied once $F = 4\left(\frac{2-a}{d}\right)$ since $-\frac{4v(t)}{v(t)} = 4\left(\frac{2-a}{d}\right)$. Since

$$
V(D)\sum_{i=1}^{p} w(x_i^2)x_i^2 = \frac{\sum_{i=1}^{p}(x_i^2)^{\frac{d}{2}}(\frac{2-a}{d})}{\sum_{i=1}^{p}(x_i^2)^{\frac{d}{2}}(\frac{2-a}{d})}
$$

the constant $E$ which satisfies condition iv can be broken into 2 cases.

Case 1: $\frac{2-a}{d} > 1$. Since

$$
\left[\sum_{i=1}^{p}(x_i^2)^{\frac{d}{2}}(\frac{2-a}{d})\right] \geq \sum_{i=1}^{p}(x_i^2)^{\frac{2-a}{d}} \quad (2.52)
$$

we can choose $E$ to be 1.

Case 2: $\frac{2-a}{d} < 1$.

From Jensen’s inequality for a concave functions (i.e $c(x)$ that $c(EX) \geq Ec(X)$)

$$
\left[\sum_{i=1}^{p}(x_i^2)^{\frac{d}{2}}(\frac{2-a}{d})\right] \geq \sum_{i=1}^{p}(x_i^2)^{\frac{2-a}{d}} \quad (2.53)
$$
which implies
\[
\frac{\sum_{i=1}^{p} x_i^2 \left( \frac{2-a}{d} \right)}{\left( \sum_{i=1}^{p} x_i^2 \frac{d}{2} \right)^\left( \frac{2-a}{d} \right)} \leq \left( \frac{1}{p} \right)^\left( \frac{2-a}{d} \right) - 1
\]
so that \( E = p^{1-\left( \frac{2-a}{d} \right)} \) can be chosen to satisfy condition iv. From Case 1 and 2 then \( E = \max\{1, p^{1-\left( \frac{2-a}{d} \right)}\} \) will satisfy condition iv.

The substitution of \( A, B, E, \) and \( F \) into expression (2.45) of Theorem 2.3 yields the inequality
\[
0 \leq \text{cmax}\{1, p^{1-\left( \frac{2-a}{d} \right)}\} \leq 4p^{\left( \frac{1-a}{2} \right)} - 4\left( \frac{2-a}{d} \right)\left( \frac{d}{2} \right) = 2(p - 2 - a(p - 1)).
\]
which is equivalent to the result of Maruyama[22].

For the risk to exist it is sufficient that \( p \geq 3 \). This follows from expressions (2.52) and (2.53) as
\[
E[\|g(X)\|^2] = E_\theta \left\| \frac{x_i^2}{\left( \sum_{i=1}^{p} x_i^2 \frac{d}{2} \right)^\left( \frac{1-2a}{d} \right)\left( x_i^2 \right)^a} \right\|^2 \leq KE_\theta \left[ \frac{\sum_{i=1}^{p} x_i^2 d^{d-1}}{\left( \sum_{i=1}^{p} x_i^2 \frac{d}{2} \right)^2} \right] (2.54)
\]
for some positive constant \( K \) where the change of variables \( d' = 2 - a \) in (2.54) is used. Expression (2.54) is equivalent to expression (2.32) in Example 2.1 which is bounded for \( p \geq 3 \).

Theorem 2.1 implies the positive-part version of the estimator \( \delta_M \) will dominate \( \delta_M \) since the density of \( X \) is symmetric, unimodal, and increasing in each coordinate separately for each fixed value of the others, and \( \delta_M \) is of the form
\[
\delta_M(x) = (1 - h_i(X))x_i = (1 - \frac{c}{\left( \sum_{i=1}^{p} x_i^2 \frac{d}{2} \right)^\left( \frac{2-a}{d} \right)\left( x_i^2 \right)^a})x_i
\]
with \( h_i(X) \) symmetric in \( X \) for all \( i \). Therefore the positive-part version of \( \delta_M \), \( \delta_M^+ \), with \( i \)th coordinate
\[
\delta_M^+(X) = (1 - \frac{c}{\left( \sum_{i=1}^{p} x_i^2 \frac{d}{2} \right)^\left( \frac{2-a}{d} \right)\left( x_i^2 \right)^a})_+ x_i \quad (2.55)
\]
is a minimax estimator of \( \theta \) once

\[
0 \leq \text{cmax}\{1, p^{1-\left(\frac{2-a}{p}\right)}\} \leq 2(p - 2 - a(p - 1))
\]  

(2.56)

for \( p \geq 3 \) and \( 0 \leq a \leq \frac{p-1}{p-2} \).

2.2.2 Simulation Results

In this section we assume that \( X \sim N_p(\theta, I_6) \) and look at simulations of risk over subspaces of the parameter space in \( \mathbb{R}^6 \) for estimators of the form

\[
\delta^+_Z(x)_i = (1 - \frac{c}{\|x\|_2^{2-a}|x_i|^a})_+ x_i,
\]  

(2.57)

and

\[
\delta^+_M(x)_i = (1 - \frac{c}{\|x\|_2^{2-a}|x_i|^a})_+ x_i.
\]  

(2.58)

Particularly we compute the risk over the following 4 spaces:

\[
V_1 = \langle e_1 \rangle
\]  

(2.59)

\[
V_2 = \langle e_1, e_2 \rangle
\]  

(2.60)

\[
V_3 = \langle e_1, e_2, e_3 \rangle
\]  

(2.61)

\[
V_4 = \langle e_1, e_2, e_3, e_4, e_5, e_6 \rangle
\]  

(2.62)

where \( e_i \) denotes the standard basis vectors of \( \mathbb{R}^6 \). Comparisons of simulations of risk for particularly chosen estimators of the form (2.57) and (2.58) with the positive-part James-Stein estimator, \( \delta^+_B(X) = (1 - \frac{4}{\|x\|_2^2})_+ x_i \), is given to asses if thresholding each coordinate of the estimator individually yields better risk than thresholding all the coordinates simultaneously. A companion figure plotting the simulated probability that a coordinate will be selected as a function of the \( \ell_2 \) norm of the parameter space for \( V_1 \) and \( V_3 \) will be provided as well. This will serve as a means of assessing if the thresholding procedure for estimators of the form (2.57) and (2.58) can distinguish coordinates with parameter values of 0, from
those with parameter values different from 0. To end the subsection, figures plotting the simulated risk and probability of selection will be given for $p = 12$ and $p = 17$ for $\theta \in <e_1>$, where here $e_1$ is the first standard basis vector of $\mathbb{R}^{12}$ and $\mathbb{R}^{17}$ respectively. This is done to see the effect that sparsity can have on risk and variable selection.

Figure 2.1 plots the simulated risk of $\delta_{ZH}$ for two different sets of parameter choices. Depending on the particular parameters chosen the risk at the origin can vary greatly. In this case the choice of $a = .2$ seems to offer savings over the choice of $a = .5$ when the shrinkage constant, $c$, was chosen near the maximal value required for minimaxity for both estimators.

![Figure 2.1: Simulated Risk of Zhou and Hwang Estimator when p=6](image)

Figure 2.2 plots the simulated risk of the Maruyama estimator with the same parameter values for the parameter, $a$, as Figure 2.1. The shrinkage constant $c$, was chosen to be near the maximal value required for minimaxity for both estimators. The choice for the parameter $d$ has an impact on the risk. In this case when $a = .5$ and $d = .3$, the simulated risk was larger than it was for the Zhou and Hwang estimator when $a = .5$ and $2 - a = 1.5$. In the adjacent figure when $a = .2$ and $d = 1.9$ the simulated risk was similar to that of the Zhou and Hwang estimator with $a = .2$ and $2 - a = 1.8$.  

Figure 2.2: Simulated Risk of Maruyama Estimator when p=6

Figure 2.3 compares the simulated risk of the Maruyama estimator and Zhou and Hwang estimator from Figures 2.2 and 2.3 when a=.2, to the positive part estimator James -Stein estimator for parameter values in $V_1$ and $V_3$. Over the sparse vector space $V_1$ both the Zhou and Hwang estimator and Maruyama estimator dominate the positive part James -Stein estimator. In the less sparse vector space $V_3$, the estimators of Maruyama and Zhou and Hwang do better at parameter values close to the origin. For large values of the parameter space the risk of the positive-part James-Stein estimator is similar to the risks of the Maruyama estimator and the Zhou and Hwang estimator.

Figure 2.3: Risk Comparisons in $V_1$ and $V_3$
Figures 2.3, Figure 2.4 seems to suggest that in part the savings in risk over the positive-part James-Stein estimator is due to selecting coordinates with zero parameter values to be estimated as 0 with a higher probability than those with parameter values not equal to 0. In the less sparse vector space of $V_3$ the difference in probability of selecting coordinates with parameter values equal to 0 is only slightly lower than those coordinates with parameter values different from 0 for both the Zhou and Hwang estimator and the Maruyama estimator with the difference being nearly 0 at $\|\theta\|^2 = 4$. This corresponded to the risk of the Zhou and Hwang estimator and Maruyama estimator in Figure 2.3 being comparable to the positive-part James-Stein estimator at about $\|\theta\|^2 = 4$. For $\theta \in V_1$ the difference in probability is greater, which corresponded to the Zhou and Hwang Estimator and the Maruyama estimator dominating the positive-part James-Stein estimator.

![Figure 2.4: Probability of Variables Being Selected](image)

We conclude the section with Figure 2.5 and 2.6 which plots both the simulated risks and simulated probabilities of variables getting selected for by the thresholding procedure for the Zhou and Hwang estimator when $a = 0.2$ and shrinkage constant, $c$, is chosen close to the maximal value for minimaxity for $p=12$ and $p=17$ respectively. The parameter $\theta$ is chosen to lie in the sparse space

$$V_{12} = \langle e_1 \rangle \text{ for } \theta \in \mathbb{R}^{12}$$

when $p=12$ and
$V_{17} = \langle e_1 \rangle$ for $\theta \in \mathbb{R}^{17}$ when $p=17$ as a means to assess the effect that sparsity has on the probability of selection and the risk.

Figures 2.5 and 2.6 show as the dimension gets larger, the difference in the probability between variables with parameter values not equal to 0 getting selected to variables with parameter values equal to 0 getting selected by the thresholding procedure increased. This corresponded to a greater savings in risk, when compared to the estimator $\delta(X) = X$, when $p=17$ than when $p=12$.

![Figure 2.5: Zhou and Hwang Estimator for $p=12$](image1)

![Figure 2.6: Zhou and Hwang Estimator for $p=17$](image2)
2.3 Berger Class of Spherically Symmetric Distributions

This Section extends Theorems 2.2 and Theorem 2.3 to the case when the underlying distribution of $X$ is spherically symmetric with density of the form $f(\|x - \theta\|^2)$, such that

$$\frac{F(t)}{f(t)} \geq b > 0$$  \hspace{1cm} (2.63)

for all $t \geq 0$ where

$$F(t) = \frac{1}{2} \int_t^{\infty} f(x)dx$$  \hspace{1cm} (2.64)

In particular the class of estimators considered in this section are of the form

$$\delta_b(X) = X + bg(X) = \begin{pmatrix}
(1 - bh_1(X))x_1 \\
(1 - bh_2(X))x_2 \\
\vdots \\
(1 - bh_p(x_p))x_p
\end{pmatrix}$$  \hspace{1cm} (2.65)

where

$$h_i(X) = \frac{4j'(\sum_{i=1}^{p} h(x_i^2))h'(x_i^2)}{j(\sum_{i=1}^{p} h(x_i^2))}$$  \hspace{1cm} (2.66)

when extending the result of Theorem 2.2, and

$$h_i(X) = cv(\sum_{i=1}^{p} h(x_i^2))w(x_i^2)$$  \hspace{1cm} (2.67)

when extending the results of Theorem 2.3. The constant, $b$, in (2.65) satisfies (2.63). The method of extension will be similar to that of Section 2.1 where a Stein-like differential inequality is used to bound the difference in risk of the estimator $\delta_0(X) = X$ and the proposed estimator. The method of extension is used in Fourdrinier and Strawderman [11] to study generalized Bayes estimators of location for spherically symmetric distributions. If in addition the distribution of $X$ is unimodal and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, the positive part-versions of
estimator (2.65),

$$\delta_b(X)^+ = \left( (1 - bh_1(X)_+ X_1) \\
(1 - bh_2(X)_+ X_2) \\
\vdots \\
(1 - bh_p(X)_+ X_p) \right)$$

will dominate the estimator $\delta_b(X)$.

The class of $L_1$ functions satisfying (2.63) include ones for which

$$\frac{d^2}{dt^2} \log f(t) \geq 0$$

for all $t \geq 0$. This follows from the fact that $\frac{f(t)}{F(t)}$ is non-decreasing so that

$$\frac{f(t)}{F(t)} = 2 \frac{- \int_t^{\infty} \frac{f'(u)}{f(u)} f(u) du}{\int_t^{\infty} f(u) du} = 2E_t[-\frac{f'(u)}{f(u)}]$$

where the expected value is taken with respect to the density proportional to $f(u)I_{[u \geq t]}(u)$. This density has an increasing monotone likelihood ratio for the parameter $t$, since for $t_2 > t_1$ the ratio of the densities is 0 for $u \in [t_1, t_2)$ and 1 for $u \geq t_2$. This implies

$$\frac{f(t_2)}{F(t_2)} = 2E_{t_2}[-\frac{f'(u)}{f(u)}] \leq 2E_{t_1}[-\frac{f'(u)}{f(u)}] = \frac{f(t_1)}{F(t_1)}$$

by the assumption $h(x) = -\frac{f'(u)}{f(u)}$ is non-increasing and hence

$$\frac{F(t_1)}{f(t_1)} \leq \frac{F(t_2)}{f(t_2)}.$$

As a result for $t \geq 0$

$$\frac{F(t)}{f(t)} \geq \frac{F(0)}{f(0)} = b$$

which will satisfy (2.63). The Berger class includes any densities which are scale mixtures of normal distributions. Note that it is not required that $f(||x - \theta||^2)$ satisfy (2.63) but $f(t)$.

The following Lemma is the Stein-like equality that will be used. It is found in Fourdrinier and Strawderman [11]. We give the proof for completeness.

**Lemma 2.2** (Fourdrinier and Strawderman [11]). Let $g(x)$ be a weakly differentiable function from $\mathbb{R}^p$ into $\mathbb{R}^p$. Let $X$ be a random vector in $\mathbb{R}^p$ with a spherically symmetric density

\[ g(X) = \sum_{i=1}^{p} a_i X_i \]

where $a_i$ are real numbers.

Then the Stein-like equality holds if $g(x)$ is a weakly differentiable function from $\mathbb{R}^p$ into $\mathbb{R}^p$.
about the point θ in \( \mathbb{R}^p \) with density \( f(\|x - \theta\|^2) \). For \( F(t) = \frac{1}{2} \int_{t}^{\infty} f(u)du \),

\[
E_{\theta}[(X - \theta)'g(X)] = E_{\theta}\left[\frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)} \text{div}_x(g(X))\right].
\] (2.70)

**Proof.** Let \( \sigma_{r,\theta} \) denote the uniform measure over the sphere of radius \( r \) centered at \( \theta \), \( S_{r,\theta} \). Changing to spherically coordinates,

\[
E_{\theta}[(X - \theta)'g(X)] = \int_{\mathbb{R}^p} (x - \theta)'g(x)f(\|x - \theta\|^2)dx = \int_{0}^{\infty} \int_{S_{r,\theta}} (x - \theta)'g(x)d\sigma_{r,\theta}(x)f(r^2)dr.
\]

By an application of Stokes’ theorem on the inner integral,

\[
\int_{0}^{\infty} \int_{S_{r,\theta}} (x - \theta)'g(X)d\sigma_{r,\theta}(x)f(r^2)dr = \int_{0}^{\infty} \int_{B_{r,\theta}} \text{div}_x(g(x))dxrf(r^2)dr.
\]

By Fubini’s theorem,

\[
\int_{0}^{\infty} \int_{B_{r,\theta}} \text{div}_x(g(x))dxrf(r^2)dr = \int_{\mathbb{R}^p} [\int_{\mathbb{R}^p} f(\|x - \theta\|^2)dx]f(r^2)dr \int_{\mathbb{R}^p} \text{div}_x(g(x))dx.
\]

Letting \( u = r^2 \) so that by the change of variables formula

\[
\int_{\mathbb{R}^p} [\int_{\mathbb{R}^p} f(\|x - \theta\|^2)dx]f(r^2)dr \int_{\mathbb{R}^p} \text{div}_x(g(x))dx = \int_{\mathbb{R}^p} \frac{1}{2} \int_{\mathbb{R}^p} f(\|x - \theta\|^2)2rf(\|x - \theta\|^2)\text{div}_x(g(x))dx = \int_{\mathbb{R}^p} \frac{F(\|x - \theta\|^2)}{f(\|x - \theta\|^2)} \text{div}_x(g(x))f(\|x - \theta\|^2)dx =
\]

\[
E_{\theta}\left[\frac{F(\|x - \theta\|^2)}{f(\|x - \theta\|^2)} \text{div}_x(g(X))\right]
\]

establishing the result.

\[\blacksquare\]

The following Lemma gives conditions for which estimators of the form

\[
\delta_{\alpha}(X) = X + ag(X)
\] (2.71)

will be minimax for the class of spherically densities that satisfy (2.63).
Lemma 2.3. Let $X \sim SS_p(\theta)$ with all second moments existing and density $f(\|x - \theta\|^2)$ were the location vector $\theta$ in $\mathbb{R}^p$ is unknown. Let $F(t) = \frac{1}{2} \int_{-\infty}^{\infty} f(u) du$ and assume that the density of $X$ satisfies

$$\frac{F(t)}{f(t)} \geq b > 0 \text{ for all } t \geq 0.$$  

Assume that in the special case when $X \sim N_p(\theta, I_p)$ the estimator $\delta(X) = X + g(X)$ satisfies:

i) $g(X)$ is a weakly differentiable function from $\mathbb{R}^p$ into $\mathbb{R}^p$,

ii) $E_{\theta}[\|g(X)\|^2] < \infty$,

iii) $\|g(X)\|^2 + 2 div_x(g(X)) \leq 0$,

then estimator $\delta_b(X) = X + bg(X)$ is minimax for $\theta$ for quadratic loss, $\|d - \theta\|^2$.

Proof. The risk of the minimax estimator $X$ is

$$R(\theta, X) = E_{\theta}[\|X - \theta\|^2]$$

while the risk of the estimator $\delta_b(X)$ is

$$R(\theta, \delta_b) = E_{\theta}[\|X + bg(X) - \theta\|^2] =$$

$$E_{\theta}[\|X - \theta\|^2 + 2b(X - \theta)'g(X) + b^2\|g(X)\|^2].$$

Let $\triangle_{\theta}(\delta_b)$ denote the difference in risk between the estimator $X$ and the estimator $\delta_b$. Then

$$\triangle(\delta_b) = E_{\theta}[b^2\|g(X)\|^2 + 2b(X - \theta)'g(X)]. \quad (2.72)$$

Since the conditions of Lemma 2.2 are satisfied expression (2.72) is equivalent to

$$\triangle_{\theta}(\delta_b) = E_{\theta}[b^2\|g(X)\|^2 + 2b\frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)}div_x(g(X))].$$

Since $\frac{F(t)}{f(t)} \geq b > 0$, and $\|g(x)\|^2 + div_x(g(x)) \leq 0$ implying $div_x(g(x)) \leq 0$ as $\|g(X)\|^2 \geq 0$, the difference in risk satisfies

$$\triangle_{\theta}(\delta_b) \leq E_{\theta}[b^2\|g(X)\|^2 + b^22div_x(g(X))].$$
Theorem 2.4 is a straightforward consequence of Lemma 2.3 and extends the class of estimators given in Section 2.1 to the class of spherical distributions that satisfy (2.63).

**Theorem 2.4.** Let \( X \sim SS_p(\theta) \) with finite second moments, \( \dim(X) = \dim(\theta) = p \geq 3 \), and density \( f(\|x - \theta\|^2) \) satisfying

\[
\frac{F(t)}{f(t)} \geq b > 0 \quad (2.73)
\]

for all \( t \geq 0 \) where \( F(t) = \int_t^{\infty} f(u)du \). The estimators

\[
\delta_{1,b}(X) = \begin{pmatrix}
(1 - b\frac{j'(D)h'(X_1^2)}{j(D)})X_1 \\
(1 - b\frac{j'(D)h'(X_2^2)}{j(D)})X_2 \\
\vdots \\
(1 - b\frac{j'(D)h'(X_p^2)}{j(D)})X_p
\end{pmatrix} = X + bg_1(X) \quad (2.74)
\]

where \( D = \sum_{i=1}^p h(X_i^2) \), and \( h \) and \( j \) satisfying conditions i)-vii) of Theorem 2.2 with \( \frac{pABC}{2} \geq 1 \), and

\[
\delta_{2,b}(X) = \begin{pmatrix}
(1 - bcv(D)w(X_1^2))X_1 \\
(1 - bcv(D)w(X_2^2))X_2 \\
\vdots \\
(1 - bcv(D)w(X_p^2))X_p
\end{pmatrix} = X + bg_2(X) \quad (2.75)
\]

where \( D = \sum_{i=1}^p h(X_i^2) \), and \( v, w, \) and \( h \) satisfying conditions i)-xi) of Theorem 2.3 with \( cE + FB \leq 4p(A + \frac{1}{2}) \) are minimax estimators of \( \theta \) under quadratic loss. If in addition, \( f \) is unimodal, and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, the positive-part versions of the original estimators dominate the original estimators.

**Proof** For \( p \geq 3 \), it follows from the proofs of Theorem 2.2 and Theorem 2.3 that when

\[
E_\theta[b^2(\|g(X)\|^2 + 2\text{div}_x(g(X)))] \leq 0.
\]
$X \sim N_p(\theta, I_p)$ the estimators $\delta_{1,1}(X)$ and $\delta_{2,1}$ satisfy

$$\|g_i(x)\|^2 + 2\text{div}_x(g_i(x)) \leq 0$$

(2.76)

for $i = 1, 2$. Since $f$ satisfies the conditions of Lemma 2.3, it follows that $\delta_{1,b}$ and $\delta_{2,b}$ are minimax for the general case that the distribution of $X$ is spherically symmetric. If in addition $f$ is unimodal and non-increasing in each of its coordinates separately for each fixed value of the other coordinates, Theorem 2.1 implies the positive-part versions of the original estimators dominate the original estimators.

Example 2.4. (Normal with known variance) Let $X \sim N_p(\theta, \sigma^2 I_p)$ with $\sigma^2$ known. Then

$$f(X) = (2\pi \sigma^2)^{-p/2} e^{-\frac{1}{2\sigma^2} \|X-\theta\|^2}.$$ 

Setting $t = \|X - \theta\|^2$,

$$\frac{d^2}{dt^2} f(t) = 0$$

and hence $f$ is log convex. Therefore

$$\frac{F(t)}{F(0)} \geq \frac{F(0)}{F(0)} = \frac{1}{2} \int_0^\infty e^{-\frac{x}{\sigma^2}} dx = \sigma^2.$$ 

Assume now that $X \sim N_p(\theta, I_p)$ and let

$$\delta_{ZH}(x)_i = (1 - \frac{ad}{p_x} x_i^{d-2}) x_i = (1 - h_{1i}(x)) x_i$$

with $d \in (1, 2]$, $0 < a < 4[p \frac{d-1}{d} - 1]$, and $p \geq 3$,

$$\delta_S(x)_i = (1 - \frac{2dk_1 x_i^{d-2}}{(\sum_{i=1}^p |x_i|^d + c)^2}) x_i = (1 - h_{2i}(x)) x_i$$

with $d \in (1, 2]$, $p > 4da$, and $p[\frac{d-1}{d}]\left[\frac{1}{a + \frac{k_1}{c^2}}\right] > 1$ for $a > 1$, $c > 0$, $k_1 > 0$, and $d \in (1, 2]$,

and

$$\delta_M(x)_i = (1 - \frac{c}{(\sum_{i=1}^p |x_i|^d)^{\frac{2}{d} - \frac{1}{d}}}) x_i = (1 - h_{3i}(x)) x_i$$
with \( p \geq 3, d > 0, 0 \leq a \leq \frac{p-1}{p^2} \), and \( 0 \leq c \cdot \max\{1, p^{1-\frac{2-a}{a}}\} \leq 2(p-2+a(p-1)) \).

Examples 2.1 and 2.2 established that \( \delta_{ZH}(x) \) and \( \delta_S(x) \) satisfy the conditions of Theorem 2.2 and Example 2.3 established that \( \delta_M(X) \) satisfies the conditions of Theorem 2.3 so that by Theorem 2.4 estimators with \( i \)-th coordinates of the form

\[
\delta^+_{ZH,\sigma^2}(x)_i = (1 - \sigma^2 \frac{ad|x_i|^{d-2}}{\sum_{i=1}^{d} |x_i|^a}) + x_i
\]

with \( d \in (1, 2), 0 < a < 4\{p^{\frac{d-1}{d}} - 1\} \), and \( p \geq 3 \),

\[
\delta^+_S(\sigma^2)(x)_i = (1 - \frac{2dk_1\sigma^2|x_i|^{d-2}}{\sum_{i=1}^{d} |x_i|^a}) + x_i
\]

with \( d \in (1, 2), p > 4da \), and \( p\{\frac{d-1}{d}\} \left[1 - \frac{1}{a+x^{-a}}\right] > 1 \) for \( a > 1, c > 0, k_1 > 0 \), and \( d \in (1, 2) \), and

\[
\delta^+_M(\sigma^2)(x)_i = (1 - \frac{ca^2}{\sum_{i=1}^{d} |x_i|^a}) + x_i
\]

with \( p > 3, d > 0, 0 \leq a \leq \frac{p-1}{p^2}, \) and \( c \cdot \max\{1, p^{1-\frac{2-a}{a}}\} \leq 2(p-2+a(p-1)) \) are minimax estimators of \( \theta \).

Example 2.5. (Scale mixtures of Normals) Let \( X|v \sim N_p(\Theta, vI_p) \) with the distribution of \( V \) having density \( h(v) \) with \( E[V^{-p/2}] \) finite. Setting \( t = \|X - \Theta\|^2 \), the marginal distribution of \( X \) has the form

\[
f(t) = (2\pi)^{-p/2} \int_0^\infty v^{-(p/2)} e^{-\frac{1}{2v}h(v)} dv.
\]

For \( t_2 > t_1 \)

\[
\frac{v^{-p/2} e^{-\frac{1}{2v}h(v)}}{v^{-p/2} e^{-\frac{1}{2v}h(v)}} = e^{-\frac{t_2 + t_1}{2v}}
\]

which is increasing in \( v \) and hence the family of distributions with densities proportional to

\[
j_t(v) = v^{-p/2} e^{-\frac{1}{2v}h(v)}
\]

will have increasing monotone likelihood ratio in \( t \). Since

\[
\frac{f'(t)}{f(t)} = -\frac{1}{2} \int_0^\infty v^{p-2} e^{-\frac{1}{2v}h(v)} dv = -\frac{1}{2} E[V^{-1}],
\]

where the expected value is taken with respect to the density \( Kj_t(V) \), with normalizing constant \( K \), \( \frac{f'(t)}{f(t)} \) is a non-decreasing function of \( t \) as \( -\frac{1}{2} v^{-1} \) is an increasing function of \( v \) and hence for \( t_1 < t_2 \).
by the monotone likelihood ratio property of \( j_1(v) \). Expression (4.40) implies the constant \( b \) in Lemma 2.3 has the form

\[
F(0) = \frac{1}{2} \int_{0}^{\infty} e^{(p/2)h(v)} [\int_{0}^{x} e^{-\frac{x}{2y}} dy] dv = \int_{0}^{x} e^{(p/2)h(v)} dv = \frac{E[V'(p/2)+1]}{E[V'(p/2)]} = b > 0
\]  

(2.77)

Let

\[
\delta_{ZH}(x)_i = (1 - \frac{d|x_i|^d-2}{\sum_{i=1}^{d} |x_i|^d}) x_i = (1 - h_{1i}(x)) x_i
\]

with \( d \in (1, 2], 0 < a < 4[p^{d-1}/d - 1], \) and \( p \geq 3, \)

\[
\delta_{S}(x)_i = (1 - \frac{2dk_1|x_i|^d-2}{\sum_{i=1}^{d} |x_i|^d}) x_i = (1 - h_{2i}(x)) x_i
\]

with \( d \in (1, 2], p > 4da, \) and \( p[\frac{d-1}{d}]\left[\frac{1}{a+c}\right] > 1, \) for \( a > 1, c > 0, k_1 > 0, \) and \( d \in (1, 2], \)

\[
\delta_{M}(x)_i = (1 - \frac{c}{\sum_{i=1}^{d} |x_i|^d}) x_i = (1 - h_{3i}(x)) x_i
\]

with \( p > 3, d > 0, 0 \leq a \leq \frac{p-1}{p^2}, \) and \( 0 \leq c \leq \max\{1, p[1-2/p^2] \leq 2(p-2+a(p-1))\}.\)

Examples 2.1 and 2.2, established that \( \delta_{ZH}(x) \) and \( \delta_{S}(x) \) satisfy the conditions of Theorem 2.2 and Example 2.3 established that \( \delta_{M}(x) \) satisfies the conditions of Theorem 2.3. Furthermore since \( f \) is symmetric, unimodal, and non-increasing in each of its coordinates by Theorem 2.4 estimators with \( i \)th coordinate of the form

\[
\delta_{ZH,b}^+(x)_i = (1 - \frac{E[V'(p/2)+1]}{E[V'(p/2)]} \frac{ad|x_i|^d-2}{\sum_{i=1}^{d} |x_i|^d}) x_i
\]

with \( d \in (1, 2], 0 < a < 4[p^{d-1}/d - 1], \) and \( p \geq 3, \)

\[
\delta_{S,b}^+(x)_i = (1 - \frac{2dk_1 E[V'(p/2)+1] |x_i|^d-2}{\sum_{i=1}^{d} |x_i|^d}) x_i
\]

with \( d \in (1, 2], p > 4da, \) and \( p[\frac{d-1}{d}]\left[\frac{1}{a+c}\right] > 1, \) for \( a > 1, c > 0, k_1 > 0, \) and \( d \in (1, 2], \) and
\[
\delta^+_{M,b}(x)_i = (1 - \frac{E[V(-p/2)+1, \sigma^2]}{(\sum_{i=1}^p |x_i|^d)^{\frac{1}{d}} |x_i|^a}) + x_i
\]

with \(p > 3, d > 0, 0 \leq a \leq \frac{p-1}{p-2}, \) and \(c \ast \max\{1, p^{1-\frac{2-a}{d}} \leq 2(p-2+a(p-1))\}\) are minimax estimators of \(\theta.\)

Berger [4] provides additional example of densities that satisfy (2.63) so that the estimators developed in Section 2 extends to these distributions as well. These density include densities of the form

\[
f(s) = ks^n e^{-\frac{s}{2}}
\]

with \(n \geq 2\) for \(b=2\) in (2.63),

\[
f(s) = \frac{k}{\cosh(s)}
\]

with the constant \(b=1\) in (2.63),

\[
f(s) = \frac{ks}{(1+s^2)^{m+1}}
\]

with \(m > \frac{p}{4}\) and the constant \(b = \frac{1}{m}\) in (2.63) and,

\[
f(s) = \frac{ke^{-as-b}}{(1+e^{-as-b})^2}
\]

with \(b = \frac{1}{a}\) in (2.63).

### 2.4 Normal Distribution with Residual Vector

In this Section we extend the results of Section 2.1 to the case that the underlying distribution is normal with a residual vector present. In particular

\[
\begin{pmatrix}
X \\
U
\end{pmatrix} \sim N_p\left(\begin{pmatrix}
\theta \\
0
\end{pmatrix}, \sigma^2 I_{p+k}\right)
\]

with \(\text{dim}(X) = \text{dim}(\theta) = p \geq 3\) and \(\text{dim}(U) = \text{dim}(0) = k \geq 1.\) The class of estimators considered are of the form

\[
\delta_S(X, S) = X + \frac{S}{k+2} g(X)
\]

(2.79)
where $S = \|U\|^2$, and $X$ are complete sufficient statistics and $g(X)$ is given by (2.66) or (2.67).

Condition for $g$ will be given so that the risk of the estimator, $\delta_S(X,S)$ dominates the usual estimator $\delta_0(X) = X$ under quadratic loss, $\|d - \theta\|^2$. If instead the loss is given by

$$l(d, \theta) = \frac{\|d - \theta\|^2}{\sigma^2}$$

(2.80)

$\delta_S(X,S)$ will also be minimax. Just as in Section 3, a Stein-like differential equality is used to bound the difference in risk between the estimator, $\delta_0(X) = X$, and the proposed estimator. The following lemma found in Fourdrinier, Strawderman, Wells [13] gives the form of Stein Lemma used in the Section.

**Lemma 2.4.** Let $\left(\begin{array}{c} X \\ U \end{array}\right) \sim N_{p+k}(\left(\begin{array}{c} \theta \\ 0 \end{array}\right), \sigma^2 I_{p+k})$ with $\theta$ and $\sigma^2$ unknown where $\dim(X) = \dim(\theta) = p \geq 1$ and $\dim(U) = \dim(0) = k \geq 1$. Let $S = \|U\|^2$, and $h(x,s)$ be a function from $\mathbb{R}^p \times \mathbb{R}^+ \rightarrow \mathbb{R}$ such that $\frac{uh(x,s)}{\|u\|^2}$ is weakly differentiable in $u$ for every $x$. Then

$$E(\theta', 0', \sigma^2)[\frac{h(X,S)}{\sigma^2}] = E(\theta', 0', \sigma^2)[\frac{U'Uh(X,S)}{\sigma^2\|U\|^2}]$$

(2.81)

**Proof.** Since

$$E(\theta', 0', \sigma^2)[\frac{h(X,S)}{\sigma^2}] = E(\theta', 0', \sigma^2)[\frac{U'Uh(X,S)}{\sigma^2\|U\|^2}]$$

expression (2.81) is equivalent to

$$(2\pi\sigma^2)^{-(p+k)} \int_{\mathbb{R}^p} \frac{u'uh(x,s)}{\sigma^2\|u\|^2} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^p (x_i - \theta_i)^2 + \sum_{i=1}^k u_i^2)} d\mu =$$

(2.82)

$$\int_{\mathbb{R}^p} (2\pi\sigma^2)^{-\frac{p+k}{2}} \int_{\mathbb{R}^k} \frac{u'uh(x,\|u\|)}{\sigma^2\|u\|^2} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^k u_i^2)} e^{-\frac{1}{2\sigma^2}(\sum_{i=1}^p (x_i - \theta_i)^2)} dx,$$

(2.83)

were (2.83) follows from (2.82) by Fubini's theorem. Since $x$ can be viewed as a constant in the inner integral of expression (2.83), let

$$\frac{uh(x,\|u\|)}{\sigma^2\|u\|^2} = \frac{g(x)}{\sigma^2}$$
so that
\[
\int_{\mathbb{R}^k} u \frac{uh(x,||u||^2)}{\sigma^2||u||^2} (2\pi\sigma^2)^{-\frac{k}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^k u_i^2} du = E_{\sigma^2}[U^t g_x^*(U)] \tag{2.84}
\]
where the expectation in (2.84) is taken with respect to a \(N_k(0,\sigma^2 I_k)\) distribution. By assumption \(g_x^*(U)\) is weakly differentiable with respect to \(u\) for every \(x\) so that by Stein's lemma
\[
E_{\sigma^2}[U^t g_x^*(U)] = E_{\sigma^2}[div_u(g_x^*(U))] = E_{\sigma^2}[div_u(Uh(X,S))].
\]
Since
\[
\frac{\partial}{\partial x_i} u_i h(x,||u||^2) = h(x,s) + u_i \left[\frac{||u||^2}{s} \frac{\partial}{\partial s} h(x,s) - h(x,s) \frac{2u_i}{||u||^2} \right],
\]
\[
div_u \left( \frac{uh(x,S)}{S} \right) = (k-2) \frac{h(x,s)}{s} + 2 \frac{\partial}{\partial s} h(x,s). \tag{2.85}
\]
Using expression (2.85) in (2.83) results in
\[
E_{(\theta',0)^\prime,\sigma^2} \left[ \frac{h(X,S)}{\sigma^2} \right] =
E\left[ E[div_u \left( \frac{Uh(X,S)}{\sigma^2 S} \right) | X = x] \right] =
E\left[ E\left[ (k-2) \frac{h(x,s)}{s} + 2 \frac{\partial}{\partial s} h(x,s) | X = x \right] \right] =
E_{(\theta',0)^\prime,\sigma^2} \left[ (k-2) \frac{h(X,S)}{S} + 2 \frac{\partial}{\partial s} h(X,S) \right]
\]
\[
\square
\]
For estimators of the form
\[
\delta_S(X,S) = X + \frac{S}{k+2} g(X) \tag{2.86}
\]
Lemma 2.5 gives conditions on the function \(g\), so that \(\delta_S(X,S)\) will be minimax for the loss given by (2.79).

**Lemma 2.5.** Let \(\begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k} \left( \begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k} \right)\) with \(\dim(X) = \dim(\theta) = p \geq 3\), \(\dim(U) = \dim(0) = k \geq 1\), and \(\sigma^2\) unknown. Let \(S = ||U||^2\), and \(\delta_S(X) = X + \frac{S}{k+2} g(X)\) be an estimator for \(\theta\) such that
i) $g(x)$ is weakly differentiable with $E[\|g(X)\|^2] < \infty$, 

ii) $\|g(X)\|^2 + 2\text{div}_X(g(X)) \leq 0$ for almost every $x$,

then the estimator $\delta_S(X)$ will be minimax for the loss $l(d, \theta) = \frac{\|d-\theta\|^2}{\sigma^2}$.

**Proof.** The risk of the estimator $\delta_S(X)$, $R(\theta, \delta_S(X))$, is expressible as

$$E[\frac{1}{\sigma^2}[\|X + \frac{S}{k+2}g(X) - \theta\|^2]] =$$

$$E[\frac{1}{\sigma^2}(\|X - \theta\|^2) + \frac{S^2}{(k+2)^2}\|g(X)\|^2] + \frac{2S}{(k+2)}(X - \theta)'g(X)].$$

(2.87)

The condition $E[\|g(X)\|^2]$ is sufficient for $R(\delta, \theta_S(X)) < \infty$ as the independence of $X$ and $S$ imply the term

$$E[\frac{S^2\|g(X)\|^2}{\sigma^2(k+2)^2}] = E[\frac{S^2}{\sigma^2(k+2)^2}]E[\|g(X)\|^2] < \infty$$

(2.88)

and the term

$$E[\frac{2S}{\sigma^2(k+2)}(X - \theta)'g(X)] = E[\frac{2S}{\sigma^2(k+2)}]E[(X - \theta)'g(X)] \leq$$

(2.89)

$$E[\frac{2S}{\sigma^2(k+2)}](E[\|X - \theta\|^2])^{\frac{1}{2}}(E_{\theta, \sigma^2}[\|g(X)\|^2])^{\frac{1}{2}} < \infty$$

(2.90)

were (2.90) uses the Cauchy-Schwarz inequality.

To establish minimaxity it is sufficient to show

$$E[\frac{S^2\|g(X)\|^2}{\sigma^2(k+2)^2}] + \frac{2S(X - \theta)'g(X)}{\sigma^2(k+2)} \leq 0$$

(2.91)

as the estimator $\delta_0(X) = X$ is a minimax estimator of $\theta$. From the independence of $X$ and $S$, the term

$$E[\frac{2S}{\sigma^2(k+2)}(X - \theta)'g(X)] = E[\frac{2S}{(k+2)}]E[\frac{1}{\sigma^2}(X - \theta)'g(X)] =$$


\[ E\left[ \frac{2S}{(k+2)} \right] E[\text{div}_x(g(X))] = E\left[ \frac{2S}{(k+2)} \text{div}_x(g(X)) \right]. \] (2.92)

were Stein’s Lemma is used in (2.92). Letting \( h(X, S) = \frac{S^2}{(k+2)^2} \|g(X)\|^2 \) the other term in (2.91),

\[
E\left[ \frac{S^2}{\sigma^2(k+2)^2} \|g(X)\|^2 \right] = E\left[ \frac{h(X, S)}{\sigma^2} \right] = \\
E\left[ (2 \frac{\partial}{\partial S} h(X, S) + (k - 2) \frac{h(X, S)}{S} \right] = \\
E\left[ (2 \frac{2S}{(k+2)^2} \|g(X)\|^2 + (k - 2) \frac{S^2}{(k+2)^2} \|g(X)\|^2 \right] = \\
E[\|g(X)\|^2 \left( \frac{(k + 2)S}{(k+2)^2} \right)] \quad (2.93)
\]

were (2.93) follows from Lemma 2.4. Therefore

\[ R(\theta, \delta_S(X)) = p + E\left[ \frac{\sigma^2 S}{(k+2)} \|g(X)\|^2 + 2\text{div}_x(g(X)) \right] \leq p \]

by the assumption \( \|g(X)\|^2 + 2\text{div}_x(g(X)) \leq 0 \) (a.e.) so that \( \frac{\sigma^2 S}{(k+2)} \|g(X)\|^2 + 2\text{div}_x(g(X)) \leq 0 \) (a.e.) since \( \frac{\sigma^2 S}{(k+2)} \geq 0 \).

A straightforward consequence of Lemma 2.5 is Theorem 2.5 which provides the extension to the results of Section 2.1 from the Gaussian case to the Gaussian case with unknown scale and residual vector for estimators given by (2.79).

**Theorem 2.5.** Let \( \begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k} \left( \begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k} \right) \) where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = \text{dim}(0) = k \geq 1 \). Let \( S = \|U\|^2 \), the estimators

\[
\delta_{1,S}(X)^+ = \begin{pmatrix}
(1 - \frac{S}{(k+2)} \frac{4j(D)h'(X_1^2)}{j(D)}) + X_1 \\
(1 - \frac{S}{(k+2)} \frac{4j(D)h'(X_2^2)}{j(D)}) + X_2 \\
\vdots \\
(1 - \frac{S}{(k+2)} \frac{4j(D)h'(X_p^2)}{j(D)}) + X_p
\end{pmatrix} \quad (2.95)
\]
where \( D = \sum_{i=1}^{p} h(X_i^2) \), and \( h \) and \( j \) satisfy conditions i)-vii) of Theorem 2.2 with \( \frac{pABC}{2} \geq 1 \), and

\[
\delta_{2,S}(X)^+ = \begin{pmatrix}
(1 - \frac{s}{(k+2)} cv(D)w(X_1^2)) + X_1 \\
(1 - \frac{s}{(k+2)} cv(D)w(X_2^2)) + X_2 \\
\vdots \\
(1 - \frac{s}{(k+2)} cv(D)w(X_p^2)) + X_p
\end{pmatrix}
\]  

(2.96)

where \( D = \sum_{i=1}^{p} h(X_i^2) \), and \( v, w, \) and \( h \) satisfy conditions i)-xi) of Theorem 2.3 with \( cE + FB \leq 4p(A + \frac{1}{2}) \) are minimax estimators of \( \theta \) for loss

\[
l(d, \theta) = \frac{\|d - \theta\|^2}{\sigma^2}
\]

(2.97)

**Proof.** For \( p \geq 3 \), it follows from the proofs of Theorem 2.2 and Theorem 2.3 that when \( X \sim N_p(\theta, I_p) \) the estimators \( \delta_{1,1}(X) \) and \( \delta_{2,1} \) satisfy

\[
\|g_i(x)\|^2 + 2\text{div}_x(g_i(x)) \leq 0
\]

(2.98)

for \( i = 1, 2 \). Therefore Lemma 2.5 implies \( \delta_{1,\frac{s}{(k+2)}} \) and \( \delta_{2,\frac{s}{(k+2)}} \) are minimax estimators. Furthermore since the distribution of \( X \) and the estimators satisfy the conditions of Theorem 2.1, the positive-part versions of the estimators dominate the original estimators.

### 2.5 Spherically Symmetric Distributions with Residual

This Section extends the results of Section 2.1 to the case of a general spherically symmetric distribution with a residual vector. In particular

\[
\begin{pmatrix}
X \\
U
\end{pmatrix} \sim SS_{p+k}(\begin{pmatrix}
\theta \\
0
\end{pmatrix})
\]

(2.99)

with \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = \text{dim}(0) = l \geq 1 \), and loss

\[
l(d, \theta) = \frac{\|d - \theta\|^2}{E[\|U\|^2]}
\]

(2.100)
The class of estimators considered are of the form

\[ \delta_S(X, S) = X + \frac{S}{k+2}g(X) \]  \hspace{1cm} (2.101)

where \( S = \|U\|_2^2 \), and \( g \) is of the form (2.66) or (2.67). The method of extension will make use of the results in Section 4 to establish a Stein-like equality that will be used to bound the difference in risk between the minimax estimator, \( \delta_0(X) = X \), and the proposed estimator.

The following Lemma is found in Fourdriner, Strawderman, and Wells [10]. It’s proof is presented for completeness.

**Lemma 2.6.** (Fourdrinier, Strawderman, and Wells [10]) Suppose \( \begin{pmatrix} X \\ U \end{pmatrix} \sim S\mathbb{S}_{p+k} \begin{pmatrix} \theta \\ 0 \end{pmatrix} \) with \( \dim(X) = \dim(\theta) = p \) and \( \dim(U) = \dim(0) = k \geq 1 \). Assume \( g(X) \) is a weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \) such that

\[ E[(X - \theta)'g(X)] = \sigma^2 E[\text{div}_x(g(X))] \]

for all \( \sigma^2 \), where the expectations exist for the special case where \( \begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k} \begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k} \).

Then

\[ E[\|U\|^2(X - \theta)'g(X)] = \frac{1}{(k+2)} E[\|U\|^4 \text{div}_x(g(X))]. \]

**Proof.** Let \( \begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k} \begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k} \) with \( \sigma^2 \) unknown. Then \( X \sim N_p(\theta, \sigma^2 I_p) \) is independent of \( \|U\|^2 \sim \sigma^2 \chi^2_k \) so that

\[ E[\|U\|^2(X - \theta)'g(X)] = E[\|U\|^2] E[(X - \theta)'g(X)] = k\sigma^2 E[\sigma^2 \text{div}_x(g(X))] = E[\sigma^4 \text{div}_x(g(X))] \]

\[ = E[\|U\|^4 \text{div}_x(g(X))] \]

as \( E[\|U\|^2] = k\sigma^2 \), and \( E[(\|U\|^2)^2] = k(k+2)\sigma^4 \).

For each fixed value of \( \theta \), the statistic \( \|X - \theta\|^2 + \|U\|^2 \) is a complete and sufficient statistic for \( \sigma^2 \), so that by completeness...
\[ E[E[||U||^2(X - \theta)'g(X)||X - \theta||^2 + ||U||^2 = R^2]] = \]
\[ E_{(\theta', \sigma^2)}[||U||^2(X - \theta)'g(X)] = \]
\[ \frac{1}{(k + 2)} E[||U||^{4} \text{div}_x(g(X))] = \]
\[ \frac{1}{(k + 2)} E[\mathbb{E}[\|U\|^4 \text{div}_x(g(X))]|X - \theta||^2 + ||U||^2 = R^2]] \]
so that
\[ E[||U||^2(X - \theta)'g(X)||X - \theta||^2 + ||U||^2 = R^2] = \]
\[ \frac{1}{(k + 2)} E[||U||^{4} \text{div}_x(g(X))]|X - \theta||^2 + ||U||^2 = R^2] \] (2.102)
almost everywhere.

Since the conditional distribution of \( \begin{pmatrix} X \\ U \end{pmatrix} \) given \( ||X - \theta||^2 + ||U||^2 = R^2 \) is the uniform distribution on a sphere centered at \( \begin{pmatrix} \theta \\ 0 \end{pmatrix} \) of radius \( R \), expression (2.102) gives the equality of expectations for such uniform distributions.

In the general case where \( \begin{pmatrix} X \\ U \end{pmatrix} \sim \mathcal{S}_{p+k}(\theta, 0) \) the conditional distribution of \( \begin{pmatrix} X \\ U \end{pmatrix} \) given \( ||X - \theta||^2 + ||U||^2 = R^2 \) is also a uniform distribution on the sphere of radius \( R \) centered at \( \begin{pmatrix} \theta \\ 0 \end{pmatrix} \) and so
\[ E[||U||^2(X - \theta)'g(X)||X - \theta||^2 + ||U||^2 = R^2] = \]
\[ \frac{1}{(k + 2)} E[||U||^{4} \text{div}_x(g(X))]|X - \theta||^2 + ||U||^2 = R^2] \]
as well. Taking the expectation with respect to the radial distribution squared of the above expectations establishes the result.
For estimators of the form (2.101) and loss given by 2.100, Lemma 2.7 gives conditions on \( g \) for minimaxity.

**Lemma 2.7.** Let \( \begin{pmatrix} X \\ U \end{pmatrix} \sim SS_{p+k}(\theta, 0) \) where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = \text{dim}(0) = k \geq 1 \). Suppose all second moments exist. Let \( \delta \|U\|^2 \) be an estimator of \( \theta \). Assume that \( g(X) \) is such that

i) \( g(X) \) is a weakly differentiable function with \( E[\|g(X)\|^2] < \infty \),

ii) \( \|g(X)\|^2 + 2 \text{div}_x(g(X)) \leq 0 \) a.e.,

then the estimator \( \delta \|U\|^2 \) is minimax for the loss

\[
 l(d, \theta) = \frac{\|d - \theta\|^2}{E[\|U\|^2]}.
\]

(2.103)

**Proof.** The risk of the estimator \( \delta \|U\|^2 \):

\[
 R(\theta, \delta \|U\|^2) = E\left[\frac{1}{E[\|U\|^2]}\|X + \frac{\|U\|^2}{(k+2)^2}g(X) - \theta\|^2\right] =
\]

\[
 \frac{1}{E[\|U\|^2]} E[\|X - \theta\|^2] + \frac{1}{E[\|U\|^2]} E\left[\frac{\|U\|^4}{(k+2)^2}\|g(X)\|^2 + 2\frac{\|U\|^2}{(k+2)}(X - \theta)'g(X)\right].
\]

Since the estimator \( X \) is a minimax for \( \theta \), the minimaxity of \( \delta \|U\|^2 \) follows once

\[
 E\left[\frac{\|U\|^4}{(k+2)^2}\|g(X)\|^2 + 2\frac{\|U\|^2}{(k+2)}(X - \theta)'g(X)\right] \leq 0.
\]

(2.104)

By Lemma 2.6, the term

\[
 E[2\frac{\|U\|^2}{(k+2)}(X - \theta)'g(X)] = E\left[\frac{\|U\|^4}{(k+2)^2}2\text{div}_x(g(X))\right]
\]

so that the expectation in (2.104) becomes

\[
 E\left[\frac{\|U\|^4}{(k+2)^2}\|g(X)\|^2 + \frac{\|U\|^4}{(k+2)^2}2\text{div}_x(g(X))\right] =
\]

\[
 E\left[\frac{\|U\|^4}{(k+2)^2}(\|g(X)\|^2 - 2\text{div}_x(g(X)))\right] \leq 0
\]
as $\|U\|^4 / (k+2)^2 \geq 0$ and $\|g(X)\|^2 + 2 \text{div}_x(g(X)) \leq 0$ a.e. by assumption establishing the result.

\[\blacksquare\]

Theorem 2.6 is a straightforward consequence of Lemma 2.7 and extends the class of estimators given in Section 2.1 to the the class of generally spherically symmetric distributions with residual vectors. Since the proof is straightforward and similar to Theorem 2.5 its proof is omitted.

**Theorem 2.6.** Let \( \begin{pmatrix} X \\ U \end{pmatrix} \sim SS_p+k(\theta,0) \) where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = \text{dim}(0) = k \geq 1 \) such that all second moments exist. The estimators

\[
\delta_1,\|U\|^2 / (k+2) (X) = \begin{pmatrix} (1 - \|U\|^2 / (k+2) j(D) h'(X^2_j) X_1 \\ (1 - \|U\|^2 / (k+2) j(D) h'(X^2_j) X_2 \\ \vdots \\ (1 - \|U\|^2 / (k+2) j(D) h'(X^2_j) X_p \end{pmatrix}
\]

(2.105)

where \( D = \sum_{i=1}^p h(X_i^2) \), and \( h \) and \( j \) satisfying conditions i)-vii) of Theorem 2.2 with \( \sum_{i=1}^p h(X_i^2) / (k+2)^2 \geq 1 \), and

\[
\delta_2,\|U\|^2 / (k+2) (X) = \begin{pmatrix} (1 - \|U\|^2 / (k+2) c(D) w(X_1^2) X_1 \\ (1 - \|U\|^2 / (k+2) c(D) w(X_2^2) X_2 \\ \vdots \\ (1 - \|U\|^2 / (k+2) c(D) w(X_p^2) X_p \end{pmatrix}
\]

(2.106)

where \( D = \sum_{i=1}^p h(X_i^2) \), and \( v, w, \) and \( h \) satisfy conditions i)-xi) of Theorem 2.3 with \( cE + FB \leq 4p(A + \frac{1}{2}) \) are minimax estimators of \( \theta \) under loss

\[
l(d,\theta) = \frac{\|d - \theta\|^2}{\sigma^2}.
\]

(2.107)

If in addition \( \begin{pmatrix} X \\ U \end{pmatrix} \) has density \( f \), that is unimodal and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, the positive-part versions of the original estimators dominate the original estimators.
2.6 Summary and Conclusion

A class of Psuedo-Bayes estimators, and a general class of estimators were developed that were minimax under quadratic loss, $\|d - \theta\|^2$, and whose positive-part versions allowed adaptive selection of subsets of the coordinates to be selected as 0 when $X \sim N_p(\theta, \sigma^2 I_p)$ and $\sigma^2$ was known. Extensions of these results were given for certain classes of spherically symmetric distributions (which included scale mixtures of normal distributions), and the general class of spherically symmetric distributions with residual vector and unknown scale. The class of estimators considered gave slight extensions to the result of Zhou and Hwang [31] and Maruyama [22]. While analytic comparison of the risks of the usual positive-part James-Stein estimator and the estimator developed in Section 2 seem difficult, numerical studies in Section 2.2 indicate that classes of these estimators seem to dominate the James-Stein positive-part estimator for sparse models. It would also be interesting to develop analogous results for the case $X \sim N(\theta, \Sigma)$ with $\Sigma$ known.
Chapter 3

Minimaxity of Shrinkage Estimators Under Concave Loss

3.1 Introduction

In this Chapter we study minimax shrinkage estimators of a location vector for certain classes of spherically symmetric distributions under concave loss. In the most basic case we observe

\[ X \sim \mathcal{N}_p(\theta, \sigma^2 I_p) \quad (3.1) \]

with \( \sigma^2 \) known and consider estimation of the location vector \( \theta \in \mathbb{R}^p \) under a loss function of the form

\[ l(d, \theta) = a \|d - \theta\|^2 + \int_0^\infty (1 - e^{-b \|d - \theta\|^2}) dh(b) \quad (3.2) \]

where \( a \geq 0 \) and \( h \) is any non-negative monotonic increasing right continuous function with \( \int_0^\infty dh(b) < \infty \). The class of shrinkage estimators considered is the class of Baranchik-type estimators

\[ \delta_{a,r}(X) = (1 - \frac{a \sigma^2 r(\|X\|^2)}{\|X\|^2}) X. \quad (3.3) \]

The basic result of Section 2 shows that estimator given by (3.3) is minimax for \( p \geq 3 \),
provided

i) \( r'(u) \geq 0, \)

ii) \( 0 \leq r(u) \leq 1 \) and,

iii) \( 0 \leq a \leq 2(p - 2) \)

with strict domination of the usual minimax estimator \( \delta_0(X) = X \) if either iii) is strict or ii) is strict on a set of positive measure. Interestingly, this is the same set of condition for minimaxity (domination) under quadratic loss

\[
l(d, \theta) = \|d - \theta\|^2. \tag{3.4}\]

In almost all results in the literature, (e.g. Brandwein and Strawderman [8], [9] and Kubokawa, Marchand, and Strawderman [18]), the maximum value of the shrinkage constant, \( a \), for the minimaxity of (3.3) is strictly smaller for classes of concave losses which are functions of quadratic loss. An exception is the work of Ghosh, Mergel, and Datta [14] which established for the loss function,

\[
l(d, \theta) = 1 - e^{-\frac{b}{2\sigma^2}\|d - \theta\|^2} \tag{3.5}\]

where \( b > 0 \), the maximum value for the shrinkage constant, \( a \), in (3.3) for minimaxity is also \( 2(p-2) \).

Extensions to the unknown scale case will be given. In particular

\[
\begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \tag{3.6}
\]

where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \) and \( \text{dim}(U) = \text{dim}(0) = k \geq 1 \). Here \( \|U\|^2 = S \) and \( X \) are complete sufficient statistics for \( \theta \) and \( \sigma^2 S \sim \sigma^2 \chi_k^2 \). The loss studied is (3.2), and the class of estimators studied are

\[
\delta(X, U) = (1 - \frac{\|U\|^2 ar(\frac{\|X\|^2}{\|U\|^2})}{(m + 2)\|X\|^2})X \tag{3.7}
\]
Section 3 establishes that if

\[ i) \ 0 \leq a \leq 2(p-2), \]
\[ ii) \ r'(u) \geq 0, \] and
\[ iii) \ 0 \leq r(u) \leq 1 \]

in (3.7), estimator (3.7) will dominate the usual estimator \( \delta_0(X) = X \) with strict domination if either i) or ii) is strict. Replacing \( \|d-\theta\|^2 \) with \( \frac{\|d-\theta\|^2}{\sigma^2} \) in (3.2) will result in the estimator \( X \) being minimax and estimator (3.7) dominating the estimator \( X \) once \( 0 \leq a \leq 2(p-2) \), \( r'(u) \geq 0 \), and \( 0 \leq r(u) \leq 1 \).

Section 3 also studies the general class of scale mixtures of normal distributions when a residual vector is present. In particular

\[
\begin{pmatrix} X \\ U \end{pmatrix} | \sigma^2 \sim N_{p+k}( \begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \] 
\tag{3.8}
\]

with \( \sigma^2 \sim F(\sigma^2) \) where \( F(\cdot) \) is a distribution function on \( (0, \infty) \) and \( E[\sigma^2] \) and \( E[\frac{1}{\sigma^2}] \) are both finite. We show, as in the normal case, that the estimator (3.7) improves upon the usual estimator \( X \) under loss (3.2), once \( r'(u) \geq 0, \ 0 \leq r(u) \leq 1, \) and \( 0 \leq a \leq 2(p-2) \). This result also gives improvement over \( X \) under the same set of condition as in the case of quadratic loss \( \|d-\theta\|^2 \). In addition a general result for the general class of spherically symmetric distributions with residuals and densities of the form

\[
\frac{1}{\sigma^p} f\left( \frac{1}{\sigma^2}(\|x-\theta\|^2 + \|u\|^2) \right), \] 
\tag{3.9}
\]

for estimators of location the form

\[
\delta(X, U) = X + \frac{\|U\|^2}{k+2} g(X) \] 
\tag{3.10}
\]

will be given for any non-negative concave differentiable loss function \( l(t) \), where \( t = \|d-\theta\|^2 + \|U\|^2 \). The conditions for minimaxity being
i) \( g(X) \) is a weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \)

ii) \( E_\theta[\|g(X)\|^2] < \infty \)

iii) \( \|g(X)\|^2 + 2\text{div}_x(g(X)) < 0 \) a.e.

are equivalent to those under quadratic loss \( \|d - \theta\|^2 \).

A natural question to ask is if the results of Section 2 (the known scale case) extend to the general class of scale mixtures of normal distributions. In Section 4 we present some numerical simulations indicating this is not the case.

Section 5 has some concluding remarks.

### 3.2 Normal Theory with Known Scale

Section 2 establishes the minimaxity of Baranchik-type estimators of the form

\[
\delta_{a,r}(X) = (1 - a\sigma^2 r(\frac{\|X\|}{\sigma^2})) X
\]  

(3.11)

for classes of losses

\[
l(d, \theta) = a\|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|d-\theta\|^2}) dh(b)
\]

(3.12)

where \( a \geq 0 \) and \( h \) any non-negative monotonically increasing right continuous continuous function satisfying \( \int_0^\infty dh(b) < \infty \) when \( X \sim N_p(\theta, \sigma^2 I_p) \). Estimator (3.11) will be minimax under loss (3.12) once

i) \( 0 \leq a \leq 2(p-2) \),

ii) \( 0 \leq r \leq 1 \),

iii) \( r'(u) \geq 0 \).

This result extends the result of Ghosh, Mergel, and Datta [14] from the class of losses of the form

\[
1 - e^{-\frac{b\|d-\theta\|^2}{\sigma^2}}
\]

(3.13)
Previous results in the literature have studied minimaxity for estimators of location vectors for spherically symmetric distributions using more general classes of concave loss. In particular Kubokawa, Marchand, and Strawderman [18] have studied conditions for domination of the “usual” estimator \( \delta_0(X) = X \) for estimators of the form

\[
\delta(X) = X + g(X)
\]

(3.14)

for the class of \( p \)-dimension spherically symmetric distributions with densities

\[
f(\|X - \theta\|^2).
\]

(3.15)

There results imply for \( X \sim N_p(\theta, \sigma^2 I_p) \) estimators given by (3.11) will be minimax for a maximum value of the constant, \( a \), less than the maximum value of the constant, \( a \), of 2(p-2) under quadratic loss. The following Theorem found in Kubokawa, Marchand, and Strawderman [18] gives the relevant condition for domination of the estimator \( \delta_0(X) = X \) for estimators given by (3.14) when \( X \) has density given by (3.15). Its proof is presented for completeness and an immediate Corollary is given so that the result can be directly applied to estimators given by (3.11) when \( X \sim N_p(\theta, \sigma^2 I_p) \) and loss given by

\[
l(d, \theta) = 1 - e^{-b\|d - \theta\|^2},
\]

(3.16)

a subset in the class of losses consider in Theorem 3.1 in Example 3.1. Example 3.1 will establish how much the shrinkage value of \( a \) must be modified to ensure minimaxity under Theorem 3.1.

**Theorem 3.1.** *(Kubokawa, Marchand, Strawderman [18])* Suppose that \( X \) is a \( p \)-dimensional spherically symmetric distribution with density

\[
f(\|x - \theta\|^2).
\]

Let \( l(t) \) be a concave, non-negative, non-decreasing, and differentiable loss function, and let

\[
\delta(X) = X + g(X)
\]
Then $\delta$ will dominate $X$ under loss $l(t)$, if $\delta$ dominates $X$ under square error loss for a location family with density $f^*(\|x - \theta\|^2)$ which is proportional to $f(\|x - \theta\|^2)l'(\|x - \theta\|^2)$.

Proof. By the concavity of $l(t)$

$$
\frac{l(t+h) - l(t)}{h} \leq \lim_{h \to 0} \frac{l(t+h) - l(t)}{h} = l'(t)
$$

(3.17)

so that

$$
l(t+h) \leq l(t) + hl'(t).
$$

Setting $t = \|X - \theta\|^2$ and $h = \|g(X)\|^2 + 2(X - \theta)'g(X)$ in (3.2),

$$
l(\|X - \theta\|^2 + \|g(X)\|^2 + 2(X - \theta)'g(X)) = l(\|\delta - \theta\|^2) \leq
$$

$$
l(\|X - \theta\|^2) + l'(\|X - \theta\|^2)\|g(X)\|^2 + 2(X - \theta)'g(X)]
$$

hence,

$$
R(\delta, \theta) \leq R(X, \theta) + E_{\theta}[l'(\|X - \theta\|)[\|g(X)\|^2 + 2(X - \theta)'g(X)]].
$$

(3.18)

If $R(\delta, \theta) \leq R(X, \theta)$ under square error loss assuming the underlying density is $f^*$, then

$$
E_{\theta}[\|g(X)\|^2 + 2(X - \theta)'g(X)] \leq 0
$$

for all $\theta$ so that

$$
E_{\theta}[l'(\|X - \theta\|^2)[\|g(X)\|^2 + 2(X - \theta)'g(X)]] \leq 0
$$

(3.19)

in (3.18) where the expectation is taken with respect to the density $f(\|x - \theta\|^2)$. Hence $R(\delta, \theta) \leq R(X, \theta)$. ■

The following Corollary is a direct consequence of Theorem 3.1. It uses the techniques found in Fourdiner and Strawderman [11] to provided sufficient conditions for estimators of the form

$$
\delta(X) = X + g(X)
$$

to be minimax for the classes of concave loss function considered by Theorem 3.1. It will be used to analyze sufficient conditions for minimaxity for estimators given by (3.11) when the loss is given by (3.16).
Corollary 3.1. Let $X$ have a spherically symmetric distribution with density $f(\|X - \theta\|^2)$. Assume that $l(t)$ is a concave, non-decreasing, non-negative differentiable loss function and let:

$$f^*(t) = f(t)l'(t), \text{ and}$$
$$F^*(t) = \frac{1}{2} \int_t^\infty f(u)l'(u)du$$

Assume that:

$$\frac{F^*(t)}{F^*(t)} \geq b > 0 \text{ for all } t.$$

If $\delta(X) = X + g(X)$ is such that in the case when $X \sim N_p(\theta, I_p)$ the estimator $\delta$ of $\theta$ is such that:

$$\|g(X)\|^2 + 2\text{div}_X(g(X)) \leq 0,$$

then estimator $\delta_a(X) = X + ag(X)$ for $0 < a < b$ will dominate $\delta_0(X) = X$ under the concave loss $l(t)$.

Proof. Setting $f^{**}(\|X - \theta\|^2) = l'(\|x - \theta\|^2)f(\|x - \theta\|^2)$ and $f^*(\|x - \theta\|^2) = kl'(\|x - \theta\|^2)f(\|X - \theta\|^2)$, where $k$ is chosen to be the constant of proportionality

$$k = \frac{1}{\int_{\mathbb{R}^p} l'(\|x - \theta\|^2)f(\|x - \theta\|^2)dx}.$$

By the spherically symmetry of the density $f^*(\|x - \theta\|^2)$

$$E_\theta[l'(\|X - \theta\|^2)(X - \theta)'g(X)] =$$

$$\int l'(\|X - \theta\|^2)f(\|x - \theta\|^2)(x - \theta)'g(x)dx =$$

$$\frac{1}{k}E_\theta[(X - \theta)'g(X)] = \frac{1}{k}E_\theta[F^*(\|x - \theta\|^2)\text{div}_X(g(X))] =$$

$$E[f^{**}(\|x - \theta\|^2)\text{div}_X(g(X))],$$

where the expectation in (3.20) is taken with respect to the density $f(\|x - \theta\|^2)$ and the expectation in (3.21) is taken with respect to the density $f^*(\|x - \theta\|^2)$. By inequality (3.18), the risk of the estimator $\delta_b(X)$ under the concave loss function $l(t)$ as follows:
\[ R(\delta_b, \theta) \leq R(X, \theta) + E_\theta[l'(\|X - \theta\|^2)[b^2\|g(X)\|^2] + 2b(X - \theta)'g(X)] \leq \\
R(X, \theta) + EF^{**}[b^2\|g(X)\|^2 + 2b\frac{F^*(\|x-\theta\|^2)}{\|X-\theta\|^2}div_X(g(X))]] \leq \\
R(X, \theta) + b^2EF^{**}[\|g(X)\|^2 + 2div_X(g(X))]. \tag{3.22} \]

as \( \frac{F^*(t)}{f^*(t)} \geq b > 0 \) and \( div_X(g(X)) \leq 0 \). By the assumption \( \|g(X)\|^2 + 2div_X(g(X)) \leq 0 \), \( EF^{**}[\|g(X)\|^2 + 2div_X(g(X))] \leq 0 \). Hence \( R(\delta_b, \theta) \leq R(X, \theta) \). \( \blacksquare \)

**Example 3.1.** Let \( X \sim N_p(\theta, \sigma^2 I_p) \) with \( p \geq 3 \) and \( \sigma^2 \) known. Let \( \delta(X) \) be a James-Stein Estimator of \( \theta \) of the form

\[ \delta(x) = (1 - \frac{a}{\|X\|^2})X = X + g(X). \tag{3.23} \]

Under square error loss James and Stein showed that \( \delta(X) \) is a minimax estimator of \( \theta \) as long as

\[ 0 \leq a \leq 2(p - 2)\sigma^2. \tag{3.24} \]

Under the concave loss

\[ l(t) = 1 - e^{-ct} \]

an application of Corollary 3.1 results in the minimaxity of estimator (3.23) once the constant, \( a \), satisfies

\[ 0 \leq a \leq 2(p - 2)\sigma^2\left(\frac{1}{1 + 2\sigma^2c}\right). \tag{3.25} \]

This follows from Corollary 3.1 as

\[ \frac{F^*(t)}{f^*(t)} \leq \frac{F^*(0)}{f^*(0)} = \int_0^{\infty} e^{-\frac{(1+2\sigma^2c)}{2}\sigma^2} du = \sigma^2\left(\frac{1}{1 + 2\sigma^2c}\right) = b \tag{3.26} \]

where

\[ f^*(t) = f(t)\frac{d}{dt}l(t) = c\sigma^{-p}(2\pi)^{-\frac{p}{2}}e^{-\frac{(1+2\sigma^2c)}{2\sigma^2}t} \tag{3.27} \]
and

\[ F^*(t) = \frac{1}{2} \int_t^\infty c^\sigma (2\pi)^{-\frac{p}{2}} e^{-(\frac{1+2\sigma^2 c^2}{2\sigma^2})u} du \]  

(3.28)

with \( t = \|x - \theta\|^2 \) in (3.26)-(3.28), and in the case when \( X \sim N_p(\theta, I_p) \) estimator (3.23) will satisfy

\[ \|g(x)\|^2 + 2 \text{div}_x (g(x)) \leq 0 \]  

(3.29)

once \( 0 \leq a \leq 2(p - 2) \) so that the estimator

\[ \delta(X) = (1 - \frac{a}{\|X\|^2}) X \]

for

\[ 0 \leq a \leq 2(p - 2)b = 2(p - 2)\sigma^2(\frac{1}{1+2\sigma^2 c^2}) \]

is minimax for loss \( l(t) = 1 - e^{-ct} \) and \( X \sim N_p(\theta, \sigma^2 I_p) \).

### 3.2.1 Minimaxity of the estimator X

In this subsection we establish the minimaxity of the “usual” estimator \( \delta_0(X) = X \) for losses of the form

\[ l(d, \theta) = a\|d - \theta\|^2 + \int_0^\infty dg(b) - \int_0^\infty e^{-b\|d - \theta\|^2} dh(b). \]  

(3.30)

The minimaxity of X will be established, as in Ghosh, Mergel, and Datta [14], by developing the Bayes estimators for a least favorable sequence of priors. Once the minimaxity of X is established, conditions under which the estimator (3.11) dominates the estimator X for the loss (3.30) will be given in the subsequent subsection. The following Lemma found in Ghosh, Mergel, and Datta [14] is used to determine the Bayes estimator for normal priors. Its proof is presented for completeness.

**Lemma 3.1.** (Ghosh, Mergel, and Datta [14]) Let \( X_1 \sim N_p(\mu_1, \Sigma_1) \) with density \( f_1(X) \) and let \( X_2 \sim N_p(\mu_2, \Sigma_2) \) with density \( f_2(X) \) with \( \Sigma_1 \) and \( \Sigma_2 \) positive definite. Then for \( a_1 > 0 \) and \( a_2 > 0 \)
\[ \int_{\mathbb{R}^p} f_1(x)^{a_1} f_2(x)^{a_2} \, dx = \]
\[ (2\pi)^{\frac{p(1-a_1-a_2)}{2}} |\Sigma_1|^{-\frac{1-a_1}{2}} |\Sigma_2|^{-\frac{1-a_2}{2}} \]
\[ |a_1 \Sigma_2 + a_2 \Sigma_1|^{-\frac{1}{2}} e^{-\frac{a_1 a_2}{2} (\mu_1 - \mu_2)' (a_1 \Sigma_2 + a_2 \Sigma_1)^{-1} (\mu_1 - \mu_2)} . \]

**Proof.**

\[ \int_{\mathbb{R}^p} f_1(x)^{a_1} f_2(x)^{a_2} \, dx = \]
\[ \int_{\mathbb{R}^p} (2\pi)^{-\frac{p(1-a_1-a_2)}{2}} |\Sigma_1|^{-\frac{a_1}{2}} |\Sigma_2|^{-\frac{a_2}{2}} e^{\frac{1}{2} [(x - \mu_1)' a_1 (\Sigma_1)^{-1} (x - \mu_1) + (x - \mu_2)' a_2 (\Sigma_2)^{-1} (x - \mu_2)]} \, dx. \quad (3.31) \]

Setting \( A = a_1 \Sigma_1^{-1} \) and \( B = a_2 \Sigma_2^{-1} \) the quadratic form in expression (3.31) is expressible as

\[ [(x - \mu_1)' a_1 (\Sigma_1)^{-1} (x - \mu_1) + (x - \mu_2)' a_2 (\Sigma_2)^{-1} (x - \mu_2)] = \]
\[ x'(A + B)x - 2x' A \mu_1 - 2x' B \mu_2 + \mu_1' A \mu_1 + \mu_2' B \mu_2 = \]
\[ x'(A + B)x - 2x'(A + B)(A + B)^{-1}(A \mu_1 + B \mu_2) + \mu_1' A \mu_1 + \mu_2' B \mu_2. \quad (3.32) \]

Setting \( w = (A + B)^{-1}(A \mu_1 + B \mu_2) \) and completing the square in (3.32) results in

\[ x'(A + B)x - 2x'(A + B)w + w'(A + B)w - [w'(A + B)w - \mu_1' A \mu_1 - \mu_2' B \mu_2] = \]
\[ (x - w)'(A + B)(x - w) - [w'(A + B)w - \mu_1' A \mu_1 - \mu_2' B \mu_2]. \quad (3.33) \]

Since

\[ w'(A + B)w - \mu_1' A \mu_1 - \mu_2' B \mu_2 = \]
\[ (A \mu_1 + B \mu_2)'(A + B)^{-1}(A \mu_1 + B \mu_2) - \mu_1' A \mu_1 - \mu_2' B \mu_2 = \]
\[ ((A + B) \mu_1 + B(\mu_2 - \mu_1))'(A + B)^{-1}((A + B) \mu_1 + B(\mu_2 - \mu_1)) - \mu_1' A \mu_1 - \mu_2' B \mu_2 = \]
\[ \mu_1'(A + B) \mu_1 + (\mu_2 - \mu_1)' B'(A + B)^{-1} B(\mu_2 - \mu_1) + 2 \mu_1' B(\mu_2 - \mu_1) - \mu_1' A \mu_1 - \mu_2' B \mu_2 = \]
\( \mu'_1 A \mu_1 + \mu'_1 B \mu_1 + (\mu_2 - \mu_1)' B (A + B)^{-1} B (\mu_2 - \mu_1) + 2 \mu'_1 B \mu_2 - 2 \mu'_1 B \mu_1 - \mu'_1 A \mu_1 - \mu'_2 B \mu_2 = \\
(\mu_2 - \mu_1)' B (A + B)^{-1} B (\mu_2 - \mu_1) - \mu_2' B \mu_1 - \mu'_1 B \mu_1 + 2 \mu'_1 B \mu_2 = \\
(\mu_2 - \mu_1)' B (A + B)^{-1} B (\mu_2 - \mu_1) - (\mu_2 - \mu_1)' B (\mu_2 - \mu_1) = \\
- (\mu_2 - \mu_1)' [B - B (A + B)^{-1} B] (\mu_2 - \mu_1) \\
(3.34) \\

Substitution of (3.34) into (3.33), result in expression (3.31) being equivalent to 

\[
(2\pi)^{-\rho(a_1 + a_2) - \frac{\rho}{2}} |\Sigma_1|^{-\frac{\rho}{2}} |\Sigma_2|^{-\frac{\rho}{2}} e^{-\frac{\rho}{2} (\mu_1 - \mu_2)' [B - B (A + B)^{-1} B] (\mu_1 - \mu_2)} * \\
\int_{\mathbb{R}^p} e^{-\frac{1}{2} |(x - w)' (A + B) (x - w)|} dx. \tag{3.35}
\]

Evaluating expression (3.35)

\[
\int_{\mathbb{R}^p} e^{-\frac{1}{2} |(x - w)' (A + B) (x - w)|} dx = \\
(2\pi)^{\frac{\rho}{2} (|A + B|^{-\frac{1}{2}})} = \\
(2\pi)^{\frac{\rho}{2}} (|\Sigma_1|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}} |(a_1 \Sigma_2 + a_2 \Sigma_1)|^{-\frac{1}{2}}) \tag{3.36}
\]

since

\[
\Sigma_1 (A + B) \Sigma_2 = a_1 \Sigma_2 + a_2 \Sigma_1, \tag{3.37}
\]

implies

\[
|A + B|^{-\frac{1}{2}} = |\Sigma_1|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}} |(a_1 \Sigma_2 + a_2 \Sigma_1)|^{-\frac{1}{2}}. \tag{3.38}
\]

Furthermore the term \( e^{-\frac{\rho}{2} (\mu_1 - \mu_2)' [B - B (A + B)^{-1} B] (\mu_1 - \mu_2)} \) in (3.35) is expressible as

\[
e^{-\frac{a_1 a_2}{2} (\mu_1 - \mu_2)' (a_1 \Sigma_2 + a_2 \Sigma_1)^{-1} (\mu_1 - \mu_2)} \tag{3.39}
\]

from
\[ [B - B(A + B)^{-1}B][B^{-1}(A + B)A^{-1}] = I_p \]

implying
\[ e^{-\frac{1}{2}(\mu_1 - \mu_2)'}[B - B(A+B)^{-1}B](\mu_1 - \mu_2) = e^{-\frac{1}{2}(\mu_1 - \mu_2)'}(B^{-1}(A+B)A^{-1})^{-1}(\mu_1 - \mu_2). \]

Substitution of (3.36) and (3.39) into (3.35) results
\[
(2\pi)^{-\frac{p(a_1 + a_2)}{2}} |\Sigma_1|^{-\frac{a_1}{2}} |\Sigma_2|^{-\frac{a_2}{2}} e^{-\frac{1}{2}(\mu_1 - \mu_2)'} [B - B(A+B)^{-1}B](\mu_1 - \mu_2) \times
\]
\[
\int_{\mathbb{R}^p} e^{-\frac{1}{2}[(x-w)'(A+B)(x-w)]} dx =
\]
\[
(2\pi)^{\frac{p}{2}} (|\Sigma_1|^{\frac{1}{2}} |\Sigma_2|^{\frac{1}{2}}) (a_1 \Sigma_2 + a_2 \Sigma_1) |\Sigma_1|^{-\frac{1}{2}}
\]

\[ \Box \]

Theorem 2 specifies the form of the Bayes estimator when \( X|\Theta \sim N_p(\theta, \sigma^2 I_p) \) and the prior \( \Theta \sim N_p(\mu, \tau^2 I_p) \), when the loss is given by (3.30). It extends the results of Gosh, Mergel, and Datta [14] which computed the Bayes estimator when \( X|\Theta \sim N_p(\theta, \sigma^2 I_p) \) with prior \( \Theta \sim N_p(\mu, \tau^2 I_p) \) and loss
\[ l(d, \theta) = 1 - e^{-\frac{b}{\tau^2}e^{-b\|d - \theta\|^2}} \]
for \( b > 0 \).

**Theorem 3.2.** Let \( X|\Theta \sim N_p(\theta, \sigma^2 I_p) \) with \( \sigma^2 \) known and with \( \Theta \sim N_p(\mu, \tau^2 I_p) \). For \( b > 0 \), \( a \geq 0 \) and \( h \), a non-negative monotonic increasing right-continuous function with \( \int_0^\infty df(b) < \infty \), the Bayes estimator under the loss \( l(\|\delta - \theta\|^2) = a\|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|\delta - \theta\|^2})dh(b) \) is the posterior mean.

**Proof.** The posterior distribution \( \theta|X \sim N_p(\mu_{\sigma^2}/\sigma^2_{\sigma^2 + \tau^2}X + \sigma^2_{\sigma^2 + \tau^2} \mu, \sigma^2_{\sigma^2 + \tau^2} I_p) \). The Bayes estimator is obtained by minimizing the quantity \( E[a\|\theta - d\|^2 - \int_0^\infty e^{-b\|\theta - d\|^2} dh(b)|X] \) with respect to \( d \).
\[
\min_d \{E[a\|\theta - d\|^2 - \int_0^\infty e^{-b\|\theta - d\|^2} dh(b)|X] \} \geq (3.41)
\]
\[ a \ast \min_d \{ E[\|\theta - d\|^2 | X] \} - \max_d \{ E[\int_0^\infty e^{-b\|\theta - d\|^2} dh(b) | X] \}. \] (3.42)

Under quadratic loss, \( \|d - \theta\|^2 \), the minimum value of \( E[\|d - \theta\|^2 | X] \) is achieved for \( d \) equal to the posterior mean,

\[ d = \left( \frac{\sigma^2}{\sigma^2 + \tau^2} \right) X + \left( \frac{\sigma^2}{\sigma^2 + \tau^2} \right) \mu. \]

The maximum value of \( E[\int_0^\infty e^{-b\|\theta - d\|^2} dh(b) | X] \) can be found by applying Lemma 3.1 to

\[ E[\int_0^\infty e^{-b\|\theta - d\|^2} dh(b) | X] = \int_{\mathbb{R}^p} \int_0^\infty e^{-b\|\theta - d\|^2} dh(b) \left( \frac{\sigma^2 + \tau^2}{2\pi\tau^2\sigma^2} \right)^{\frac{p}{2}} e^{-\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} \|\theta - (\frac{\tau^2}{\tau^2 + \sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu)\|^2} d\theta = \int_0^\infty \int_{\mathbb{R}^p} e^{-b\|\theta - d\|^2} \left( \frac{\sigma^2 + \tau^2}{2\pi\tau^2\sigma^2} \right)^{\frac{p}{2}} e^{-\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} \|\theta - (\frac{\tau^2}{\tau^2 + \sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu)\|^2} d\theta dh(b), \] (3.43)

\[ \int_0^\infty \int_{\mathbb{R}^p} e^{-b\|\theta - d\|^2} \left( \frac{\sigma^2 + \tau^2}{2\pi\tau^2\sigma^2} \right)^{\frac{p}{2}} e^{-\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} \|\theta - (\frac{\tau^2}{\tau^2 + \sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu)\|^2} d\theta dh(b) = \int_0^\infty (\frac{\pi}{b})^{\frac{p}{2}} \int_{\mathbb{R}^p} e^{-\frac{1}{2\pi(\tau^2 + \sigma^2)} \|\theta - d\|^2} \left( \frac{\sigma^2 + \tau^2}{2\pi\tau^2\sigma^2} \right)^{\frac{p}{2}} e^{-\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} \|\theta - (\frac{\tau^2}{\tau^2 + \sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu - d)\|^2} d\theta dh(b) = \int_0^\infty (\frac{\pi}{b})^{\frac{p}{2}} (2\pi)^{\frac{p}{2}} \frac{2b(\tau^2 + \sigma^2)}{2b(\tau^2 + \sigma^2 + \tau^2 + \sigma^2)} \frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} \mu - d)^{\frac{p}{2}} (e^{\frac{1}{2} \frac{2b(\tau^2 + \sigma^2)}{2b(\tau^2 + \sigma^2 + \tau^2 + \sigma^2)} (\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu - d)(\frac{\sigma^2 + \tau^2}{2\tau^2\sigma^2} X + \frac{\sigma^2}{\tau^2 + \sigma^2} \mu - d)}) dh(b) \]

which is maximized for all \( b > 0 \) once \( d = \left( \frac{\tau^2}{\sigma^2 + \tau^2} \right) X + \left( \frac{\sigma^2}{\sigma^2 + \tau^2} \right) \mu \). When \( d = \left( \frac{\tau^2}{\sigma^2 + \tau^2} \right) X + \left( \frac{\sigma^2}{\sigma^2 + \tau^2} \right) \mu \), expression (3.41) equals (3.42), so that the Bayes estimator is the posterior mean.

\[ \Box \]

Theorem 3.3 establishes the minimaxity of the estimator \( \delta_0(X) = X \) for losses of the form (3.30). The result is a consequence of Theorem 1.3 applied to the least favorable sequence of priors, \( \{ N_p(0, nI_p) \}_{n=1}^\infty \).

**Theorem 3.3.** Let \( X \sim N_p(\theta, \sigma^2 I_p) \) with \( \sigma^2 \) known. Then the estimator \( X \) will be a minimax estimator for \( \theta \) under the loss \( l(\|\delta - \theta\|^2) \) given in Theorem 3.2.

**Proof.** By Tonelli’s Theorem
\[ R(X, \theta) = E[a\|X - \theta\|^2 + \int_0^\infty dh(b) - \int_0^\infty e^{-b\|X-\theta\|^2}dh(b)] = \]
\[ aE[\|X - \theta\|^2] + \int_0^\infty dh(b) - \int_0^\infty E[e^{-b\|X-\theta\|^2}]dh(b). \]

By the change of variables \( Y = \|X - \theta\|^2 \), \( Y \sim \sigma^2 \chi^2_p \)
\[ E[e^{-b\|X-\theta\|^2}] = \int_0^\infty e^{-by^2} \frac{y^p - 1}{\Gamma(p)} \frac{dy}{2^p} = \]
\[ \int_0^\infty e^{-y(2b\sigma^2 + 1)} y^{p-1} \frac{dy}{\Gamma(p)2^p} = \frac{\Gamma(p/2)}{(2b\sigma^2 + 1)^{p/2}} \]
\[ \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} \] (3.45)

and,
\[ aE[\|X - \theta\|^2] = aE[Y] = a\sigma^2 \] (3.46)

so that the \( R(X, \theta) \) is expressible as
\[ a\sigma^2 + \int_0^\infty - \int_0^\infty \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} dh(b). \] (3.47)

The Bayes risk \( r_\Lambda \) of the Bayes estimator \( \delta_\Lambda \) will satisfy
\[ R_\Lambda = \int R(\delta_\Lambda, \theta)d\Lambda(\theta) \geq \int R(X, \theta)d\Lambda(\theta) = a\sigma^2 + \int_0^\infty dh(b) - \int_0^\infty \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} dh(b) \]

Let the sequence of priors \( N_p(0, nI_p) \) indexed by n be denoted by the sequence \( \pi_n \). Then by Theorem 3.2 the Bayes estimator for the prior \( \pi_n \) will be
\[ \delta_{\pi_n}(X) = \frac{n}{\sigma^2 + n}X. \]

with Bayes Risk
\[ R_{\pi_n} = p\sigma^2 \left( \frac{n}{\sigma^2 + n} \right) + \int_0^\infty dh(b) - E\theta[\int_0^\infty E[e^{-b\|\theta-\frac{n}{\sigma^2 + n}X\|}d\theta]] = \]
\[ a\sigma^2 \left( \frac{n}{\sigma^2 + n} \right) + \int_0^\infty dh(b) - \int_0^\infty \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} dh(b). \]

as \( \Theta|X \sim N_p\left( \frac{n}{\sigma^2 + n}X, \left( \frac{n}{\sigma^2 + n} \right)\sigma^2 I_p \right) \).

\[ \lim_{n \to \infty} R_{\pi_n} \to p\sigma^2 + \int_0^\infty dh(b) - \int_0^\infty \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} dh(b) = R(X, \theta) \]
and \( R_{\pi_n} < p\sigma^2 + \int_0^\infty dh(b) - \int_0^\infty \left( \frac{1}{2b\sigma^2 + 1} \right)^{\frac{p}{2}} dh(b) \) for every n, so that \( \pi_n \) is a least favorable sequence and X is a minimax estimator.

\[■\]
3.2.2 Minimax Estimators

In this subsection we study the minimaxity of estimators of the form

$$\delta_{a,r} = (1 - \frac{a \sigma^2 r(\|X\|^2)}{\|X\|^2})X$$

(3.48)

for the class of loss functions given by (3.12) when $X \sim N_p(\theta, \sigma^2 I_p)$ with $\sigma^2$ known. The results extend those of Ghosh, Mergel, and Data [14] which gave the conditions

i) $0 < a < 2(p - 2)$

ii) $0 < r \leq 1$

iii) $r'(u) \geq 0$

for which estimators given by (3.48) will strictly dominate the estimator $\delta_0(X) = X$ for the class of losses given by

$$l(d, \theta) = 1 - e^{\beta \sigma^2 \|d - \theta\|^2}$$

(3.49)

where $\beta > 0$. The change of variables $b = \frac{\beta}{\sigma^2}$ results in the following Lemma and Theorem which are found in Gosh, Mergel, and Datta [14]. The proofs are presented for completeness in Appendix A.

**Lemma 3.2.** (Gosh, Mergel, and Datta [14]) Let $X \sim N_p(\theta, \sigma^2 I_p)$ with $\sigma^2$ known and let $\delta(X)$ be an estimator for $\theta$ of the form $\delta_r(X) = (1 - \sigma^2 \tau(\|X\|^2))X$, then

$$E_{\theta}[e^{-b \|\delta_r(X) - \theta\|^2}] =$$

$$(2b \sigma^2 + 1) - \frac{2}{\sum_{r=0}^{\infty} e^{-\left(b + \frac{1}{2 \sigma^2}\right)\|\theta\|^2}\left(\frac{b + \frac{1}{2 \sigma^2}}{r!}\right)^r I_b(r)}$$

where

$$I_b(r) = \int_0^\infty \left[1 - \frac{b \sigma^2}{t - \tau\left(\frac{2t}{2b \sigma^2 r} + 1\right)}\right]^{2r} t^{r + \frac{b}{2} - 1} e^{-t\left(\frac{2t}{2b \sigma^2 r} + 1\right) + 2b \sigma^2 \tau\left(\frac{2t}{2b \sigma^2 r}\right)} dt$$
Theorem 3.4. (Gosh, Mergel, and Datta [14]) Let \( X \sim N_p(\theta, \sigma^2 I_p) \) with \( \sigma^2 \) known, and let \( \delta_\tau(X) \) be an estimator for \( \theta \), such that \( \tau \) satisfies:

i) \( 0 \leq \tau(u) \leq 2(p - 2) \)

ii) \( \tau \) is differentiable with a non negative derivative

then the estimator \( \delta_\tau \) will be minimax estimator of location under the loss function \( l(t) = 1 - e^{-bt} \) for \( b > 0 \), and \( p \geq 3 \).

Lemma 3.3 and Theorem 3.5 are the main results of this section. They extend the above result to losses of the form (3.12).

Lemma 3.3. Let \( h \) be any non-negative monotonically increasing right-continuous function such that \( \int_0^\infty dh(b) < \infty \). Suppose that \( \delta(X) \) is an estimator for which

i) \( E[\|d - \theta\|^2] < E[\|X - \theta\|^2] \) under square error loss,

ii) \( E_\theta[e^{-b\|X - \theta\|^2}] \leq E_\theta[e^{-b\|\delta - \theta\|^2}] \) \( \forall b > 0 \) and \( \forall \theta \)

then for the class of losses given by

\[
l(d, \theta) = a\|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|d - \theta\|^2})dh(b),
\]

(3.50)

\[
R(\delta, \theta) \leq R(X, \theta)
\]

Proof.

\[
R(\delta, \theta) - R(X, \theta) =
\]

\[
E_\theta[l(\|\delta(X) - \theta\|^2) - l(\|X - \theta\|^2)] =
\]

\[
aE_\theta[\|\delta(X) - \theta\|^2 - \|X - \theta\|^2] - E_\theta[\int_0^\infty \{e^{-b\|\delta(X) - \theta\|^2} - e^{-b\|X - \theta\|^2}\}dh(b)] =
\]

\[
aE_\theta[\|\delta(X) - \theta\|^2 - \|X - \theta\|^2] - \int_0^\infty E_\theta[\{e^{-b\|\delta(X) - \theta\|^2} - e^{-b\|X - \theta\|^2}\}]dh(b)
\]

(3.51)

were the interchange of the integral and expected value is valid as \( E_\theta \int_0^\infty e^{-b\|\delta - \theta\|^2}dh(b) \leq \int_0^\infty dh(b) < \infty \). Since \( E_\theta[\|\delta - \theta\|^2] \leq E_\theta[\|X - \theta\|^2] \) and \( E_\theta[e^{-b\|\delta - \theta\|^2}] \geq E_\theta[e^{-b\|X - \theta\|^2}] \) \( \forall b > 0 \), (3.51) is non positive so that \( R(\delta, \theta) \leq R(X, \theta) \).
Theorem 3.5. Let \( X \sim N_p(\theta, \sigma^2 I_p) \) for \( p \geq 3 \). Let \( h \) be any non-negative monotonic increasing right-continuous function such that \( \int_0^\infty dh(b) < \infty \). Let \( \delta_{a,r}(X) \) be an estimator of \( \theta \) of the form:

\[
\delta_{a,r}(X) = (1 - \frac{\sigma^2 ar(\|X\|^2)}{\|X\|^2})X
\]

Then for \( c > 0 \), \( \delta_{a,r} \) will be a minimax estimator of \( \theta \) for each loss of the form:

\[
l(d, \theta) = c\|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|d - \theta\|^2})dh(b)
\]

once

i) \( 0 \leq a \leq 2(p - 2) \),

ii) \( 0 \leq r \leq 1 \),

iii) \( r'(t) \geq 0 \)

with strict domination over \( \delta_0(X) = X \) if either i) or iii) is strict.

Proof. If the constant, \( a \), satisfies i) and the function, \( r \), satisfies ii) and iii) then \( \delta_{a,r} \) is a minimax estimator under quadratic loss, \( \|d - \theta\|^2 \), so that \( R(\delta_{a,r}, \theta) \leq R(X, \theta) \) under quadratic loss with strict inequality if either i) or iii) is strict.

Letting \( ar(u) = \tau(\frac{u}{\sigma^2}) \), conditions i)-iii) implies \( 0 \leq \tau \leq 2(p - 2) \), and \( \tau'(u) \geq 0 \), so that Theorem 3.4 implies for any \( b > 0 \),

\[
E[1 - e^{-b\|\delta_{a,r}(X) - \theta\|^2}] \leq E[1 - e^{-b\|X - \theta\|^2}]
\]

where (3.54) is strict if either i) or iii) is strict. Therefore the assumption of Corollary 3.3 are satisfied so that \( \delta_{a,r} \) is a minimax estimator of \( \theta \) for the class of losses given by (3.53), since \( X \) is also a minimax estimator for the class of losses given by (3.53) and \( R(\delta_{a,r}, \theta) \leq R(X, \theta) \).

The following example illustrates Theorem 3.5 when \( h(\cdot) \) is a Gamma\((k_1, k_2)\) distribution.
Example 3.2. Let $X \sim N_p(\theta, \sigma^2 I_p)$ with $\sigma^2$ known, and $\delta$ be an estimator of $\theta$ of the form
\[
\delta(X) = (1 - \frac{a\sigma^2}{\|X\|^2})X.
\] (3.55)

$\delta$ will be a minimax estimator for the loss,
\[
l(d, \theta) = c\|d - \theta\|^2 + 1 - (k_2\|d - \theta\|^2 + 1)^{-k_1}
\] (3.56)

where $c \geq 0$ and $k_1, k_2 > 0$, once

i) $0 \leq a \leq 2(p - 2)$.

This follows from Theorem 3.5 by setting $h(b)$ by the cumulative distribution function of a Gamma distribution with location parameter $k_1$ and scale parameter $k_2$. Letting $t = \|d - \theta\|^2$, loss (3.12) is equivalent to

\[
l(t) = ct + 1 - \frac{1}{\Gamma(k_1)k_2^{k_1}} \int_0^{\infty} e^{-bt} e^{-\frac{b}{k_2}} b^{k_1-1} db =
\]

\[
ct + 1 - \frac{1}{\Gamma(k_1)k_2^{k_1}} \int_0^{\infty} e^{-b(t + \frac{1}{k_2})} b^{k_1-1} db =
\]

\[
ct + 1 - \left(\frac{k_1}{k_2} t + 1\right)^{-k_1}
\]

By Theorem 3.5 estimator (3.55) is minimax once $0 \leq a \leq 2(p - 2)$, as the function $r(\|X\|^2) = 1$ satisfies $r'(u) \geq 0$ and $0 \leq r(u) \leq 1$ for all $u \geq 0$.

3.3 Extensions

In this Section we extend the result of Section 2, initially to the case that the underlying distribution is normal with an unknown scale and a residual vector, and then to the case that the underlying distribution is in the class of spherically symmetric scale mixture of normal distributions with a residual vector present, and find classes of estimators that dominate $\delta_0(X) = X$ with loss,

\[
l(d, \theta) = c\|d - \theta\|^2 + \int_0^{\infty} (1 - e^{-b\|d - \theta\|^2})dh(b)
\] (3.57)
that also dominate $\delta_0(X) = X$ under quadratic loss, $\|d - \theta\|^2$. Replacing $\|d - \theta\|^2$ in (3.57) with $\frac{\|d - \theta\|^2}{\sigma^2}$ the estimator, $\delta_0(X) = X$, will have constant risk and be minimax and thus dominating estimators will be minimax.

Section 3.1 gives a preliminary result from Gosh and Mergel [15] that will be used in extending the results of Section 2.

Section 3.2 deals with the case that

$$\begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k})$$

with $\sigma^2$ unknown, $\text{dim}(X) = \text{dim}(\theta) = p \geq 3$, and $\text{dim}(U) = \text{dim}(0) = k \geq 1$, and $\delta_{a,r}(X,S)$ is an estimator of $\theta$ of the form

$$\delta_{a,r}(X,S) = (1 - \frac{S\text{ar}(\frac{\|X\|^2}{S})}{(k + 2)\|X\|^2})X$$  \hspace{1cm} (3.58)

where $S = \|U\|^2$ in (3.58). The estimator $\delta_{a,r}(X,S)$ will dominate the estimator $\delta_0(X) = X$ for the class of losses given by (3.57) once

i) $0 \leq a \leq 2(p - 2)$, ii) $r'(u) \geq 0 \forall u > 0$, iii) $0 \leq r(u) \leq 1 \forall u > 0$,

where conditions i-iii) are equivalent conditions for domination of the estimator $\delta_0(X) = X$ under quadratic loss.

Section 3.3 deals with the case

$$\begin{pmatrix} X \\ U \end{pmatrix} |\sigma^2 \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k})$$  \hspace{1cm} (3.59)

with $\sigma^2 \sim F(\sigma^2)$, where $F$ is a distribution function on $(0, \infty)$ such that $E[\sigma^2] < \infty$ and $E[\frac{1}{\sigma^2}] < \infty$. The estimator $\delta_{a,r}$ given by (3.58) will dominate $X$ under loss (3.57). Under quadratic loss the results of Fourdrinier, Strawderman, and Wells [10], which apply to the general class of spherically symmetric distributions with residuals, imply that estimator $\delta_{a,r}$.
given by (3.58) will dominate the estimator \( \delta_0(X) = X \) if i-iii) are satisfied.

Section 3.4 gives results for generally spherically symmetric distributions for general concave loss functions. In particular when the distribution of \( \begin{pmatrix} X \\ U \end{pmatrix} \sim SS_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}) \) with \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3, \text{dim}(U) = \text{dim}(0) = k \geq 1 \) and density of the form

\[
\frac{1}{\sigma^p} f\left(\frac{1}{\sigma^2}(\|x - \theta\|^2 + \|u\|^2)\right),
\] (3.60)

and \( \delta(X) \) is an estimator of \( \theta \) of the form

\[
\delta(X, \|U\|^2) = X + \frac{\|U\|^2}{(k + 2)} g(X)
\] (3.61)

with \( E[g(X)] < \infty \) and \( g \) a weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \), the results of Fourdinier and Strawderman imply that for \( l(t) \), any non-negative differentiable concave loss function where \( t = \|d - \theta\|^2 + \|U\|^2 \), once \( g(X) \) satisfies \( \|g(X)\|^2 + 2\text{div}_x(g(X)) < 0 \), \( R(\delta, \theta) \leq R(X, \theta) \). These conditions are the equivalent conditions for domination of the estimator \( \delta_0(X) \) under quadratic loss, \( \|d - \theta\|^2 \).

### 3.3.1 Preliminaries

This subsection gives the results of Gosh and Mergel [15] which gave conditions in which the estimator

\[
\delta_\tau(X, S) = (1 - \frac{S_T(\|X\|^2)}{S_T(\|X\|^2) + (k + 2)}X)
\] (3.62)

where \( S = \|U\|^2 \), dominates the estimator \( X \) for the loss

\[
l(d, \theta) = 1 - e^{-\frac{\|d - \theta\|^2}{\sigma^2}}
\] (3.63)

when \( \begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \) with \( \sigma^2 \) unknown, \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \), and \( \text{dim}(U) = \text{dim}(0) = k \geq 1 \). It is used in Theorem 3.6 which studies the normal case with
unknown scale and a residual vector, for losses given by (3.57). It’s stated using the loss function

\[ l(d, \theta) = 1 - e^{-b\|d-\theta\|^2} \]  

(3.64)

for \( b > 0 \) instead which is remarked upon in their proof. For completeness the proof is presented in Appendix B.

**Lemma 3.4.** (Gosh and Mergel [15]) Let \( \begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \) with \( \sigma^2 \) unknown, \( \dim(X) = \dim(\theta) = p \geq 3 \), and \( \dim(U) = \dim(0) = k \geq 1 \). Let \( \delta_\tau(X, S) \) where \( S = \|U\|^2 \), be an estimator of \( \theta \) of the form

\[ \delta_\tau(X, S) = (1 - \frac{S\tau(\|X\|^2)}{(k+2)\|X\|^2})X. \]  

(3.65)

If

i) \( 0 \leq \tau(u) \leq 2(p-2) \) for \( u > 0 \),

ii) \( \frac{d}{du} \tau(u) \geq 0 \) for \( u > 0 \)

then

\[ R(\delta_\tau, \theta) \leq R(X, \theta) \]

for the loss

\[ l(d, \theta) = 1 - e^{-b\|d-\theta\|^2} \]  

(3.66)

for \( b > 0 \).

### 3.3.2 Normal with Scale Unknown

Theorem 3.6 extends the class of losses for which \( \delta_{a,r} \) given in Lemma 3.4 will dominate the estimator \( \delta_0(X) = X \) , to the class of losses given in (3.57).
Theorem 3.6. Let \( \begin{pmatrix} X \\ U \end{pmatrix} \) be distributed as in Lemma 3.4. Then for the class of losses,

\[
l(d, \theta) = c \|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|d - \theta\|^2}) dh(b)
\] (3.67)

where \( c \geq 0 \) and \( h \) is non-negative monotonically increasing right-continuous function with \( \int_0^\infty dh(b) < \infty \), the estimator

\[
\delta_{a,r}(X, S) = (1 - \frac{aSr(\|X\|^2)}{S(k + 2)\|X\|^2})
\] (3.68)

will dominate the estimator \( X \) once

i) \( 0 \leq a \leq 2(p - 2) \),

ii) \( \frac{dr}{du}(u) \geq 0 \) for \( u > 0 \),

iii) \( 0 \leq r(u) \leq 1 \) for \( u > 0 \)

with strict domination if the inequality i) or ii) is strict.

Proof. Let \( ar(u) = \tau(u) \). If conditions i-iii) are satisfied from Theorem 3.6 then \( \tau(u) \) satisfies conditions i) and ii) from Lemma 3.4. For \( b > 0 \),

\[
E[e^{-b\|X - \theta\|^2}] \leq E[e^{-b\|\delta_{a,r} - \theta\|^2}]
\] (3.69)

where the inequality in (3.69) follows from Lemma 3.4 with strict inequality if either i) or ii) is strict. Under the quadratic loss, \( \|d - \theta\|^2 \), once conditions i-iii) are satisfied,

\[
R(\delta_{a,r}, \theta) \leq R(X, \theta)
\]

with strict inequality when either i) or ii) is strict. Therefore Lemma 3.3 implies that \( \delta_{a,r}(X, S) \) will dominate \( \delta_0(X) = X \) when the loss is given by (3.67) once conditions i-iii) are satisfied. ■

Remark: If \( \|d - \theta\|^2 \) is replaced by \( \frac{\|d - \theta\|^2}{\sigma^2} \) in (3.67) the estimator \( \delta_\tau \) still dominates the estimator \( X \), and since \( X \) is minimax the dominating estimator is also minimax.
3.3.3 Scale Mixtures of Normals with Unknown Scale

Theorem 3.7 gives conditions for which estimators of the form (3.58) dominate the estimator \( \delta_0(X) = X \) for losses of the form (3.57) when the underlying distribution \( \begin{pmatrix} X \\ U \end{pmatrix} | \sigma^2 \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \) with \( \sigma^2 \sim F(\sigma^2) \), where \( F \) is a distribution function on \((0, \infty)\) having density \( v(\sigma^2) \).

**Theorem 3.7.** Let \( \begin{pmatrix} X \\ U \end{pmatrix} | \sigma^2 \sim N_{p+m}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+m}) \) where \( \text{dim}(X) = \text{dim}(\theta) = p \geq 3 \), \( \text{dim}(U) = \text{dim}(0) = m \geq 1 \), and the distribution of \( \sigma^2 \) having density \( v(\sigma^2) \). Let \( c \geq 0 \), and \( h \) be any non-negative monotonic increasing right-continuous function with \( \int_0^\infty dh(b) < \infty \), and loss

\[
l(d, \theta) = c\|d - \theta\|^2 + \int_0^\infty (1 - e^{-b\|d - \theta\|^2})dh(b).
\]

(3.70)

Then for an estimator \( \delta_\tau \) of \( \theta \) of the form

\[
\delta_\tau(X, U) = (1 - \frac{\|U\|^2 \tau(\frac{\|X\|^2}{(m+2)\|U\|^2})}{(m+2)\|X\|^2})X,
\]

(3.71)

\( R(X, \theta) > R(\delta_\tau, \theta) \) uniformly in \( \theta \) provided

i) \( 0 < \tau(u) < 2(p - 2) \) for \( u > 0 \),

ii) \( \frac{d}{du}\tau(u) \geq 0 \) for \( u > 0 \).

**Proof** By Theorem 3.6, conditioning on \( \sigma^2 \), the difference in risk between the estimator \( \delta_\tau \) and \( X \) satisfies

\[
E[\|c\|\delta_\tau - \theta\|^2 + \int_0^\infty (1 - e^{-b\|\delta_\tau - \theta\|^2})dh(b)]|\sigma^2| -

E[\|c\|X - \theta\|^2 + \int_0^\infty (1 - e^{-b\|X - \theta\|^2})dh(b)]|\sigma^2| \leq 0
\]
for all \( \sigma^2 > 0 \). Therefore taking the expectation of the above establishes the result.

\[ \square \]

**Remark:** If \( \|d - \theta\|_2^2 \) is replaced by \( \frac{\|d - \theta\|_2^2}{\sigma^2} \) in (3.70) the estimator \( \delta_\tau \) still dominates the estimator \( X \), and since \( X \) is minimax the dominating estimator is also minimax.

### 3.3.4 A General Result for Sypherically Symmetric Distributions with a Residual

A more general result holds for the class of spherically symmetric distributions with residual vectors having densities of the form \( f(\|x - \theta\|_2^2 + \|u\|_2^2) \) with \( \dim(X) = \dim(\theta) = p \geq 3 \) and \( \dim(U) = k \geq 1 \), for estimators of the form

\[
\delta(X, \|U\|_2^2) = X + \frac{\|U\|_2^2}{k + 2} g(X),
\]

(3.72)

for any concave non-negative differentiable loss function, \( l(t) \). However the result requires \( t \) to be of the form

\[
t = \|d - \theta\|_2^2 + \|u\|_2^2.
\]

(3.73)

The proof is a consequence of applying the results of Foudinier and Strawderman [30] to Theorem 3.1.

**Theorem 3.8.**

Let

\[
\begin{pmatrix}
X \\
U
\end{pmatrix} \sim SS_{p+k}(\begin{pmatrix}
\theta \\
0
\end{pmatrix})
\]

with density \( f(\|X - \theta\|_2^2 + \|U\|_2^2) \). Let

\[
\delta(X, U) = X + \frac{\|U\|_2^2}{k + 2} g(X)
\]

be an estimator for \( \theta \) such that

i) \( g(X) \) is weakly differentiable function from \( \mathbb{R}^p \) into \( \mathbb{R}^p \),

ii) \( E_\theta[\|g(X)\|_2^2] < \infty \),

iii) \( \|g(X)\|_2^2 + 2 \text{div}_x(g(X)) < 0 \) (a.e.)
Then for \(l(t)\), any non-negative differentiable loss function in \(t\) where \(t = \|\delta - \theta\|^2 + \|U\|^2\), \(R(\delta, \theta) < R(X, \theta)\).

**Proof.** By the concavity of \(l(t)\)

\[
l(t + h) \leq l(t) + hl'(t)
\]

so that

\[
l(\|\delta(X, U) - \theta\|^2 + \|U\|^2) \leq
\]

\[
= l(\|X - \theta\|^2 + \|U\|^2) + l'(\|X - \theta\|^2 + \|U\|^2)(\frac{\|U\|^4}{(k + 2)^2}\|g(X)\|^2 + \|U\|^2 \frac{k}{k + 2}(x - \theta)'g(X)).
\]

(3.74)

Letting \(f^*(\|x - \theta\|^2 + \|u\|^2) = \frac{1}{k}l'(\|x - \theta\|^2 + \|u\|^2)f(\|x - \theta\|^2 + \|u\|^2)\) where \(k\) is the constant of proportionality of density \(f^*\), expression (3.74) implies

\[
R(\delta, \theta) \leq R(X, \theta) + kE[\|U\|^4 \|g(X)\|^2 + \|U\|^2 \frac{k}{k + 2}(x - \theta)'g(X)],
\]

(3.75)

where the density of the expectation in (3.75) is taken with respect to the density \(f^*\). Since \(f^*\) is a spherically symmetric distribution expression the results of Fourdrinier, Strawderman, and Wells [11] imply (3.75) is equivalent to

\[
R(\delta, \theta) \leq R(X, \theta) + kE[(\frac{\|U\|^4}{(k + 2)^2})(\|g(X)\|^2 + div_x(g(X)))] < R(X, \theta).
\]

(3.76)

Since \(\frac{\|U\|^4}{(k + 2)^2} > 0\) and and by assumption iii) \(\|g(X)\|^2 + 2div_x(g(X)) < 0\) a.e.,

\[
kE[\frac{\|U\|^4}{(k + 2)^2}(\|g(X)\|^2 + div_x(g(X)))] < 0
\]

\[\square\]

### 3.4 Simulation Results

In this Section we investigate via simulations the behavior of the risk of the James-Stein estimator

\[
\delta_{JS,c}(X) = (1 - \frac{c}{\|X\|^2})X
\]

(3.77)
when the underlying distribution $X$, is a scale mixture of normal distributions without a residual, and the loss is given by

$$l(d, \theta) = 1 - e^{-b\|d-\theta\|^2}. \quad (3.78)$$

In Section 2 it was shown for $X \sim N_p(\theta, \sigma^2 I_p)$ with $p \geq 3$ and $\sigma^2$ known, the class of shrinkage estimators

$$\delta_{a,r}(X) = (1 - a\sigma^2 r(\|X\|^2) / \|X\|^2) (3.79)$$

will dominate the estimator $\delta_0(X) = X$ for loss functions of the form

$$l(d, \theta) = c\|d-\theta\|^2 + \int_0^{\infty} (1 - e^{-b\|d-\theta\|^2}) dh(b) \quad (3.80)$$

once: i) $r'(u) > 0$, ii) $0 \leq r(u) \leq 1$, and iii) $0 \leq a \leq 2(p - 2)$. Interestingly for the class of estimators given by (3.79), the same set of values for the shrinkage constant, $a$, as condition iii) gives domination of the minimax estimator $\delta_0(X) = X$ under quadratic loss, $\|d - \theta\|^2$.

In Section 3 when $\sigma^2$ was unknown and a residual vector, $U$, is present, i.e.

$$\begin{pmatrix} X \\ U \end{pmatrix} \sim N_{p+k}(\begin{pmatrix} \theta \\ 0 \end{pmatrix}, \sigma^2 I_{p+k}) \quad (3.81)$$

with $\text{dim}(X) = \text{dim}(\theta) = p \geq 3$ and $\text{dim}(U) = \text{dim}(0) = k \geq 1$, the class of shrinkage estimators of the form

$$\delta_{a,r}(X, U) = (1 - \frac{\|U\|^2 ar(\|X\|^2)}{(k + 2)\|U\|^2}) \quad (3.82)$$

was shown to dominate the estimator $\delta_0(X) = X$ for losses of the form (3.80) once i) $r'(u) > 0$, ii) $0 \leq r(u) \leq 1$, and iii) $0 \leq a \leq 2(p - 2)$. The same set of conditions on the shrinkage constant, $a$, as iii) gives domination over the estimator $\delta_0(X) = X$ under quadratic loss as well. Furthermore for the general class of scale mixture of normal distributions when
a residual vector is present, i.e.

\[
\begin{pmatrix} X \\ U \end{pmatrix} | \sigma^2 \sim N_{p+k}(\theta, \sigma^2 I_{p+k})
\]

(3.83)

with \(\sigma^2 \sim F(\sigma^2)\) where \(F(\sigma^2)\) is a distribution function on \((0, \infty)\) with \(E[\sigma^2]\) and \(E[\frac{1}{\sigma^2}]\) both finite, estimators of the form (3.82) will dominate the estimator \(\delta_0(X) = X\) once i) \(r'(u) > 0\), ii) \(0 \leq r(u) \leq 1\), and iii) \(0 \leq a \leq 2(p-2)\). These conditions also give domination of \(\delta_0(X) = X\) under quadratic loss.

From Section 2 and Section 3 one might suspect that for the class of estimators given by (3.77), which is a subclass of the estimators given in (3.79), once \(\delta_{JS}\) dominates the risk of \(X\) under quadratic loss it will dominate the risk of \(X\) for the family of losses given by (3.78). Our simulation indicate this is not true in general.

Specifically we assume

\[
X|V \sim N_p(\theta, V I_p)
\]

(3.84)

with \(p = 6\), where \(V\) has an inverse gamma distribution with density

\[
h(v) = \frac{1}{\Gamma(a)s^a} v^{-a-1} e^{-\frac{1}{sv}}
\]

(3.85)

and loss

\[
l(d, \theta) = 1 - e^{-b\|d-\theta\|^2}
\]

(3.86)

for \(b > 0\). We plot the risk of the estimator \(\delta_{JS}\) given by (3.77) when \(c\) is chosen to be near the upper bound established by Strawderman [29] for minimaxity under square error loss to see if it is minimax using the loss given by (3.86). In particular for scale mixtures of normal distributions, Strawderman [29] established that estimators given by (3.77) will be minimax for square error loss once

\[
0 \leq c \leq \frac{2(p-2)}{E[\frac{1}{V}]}
\]

(3.87)
and thus when the mixture distribution $V$ has an inverse gamma distribution given by (3.85), the estimator $\delta_{JS}$ will be minimax under square error loss once

$$0 \leq c \leq \frac{2(p-2)}{E[1/V]} = \frac{2(p-2)}{as} \quad (3.88)$$

as $\frac{1}{V} \sim \Gamma(a, s)$.

Figure 3.1 plots the risks under square error for the case that the mixture distribution is given by an inverse gamma distribution with parameters $a=3, s=1$ and $a=5, s=2$ respectively versus the risk of the minimax estimator $X$. The James-Stein estimators considered are

$$\delta_{JS} = (1 - \frac{2.66}{\|X\|_2})X$$

when $a=3, s=1$ as $\frac{2(6-2)}{3} = \frac{8}{3} \approx 2.66$ and

$$\delta_{JS} = (1 - \frac{8}{\|X\|_2})X$$

when $a=5$ and $s=2$. The simulation indicate the risk of the James-Stein estimators are non-monotonic for values of $c$ close to $\frac{2(p-2)}{E[1/V]}$ and are asymptotic to the risk of $X$.

Figure 3.1: Risk of James-Stein Estimator (3.77) for Mixture Distribution (3.85) with Shape Parameter, $a$, and Rate Parameter, $s$, Under Quadratic Loss, $l(d, \theta) = \|d - \theta\|^2$.

Figure 3.2 plots the risk of the estimator
\[ \delta_{JS} = (1 - \frac{2.66}{\|X\|^2})X \]

verses the risk of the minimax estimator \(X\) using loss

\[ l(d, \theta) = 1 - e^{-b\|d - \theta\|^2} \]

when the mixture has parameters \(a=3\) and \(s=1\), and when the loss has parameters \(b = 4, 2, 1, \) and \(0.5\) respectively. The simulation indicate that when \(c=2.66\) the estimator \(\delta_{JS}\) is minimax for \(b=0.5\). For sufficiently large values of \(\|\theta\|^2\) when \(b=4, 2, 1\) the risk of the estimator is greater than the Risk of \(X\) so that \(\delta_{JS}\) is not minimax indicating the conjecture is false.

![Figure 3.2: Simulated Risk of James-Stein Estimator (3.77) for Various b for Mixture Parameters (3.85) \(a = 3, s = 1\) and Loss \(l(t) = 1 - e^{-bt}\).](image)

Figure 3.3 plots the risk of the estimator

\[ \delta_{JS}(X) = (1 - \frac{8}{\|X\|^2})X \]

using loss (3.86) with mixture parameters \(a=5\), and \(s=2\) and when the loss has parameter values \(b=10, 4, 2,\) and \(1\) respectively. When \(b=10\) and \(4\) and for sufficient large values \(\|\theta\|^2\), the risk of the estimator \(\delta_{JS}\) has larger risk than the estimator \(\delta_0(X) = X\) and is thus not minimax. However this does not seem to be the case when \(b=2\) and \(1\) as the risk of the estimator dominates the risk of \(X\).
Figure 3.3: Simulated Risk of James-Stein Estimator (3.77) for Various b for Mixture Parameters (3.85) $a = 5$, $s = 2$ and Loss, $l(t) = 1 - e^{-bt}$.

To investigate how the tail behavior of the mixture might affect the risk, Figure 3.4 plots the risk of various James-Stein estimators when the mixture has fixed shape parameter $a=3$ with varying rate parameters, $s$. Each James-Stein estimators plotted has a shrinkage constant, $c$, that is minimax under quadratic loss. More specifically they were chosen to be at or near the value of $\frac{2(p-2)}{a^2}$. For $s=10$ and $s=4$ the James-Stein estimator appears to be minimax. For the distributions with heavier tails ($s=.5$ and $s=1$ ) , the risk of the estimator $X$ seem to be less than the risk of $\delta_{JS}$ for sufficiently large $\|\theta\|^2$.

Figure 3.4: Simulated Risk of James-Stein Estimator (3.77) for Various Mixture Parameter (3.85) $s$, $a=3$ and Loss , $l(t) = 1 - e^{-2t}$
Similarly in Figure 3.5 the mixture parameter, $a$, was varied and the rate parameter, $s$, was set to be 1. The loss parameter, $b$, was chosen to be .5. When $a=3$ and $s=1$ the risk of $\delta_{JS}$ with $c=2.66$ appears to be minimax. For the lighter tail mixtures ($a=5$ and 10) the estimator $\delta_{JS}$ also appears to be minimax.

Figure 3.5: Simulated Risk of James-Stein Estimator (3.77) for Various Mixture Parameter (3.85) $a$, $s=1$ and Loss, $l(t) = 1 - e^{-5t}$

Figures 3.6 investigates the behavior of the parameter $b$ on the risk of the James-Stein estimator with the shrinkage factor of $c = .8$ when the mixture parameters are $a = 10$ and $s = 1$. When $b = .5$ and $b = 1$ the risk of the James-Stein estimator approaches the risk of $X$ from below, while for $b = 10$ and $b = 20$ the risk of the James-Stein estimator approaches the risk of $X$ from above. This is similar to the results of Figures 3.3 and 3.4, where there exist $b_0 > 0$ such that for $b > b_0$ the risk of the James-Stein estimator approached the risk of $X$ from above and for $b < b_0$ the risk of the James-Stein estimator approaches the risk of $X$ from below.
Figure 3.6: Simulated Risk of James-Stein Estimator (3.77) for Varying b, for Mixture Parameters (3.85) $a = 10, s = 1$ and Loss, $l(t) = 1 - e^{-bt}$

Figure 3.7 also investigates the behavior of the parameter b on the risk of the James-Stein estimator with the shrinkage factor of $c = .53$ when the mixture parameters are $a = 15$ and $s = 1$. Just as in figure 3.6 when $b = .5$ and $b = 1$, the risk of the James-Stein approaches the risk of X from below, while for $b = 10$ and $b = 20$ the risk of the James-Stein estimator approaches the risk of X from above.

Figure 3.7: Simulated Risk of James-Stein Estimator (3.77) for Varying b for Mixture Parameters (3.85) $a = 15, s = 1$ and Loss, $l(t) = 1 - e^{-bt}$

Based on the simulation in this Section it appears that for for a given mixture with shape parameter $a = a_0$ and rate parameter, $s = s_0$, there exist a $b^*$ such that for $b > b^*$ the
James-Stein estimator with the shrinkage constant $c = \frac{2(p-2)}{a_0 s_0}$ has risk that starts off below the risk of $X$ at the origin, and asymptotically approaches the risk of $X$ from above. For $b < b^*$ the James-Stein estimator uniformly dominates the risk of $X$. Given $b_0 < b^*$ and rate parameter $s = s_0$, the risk of the James-Stein estimator will be uniformly less than the risk of $X$ for all values of the shape parameter, $a > a_0$ when the shrinkage constant $c = \frac{2(p-2)}{a_0 s_0}$. Similarly given $b_0 < b^*$ and shape parameter $a = a_0$, the risk of the James-Stein estimator will be uniformly less than the risk of $X$ for all values of the scale parameter, $0 < s < s_0$ when the shrinkage constant $c = \frac{2(p-2)}{a_0 s}$.

3.5 Conclusion.

For the large class of bounded concave loss functions which are functions of quadratic loss, $\|d - \theta\|^2$, given in Section 2, we establish conditions for which classes of Baranchik-type estimators are minimax and uniformly dominate the risk of $X$ for the Gaussian case. The condition for minimaxity on the shrinkage factor, $a$, are identical to those for quadratic loss. Extensions to when the scale is unknown and a residual vector is present, and to the case where the underlying density is a scale mixture with residual present are given as well for the same class of losses. Further investigation is needed in the case that the underlying density is a scale mixture without a residual vector present in order to determine conditions for which Baranchik-type estimators will be minimax for the classes of losses given in Section 2.
Chapter 4

Combining Unbiased and Possibly Biased Estimators

4.1 Introduction

This chapter considers estimation of a mean vector $\theta$ when both an unbiased estimator $X$ and a possibly biased estimator $Y$ are available. A convenient basic model is

$$
\begin{pmatrix}
X \\
Y
\end{pmatrix}
\sim
N_{2p}\left(\begin{pmatrix}
\theta \\
\theta + \eta
\end{pmatrix}, \Sigma\right)
$$

(4.1)

where $\Sigma$ is a known $2p \times 2p$ positive definite covariance matrix. Here $\text{dim}(X) = \text{dim}(Y) = \text{dim}(\theta) = \text{dim}(\eta) = p$. We consider estimation of $\theta$ under the loss function

$$
L(\theta, d) = (d - \theta)'Q(d - \theta) = \|d - \theta\|_Q
$$

(4.2)

where $Q$ is a symmetric positive definite matrix in (4.2). The usual estimator $\delta_0(X, Y) = X$ is minimax, minimum risk equivariant under the location group, best unbiased, and admissible if $p=1$ or 2. $X$ is however, inadmissible if $p \geq 3$ and we study the possibility of improving (uniformly) on $X$ by combining it with $Y$. 
The form of the combined estimator we consider is given by

$$
\delta(X, Y) = Z + (I_p - Br(\|X - Z\|^2_{Q^*})) (X - Z)
$$

(4.3)

where $B$ is a $p \times p$ nonsingular matrix, $Q^*$ is a $p \times p$ positive definite matrix, and $Z$ is a linear combination of $X$ and $Y$ (all to be specified).

Section 2 studies this normal model and gives conditions on $r(\cdot)$, $Q^*$, $B$, and $Z$ so that (4.3) dominates the estimator $\delta_0(X) = X$ under loss (4.2), when $p \geq 3$. The basic results extend those of Green and Strawderman [16] for the case that $\Sigma = \sigma^2 I_{2p}$, $Q^* = Q = I_p$, and $Z = Y$.

A main result of Section 2 is the following: Letting $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $B = Q^{-1} \Sigma_{11}^{-1}$, $Q^* = \Sigma_{11}^{-1} Q^{-1} \Sigma_{11}^{-1}$, and $Z = C(Y - \Sigma_{21} \Sigma_{11}^{-1} X)$ for $C$ an arbitrary non-singular $p \times p$ matrix leads to an improved estimator provided $r'(\cdot) \geq 0$ and $0 < r(\cdot) < 2(p - 2)$. If $r(t) \equiv a$, then $a = p - 2$ is the uniformly best choice. The choice $C = (I_p - \Sigma_{21} \Sigma_{11}^{-1})^{-1}$ is convenient in that the risk difference between $X$ and the estimator (4.3) will depend only on the bias $\eta$ of the estimator $Y$.

Section 3 extends results to certain non-normal models. Section 3.1 considers scale mixture of normal models, while section 3.2 considers elliptically symmetric models corresponding to spherical models originally considered by Berger [4], i.e.,

$$
\begin{pmatrix} X \\ Y \end{pmatrix} \sim f(\left\| \begin{pmatrix} x - \theta \\ y - (\theta + \eta) \end{pmatrix} \right\|^2_{\Sigma^{-1}})
$$

where, for $F(t) = \frac{1}{2} \int_t^\infty f(u) du$, $F(t)/f(t) \geq b > 0$. In each case suitable conditions are given so that the estimator (4.3) dominates $\delta_0(x) = X$ and is minimax.

Section 4 considers normal models where the covariance matrix is known up to a constant multiple and when there is a residual vector available to estimate scale. Section 5 generalizes the results of Section 4 to the case of a general elliptically symmetric distribu-
tion with a residual vector.

Section 6 gives a numerical study of the degree of risk improvements in certain settings, while section 7 gives some concluding remarks.

4.2 Normal Model with Known Covariance Matrix ($\Sigma$)

In this section we assume 

$$
\begin{pmatrix}
X \\
Y
\end{pmatrix}
\sim
N_{2p}(\begin{pmatrix}
\theta \\
\theta + \eta
\end{pmatrix},
\Sigma = 
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix})
$$

where $dim(X) = dim(Y) = p \geq 3$, $\Sigma$ known and positive definite, and the loss is given by 

$$
L_Q(\theta, d) = (d - \theta)'Q(d - \theta) = \|d - \theta\|_Q^2
$$

with $Q > 0$ and symmetric. Theorem 4.1 develops a class of estimators $\delta(X, Y)$ of $\theta$ that will be minimax under the loss $L_Q(\theta, d)$. More specifically the class of estimators considered will be of the form:

$$
\delta(X, Y) = Z + (I_p - \frac{Q^{-1}\Sigma_{11}^{-1}r(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2})(X - Z)
$$

(4.4)

for $Z = B(Y - AX)$ in (4.4), where the specific conditions on $Q^*$, $B$, $A$, and $r(\cdot)$ are given in Theorem 4.1 to ensure minimaxity of $\delta(X, Y)$. The estimators in (4.4) can be seen as a generalization of the class of estimators

$$
\delta_n(X, Y) = Y + (1 - \frac{a\sigma^2}{\|X - Y\|^2})(X - Y)
$$

(4.5)

studied in Green and Strawderman [16] which considers minimax estimation of $\theta$ under square error loss ($L_I(\theta, d)$) when $X$ and $Y$ are independent $p$-variate normal random vectors, with $E[X] = \theta$, $E[Y] = \theta + \eta$, and covariance matrices $\sigma^2 I_p$ and $\tau^2 I_p$ ($\sigma^2$, $\tau^2$ known) respectively.

The following Lemmas will be used to establish the finiteness of the risk for the class of estimators given by expression (4.4), and to establish a convenient form for the risk. Lemma
4.1 gives the form of Stein’s Lemma that will be used.

**Lemma 4.1.** Let $X \sim N_p(\theta, \Sigma)$ where $\Sigma$ is non singular. Let $h(x)$ be a weakly differentiable function from $\mathbb{R}^p \rightarrow \mathbb{R}^p$ with $E[\|h(X)\|^2] < \infty$, then:

$$E[\Sigma^{-1}(X - \theta)'h(X)] = E[\text{div}_x(h(X))].$$

(4.6)

Lemma 4.2 can be seen as a slight generalization of the bounds found for $E\left[\frac{1}{X'X}\right]$ when $Z \sim N_p(\theta, I_p)$, found in Green and Strawderman [16]. For estimators of the form (4.4), Lemma 4.2 will be used to establish upper and lower bounds for the risk.

**Lemma 4.2.** Let $X \sim N_p(\mu, \Sigma)$ with $\Sigma > 0$. Then $E_{\mu}\left[\frac{1}{X'X}\right] < \infty$ provided $p \geq 3$. Furthermore denoting $\Lambda$ as the diagonal matrix of eigenvectors of $\Sigma$ and $\lambda(p)$ and $\lambda(1)$ as the largest and smallest eigenvectors of $\Sigma$ respectively, for $p = 3$:

$$\max\left(\frac{1}{\text{tr}(\Lambda) + \mu'\mu}, \frac{1}{\lambda(p)}\left[\frac{1}{p-2+\mu'\Sigma^{-1}\mu}\right]\right) \leq E\left[\frac{1}{X'X}\right] \leq \frac{1}{\chi_1^2}\left[\frac{1}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)}\right],$$

and for $p \geq 4$:

$$\max\left(\frac{1}{\text{tr}(\Lambda) + \mu'\mu}, \frac{1}{\lambda(p)}\left[\frac{1}{p-2+\mu'\Sigma^{-1}\mu}\right]\right) \leq E\left[\frac{1}{X'X}\right] \leq \min\left(\frac{1}{\chi_1^2}\left[\frac{1}{p+4+\mu'\Sigma^{-1}\mu}\right], \frac{1}{\chi_1^2}\left[\frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)}\right]\right).$$

**Proof:** Let $Z \sim N_p(\mu, I_p)$, then Stapleton [26, Definition 2.5.1] characterizes the density of $Y = Z'Z \sim \chi^2_p(\mu'\mu)$ as

$$f(y|p, \mu') = \sum_{k=0}^{\infty} P_k\left(\frac{\mu'\mu}{2}\right) f_{p+2k}(y),$$

where $P_k$ is the poisson density, $e^{-\frac{\mu'\mu}{2}}\frac{\left(\frac{\mu'\mu}{2}\right)^k}{k!}$, and $f_{p+2k}$ is the density of a $\chi^2_{p+2k}$, $\frac{y^{p+2k-1}e^{-\frac{y}{2}}}{\Gamma\left(p+2k\right)2^{p+2k}}$, so that $Y$ is a Poisson mixture of central $\chi^2$ densities. Supposing that $p \geq 3$, by Tonelli’s theorem.
\[ E[\frac{1}{Y}] = \sum_{k=0}^{\infty} P_k(\frac{\mu' \mu}{2}) E[\frac{1}{\chi_{p+2k}^2}] = \]
\[ \sum_{k=0}^{\infty} P_k(\frac{\mu' \mu}{2}) \frac{1}{p+2k-2} = E_{\mu' \mu} \left[ \frac{1}{p+2k-2} \right] \tag{4.7} \]

where the expectation in (4.7) is taken with respect to a Poisson distribution with parameter \( \frac{\mu' \mu}{2} \). In order for the expectation of the random variable \( \frac{1}{Y} \) to exist, \( p \geq 3 \) is necessary since for \( k = 0 \), \( E[\frac{1}{\chi_p^2}] < \infty \) implies \( p \geq 3 \). Furthermore for \( p \geq 3 \) and \( l > 0 \),
\[
E[\frac{1}{\chi^2 + 2k}] < \frac{1}{p^2} \]
so that
\[
E[\frac{1}{Y}] < \sum_{k=0}^{\infty} P_k(\frac{\mu' \mu}{2}) [\frac{1}{p^2}] = \frac{1}{p^2}
\]
which is finite. By the convexity of the function \( g(k) = \frac{1}{p+2k-2} \), Jensen’s inequality implies
\[
\frac{1}{p-2 + \mu' \mu} = \frac{1}{p-2 + 2Ek} \leq E[\frac{1}{p+2k-2}] \leq E[\frac{1}{p-2}] = \frac{1}{p-2}. \tag{4.8}
\]

To get a more refined upper bound for \( E[\frac{1}{Z^2}] \), the result of Casella and Hwang [6] gives the upper bound
\[
E_{\mu' \mu} \left[ \frac{1}{Z^2} \right] \leq \frac{1}{p-2} \left( \frac{p+2}{p+2+\mu' \mu} \right) \tag{4.9}
\]
for \( p \geq 3 \). When \( p > 3 \) further refinement on the upper bound for the expectation in expression (4.9) are given in Green and Strawderman [16] by expressing the term in the expectation of expression (4.7) as
\[
\frac{1}{p+2k-2} = \frac{1}{p+2(k+1)-4} = \frac{1}{k+1+2}.
\]
By concavity of the function \( g(x) = \frac{x}{(p-4)x+2} \) (as long as \( p \geq 4 \)),
\[
E_{\mu' \mu} \left[ \frac{1}{p+2k-2} \right] \leq \frac{E[\frac{1}{k+1}]}{E[\frac{1}{k+1}+2]}. 
\]
Since
\[
E[\frac{1}{k+1}] = \sum_{k=0}^{\infty} \frac{1}{k+1} \frac{e^{-\mu' \mu} \frac{\mu' \mu}{x}^{k}}{k!} =
\]
\[
\frac{e^{-\mu'}}{\left(\mu'\right)^2} \sum_{k=1}^{\infty} \frac{\left(\mu'\right)^k}{k!} = \frac{2}{\mu'} \left(1 - e^{-\mu'}\right) \leq \frac{2}{\mu'}
\]

and since \(g(x)\) is monotonic increasing in \(x\),

\[
E[\frac{1}{p+2k-2}] \leq \frac{E[\frac{1}{k+1}]}{E[\frac{1}{k+1}]+2} \leq \frac{1}{p-4+\mu'\mu}.
\]

This yields the result

\[
E[\mu'\mu][\frac{1}{Z'Z}] \leq \frac{1}{p-4+\mu'\mu}.
\] (4.10)

Combining the bounds from expression (4.10) and (4.9), the following expression was established in Green and Strawderman (1991) [16] when \(p \geq 4\)

\[
\frac{1}{p-2+\mu'\mu} \leq E[\mu'\mu][\frac{1}{Z'Z}] \leq \min\left(\frac{p+2}{(p-2)(p+2+\mu'\mu)}\right) \frac{1}{p-4+\mu'\mu}.
\] (4.11)

Suppose now that \(X \sim N_{p}(\mu, \Sigma)\) and let \(Y = \Sigma^{-\frac{1}{2}}X \sim N_{p}(\Sigma^{-\frac{1}{2}}\mu, I_{p})\). For any orthogonal matrix \(U\) with rank \(p\), \(Z = U'Y \sim N_{p}(U'\Sigma^{-\frac{1}{2}}\mu, I_{p})\). By using the spectral value decomposition of \(\Sigma\) to get

\[P'\Sigma P = \Lambda = \text{Diag}(\lambda_1, \lambda_2, ..., \lambda_p)\]

where \(P\) is the orthogonal matrix of eigenvectors of \(\Sigma\) and \(\Lambda\) is the diagonal matrix of eigenvalue of \(\Sigma\), and denoting \(\lambda_1(1)\) and \(\lambda_1(p)\) as the largest and smallest eigenvalues of \(\Sigma\), the following inequalities exist:

\[\lambda_1(1)E[Z'Z] \leq E[X'X] = E[X'\Sigma^{-\frac{1}{2}}PP'\Sigma P'\Sigma^{-\frac{1}{2}}X] = E[Z'\Lambda Z] \leq \lambda_1(p)E[Z'Z]\]

with \(Z = P'\Sigma^{-\frac{1}{2}}X\). Noting that the function \(\phi(x) = \frac{1}{x}\) is convex for \(x > 0\),

\[\max\left(\frac{1}{\lambda_{p}}E[\frac{1}{Z'Z}], \frac{1}{\text{tr}(\Lambda)}+\mu'\mu\right) \leq E[\frac{1}{X'X}] = E[\frac{1}{X'\Sigma^{-\frac{1}{2}}X}] \leq \frac{1}{\lambda_1(1)}E[\frac{1}{Z'Z}] \leq \infty\]

provided \(p \geq 3\), where \(Z \sim N_{p}(P'\Sigma^{-\frac{1}{2}}\mu, I_{p})\).

From expression (4.11) then, when \(p \geq 4\)

\[\max\left(\frac{1}{\text{tr}(\Lambda)}+\mu'\mu, \frac{1}{\lambda_1(p)} \left(\frac{1}{p-2+\mu'\mu}\right)\right) \leq E[\frac{1}{X'X}] \leq \infty\] (4.12)
\[ \min \left( \frac{1}{\lambda(1)} \left[ \frac{1}{p-4+\mu'\Sigma^{-1}\mu} \right], \frac{1}{\lambda(1)} \left[ \frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)} \right] \right), \]  
\tag{4.13} 

and for \( p \geq 3 \)

\[ \max \left( \frac{1}{\lambda(p)+\mu'\mu}, \frac{1}{\lambda(p)} \left[ \frac{1}{p-2+\mu'\Sigma^{-1}\mu} \right] \right) \leq E \left[ \frac{1}{XX} \right] \leq \frac{1}{\lambda(1)} \left[ \frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)} \right]. \]

\[ \square \]

**Remark 1:**

For \( p \geq 4 \), \( \mu'\Sigma^{-1}\mu \geq \frac{p}{2} + 1 \) implies

\[ \min \left( \frac{1}{\lambda(1)} \left[ \frac{1}{p-4+\mu'\Sigma^{-1}\mu} \right], \frac{1}{\lambda(1)} \left[ \frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)} \right] \right) = \frac{1}{\lambda(1)} \left( \frac{1}{p-4+\mu'\Sigma^{-1}\mu} \right) \]

while \( \mu'\Sigma^{-1}\mu < \frac{p}{2} + 1 \) implies

\[ \min \left( \frac{1}{\lambda(1)} \left[ \frac{1}{p-4+\mu'\Sigma^{-1}\mu} \right], \frac{1}{\lambda(1)} \left[ \frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)} \right] \right) = \frac{1}{\lambda(1)} \left( \frac{p+2}{(p-2)(p+2+\mu'\Sigma^{-1}\mu)} \right). \]

**Remark 2:**

For \( p \geq 3 \), a necessary and sufficient condition for

\[ \max \left( \frac{1}{\lambda(p)+\mu'\mu}, \frac{1}{\lambda(p)} \left[ \frac{1}{p-2+\mu'\Sigma^{-1}\mu} \right] \right) = \frac{1}{\lambda(p)} \left( \frac{1}{p-2+\mu'\Sigma^{-1}\mu} \right) \]

is \( \lambda(p)(p-2) + \lambda(p)\mu'\Sigma^{-1}\mu > \text{tr}(\Lambda) + \mu'\mu \). This will be satisfied if \( \lambda(p)(p-2) > \text{tr}(\Lambda) \).

Lemma 4.3 will be used to get a convenient expression for the risk of the proposed estimator in order to apply Lemma 4.1.

**Lemma 4.3.** Let \( g : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}^p \) and let \( \delta \) be of the form \( \delta(X,Y) = X + g(X,Y) \). Let \( Q > 0 \) be a symmetric \( pxp \) matrix. Then

\[ L_Q(\delta,\theta) = \text{Tr}(Q(X - \theta)(X - \theta)' + g'(X,Y)Qg(X,Y) + 2(X - \theta)'Qg(X,Y)). \]

**Proof:** For ease of notation set \( g = g(X,Y) \). Expanding \( L_Q(\delta,\theta) \) yields

\[ (X + g - \theta)'Q(X + g) = ((X - \theta) + g)'Q((X - \theta) + g) = \text{tr}(((X - \theta) + g)'Q((X - \theta) + g)) = \]
where (4.14) follows since $a^t Q b = b^t Q a$ for symmetric $Q$ and $a, b$ vectors in $\mathbb{R}^p$.

Lemma 4.4 is used as a means to calculate the divergence term for the proposed estimator (4.4). A version of the statement can be found in Mardia, Kent, and Bibby [21, Appendix A.9].

**Lemma 4.4.** For a symmetric positive definite matrix $A$,

$$\frac{d}{dx} x' Ax = \nabla_x x' Ax = 2Ax.$$

**Proof.**

Denoting the $(i, j)^{th}$ element of $A$ as $a_{ij}$ the expansion of $x'Ax$ is

$$\sum_{j=1}^{p} \sum_{k=1}^{p} a_{kj} x_k x_j = \sum_{k=1}^{p} a_{ki} x_k x_i + \sum_{j \neq i}^{p} \sum_{k=1}^{p} a_{kj} x_k x_j$$

so that

$$\frac{\partial}{\partial x_i} x' Ax = \frac{\partial}{\partial x_i} \left[ \sum_{k=1}^{p} a_{ki} x_k x_i + \sum_{j \neq i}^{p} \sum_{k=1}^{p} a_{kj} x_k x_j \right] =$$

$$2a_{ii} x_i + \sum_{k \neq i} a_{ki} x_k + \sum_{j \neq i} a_{ij} x_j =$$

$$2a_{ii} x_i + \sum_{k \neq i} a_{ki} x_k + \sum_{j \neq i} a_{ij} x_j =$$

$$2a_{ii} x_i + \sum_{k \neq i} a_{ki} x_k + \sum_{j \neq i} a_{ij} x_j =$$

$$2a_{ii} x_i + \sum_{j \neq i} a_{ij} x_j + \sum_{j \neq i} a_{ij} x_j = 2 \sum_{i=1}^{p} a_{ij} x_j$$

where the symmetry of $A$ was used in the equivalence of equations (4.15) and (4.16). The result follows since the $i^{th}$ entry in $2Ax$ is $2 \sum_{j=1}^{p} a_{ij} x_j = \frac{\partial}{\partial x_i} x' Ax$. 
The following Theorem presents the main result of this Section. It will establish the conditions necessary so that estimators of the form (4.4) will be minimax.

**Theorem 4.1.** Let
\[ \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \]
for \( p > 2 \) and \( \Sigma_{11} > 0 \), and \( l_Q(\delta, \theta) = (\delta - \theta)^\prime Q(\delta - \theta) \) for \( Q \) a symmetric positive definite matrix. Let
\[ \delta(X, Y) = X + g(X, Y) = Z + (I - \frac{Q^{-1}\Sigma_{11}^{-1}r(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2})(X - Z) \]
be an estimator for \( \theta \) where \( Z = B(Y - AX) \), \( A = \Sigma_{21}\Sigma_{11}^{-1}, |B| \neq 0 \), and \( Q^* = \Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1} \).

A sufficient condition for \( \delta(X, Y) \) to be minimax is

i) \( 0 \leq r(t) \leq 2(p - 2) \) for \( t \geq 0 \),

ii) \( r(t) \) is non-decreasing in \( t \) for \( t > 0 \),

with strict domination over \( X \) provided the inequalities in i) are strict or \( r \) is strictly increasing on a set of positive measure.

**Proof.** Let \( Z = B(Y - AX) \). The estimator \( \delta(X, Y) \) is of the form
\[ \delta(X, Y) = X + g(X, Y) \]
with
\[ g(X, Y) = -\frac{Q^{-1}\Sigma_{11}^{-1}r(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2}(X - Z) \]
(4.17)
so that
\[ g'Qg = \frac{r^2(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2}, \]
(4.18)
and
\[ 2(X - \theta)'Qg = -2(\Sigma_{11}^{-1}(X - \theta))'r(\|X - Z\|_Q^2)(X - Z) \]
(4.19)

From Lemma 4.3 and expressions (4.18) and (4.19), the difference in risk between the estimator \( \delta \) and \( X \), \( \Delta(\delta, X, \theta, \eta) \), is
\[ E[\frac{r^2(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2} - 2(\Sigma_{11}^{-1}(X - \theta))'r(\|X - Z\|_Q^2)(X - Z)] \]
(4.20)
A sufficient condition for $\triangle(\delta, X, \theta, \eta)$ to exist is $p \geq 3$ since by assumption i)

$$E[r^2(\|X-B(Y-AX)\|^2_{Q})_{\|\cdot\|_{Q^*}}] \leq 4(p-2)^2E[\frac{1}{\|X-B(Y-AX)\|^2_{Q^*}}] =$$

$$4(p-2)^2E[\frac{1}{W'W}] = 4(p-2)^2E[\frac{1}{V'V}] < \infty$$  \hspace{1cm} (4.21)

by Lemma 4.2 as V is a p variate normal distribution where the change of variables $W = X - B(Y - AX)$ and $V = Q^*\hat{\xi}W$ was used in (4.21). Furthermore by the Cauchy Schwarz inequality and the transformation $W = X - B(Y - AX)$, the other term in (4.20) has finite expected value as

$$E[(\Sigma_{11}^{-1}(X - \theta)']r\frac{\|X-B(Y-AX)\|^2_{Q^*}}{\|X-B(Y-AX)\|_{Q^*}}(X - B(Y - AX))] \leq$$

$$E^{\frac{1}{2}}[(X - \theta)\Sigma_{11}^{-1}(X - \theta)]E^{\frac{1}{2}}[\frac{a^2(W'Q^*W)(W'W)}{(W'Q^*W)^2}] \leq$$

$$E^{\frac{1}{2}}[(X - \theta)\Sigma_{11}^{-1}(X - \theta)]E^{\frac{1}{2}}[\frac{(4(p-2)^2\nu(\rho))}{(W'Q^*W)}] < \infty$$

where $\nu(\rho)$ is the largest eigenvalue of $Q^*$. Since each term in $\triangle(\delta, X, \theta, \eta)$ exist a sufficient condition for the risk of the $\delta(X, Y)$ to exist is $p \geq 3$.

Conditioning on the random variable $Z = B(Y - AX)$ gives $\triangle(\delta, X, \theta, \eta)$ as

$$E[E[\frac{r^2(\|X - \theta_0\|^2_{Q^*})}{\|X - \theta_0\|^2_{Q^*}} - 2(\Sigma_{11}^{-1}(X - \theta)')r(\|X - \theta_0\|^2_{Q^*})(X - \theta_0)) | Z = \theta_0]] \hspace{1cm} (4.22)$$

Setting $h_{\theta_0}(X) = \frac{r(\|X - \theta_0\|^2_{Q^*})(X - \theta_0)}{\|X - \theta_0\|^2_{Q^*}}$ inside of (4.22) and noting that since $A = \Sigma_{21}\Sigma_{11}^{-1}$, $Cov(X, Y - AX) = 0$ implies the independence of $X$, by the weak differentiability of $h_{\theta_0}(x)$ Lemma 4.1 implies

$$E[E[\Sigma_{11}^{-1}(X - \theta)'h_{\theta_0}(X) | Z = \theta_0]] =$$

$$E[E[div(h_{\theta_0}(X)) | Z = \theta_0]]$$

Expression (4.22) can therefore be expressed as

$$E[E[\frac{r^2(\|X - \theta_0\|^2_{Q^*})}{\|X - \theta_0\|^2_{Q^*}} - 2div_x(\frac{r(\|X - \theta_0\|^2_{Q^*})(X - \theta_0)}{\|X - \theta_0\|^2_{Q^*}})) | Z = \theta_0]] \hspace{1cm} (4.23)$$
Setting \( q^*_{ij} \) as the \((i, j)^{th}\) entry in \( Q^* \) and noting that

\[
\frac{\partial}{\partial x_i} r(\|x - \theta_0\|_{Q^*}^2)(x_i - \theta_{0i}) \left\| x - \theta_0 \right\|_{Q^*}^2 = (\text{by an application of Lemma 4.4 on } \|x - \theta_0\|_{Q^*}^2)
\]

\[
\|x - \theta_0\|_{Q^*}^2 [2r(\|x - \theta_0\|_{Q^*}^2) \sum_{j=1}^p q^*_{ij} (x_j - \theta_{0j})(x_i - \theta_{0i}) + r(\|x - \theta_0\|_{Q^*}^2)] - \frac{2r(\|x - \theta_0\|_{Q^*}^2) \sum_{j=1}^p q^*_{ij} (x_j - \theta_{0j})(x_i - \theta_{0i})}{(\|x - \theta_0\|_{Q^*}^2)^2}
\]

the conditional expectation in expression (4.23) can be rewritten after substitution of the divergence term in (4.23) as

\[
E\left[ -\frac{r^2(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} - 2\frac{(p - 2)r(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} + 2r(\|X - \theta_0\|_{Q^*}^2) \right] \leq (4.24)
\]

(by assumption ii) \( E\left[ -\frac{r^2(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} - 2\frac{(p - 2)r(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} \right] =
\]

\[
E\left[ -\frac{r^2(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} - 2\frac{(p - 2)r(\|X - \theta_0\|_{Q^*}^2)}{\|X - \theta_0\|_{Q^*}^2} \right] \leq 0
\]

(by assumption i). Thus \( \Delta(\delta, X, \theta, \eta) \leq 0 \) upon taking the expected value of (4.24) establishing the result.

\[
\square
\]

The following Corollary is a direct consequence of Theorem 4.1. It will provide an easily analyzable class of minimax estimators of the form given in (4.4) for which bounds on the risk can be developed.

**Corollary 4.1.** Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \) for \( p > 2 \) where \( \Sigma_{11} > 0 \). Let \( l_Q(\delta, \theta) = (\delta - \theta)'Q(\delta - \theta) \) for \( Q \) a symmetric positive definite matrix. For a non-singular matrix, \( B \), and \( A = \Sigma_{21} \Sigma_{11}^{-1} \), let
\[ \delta(X, Y) = X + g(X, Y) = \]
\[ B(Y - AX) + (I - \frac{aQ^{-1}\Sigma_{11}^{-1}}{(X - B(Y - AX))'\Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1}(X - B(Y - AX))})[X - B(Y - AX)] \]

be an estimator for \( \theta \). Then a sufficient condition for \( \delta(X, Y) \) to be minimax is \( a \in [0, 2(p - 2)] \).

**Proof.** The result is immediate from Theorem 4.1 by setting \( r(t) = a \). Since \( r(t) \) is non-decreasing, non-negative, and bounded by \( 2(p - 2) \) the result follows.

\[ \square \]

For estimators of the form given in Corollary 4.1 the following Theorem presents bounds that can be easily established using Lemma 4.2. This can serve as a means of assessing how much savings in risk can be accomplished by using the proposed combination estimator.

**Theorem 4.2.** Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \), \( \delta(X, Y) \), and \( L_Q(\delta, \theta) \) be as in Corollary 4.1 with \( p > 2 \) and \( a \in [0, 2(p - 2)] \). Let \( Q^* = \Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1} \), \( \Lambda \) as the diagonal matrix of eigenvalues of \( [\Sigma_{11} + B(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})B']Q^* \) with largest and smallest eigenvalues \( \lambda_{(p)} \) and \( \lambda_{(1)} \) respectively,

\[ \mu_{\theta\eta} = [I - B(I - \Sigma_{21}\Sigma_{11}^{-1})]\theta - B\eta, \]

and,

\[ Q_1(\mu_{\theta\eta}) = \mu_{\theta\eta}'[\Sigma_{11} + B(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})B']^{-1}\mu_{\theta\eta}. \]

For \( p = 3 \)

\[ \text{tr}(Q\Sigma_{11}) - \frac{a(2(p - 2) - a)}{\lambda_{(1)}} \frac{p+2}{(p-2)(p+2)+Q_1(\mu_{\theta\eta})} \leq R(\delta, \theta, \eta) \leq \]

\[ \text{tr}(Q\Sigma_{11}) - [a(2(p - 2) - a)]\max\left\{ \frac{1}{\text{tr}(\Lambda)'[\mu_{\theta\eta}]Q^*(\mu_{\theta\eta})} \lambda_{(p)}^{-1} \frac{1}{p-2+Q_1(\mu_{\theta\eta})} \right\}. \]

For \( p \geq 4 \) the following inequality holds:

\[ \text{tr}(Q\Sigma_{11}) - \frac{a(2(p - 2) - a)}{\lambda_{(1)}} \min\left\{ \frac{p+2}{(p-2)(p+2)+Q_1(\mu_{\theta\eta})}, \frac{1}{p-4+Q_1(\mu_{\theta\eta})} \right\} \leq \]
\[ R(\delta, \theta, \eta) \leq \]
\[ \text{tr}(Q\Sigma_{11}) - [a(2(p - 2) - a)] \max \left\{ \frac{1}{\text{tr}(\Lambda + (\mu_\eta)Q^*(\mu_\eta))}, \frac{1}{\lambda_{(p)} [\mu_{-2+Q_1(\mu_\eta)}]} \right\} \]

**Proof.** Let \( Z = B(YAX) \). Using expression (4.23) the risk of the estimator \( \delta \), \( R(\delta, \theta, \eta) \) can be expressed as

\[ R(X, \theta) = E[|X - \theta_0|_Q^2(\|(X - \theta_0)\|_Q^2)] = \] (4.25)

\[ \text{tr}(Q\Sigma_{11}) + E[|X - \theta_0|_Q^2(\|(X - \theta_0)\|_Q^2)]Z = \theta_0] = \] (4.26)

\[ \text{tr}(Q\Sigma_{11}) - a(2(p - 2) - a)E[ \frac{1}{\|X - Z\|_Q^2}] = \] (4.27)

Since \( Q > 0 \) and symmetric, \( Q^* = \Sigma^{-11}Q^{-1}\Sigma_{11}^{-1} \) has a symmetric square root denoted by \( Q^*_1 \). Let

\[ \Sigma^* = \Sigma_{11} + B(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})B', \] (4.28)

Making the change of variables

\[ W = X - B(Y - AX) \sim N_p(\mu_\eta, \Sigma^*), \] (4.29)

and,

\[ V = Q^*_1W \sim N_p(Q^*_1\mu_\eta, Q^*_1\Sigma^*Q^*_1) \] (4.30)

in (4.27) implies

\[ R(\delta, \theta, \eta) = \text{tr}(Q\Sigma_{11}) - a(2(p - 2) - a)E[ \frac{1}{V^2}]\] (4.31)

Since \( V \) has a multivariate normal distribution, whose parameters are given in (4.30), an application of Lemma 4.2 to \( E[ \frac{1}{V^2}] \) in (4.31) implies the result since any eigenvalue of \( Q^*_1 [\Sigma_{11} + B(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})B']Q^*_1 \) is also an eigenvalue of \( [\Sigma_{11} + B(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})B']Q^* \).
Remark 1: From (4.31) the uniformly optimal choice of a is \( a = p - 2 \) as \( E\left[\frac{1}{V^T V}\right] > 0 \).

Remark 2: Setting \( B = (I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} \) implies that \( \triangle(\delta, X, \theta, \eta) \) depends only on the parameter \( \eta \). This is a consequence of expression (4.31) since \( E\left[\frac{1}{V^T V}\right] \) depends only on the mean of the random variable \( V \), which will be \(-(I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} \eta \).

We conclude the section by providing examples of shrinkage combination estimators for certain covariance structures and for certain quadratic loss.

Example 4.1. (Green and Strawderman [16])

In this example we recreate the result found in [16] using the generalized Theorem 1.

Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} \sigma^2 I & 0 \\ 0 & \tau^2 I \end{pmatrix}) \) with \( p \geq 3 \). Let \( L_1(d, \theta) = (d - \theta)'(d - \theta) \). From Theorem 4.1 any estimator of the form

\[
\delta(X, Y) = BY + (I_p - \frac{1}{\sigma^2}I_p \frac{r(\|X - BY\|^2)}{\|X - BY\|^2} \frac{1}{\sigma^4}I_p)(X - BY)
\]

with \( B \) a non-singular matrix, \( 0 \leq r(t) \leq 2(p - 2) \) and \( r(t) \) non decreasing for \( t > 0 \) will be a minimax estimator for \( \theta \) under the loss \( L_1 \). Setting \( B = I \) and \( r(t) = p - 2 \) yields the following minimax estimator for \( \theta \):

\[
\delta_1(X, Y) = Y + (1 - \frac{\sigma^2(p - 2)}{\sqrt{\|X - Y\|^2}})(X - Y).
\]

From theorem 4.2 the bounds on the risk for \( p = 3 \) satisfy

\[
p\sigma^2 - \frac{\sigma^4(p - 2)(p + 2)}{(\sigma^2 + \tau^2)(p + 2) + \eta' \eta} \leq E[(\delta_1(X, Y) - \theta)'(\delta_1(X, Y) - \theta)] \leq
\]

\[
p\sigma^2 - \frac{\sigma^4(p - 2)^2}{(\sigma^2 + \tau^2)(p - 2) + \eta' \eta}.
\]

For \( p \geq 4 \) the bounds become

\[
p\sigma^2 - (p - 2)^2 \sigma^4 \min \left\{ \frac{(p + 2)}{(p - 2)(p + 2) + \eta' \eta}, \frac{1}{p - 4 + \frac{\eta' \eta}{\sigma^2 + \tau^2}} \right\} \leq E[(\delta_1(X, Y) - \theta)'(\delta_1(X, Y) - \theta)] \leq
\]

\[
p\sigma^2 - \frac{(p - 2)^2 \sigma^4}{(\sigma^2 + \tau^2)} \min \left\{ \frac{(p + 2)}{(p - 2)(p + 2) + \eta' \eta}, \frac{1}{p - 4 + \frac{\eta' \eta}{\sigma^2 + \tau^2}} \right\}.
\]
\[ p\sigma^2 - \frac{\sigma^4(p-2)^2}{(\sigma^2 + \tau^2)(p-2)^2 + \eta \bar{\gamma}}. \]

The positive-part version of estimator \( \delta_1(X,Y) \) given by:
\[ \delta_1^+(X,Y) = Y + \max \{ 0, (1 - \frac{\sigma^2(p-2)}{\|X-Y\|^2}) \} (X - Y) = Y + (1 - \frac{\sigma^2 r(\frac{1}{\sigma^2} \|X-Y\|^2)}{\|X-Y\|^2})(X - Y) \]

with
\[ r(t) = \begin{cases} t\sigma^2 & 0 < t < \frac{(p-2)}{\sigma^2} \\ (p-2) & t \geq \frac{(p-2)}{\sigma^2} \end{cases} \]

will also be minimax since \( r(t) \) is a monotonic increasing function bounded between 0 and \( 2(p-2) \). Furthermore since the conditional distribution of \( X \) given \( Y \) has density \( f \) that is symmetric, unimodal, and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, and the estimator \( \delta_1(X,Y) \) has \( i^{th} \) coordinate of the form:
\[ \delta_{1i}(x,y) = y_i + (1 - h(x-y))(x_i - y_i) \]

where \( h(t) \) is a symmetric function in \( t \), the risk of \( \delta_1^+(X,Y) \) will dominate the risk of \( \delta_1(X,Y) \) under the the loss \( L_I(d,\theta) \).

Example 4.2. \textit{(Normal model with diagonal covariance)}

Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p} \left( \begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} \right) \) where \( D_1 = \text{diag}(\sigma_1^2, ..., \sigma_p^2) \) and \( D_2 = \text{diag}(\tau_1^2, ..., \tau_p^2) \) with \( p \geq 3 \). Consider the loss functions \( L_I(d,\theta) = (d - \theta)'(d - \theta) \) and \( L_{D_1^{-1}}(d,\theta) = (d - \theta)'D_1^{-1}(d - \theta) \). For the loss \( L_I(d,\theta) \), Theorem 4.1 states that any estimator of the form
\[ \delta(X,Y) = BY + (I_p - \frac{D_1^{-1}r(\|X-BY\|^2)}{\|X-BY\|^2})(X - BY) \]

will be minimax provided \( B \) is non-singular, \( 0 \leq r(t) \leq 2(p-2) \), and \( r(t) \) is non-decreasing for \( t \geq 0 \). Setting \( B = I_p \) and \( r(t) = p - 2 \) for \( t \geq 0 \) results in the following minimax estimator for \( \theta \):
\[ \delta_1(X,Y) = Y + (I_p - \frac{D_1^{-1}(p-2)}{\|X-Y\|^2})(X - Y) \]

whose \( i^{th} \) component is given by
$$\delta_{1i} = y_i + (1 - \frac{(p-2)\sigma_i^2}{\sum_{i=1}^{p} \frac{(2i-1)\sigma_i^2}{\sigma_i^2}})(x_i - y_i).$$

Let \( D^* = \text{diag}(\frac{\sigma_1^2+\tau_1^2}{\sigma_1^2}, \ldots, \frac{\sigma_p^2+\tau_p^2}{\sigma_p^2}) = [D_1 + D_2]D_1^{-2} \), with \( d^*_p \) and \( d^*_p \) as the largest and smallest eigenvalues of \( D^* \) respectively. Theorem 4.2 implies if \( p = 3 \) the risk of the estimator \( \delta_1(X,Y) \) satisfies

$$\sum_{i=1}^{p} \sigma_i^2 - (p-2)\frac{\min\{ \frac{p+2}{(p-2)((p+2)+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2})} \}}{\sum_{i=1}^{p} \frac{\sigma_i^2+\tau_i^2}{\sigma_i^2}} \leq E[(\delta_1(X,Y) - \theta)'(\delta_1(X,Y) - \theta)] \leq \sum_{i=1}^{p} \sigma_i^2 - (p-2)^2\frac{\max\{ \frac{p+2}{p-2+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}} \}}{\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}}.$$  

If \( p \geq 4 \) the following bounds for the risk apply:

$$\sum_{i=1}^{p} \sigma_i^2 - (p-2)\frac{\min\{ \frac{p+2}{p-2+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}} \}}{p-4+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}} \leq E[(\delta_1(X,Y) - \theta)'(\delta_1(X,Y) - \theta)] \leq \sum_{i=1}^{p} \sigma_i^2 - (p-2)^2\frac{\max\{ \frac{p+2}{p-2+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}} \}}{p-4+\sum_{i=1}^{p} \frac{\eta_i^2}{\sigma_i^2+\tau_i^2}}.$$  

Since the conditional distribution of \( X \) given \( Y \) has a density \( f \) that is symmetric, unimodal, and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, and the estimator \( \delta_1(X,Y) \) has \( i^{th} \) coordinate of the form:

$$\delta_{1i}(x,y) = y_i + (1 - h_i(x - y))(x_i - y_i)$$

where \( h_i(t) \) is symmetric in \( t \) for all \( i \), the positive part version of \( \delta_1(X) \), i.e. \( \delta_1^+(X) \) whose \( i^{th} \) coordinate is given by:

$$\delta_{1i}^+ = y_i + \max\{0, (1 - \frac{(p-2)\sigma_i^2}{\sum_{i=1}^{p} \frac{(2i-1)\sigma_i^2}{\sigma_i^2}})(x_i - y_i)\}$$

will dominate the risk of \( \delta_1(X) \) and hence will be a minimax estimator.

For the loss function \( L_{D_1^{-1}}(d,\theta) \), Theorem 4.1 implies any estimator of the form:

$$\delta(X,Y) = BY + (I - \frac{r(\|X-BY\|^2_{D_1^{-1}})}{\|X-BY\|^2_{D_1^{-1}}}) (X - BY) = BY + (1 - \frac{r(\|X-BY\|^2_{D_1^{-1}})}{\|X-BY\|^2_{D_1^{-1}}}) (X - BY)$$
will be minimax provided B is a non singular matrix, 0 ≤ r(t) ≤ 2(p − 2) and r(t)
non-decreasing for t > 0. Setting B = I_p and a(t) = p − 2 yields the estimator

\[ \delta_2(X, Y) = Y + (1 - \frac{(p-2)}{\sum_{i=1}^{p} \frac{\|x_i - y_i\|}{\sigma_i^2}})(X - Y). \]

Let \( D = \text{diag}(\frac{\sigma_1^2 + \tau_1^2}{\sigma_1^2}, \ldots, \frac{\sigma_p^2 + \tau_p^2}{\sigma_p^2}) = [D_1 + D_2]D_1^{-1} \), with \( d(p) \) and \( d(1) \) as the largest and
smallest eigenvalues of \( D \). According to Theorem 4.2, for \( p = 3 \) the bounds for the risk of estimator \( \delta_2(X, Y) \) satisfy

\[ p - \frac{(p-2)^2}{d(1)} \left[ \frac{p+2}{(p-2)[(p+2)]} + \frac{1}{\sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}} \right] \leq E[(\delta_2(X, Y) - \theta)'D^{-1}(\delta_2(X, Y) - \theta)] \leq \]

\[ p - (p - 2)^2 \max \left\{ \frac{1}{\sum_{i=1}^{p} \frac{\sigma_i^2 + \tau_i^2 + \sigma_i^2}{\sigma_i^2}}, \frac{d(p)}{p-2 + \sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}} \right\} \]

For \( p \geq 4 \) the following bounds for the risk of estimator \( \delta_2 \) apply:

\[ p - \frac{(p-2)^2}{d(1)} \min \left\{ \frac{p+2}{(p-2)[(p+2)]} + \frac{1}{\sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}} \right\} \leq E[(\delta_2(X, Y) - \theta)'D_1^{-1}(\delta_2(X, Y) - \theta)] \leq \]

\[ p - (p - 2)^2 \max \left\{ \frac{1}{\sum_{i=1}^{p} \frac{\sigma_i^2 + \tau_i^2 + \sigma_i^2}{\sigma_i^2}}, \frac{d(p)}{p-2 + \sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma_i^2 + \tau_i^2}} \right\}. \]

The positive-part version of \( \delta_2(X, Y) \) given by

\[ \delta_2^+(X, Y) = Y + \max \left\{ 0, (1 - \frac{(p-2)}{\sum_{i=1}^{p} \frac{\|x_i - y_i\|}{\sigma_i^2}}) \right\}(X - Y) = \]

\[ Y + (1 - \frac{r(\|X - Y\|_{D_1^{-1}}^2)}{\|X - Y\|_{D_1^{-1}}^2})(X - Y) \]

with

\[ r(t) = \begin{cases} 
  t & 0 < t < p - 2 \\
  p - 2 & t \geq p - 2 
\end{cases} \]

will be minimax as well since \( r(t) \) is bounded by \( 2(p-2) \) and monotonic increasing by Theo-
rem 4.1. Furthermore since the conditional distribution of \( X \) given \( Y \) has density \( f \) that
is symmetric, unimodal, and non-increasing in each of the coordinates separately for each
fixed value of the other coordinates, and the estimator \( \delta_2(X, Y) \) has its \( i^{th} \) coordinate of the form
\[ \delta_2(x, y)_i = y_i + (1 - h(x - y))(x_i - y_i) \]

where \( h(t) \) is symmetric in \( t \), the risk of \( \delta_2^+(X, Y) \) will dominate that of \( \delta_2(X, Y) \).

**Example 4.3.** (Normal model with covariance \( \Sigma \))

Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}) \). Suppose the loss function is of the form \( L_{\Sigma_{11}^{-1}}(d, \theta) \).

Theorem 4.1 states that any estimator of the form

\[ \delta(X, Y) = Z + (1 - \frac{r(\|X - Z\|_{\Sigma_{11}^{-1}}^{p-2})}{\|X - Z\|_{\Sigma_{11}^{-1}}^{2}})(X - Z) \]

where \( Z = B(Y - \Sigma_{21}\Sigma_{11}^{-1}X) \) with \( B \) non-singular, and \( 0 \leq r(t) \leq 2(p-2) \) and non-decreasing in \( t \) will be a minimax estimator for \( \theta \). Setting \( r(t) = p - 2 \) and \( B = (I_p - \Sigma_{21}\Sigma_{11}^{-1})^{-1} \) yields the following minimax estimator for \( \theta \):

\[ \delta_1(X, Y) = Z + (1 + \frac{p-2}{\|X - Z\|_{\Sigma_{11}^{-1}}^{2}})(X - Z). \]

with \( Z = (I_p - \Sigma_{21}\Sigma_{11}^{-1})^{-1}(Y - \Sigma_{21}\Sigma_{11}^{-1}X) \). Setting \( \Lambda \) as the diagonal matrix of eigenvalues of

\[ \left[ \Sigma_{11} + (I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})(I - \Sigma_{21}\Sigma_{11}^{-1})^{-1} \right] \Sigma_{11}^{-1} \]

with maximum and minimum eigenvalues \( \lambda_{(p)} \) and \( \lambda_{(1)} \) respectively,

\[ Q_1(\eta) = \eta'((I - \Sigma_{21}\Sigma_{11}^{-1})^{-1})'\Sigma_{11}^{-1}(I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}\eta, \]

and

\[ Q_2(\eta) = \eta'\Sigma^*\eta, \]

where

\[ \Sigma^* = B'(\Sigma_{11} + (I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})(I - \Sigma_{21}\Sigma_{11}^{-1})^{-1})'^{-1}B, \]

with \( B = (I - \Sigma_{21}\Sigma_{11}^{-1})^{-1} \), and applying Theorem 4.2 yields the following bounds for the risk of \( \delta_1(X, Y) \) if \( p = 3 \):

\[ p - \frac{(p-2)^2}{\lambda_{(1)}} \left[ \frac{p+2}{(p-2)[p+2+Q_2(\eta)]} \right] \leq \]

\[ E[(\delta_1(X, Y) - \theta)'\Sigma_{11}^{-1}(\delta_1(X, Y) - \theta)] \leq \]
\[ p - (p - 2)^2 \max\left\{ \frac{1}{\text{tr}(\Lambda) + Q_1(\eta)}, \frac{1}{\lambda(p)} \right\} \].

For \( p \geq 4 \) the following bounds apply:

\[ p - \frac{(p-2)^2}{\lambda(1)} \min\{ \frac{p+2}{(p-2)(p+2) + Q_2(\eta)}, \frac{1}{p-4 + Q_2(\eta)} \} \leq E[(\delta_1(X,Y) - \theta)^\prime \Sigma_{11}^{-1} (\delta_1(X,Y) - \theta)] \leq p - (p - 2)^2 \max\left\{ \frac{1}{\text{tr}(\Lambda) + Q_1(\eta)}, \frac{1}{\lambda(p)} \right\} . \]

The positive-part version of \( \delta_1(X,Y), \delta_1^+(X,Y) \) given by

\[ Z + \max\{0, 1 - \frac{p-2}{\|X-Z\|_{\Sigma_{11}^{-1}}}\}(X - Z) = \]

\[ Z + (1 - \frac{r(\|X-Z\|_{\Sigma_{11}^{-1}})}{\|X-Z\|_{\Sigma_{11}^{-1}}})(X - Z) \]

where

\[ r(t) = \begin{cases} t & 0 \leq t \leq p - 2 \\ p - 2 & t \geq p - 2 \end{cases} . \]

will be a minimax estimator of \( \theta \) as well since \( r(t) \) is monotonic increasing function in \( t \) and \( 0 \leq a(t) \leq 2(p - 2) \) for \( t \geq 0 \). Making the change of variables \( W = \Sigma_{11}^{-\frac{1}{2}} X \), the risk of the estimator \( \delta_1(X) \) satisfies

\[ E[(\delta_1 - \theta)^\prime \Sigma_{11}^{-1} (\delta_1 - \theta)] = E[|1 - \frac{(p-2)}{\|W - \Sigma_{11}^{-\frac{1}{2}} \theta_0\|^2}|(W - \Sigma_{11}^{-\frac{1}{2}} \theta_0) - \Sigma_{11}^{-\frac{1}{2}} (\theta - \theta_0)|^2 |Z = \theta_0|] . \]  (4.32)

Since the random variable \( W \) is independent of the random variable \( Z \), the distribution of \( W \) has a density \( f \), which is symmetric, unimodal, and non-increasing in each of the coordinates separately for each fixed value of the other coordinates, and the \( i^{th} \) coordinate of

\[ \delta_1^*(W, \theta_0) = (1 - \frac{(p-2)}{\|W - \Sigma_{11}^{-\frac{1}{2}} \theta_0\|^2})(W - \Sigma_{11}^{-\frac{1}{2}} \theta_0) \]

is of the from

\[ \delta_1^*(W, \theta_0)_i = (1 - \frac{p-2}{h(W - \Sigma_{11}^{-\frac{1}{2}} \theta_0)})(w_i - \theta_0^*_i) \]

where \( \theta_0^*_i \) is the \( i^{th} \) coordinate of \( \Sigma_{11}^{-\frac{1}{2}} \theta_0 \) with \( h(t) \) symmetric in \( t \), the positive-part version of \( \delta_1^* \), \( \delta_1^{*+} \) will have a smaller conditional expected value in (4.32) than \( \delta \), and hence the positive-part estimator \( \delta_1^+ \) will dominate \( \delta_1 \).
4.3 Non Normal Models

4.3.1 Scale Mixtures of Normals

In this section we extended the result of Section 2 to scale mixtures of multivariate normal distributions and then to a more general class of elliptically symmetric distributions. Section 3.1 considers the case of scale mixtures of normal distributions. The results in Section 3.1 can be viewed as an extension of the results found in Strawderman [29] that studied classes of minimax estimators under square error loss for scale mixtures of normal distributions. In Section 3.2 we will extend the result of Section 3.1 to a wider class of elliptically symmetric distributions using the techniques found in Fourdrinier and Strawderman [10]. For both Sections 3.1 and 3.2, the estimators considered will be of the form:

\[ \delta(X, Y) = X - \frac{Q^{-1} \Sigma_{11}^{-1} \sigma r(\|X - Z\|_Q^2)}{\|X - Z\|_Q^2} (X - Z) \]  

(4.33)

where

\[ Z = B(Y - AX) \]

with B non-singular and \( A = \Sigma_2 \Sigma_{11}^{-1} \), and

\[ Q^* = \Sigma_{11}^{-1} Q^{-1} \Sigma_{11}^{-1}. \]  

(4.34)

Specific attention will be paid to conditions and bounds for \( r(t) \) and \( a \) to ensure minimaxity. As a consequence, just as in Section 2, \( \delta(X, Y) \) will improve the estimation of the mean by using the auxiliary information when compared to the estimator \( \delta_0(X) = X \). While the result in Section 3.2 applies to more general distributions, the bounds for minimaxity attained will be tighter in Section 3.1, however the class of functions for \( r(t) \) will be slightly more general in Section 3.2. A direct comparison can be found in Example 5 for the case of scale mixtures of multivariate normal distributions. To that end Lemma 4.5 will be used to establish a correlation inequality that will be used to bound \( \Delta(X, \delta, \theta, \eta) \).

**Lemma 4.5.** Let \( X \sim N_p(\mu, \sigma^2 \Sigma) \), then \( \frac{\|X\|_Q^2}{\sigma^2} \), where \( Q \) is a symmetric positive definite matrix, is stochastically decreasing in \( \sigma^2 \).
Proof.

Let $Y = \frac{1}{\sigma} \Sigma^{-\frac{1}{2}} X$, and $Z = P'Y$ where $P'\Sigma^{\frac{1}{2}} Q \Sigma^{\frac{1}{2}} P = \Lambda$ where $\Lambda$ is the diagonal matrix of eigenvalues and $P$ is the associated column matrix of eigenvectors of $\Sigma^{\frac{1}{2}} Q \Sigma^{\frac{1}{2}}$. Then $Y \sim N_p(\frac{1}{\sigma} \Sigma^{-\frac{1}{2}} \mu, I)$ and $Z \sim N_p(\frac{1}{\sigma} P' \Sigma^{-\frac{1}{2}} \mu, I)$ so that:

$$E[\frac{X'QX}{\sigma^2}] = E[Y'\Sigma^{\frac{1}{2}} Q \Sigma^{\frac{1}{2}} Y] = E[(P'Y) P' \Sigma^{\frac{1}{2}} Q \Sigma^{\frac{1}{2}} P (P'Y)] =$$

$$E[Z'\Lambda Z] = E[\sum_{i=1}^p \lambda_i Z_i^2]$$

Now $\{Z_i\}_{i=1}^p$ is an independent collection of random variables where each

$$Z_i \sim N(\frac{\mu_i^2}{\sigma^2}, 1)$$

where $\mu^* = P' \Sigma^{-\frac{1}{2}} \mu$, and thus $\{Z_i^2\}_{i=1}^p$ is a collection of independent random variables with

$$Z_i^2 \sim \chi_1^2(\frac{\mu_i^2}{\sigma^2} = \nu_i).$$

Since a non-central chi-squared random variable has a monotone likelihood ratio [20, Problem 7.4], $\chi_1^2(\nu_i)$ is stochastically increasing in the parameter $\nu_i$ [20, Lemma 3.4.2], and thus is stochastically decreasing in $\sigma^2$. Let

$$U(z_1^2, ..., z_p^2) = \sum_{i=1}^p \lambda_i z_i^2$$

Since $U$ is increasing in each of its coordinates, $\{Z_i^2\}_{i=1}^p$ is independent collection of random variables, and each $Z_i^2$ is stochastically decreasing in $\sigma^2$, $U(Z_1^2, Z_2^2, ..., Z_p^2) = \sum_{i=1}^p \lambda_i Z_i^2$ is stochastically decreasing in $\sigma^2$ establishing the result.

Theorem 4.3 will provide the first of the generalizations of Theorem 1 to a larger class of densities. The result is presented below.

Theorem 4.3. Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \int \left[ 2\pi \sigma^2 \Sigma^* \right]^{-\frac{1}{2}} e^{-\frac{1}{\sigma^2} \left( Z - \mu \right)' \Sigma^*^{-1} \left( Z - \mu \right)} dF(\sigma)$$
with \( \Sigma^* = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \), \( Z = \begin{pmatrix} X \\ Y \end{pmatrix} \), and \( \mu = \begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix} \), where \( F(\cdot) \) is a known c.d.f. on \((0, \infty) \), \( p > 3 \) and loss function \( L_Q(\delta, \theta) = (d - \theta)'Q(d - \theta) \) with \( Q \) symmetric and positive definite. Then estimators of the form

\[
\delta(X, Y) = X + g(X, Y) = X - \frac{Q^{-1}\Sigma^{-1}_{11}\text{ar}(\|X - B(Y - AX)\|^2_{Q^*})}{\|X - B(Y - AX)\|^2_{Q^*}} (X - B(Y - AX))
\]

where \( A = \Sigma_{21}\Sigma^{-1}_{11} \) and \( B \) is non-singular, and \( Q^* = \Sigma^{-1}_{11}Q^{-1}\Sigma^{-1}_{11} \) are minimax provided

1. \( 0 \leq r(y) \leq 1 \) for all \( y \),

2. \( \frac{r(y)}{y} \) is non-increasing in \( y \),

3. \( r(y) \) is non-decreasing in \( y \),

4. \( 0 \leq a \leq \frac{2(p-2)}{E[\sigma^2]} \)

where the expectation in iv) is taken with respect to the distribution of \( \sigma \). If either iv) is strict or ii) is strict on a set of positive measure, \( \delta(X, Y) \) dominates \( X \).

**Proof.** Let \( Z = B(Y - AX) \). By Lemma 4.3 the difference between the estimator \( \delta(X, Y) \) and the estimator \( \delta(X) = X \) is

\[
\Delta(\delta, X, \theta, \eta) = E[g'Qg + 2(x - \theta)'Qg] =
\]

\[
E\left[\frac{a^2r^2(\|X - Z\|^2_{Q^*})}{\|X - Z\|^2_{Q^*}} - 2(\Sigma^{-1}_{11}(X - \theta))'\text{ar}(\|X - Z\|^2_{Q^*})(X - Z)\right] =
\]

\[
E\left[\frac{a^2r^2(\|X - Z\|^2_{Q^*})}{\|X - Z\|^2_{Q^*}} - 2(\Sigma^{-1}_{11}(X - \theta))'\text{ar}(\|X - Z\|^2_{Q^*})(X - Z)\right]|\sigma].
\]

Since

\[
E[(\Sigma^{-1}_{11}(X - \theta))'\text{ar}(\|X - Z\|^2_{Q^*})(X - Z)]|\sigma] =
\]

\[
E[E[(\Sigma^{-1}_{11}(X - \theta))'\text{ar}(\|X - Z\|^2_{Q^*})(X - Z)]|Z = \theta_0, \sigma]|\sigma] =
\]
\[ E[E[\sigma^2 \text{div}_X(h_{\theta_0}(X))]|Z = \theta_0, \sigma|\sigma] \tag{4.35} \]

by Lemma 4.1, where the function

\[ h_{\theta_0}(X) = \frac{\ar(||X-\theta_0||_{Q^*}^2)}{||X-\theta_0||_{Q^*}^2} (X - \theta_0) \]

in (4.35) and noting that

\[ \frac{\partial}{\partial x_i} \frac{\ar(||x-\theta_0||_{Q^*}^2)(x_i - \theta_{0i})}{||x-\theta_0||_{Q^*}^2} = \]

\[ \frac{\|x-\theta_0\|^2_{Q^*}[2\ar(||x-\theta_0||_{Q^*}^2) \sum_{j=1}^p q_j^* (x_j - \theta_{0j})(x_i - \theta_{0i}) + \ar(||x-\theta_0||_{Q^*}^2)]}{(\|x-\theta_0\|^2_{Q^*})^2} - \frac{2\ar(||x-\theta_0||_{Q^*}^2) \sum_{j=1}^p q_j^* (x_j - \theta_{0j})(x_i - \theta_{0i})}{(\|x-\theta_0\|^2_{Q^*})^2} \]

so that

\[ \text{div}_X(h_\theta(X)) = \left( \frac{(p-2)\ar(||x-\theta_0||_{Q^*}^2)}{||x-\theta_0||_{Q^*}^2} \right) + 2\ar(||x - \theta_0||_{Q^*}^2), \]

\( \Delta(\delta, X, \theta, \eta) \) is expressible as

\[ E[E[\frac{a^2\ar^2(||X-Z||_{Q^*}^2)}{||X-Z||_{Q^*}^2} - \sigma^2 \frac{2(p-2)\ar(||X-Z||_{Q^*}^2)}{||X-Z||_{Q^*}^2} - 4\sigma^2 \ar(||X-Z||_{Q^*}^2)|\sigma]] \leq \]

(by assumption iii)

\[ E[E[\frac{a^2\ar^2(||X-Z||_{Q^*}^2)}{||X-Z||_{Q^*}^2}] - \sigma^2 \frac{2(p-2)\ar(||X-Z||_{Q^*}^2)}{||X-Z||_{Q^*}^2}|\sigma]), \tag{4.36} \]

Using the change of variables \( W = X - B(Y - AX) = X - Z \) for the inner expectation of (4.36), expression (4.36) satisfies

\[ E[E[\frac{a^2\ar^2(||W||_{Q^*}^2)}{||W||_{Q^*}^2} - \sigma^2 \frac{2(p-2)\ar(||W||_{Q^*}^2)}{||W||_{Q^*}^2}|\sigma]] \leq \]

(by assumption i)

\[ E[E[\frac{a^2\ar(||W||_{Q^*}^2)}{||W||_{Q^*}^2}] - \sigma^2 \frac{2(p-2)\ar(||W||_{Q^*}^2)}{||W||_{Q^*}^2}|\sigma]] = \]

\[ E[E[\frac{a^2\ar(||W||_{Q^*}^2)\sigma^2}{||W||_{Q^*}^2} (\frac{a}{\sigma^2} - 2(p - 2))|\sigma]] = \]

\[ E[(\frac{a}{\sigma^2} - 2(p - 2))E[\frac{a^2\ar(||W||_{Q^*}^2)\sigma^2}{||W||_{Q^*}^2}|\sigma]]. \]
Provided \( E[\frac{ar(||W||^2_{Q^*})\sigma^2}{||W||^2_{Q^*}}|\sigma] \) is monotonic increasing as a function of \( \sigma \), the correlation inequality implies
\[
E[(\frac{a}{\sigma^2} - 2(p-2))E[\frac{ar(||W||^2_{Q^*})\sigma^2}{||W||^2_{Q^*}}|\sigma]] \leq
\]
(as \((\frac{a}{\sigma^2} - 2(p-2)) \) is a monotonic decreasing function of \( \sigma \))
\[
E[\frac{a}{\sigma^2} - 2(p-2)]E[E[\frac{ar(||W||^2_{Q^*})\sigma^2}{||W||^2_{Q^*}}|\sigma]] \leq
\]
(by assumption iv)
\[
E[\frac{2(p-2)}{E[\frac{a}{\sigma^2}]} \frac{1}{\sigma^2} - 2(p-2)]E[\frac{ar(||W||^2_{Q^*})\sigma^2}{||W||^2_{Q^*}}|\sigma] \leq 0.
\]
To establish that \( E[\frac{ar(||W||^2_{Q^*})}{\sigma^2} | \sigma] \) is monotonic increasing, for \( \sigma_1 \leq \sigma_2 \)
\[
E[\frac{ar(\frac{\sigma_2^2||W||^2_{Q^*}}{\sigma_2^2})}{\sigma_1^2} | \sigma_1] \leq
\]
(by assumption iii)
\[
E[\frac{ar(\frac{\sigma_1^2||W||^2_{Q^*}}{\sigma_1^2})}{\sigma_1^2} | \sigma_1] \leq
\]
\[
E[\frac{ar(\frac{\sigma_2^2||W||^2_{Q^*}}{\sigma_2^2})}{\sigma_2^2} | \sigma_2]
\]
since \( g(y) = \frac{r(y)}{y} \) is a decreasing function of \( y \) by assumption ii), and \( \frac{||W||^2_{Q^*}}{\sigma^2} \) is a stochastically decreasing function of \( \sigma^2 \) by Lemma 4.5 since \( W|\sigma \sim N_p(\mu^*, \sigma^2\Sigma^*) \).
where the density of the random variable \( V \) is
\[
f(v) = 
\frac{v^{-k/2 - 1}e^{-k/2 \frac{v}{2} \pi}}{\Gamma(\frac{k}{2})}.
\]

Setting \( t = \|X - \theta, Y - (\theta + \eta)\|_2^{-1}, \) the marginal distribution of \( X, Y \) has the form:
\[
|2\pi\Sigma|^{-\frac{1}{2}} \int_0^\infty v^{-p}e^{-\frac{t^2}{2}}f(v)dv =
\frac{k^{\frac{p}{2}}}{2^{\frac{p}{2}}|2\pi\Sigma|^\frac{1}{2} \Gamma(\frac{k}{2})} \int_0^\infty v^{-(p+k/2)-1}e^{-\frac{1}{2} \frac{k}{2} \frac{v}{2} \pi}dv =
\frac{\Gamma(p+k/2)2^{p+k/2}}{(t+k)^{p+k/2}} \cdot \frac{k^{\frac{p}{2}}}{2^{\frac{p}{2}}|2\pi\Sigma|^\frac{1}{2} \Gamma(\frac{k}{2})} =
\frac{\Gamma(p+k/2)(1+\frac{k}{2})^{-(p+k/2)}}{k^{p}|\pi\Sigma|^\frac{1}{2} \Gamma(\frac{k}{2})}
\]
which is a multivariate \( t \) distribution with \( k \) degrees of freedom Arslan [1].

Theorem 4.3 allows for the calculation of the shrinkage constant, \( a \), so that estimators of the form (4.33) will be minimax estimator under the general quadratic loss \( L_Q(d, \theta) \) by providing bounds for \( a \) in terms of the expected value of the the random variable \( V^{-1} \). For \( A > 0 \)
\[
E[V^{-A}] = \frac{k^{\frac{p}{2}}}{2^{\frac{p}{2}} \Gamma(\frac{k}{2})} \cdot \frac{\Gamma(A+k/2)2^{A+k/2}}{k^{A+k/2}} =
\frac{\Gamma(A+k/2)2^A}{\Gamma(\frac{k}{2})K^A}
\]
so that by Theorem 4.3 the estimator
\[
\delta_1(X, Y) = X - \frac{Q^{-1}\Sigma_{11}^{-1}a_1}{\|X - B(Y - \Sigma_{21}\Sigma_{11}^{-1}X)\|_Q^2}(X - B(Y - \Sigma_{21}\Sigma_{11}^{-1}X)) \quad (4.37)
\]
where \( B \) is non-singular and \( Q^* = \Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1} \) in (4.37) will be minimax for loss \( L_Q(d, \theta) \) once
\[
0 \leq a_1 \leq \frac{2(p-2)}{E[V^{-1}]} = 2(p-2)
\]

Remark: The positive part version of \( \delta, \delta^+_1 \), dominates \( \delta_1 \) since mixtures of normals have symmetric unimodal marginals.
4.3.2 Berger Class

We wish to extend our estimators of the location vector \( \Theta \) to the class of elliptically symmetric distributions with density 
\[
    f \left( \sqrt{\frac{(x-\theta)^2 + (y-(\theta+\eta))^2}{\Sigma^{-1}}} \right)
\]
such that 
\[
    \frac{F(t)}{f(t)} \geq b > 0
\]
for all \( t \geq 0 \), where
\[
    F(t) = \frac{1}{2} \int_{t}^{\infty} f(x) dx.
\]

Minimax shrinkage estimators for densities of this form were first studied in Berger [4].

We note that the class of functions which satisfy (4.38) include ones for which \( f(t) \) is log convex and have a second derivative with respect to \( t \), e.g. the class of densities which satisfy
\[
    \frac{d^2}{dt^2} \log f(t) \geq 0
\]
for all \( t \geq 0 \). This follows since if \( \frac{f'(t)}{f(t)} \) is non-decreasing then
\[
    \frac{f(t)}{F(t)} = 2 \int_{t}^{\infty} \frac{f'(u)}{f(u)} f(u) du = 2E_t \left[ -\frac{f'(u)}{f(u)} \right]
\]
where the expected value is taken with respect to the density proportional to \( f(u) I_{[u \geq t]}(u) \).

This density has an increasing monotone likelihood ratio for the parameter \( t \), since for \( t_2 > t_1 \) the ratio of the densities is 0 for \( u \in [t_1, t_2) \) and 1 for \( u \geq t_2 \). This implies in particular that:
\[
    \frac{f(t_2)}{F(t_2)} = 2E_{t_2} \left[ -\frac{f'(u)}{f(u)} \right] \leq 2E_{t_1} \left[ -\frac{f'(u)}{f(u)} \right] = \frac{f(t_1)}{F(t_1)}
\]
since \( h(u) = \frac{-f'(u)}{f(u)} \) is non-increasing and hence
\[
    \frac{F(t_1)}{f(t_1)} \leq \frac{F(t_2)}{f(t_2)}.
\]

As a result for \( t \geq 0 \),
\[
    \frac{F(t)}{f(t)} \geq \frac{F(0)}{f(0)} = b
\]
which will satisfy (4.38) provided \( b = \frac{F(0)}{f(0)} > 0. \)

The class of distributions includes scale mixture of normal’s and hence the results in this subsection apply to a wider class of distributions than in section 3.1. The estimators studied in this section will be of the form

\[
\delta(X, Z) = X + Q^{-1}\Sigma_{11}^{-1}bg(X, Z)
\]

where

\[
Z = B(Y - AX)
\]

for \( B \) a non singular matrix and \( Z \) a linear function uncorrelated with \( X \). Sufficient conditions on the function \( g(X, Z) \) for minimaxity will be given in Theorem 4.4.

4. The method of analysis used will be similar to the results found in Strawderman and Fourdiernier [10]. A Stein-like differential inequality given in Lemma 4.6 will be used to bound the difference in risk between \( X \) and the class of estimators considered in (4.41).

**Lemma 4.6.** Let \( \begin{pmatrix} X \\ Y \end{pmatrix} \) belong to a distribution with density \( f(t) \) for \( t = \| x - \theta \|_{\Sigma_{11}^{-1}}^2 + \| y - (\theta + \eta) \|_{\Sigma_{22}^{-1}}^2 \). Suppose all second moments exist, \( E[\| g(X, Y) \|^2] < \infty \), and \( g(x, y) \) is a weakly differentiable function in \( x \) for all \( y \). Then

\[
E[(X - \theta)'\Sigma_{11}^{-1}g(X, Y)] = E[\text{div}_x (g(X, Y)F(\| x - \theta \|_{\Sigma_{11}^{-1}}^2 + \| y - (\theta + \eta) \|_{\Sigma_{22}^{-1}}^2))]
\]

where \( F(t) = \int_t^\infty f(t)dt \).

**Proof** Let \( \sigma_{ij}^* \) denote the \((i, j)^{th}\) entry of \( \Sigma_{11}^{-1} \). Then

\[
E[(X - \theta)'\Sigma_{11}^{-1}g(X, Y)] =
\]

\[
\int [\Sigma_{j=1}^p \int [\Sigma_{j=1}^p \sigma_{ij}^*(x_j - \theta_j)g_i(x, y)] f(\| x - \theta \|_{\Sigma_{11}^{-1}}^2 + \| y - (\theta + \eta) \|_{\Sigma_{22}^{-1}}^2)dx]dy =
\]

\[
\int [\Sigma_{j=1}^p \int g_i(x, y) - \frac{\partial}{\partial x_1}(\frac{1}{2} \int_0^\infty f(\| x - \theta \|_{\Sigma_{11}^{-1}}^2 + \| y - (\theta + \eta) \|_{\Sigma_{22}^{-1}}^2)du)dx]dy =
\]
\[
\int \sum_{i=1}^{p} \int \frac{\partial}{\partial x_i} g_i(x,y) F(t) dx dy \tag{4.42}
\]
where \( t = \|x - \theta\|_{\Sigma_1}^2 + \|y - (\theta + \eta)\|_{\Sigma_2}^2 \). Upon dividing and multiplying (4.42) by \( f(t) \)

\[
\int \int \text{div}_x (g(x,y) \frac{F(t)}{f(t)} f(t) dx dy = E[\text{div}_x (g(X,Y) \frac{F(\|x - \theta\|_{\Sigma_1}^2 + \|y - (\theta + \eta)\|_{\Sigma_2}^2)}{f(\|x - \theta\|_{\Sigma_1}^2 + \|y - (\theta + \eta)\|_{\Sigma_2}^2)}] \tag{4.43}
\]
establishing the result.

The following Theorem gives sufficient conditions for estimators of the form (4.41) to be minimax.

**Theorem 4.4.** Let the loss be given by \( L_Q(d,\theta) = \|d - \theta\|_{Q}^2 \) with \( Q \) symmetric and positive definite, and \( \begin{pmatrix} X \\ Y \end{pmatrix} \) be random variables with joint density

\[
f(\begin{pmatrix} X - \theta \\ Y - (\theta + \eta) \end{pmatrix})' \left( \Sigma^{-1} \right) \begin{pmatrix} X - \theta \\ Y - (\theta + \eta) \end{pmatrix}
\]

where \( \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \), such that \( \begin{pmatrix} X \\ Y \end{pmatrix} \) belongs to the class of distributions that satisfy (4.38) with

\[
\frac{F(t)}{f(t)} \geq b > 0 \text{ for all } t \geq 0,
\]
and where all second moments exist. Denoting the entries of \( \Sigma^{-1} \) as

\[
\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}
\]
and letting

\[
Y^* = B(Y - (\Sigma_{22})^{-1} \Sigma_{21} X) = B(Y - \Sigma_{21} \Sigma_{11}^{-1} X)
\]
for \( B \) non-singular, the estimator

\[
\delta(X,Y^*) = X + Q^{-1}\Sigma_{11}^* bg(X,Y^*)
\]

(4.44)

where

\[
\Sigma_{11}^* = \Sigma_{11} - \Sigma_{12}(\Sigma_{22}^{-1}\Sigma_{12})^{-1} = \Sigma_{11}^{-1},
\]

\[E[\|g(X,Y^*)\|^2] < \infty, \text{ and } g(x,y) \text{ is weakly differentiable in } x \text{ for all } y, \text{ is minimax for } \theta \]

provided

\[i) \|g(x,y^*)\|_{Q^*} + 2\text{div}_x(g(x,y^*)) \leq 0 \text{ for } (x,y) \text{ a.e.}\]

where \( Q^* = \Sigma_{11}^*Q^{-1}\Sigma_{11} = \Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1}. \)

**Proof.** Consider the change of variables

\[
A^* \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} I & 0 \\ -BA & B \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X \\ Y^* \end{pmatrix}.
\]

Let

\[
W = \begin{pmatrix} X \\ Y^* \end{pmatrix}
\]

and

\[
\mu = \begin{pmatrix} \theta \\ B(I - A)\theta + B\eta \end{pmatrix}
\]

. Since \( A^{-1} = \begin{pmatrix} I & 0 \\ A & B^{-1} \end{pmatrix} \), by the standard change of variable formula the density of

\[
\begin{pmatrix} X \\ Y^* \end{pmatrix}, f_{(X,Y^*)}(X,Y^*), \text{ has the form}
\]

\[
f((W - \mu)' \begin{pmatrix} I & A' \\ 0 & B^{-1} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & 0 \\ A & B^{-1} \end{pmatrix} (W - \mu)) ||B^{-1}||. \quad (4.45)
\]

Setting \( A = -\Sigma_{22}^{-1}\Sigma_{21} \) implies (4.45) is equivalent to
where $\Sigma_{22} = (B^{-1})^\prime \Sigma_{22} B^{-1}$. Therefore by Lemma 4.6

$$
\triangle(\delta, X, \theta) = E[b^2 g'(X, Y^*) \Sigma_{11}^{-1} \Sigma_{11}^* g(X, Y^*) + 2b(X - \theta)^\prime \Sigma_{11}^* g(X, Y^*)] = E[b^2 g'(X, Y^*) \Sigma_{11}^{-1} \Sigma_{11}^* g(X, Y^*) + 2b \text{div}_x(g(X, Y^*))] \frac{F(t)}{f(t)} \leq (4.46)
$$

where $t = \|X - \theta\|_{\Sigma_{11}}^2 + \|Y^* - (B(I - A)\theta + B\eta)\|_{\Sigma_{22}}^2$ in (4.46). Expression (4.47) follows from (4.46) from the assumption that $\|g(X, Y^*)\|_{\Sigma_{11}^{-1} \Sigma_{11}}^2 + 2\text{div}_x(g(x, y^*)) \leq 0$ since $\|g(X, Y^*)\|_{\Sigma_{11}^{-1} \Sigma_{11}}^2 \geq 0(Q > 0)$ and the assumption that $\frac{F(t)}{f(t)} \geq b > 0$.

We conclude the section by giving an example that compares the shrinkage constants obtained from Theorem 4.3, to those of Theorem 4.4 when the underlying distribution is a scale mixture of a normal.

**Example 4.5.** Let $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, v \Sigma)$ with $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ known, $\theta$ unknown, $p \geq 3$, and $v$ having a distribution with density $h(v)$ and $E[V^{-1}]$ finite. Let the loss function be of the form $L_Q(d, \theta) = (d - \theta)^\prime Q(d - \theta)$ with $Q$ symmetric and positive definite. According to Theorem 4.3 the estimator:

$$
\delta_1(X, Y') = X - \frac{Q^{-1} \Sigma_{11}^{-1} a_1}{\|X - Y'\|^2} (X - Y')
$$

(4.48)

where $Y' = (I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} (Y - \Sigma_{21} \Sigma_{11}^{-1} X)$, $Q^* = \Sigma_{11}^{-1} Q^{-1} \Sigma_{11}^{-1}$, and

$$
0 \leq a_1 \leq \frac{2(p - 2)}{E[V^{-1}]}
$$

(4.49)
will be minimax for \( \theta \) under loss, \( L_Q(d, \theta) \) provided \((I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}\) is non-singular.

As a comparison, Theorem 4.4 implies that estimators of the form:

\[
\delta(X, Y') = X - Q^{-1}\Sigma_{11}^{-1}bg(X, Y')
\]  

(4.50)

where

\[
Y' = (I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}(Y - \Sigma_{21}\Sigma_{11}^{-1}X)
\]

will be minimax provided

i) \( \|g(x, y')\|_Q^2 + 2\text{div}_x(g(x, y')) \leq 0 \) for all \( (x, y') \) where \( Q^* = \Sigma_{11}^{-1}Q^{-1}\Sigma_{11}^{-1} \) with

\[
E[\|g(X, Y')\|^2] \leq \infty.
\]

ii) The joint density of \( \begin{pmatrix} X \\ Y \end{pmatrix} \) is in the class of distribution which satisfy (4.38).

Setting

\[
g(X, Y') = \frac{a_2}{\|X - Y'\|_Q^2}(X - Y')
\]  

(4.51)

will satisfy condition i) provided \( 0 \leq a_2 \leq 2(p - 2) \) and \( p \geq 3 \). The marginal distribution of \( \begin{pmatrix} X \\ Y \end{pmatrix} \) also satisfies condition ii). To see this, let \( t = \| \begin{pmatrix} X - \theta \\ Y - (\theta + \eta) \end{pmatrix} \|_\Sigma^{-1} \) so that the marginal distribution of \( \begin{pmatrix} X \\ Y \end{pmatrix} \) is of the form:

\[
f(x, y) = \int_0^\infty 2\pi v \Sigma^{-1/2} e^{-\frac{1}{2}v} h(v) dv = \int_0^\infty v^{-p} e^{-\frac{1}{2}v} h(v) dv.
\]

For \( t_2 > t_1 \)

\[
\frac{v^{-p} e^{-t_2/2} h(v)}{v^{-p} e^{-t_1/2} h(v)} = e^{-\frac{t_2 - t_1}{2p}}
\]

which is increasing in \( v \) and hence the family of distributions with densities proportional to
\( j_t(v) = v^{-p}e^{-\frac{1}{2\pi}h(v)} \)

will have increasing monotone likelihood ratio in \( t \). Since,

\[
\frac{f'(t)}{f(t)} = -\frac{1}{2} \frac{\int_0^\infty v e^{-\frac{1}{2\pi}v h(v)} dv}{\int_0^\infty v^{-p}e^{-\frac{1}{2\pi}v h(v)} dv} = -\frac{1}{2} E_t[V^{-1}]
\]

where the expected value is taken with respect to the density \( K_j t(V) \) with normalizing constant \( K \), \( \frac{f'(t)}{f(t)} \) is a non-decreasing function of \( t \). Therefore for \( t_1 < t_2 \)

\[
E_{t_1}[\frac{1}{2} V^{-1}] \leq E_{t_2}[\frac{1}{2} V^{-1}].
\]

This implies that

\[
\frac{F(0)}{f(0)} = \frac{\frac{1}{2} \int_0^\infty v e^{-\frac{1}{2\pi}v h(v)} dv}{\int_0^\infty v^{-p}e^{-\frac{1}{2\pi}v h(v)} dv} = \frac{\int_0^\infty v^{-p+1}h(v) dv}{\int_0^\infty v^{-p}h(v) dv} = E[V^{-p+1}] \leq \frac{F(t)}{f(t)}
\]

for all \( t \geq 0 \), condition ii) is satisfied for

\[
b = \frac{F(0)}{f(0)} \tag{4.52}
\]

in (4.40). Therefore Theorem 4.4 implies estimator of the form

\[
\delta_2(X, Y') = X - \frac{Q^{-1} \Sigma^{-1} \Sigma_1 a_2^*}{\|X - Y'\|_{Q^*}^2} (X - Y') \tag{4.53}
\]

are minimax when

\[
0 \leq a_2^* \leq \frac{2(p - 2)E[V^{-p+1}]}{E[V^{-p}]} \tag{4.54}
\]

In order to compare the shrinkage constants \( a_1 \) of (4.48), and \( a_2^* \) of (4.53), note by the correlation inequality

\[
E[V^{-p}] = E[V^{-p+1}V^{-1}] \geq E[V^{-p+1}]E[V^{-1}]
\]

so that

\[
0 \leq \frac{E[V^{-p+1}]}{E[V^{-p}]} \leq \frac{1}{E[V^{-1}]}
\]
and thus the maximal value for $a_1$ of (4.48) will be greater than the maximal value for $a_2^*$ of (4.53) for minimaxity.

As a specific example, Let $V$ to be the mixing distribution given in Example 4 i.e.,

$$V \sim I.G.(k/2, k/2)$$

with density

$$f(v) = \frac{v^{\frac{k}{2} - 1} e^{-\frac{1}{2} (\frac{k}{2})^2}}{\Gamma(\frac{k}{2})}$$

Example 4 gives the condition that

$$0 \leq a_1 \leq \frac{2(p-2)}{E[V^{-1}]} = 2(p - 2)$$

for minimaxity of the estimator (4.48). Theorem 4.4 gives the conditions

$$0 \leq a_2^* \leq \frac{2(p-2)E[V^{-p+1}]}{E[V^{-p}]} = 2(p - 2) \cdot \frac{\Gamma(p + \frac{k}{2} - 1)2^{p-1}}{\Gamma(p + \frac{k}{2} - 1 + 1)2^p} = 2(p - 2) \cdot \frac{k}{k + 2(p-1)}$$

for minimaxity of the estimator (4.53) since for $A > 0$:

$$E[V^{-A}] = \frac{k^{\frac{k}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \cdot \frac{\Gamma(A + \frac{k}{2})2^{A + \frac{k}{2}}}{k^{A + \frac{k}{2}}} = \frac{\Gamma(A + \frac{k}{2})2^A}{\Gamma(\frac{k}{2})k^A}.$$  

Hence for small values of $k$ the maximum $a_2^*$ will be much smaller then the maximal value for $a_1$. For large $k$, they will be comparable.

### 4.4 Normal Model with Covariance Matrix Known Up To a Scale

In this section we extend the results of the previous section to the case where the underlying density is a normal distribution whose covariance matrix is known up to a scale. We require that there be a residual vector $U$ to estimate the unknown scale $\sigma$ with $U$ uncorrelated with

$$\begin{pmatrix} X \\ Y \end{pmatrix}.$$  

More specifically we assume the following model for the underlying distribution:
\[
\begin{pmatrix}
X \\
Y \\
U
\end{pmatrix}
\sim N_{p+p+k}(\begin{pmatrix}
\theta \\
\theta + \eta \\
0
\end{pmatrix}, \sigma^2 \Sigma^*) \tag{4.55}
\]

where

\[
\Sigma^* = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & 0_{pxk} \\
\Sigma_{21} & \Sigma_{22} & 0_{pxk} \\
0_{kxp} & 0_{kxp} & \Sigma_{33}
\end{pmatrix}
\tag{4.56}
\]

\(\sigma^2\) unknown, and \(\Sigma^*\) known. The loss is

\[
l(\theta, \sigma^2, d) = \frac{(d - \theta)'Q(d - \theta)}{\sigma^2} \tag{4.57}
\]

for \(Q > 0\) and symmetric. To ensure minimaxity we will require \(\text{dim}(X) = \text{dim}(Y) = p \geq 3\), and \(\text{dim}(U) = k > 1\). The method of extension makes use of a Stein-like differential equality given in Lemma 4.7, so that a straightforward comparison can be made between the risk of the estimator \(\delta(X) = X\) and the proposed minimax estimator. Similar techniques used to extended classes of minimax estimators from the normal case to the general spherically symmetric case can be found in Fourdinier, Strawderman, and Wells [11]. We proceed with a Lemma giving the exact form of the Stein-like differential equality used. Without loss of generality we assume \(\Sigma_{33} = I_{kxk}\).

**Lemma 4.7.** Let

\[
\begin{pmatrix}
X \\
Y \\
U
\end{pmatrix}
\sim N_{2p+k}(\begin{pmatrix}
\theta \\
\theta + \eta \\
0
\end{pmatrix}, \sigma^2 \Sigma^*) \text{ where } \Sigma^* = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & 0_{pxk} \\
\Sigma_{21} & \Sigma_{22} & 0_{pxk} \\
0_{kxp} & 0_{kxp} & \Sigma_{33}
\end{pmatrix}
\]

Let \(S = \|U\|^2\) and let \(h : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^+ \to \mathbb{R}\) be a function such that

i) \(E[h(X, Y, S)] < \infty\),

ii) \(u \frac{h(x, y, ||u||^2)}{||u||^2}\) is weakly differentiable in \(u\) for every \(x, y\),

then

\[
E[\frac{h(X, Y, S)}{\sigma^2}] = E[(k - 2)\frac{h(X, Y, S)}{S} + 2 \frac{\partial}{\partial S} h(X, Y, S)].
\]
Proof. Since \( \begin{pmatrix} X \\ Y \end{pmatrix} \) is independent of \( U \), the density of \( \begin{pmatrix} X \\ Y \\ U \end{pmatrix} \), \( f(x, y, u) \), can be represented as

\[
 f(x, y, u) = f_2(x, y)(2\pi\sigma^2)^{-k} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{k} u_i^2}
\]

so that

\[
 E\left[ \frac{h(X,Y,S)}{\sigma^2} \right] = E\left[ \frac{U' \cdot h(X,Y,S)}{\sigma^2 \|U\|^2} \right] = \\
 \int_{\mathbb{R}^2} \int_{\mathbb{R}^k} \frac{g_{x,y}(u)}{\sigma^2} (2\pi\sigma^2)^{-k} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{k} u_i^2} du f_2(x,y) dxdy,
\]

where

\[
 g_{x,y}(u) = \frac{U \cdot h(x,y,\|u\|^2)}{\|u\|^2},
\]

with \( i^{th} \) coordinate given by

\[
 g_{x,y}(u)_i = \frac{h(x,y,\|u\|^2)u_i}{\|u\|^2},
\]

in expression (4.58). By assumption \( h(x,y,\cdot) \) is a weakly differentiable function in \( u \) for every \((x,y)\) which implies by Stein’s lemma

\[
 \int_{\mathbb{R}^2} \int_{\mathbb{R}^k} \frac{g_{x,y}(u)}{\sigma^2}(2\pi\sigma^2)^{-k} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{k} u_i^2} du f_2(x,y) dxdy = \\
 \int_{\mathbb{R}^2} \int_{\mathbb{R}^k} \text{div}_u(g_{x,y}(u))(2\pi\sigma^2)^{-k} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{k} u_i^2} du f_2(x,y) dxdy = \\
 E\left[ \text{div}_u(g_{x,y}(u)) \right]
\]

through an application of Fubini’s theorem. Using the fact that

\[
 \frac{\partial}{\partial x_i} \frac{u_i h(x,y,\|U\|^2)}{\|u\|^2} = \frac{h(x,y,s)}{s} + u_i \left[ \|u\|^2 \frac{\partial}{\partial s} h(x,y,s) 2u_i - h(x,y,s) 2u_i \right],
\]

\[
 \text{div}_u \left( u \frac{h(x,y,\|u\|^2)}{\|u\|^2} \right) = \frac{kh(x,y,s)}{s} + 2 \frac{\partial}{\partial s} h(x,y,s) - \frac{2h(x,y,s)}{s} = \\
 (k - 2) \frac{h(x,y,s)}{s} + 2 \frac{\partial}{\partial s} h(x,y,s)
\]

so that substitution of (4.60) into (4.59) yields
\[
E\left[ \frac{h(X,Y,S)}{\sigma^2} \right] = E[(k-2)\frac{h(x,y,s)}{s} + 2\frac{\partial}{\partial s} h(x,y,s)]
\]

establishing the result.

The following Theorem will give the form of and conditions on the proposed estimator to ensure minimaxity.

Theorem 4.5. Let

\[
\begin{pmatrix}
X \\
Y \\
V
\end{pmatrix}
\sim N_{2p+k}\left(\begin{pmatrix}
\theta \\
\theta + \eta \\
0
\end{pmatrix}, \sigma^2 \Sigma^*\right),
\]

where \( p > 2, k \geq 1, \) and

\[
\Sigma^* = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & 0_{pxk} \\
\Sigma_{21} & \Sigma_{22} & 0_{pxk} \\
0_{kxp} & 0_{kxp} & \Sigma_{33}
\end{pmatrix}.
\]

Let \( \delta(X,Y,V) = X + \frac{S}{k+2} g(X,Y) \) be an estimator for \( \theta \), with \( S = \| \Sigma_{33}^{-1/2} V \|^2 \), of the form

\[
\delta(X,Y,V) = X - S \frac{Q^{-1}\Sigma_{11}^{-1} r(\|X-Z\|^2_Q)}{k+2} (X-Z)
\] (4.61)

with \( Z = B(Y - AX) \), \( B \) non-singular \( pxp \) matrix, \( A = \Sigma_{21}\Sigma_{11}^{-1} \), and \( Q^* = \Sigma_{11}^{-1} Q^{-1} \Sigma_{11}^{-1} \) for \( Q \) symmetric positive definite. The estimator \( \delta(X,Y,V) \) is minimax under loss (4.57) \( l(\theta, \sigma^2, d) = \frac{(d-\theta)'Q(d-\theta)}{\sigma^2} \) provided

i) \( 0 < r(t) < 2(p-2) \) for all \( t > 0 \),

ii) \( r(t) \) is non-decreasing in \( t \) for all \( t > 0 \).

Proof. Let \( U = \Sigma_{33}^{-1/2} V \). By an application of Lemma 4.3

\[
\Delta(X, \delta, \theta) = E\left[ \frac{1}{\sigma^2} \left( \frac{S^2}{(k+2)^2} g'Qg + 2\frac{S}{k+2} (X-\theta)'Qg(X,Y) \right) \right]
\] (4.62)
where the expectation in (4.62) is taken with respect to a $N_{2p+k}(\begin{pmatrix} \theta \\ \theta + \eta \\ 0 \end{pmatrix}, \sigma^2 \Sigma^{**})$ distribution, where
\[
\Sigma^{**} = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & 0_{pxk} \\
\Sigma_{21} & \Sigma_{22} & 0_{pxk} \\
0_{kpx} & 0_{kpx} & I_{k+k}
\end{pmatrix}.
\]

Since
\[
g(X,Y) = -\frac{Q^{-1} \Sigma^{-1} r(\|X-Z\|_{Q^*}^2)}{\|X-Z\|_{Q^*}^2} (X - Z),
\]
expression (4.62) is equivalent to
\[
E\left[ \frac{S^2}{\sigma^2(k+2)} \frac{r^2(\|X-Z\|_{Q^*}^2)}{\|X-Z\|_{Q^*}^2} \right] - \frac{S}{\sigma^2(k+2)} \left( \Sigma_{11} (X - \theta) \right)' (X - Z) = \frac{S}{\sigma^2(k+2)} E\left[ div_x(h_{\theta_0}(X)) | Z = \theta_0 \right] = E\left[ \frac{S}{k+2} E\left[ E\left[ div_x(h_{\theta_0}(X)) | Z = \theta_0 \right] \right] \right] = E\left[ \frac{S}{k+2} \left( \frac{p-2}{\|X-Z\|_{Q^*}^2} + 2r'(\|X-Z\|_{Q^*}^2) \right) \right] = E\left[ \frac{S}{k+2} \left( \frac{p-2}{\|X-Z\|_{Q^*}^2} + 2r'(\|X-Z\|_{Q^*}^2) \right) \right]
\]

where
\[
h_{\theta_0}(x) = \frac{r(\|x-\theta_0\|_{Q^*}^2)}{\|x-\theta_0\|_{Q^*}^2} (x - \theta_0)
\]
in (4.63). Furthermore setting
\[
h(x,y,s) = \frac{S^2}{(k+2)^2} \frac{r^2(\|X-B(Y-AX)\|_{Q^*}^2)}{\|X-B(Y-AX)\|_{Q^*}^2}
\]
Lemma 4.7 implies
\[
E\left[\frac{h(X,Y,S)}{\sigma^2}\right] = E\left[\frac{(k-2)S^2r^2(\|X-Z\|^2_{Q^*})}{S\|X-Z\|^2_{Q^*}}\right] + 2\frac{\partial}{\partial S}\left[\frac{S^2r^2(\|X-Z\|^2_{Q^*})}{(k+2)^2\|X-Z\|^2_{Q^*}}\right] = E\left[\frac{S}{(k+2)} \frac{r^2(\|X-Z\|^2_{Q^*})}{\|X-Z\|^2_{Q^*}}\right].
\]

so that expressions (4.64) and (4.65) imply \( \triangle(\delta, X, \theta) \) is expressible as

\[
E\left[\frac{S}{(k+2)} \frac{r^2(\|X-Z\|^2_{Q^*})}{\|X-Z\|^2_{Q^*}}\right] - 2\frac{S}{k+2} \left[\frac{(p-2)r(\|X-Z\|^2_{Q^*})}{\|X-Z\|^2_{Q^*}}\right] - 4\frac{S}{k+2}r'(\|X-Z\|^2_{Q^*}) \leq 0
\]

(by assumption ii)

\[
E\left[\frac{S}{k+2} \frac{r(\|X-Z\|^2_{Q^*})}{\|X-Z\|^2_{Q^*}}\right] (r(\|X-Z\|^2_{Q^*}) - 2(p - 2)) \leq 0
\]

since \( \frac{S}{k+2} \geq 0 \) and assumption i) on \( r(\cdot) \).

\[\Box\]

### 4.5 Elliptically Symmetric Model with Unknown Scale and Residual Vector

In this section we extended the results of Section 3.2 to the general class of elliptically symmetric distributions with unknown scale and with a residual vector. In particular we assume that the joint distribution of \( Z = \begin{pmatrix} X \\ Y \\ U \end{pmatrix} \) has density of the form

\[
(\frac{1}{\sigma})^{2p+k} f(\frac{1}{\sigma^2}(z - \mu)^*\Sigma^*(z - \mu))
\]

with \( \sigma \) unknown,

\[
\Sigma^* = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} & 0 \\ \Sigma^{12} & \Sigma^{22} & 0 \\ 0 & 0 & \Sigma^{33} \end{pmatrix},
\]

\( \dim(U) \geq 1, \dim(X) = \dim(Y) = p \geq 3, \) and

\[
\mu = \begin{pmatrix} \theta \\ \theta + \eta \\ 0 \end{pmatrix}.
\]
The loss is given by
\[ l(\theta, \sigma^2, d) = \frac{(d - \theta)' Q (d - \theta)}{\sigma^2} \quad (4.68) \]
for Q symmetric and positive definite. The method of extension is similar to that of previous sections where a Stein-like differential equality is developed to express the difference in risk between the minimax estimator X of \( \theta \) under the generalized quadratic loss \( L_Q(d, \theta) \), and estimators of the form:
\[ \delta(X, L(X, Y), S) = X + Q^{-1}C \frac{S}{k + 2}g(X, L(X, Y)) \quad (4.69) \]
where \( S = \| (\Sigma^{33})^{-\frac{1}{2}} U \|^2 \), C is a non-singular matrix, and \( L(X, Y) \) is a linear function of the form:
\[ L(X, Y) = B(Y - AX). \]
Sufficient conditions will be given in Theorem 4.6 on \( g(X, L(X, Y)) \), C, and \( L(X, Y) \) so that the difference in risk between the estimator X of \( \theta \) and \( \delta(X, L(X, Y), S) \) is non positive under the generalized quadratic loss \( L_Q(d, \theta) \). As such the combined estimator \( \delta \) will dominate X. To that end the following Lemma can be seen as a generalization of Lemma 4.7 from the normal case to the elliptically symmetric case.

**Lemma 4.8.** Let \( \begin{pmatrix} X \\ Y \\ U \end{pmatrix} \) be random variables having the joint density, \( f\left(\frac{1}{\sigma^2}(\|x - \theta\|^2_{\Sigma^{-1}1} + \|y - \eta^*\|^2_{\Sigma^{-1}2} + \|u\|^2)\right) \). Let \( h : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^k \rightarrow \mathbb{R}^+ \) be a function such that

\[ i) \quad u \frac{h(x, y, \|u\|^2)}{\|u\|^2} \text{ is a weakly differentiable function in } u \text{ for every } (x, y), \]

and

\[ ii) \quad E[h(X, Y, S)] < \infty, \]

then
\[ E\left[\frac{h(X, Y, S)}{\sigma^2}\right] = E\left[\left(\left(k - 2\right) \frac{h(x, y, s)}{S} + 2 \frac{\partial}{\partial S} h(x, y, S)\right) \frac{F(t)}{f(t)}\right]. \]

**Proof.** Let \( t = \frac{1}{\sigma^2}(\|x - \theta\|^2_{\Sigma^{-1}1} + \|y - \eta^*\|^2_{\Sigma^{-1}2} + \|U\|^2) \). Setting
\[ F(t) = \frac{1}{2} \int_0^\infty f(u) \, du, \]

\[ E\left[ \frac{h(X,Y,S)}{\sigma^2} \right] = \int_{\mathbb{R}^{p+k}} \frac{h(x,y,u)\|u\|^2}{\sigma^2} f\left( \frac{1}{\sigma^2} \|x - \theta\|_{\Sigma^{-1}}^2 + \|y - \eta^*\|_{\Sigma^{-2}}^2 + \|u\|^2 \right) \, du \, dy = \]

\[ \int_{\mathbb{R}^{p+k}} \frac{U'U h(x,y,\|u\|^2)}{\|u\|^2} \, f\left( \frac{1}{\sigma^2} \|x - \theta\|_{\Sigma^{-1}}^2 + \|y - \eta^*\|_{\Sigma^{-2}}^2 + \|u\|^2 \right) \, du \, dy. \quad (4.70) \]

The partial derivative of \( F(t) \) with respect to the \( i \)th coordinate of \( u \) is

\[ \frac{\partial}{\partial u_i} F(t) = \frac{\partial}{\partial u_i} \frac{1}{2} \int_0^\infty f(u) \, du = -f(t) \frac{u_i}{\sigma^2} \quad (4.71) \]

so that setting \( g_{x,y}(u) = \frac{u_i h(x,y,\|u\|^2)}{\|u\|^2} \) with \( i \)th coordinate

\[ g_{x,y}(u) = \frac{u_i h(x,y,\|u\|^2)}{\|u\|^2} \]

in (4.70), and using the weak differentiability of \( g \) and the expression for the partial derivative of \( F(t) \) in (4.71), expression (4.70) is equivalent to

\[ \sum_{i=1}^k \int_{\mathbb{R}^{p+k}} \frac{u_i}{\sigma^2} g_{x,y}(u)_i f\left( \frac{1}{\sigma^2} \|x - \theta\|_{\Sigma^{-1}}^2 + \|y - \eta^*\|_{\Sigma^{-2}}^2 + \|u\|^2 \right) \, du \, dy = \]

\[ \sum_{i=1}^k \int_{\mathbb{R}^{p+k}} g_{x,y}(u)_i \left( -\frac{\partial}{\partial u_i} F(t) \right) \, du \, dy = \quad (4.72) \]

\[ \sum_{i=1}^k \int_{\mathbb{R}^{p+k}} \left( \frac{\partial}{\partial u_i} g_{x,y}(u)_i \right) F(t) \, du \, dy = \quad (4.73) \]

\[ \int_{\mathbb{R}^{p+k}} \text{div}_u (g_{x,y}(u)) \frac{F(t)}{f(t)} \, f(t) \, du \, dy = \]

\[ E[\text{div}_u (g_{x,y}(u)) \frac{F(t)}{f(t)}] \quad (4.74) \]

where the equality of (4.72) to (4.73) is justified by the weak differentiability of \( g_{x,y}(u) \) for all \((x,y)\).

From the previous expression for the \( \text{div}_u (g_{x,y}(u)) \) found in (4.60)

\[ \text{div}_u (g_{x,y}(u)) = ((k - 2) \frac{h(x,y,s)}{S} + 2 \frac{\partial}{\partial S} h(x,y,s)) \]

so that (4.74) is equivalent to

\[ E\left[ ((k - 2) \frac{h(x,y,s)}{S} + 2 \frac{\partial}{\partial S} h(x,y,s)) \frac{F(t)}{f(t)} \right] \]
establishing the result.

The following Theorem gives sufficient condition so that estimators of the form (4.69) will be minimax.

**Theorem 4.6.** Let \( Z = \begin{pmatrix} X \\ Y \\ U \end{pmatrix} \) and \( \mu = \begin{pmatrix} \theta \\ \theta + \eta \\ 0 \end{pmatrix} \). Suppose that \( Z \) has a distribution whose joint density is given by \((\frac{1}{\sigma})^{2p+k} f(\frac{1}{\sigma^2}(z - \mu)^t \Sigma^*(z - \mu))\) with \( \Sigma^* = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & 0 \\ \Sigma_{12}^t & \Sigma_{22} & 0 \\ 0 & 0 & \Sigma_{33} \end{pmatrix} \). Then the estimator

\[
\delta(X, Y', S) = X + Q^{-1} \Sigma_{11*}^S \frac{S}{\Sigma_{33}} g(X, Y')
\]

where

\[
Y' = B(Y - (\Sigma_{22})^{-1} \Sigma_{12}^t X) \quad (B \text{ non singular})
\]

\[
\Sigma_{11*} = \Sigma_{11} - \Sigma_{12} (\Sigma_{22})^{-1} \Sigma_{12}^t,
\]

\[
S = \|(\Sigma_{33})^{1/2} U\|^2,
\]

and \( Q \) is a symmetric positive definite matrix, is minimax for \( \theta \) under loss \( l(\theta, \sigma^2, d) = \frac{(d - \theta)Q(d - \theta)}{\sigma^2} \), provided

i) The risk of \( \delta \) exists,

ii) \( g(X, Y') \) is weakly differentiable in \( x \) for all \( y \),

iii) \( \frac{\|u\|^2 \|g\| Q^*}{\|u\|^2} \) is weakly differentiable in \( u \) for all \( (x, y') \),

iv) \( \|g\| Q^* + 2 \text{div}_x(g(x, y')) \leq 0 \) for \( (x, y') \) a.e..

**Proof.** Consider the linear change of variables
\[
A^* \begin{pmatrix} X \\ Y \\ U \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ -BA & B & 0 \\ 0 & 0 & (\Sigma^{33})^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} X \\ Y \\ U \end{pmatrix} = \begin{pmatrix} X \\ Y' \\ U' \end{pmatrix}
\]
so that
\[
(A^*)^{-1} = \begin{pmatrix} I & 0 & 0 \\ A & B^{-1} & 0 \\ 0 & 0 & (\Sigma^{33})^{-\frac{1}{2}} \end{pmatrix}
\]
and
\[
(A^* - 1)^\prime \Sigma^*(A^* - 1) =
\begin{pmatrix} \Sigma^{11} + A^\prime \Sigma^{12'} + \Sigma^{12} A + A^\prime \Sigma^{22} A & (\Sigma^{12} + A^\prime \Sigma^{22}) B^{-1} 0 \\ (B^{-1})^\prime (\Sigma^{12'} + \Sigma^{22} A) & (B^{-1})^\prime \Sigma^{22} B^{-1} 0 \\ 0 & 0 & I \end{pmatrix}
\]
. (4.75)

Setting \( A = -(\Sigma^{22})^{-1} \Sigma^{12'} \) in (4.75) yields
\[
(A^* - 1)^\prime \Sigma^*(A^* - 1) =
\begin{pmatrix} \Sigma^{11} - \Sigma^{12} \Sigma^{22} - (\Sigma^{12})' & 0 & 0 \\ 0 & (B^{-1})^\prime \Sigma^{22} (B^{-1}) & 0 \\ 0 & 0 & I \end{pmatrix}
\]
so that by the change of variables formula, the density of the random variables \( \begin{pmatrix} X \\ Y' \\ U' \end{pmatrix} \) is
proportional to
\[
f(\|x - \theta\|^2_{\Sigma^{11}}, + \|y' - \eta'\|^2_{\Sigma^{22}}, + \|u'\|^2).
\]
(4.76)

Let \( t = \|x - \theta\|^2_{\Sigma^{11}}, + \|y' - \eta'\|^2_{\Sigma^{22}}, + \|u'\|^2 \). Since the difference in risk between \( X \) and the estimator \( \delta(X, Y', S) \) is expressible as
\[
E[\left( \frac{1}{\sigma^2} \right) \left( \frac{S^2}{(k + 2)^2} \|g(X, Y')\|^2 + 2 \frac{S}{k + 2} (X - \theta)^\prime \Sigma^* g(X, Y') \right)]
\]
(4.77)
where $Q^* = (\Sigma^{11})^{-1}Q^{-1}(\Sigma^{11})^{-1}$, an application of Lemma 4.8 to the first term in (4.77) with

$$h(X, Y', S) = \frac{S^2}{(k+2)^2} ||g(X, Y')||^2_{Q^*}$$

yields

$$E\left[\frac{S^2}{\sigma^2_{(k+2)^2}} ||g(X, Y')||^2_{Q^*}\right] =$$

$$E\left[\frac{(k-2)S||g(X, Y')||^2_{Q^*}}{(k+2)^2} + \frac{4S||g(X, Y')||^2_{Q^*}}{(k+2)^2}\right] \frac{F(t)}{f(t)},$$

(4.78)

By an application of Lemma 4.6, the second term in expectation of (4.77)

$$E\left[\frac{S}{\sigma^2_{(k+2)}}(X - \theta)'\Sigma_{11}^*g(X, Y')\right] =$$

$$E\left[(\frac{S}{k+2} div_x(g(X, Y')) \frac{F(t)}{f(t)}\right],$$

(4.79)

so that the difference in risk, $\Delta(\theta, X, \theta, \eta)$ is expressible as

$$E\left[\frac{S}{(k+2)}||g(X, Y')||^2_{Q^*} \frac{F(t)}{f(t)} + 2\frac{S}{k+2} div_x(g(X, Y')) \frac{F(t)}{f(t)}\right],$$

once (4.78) and (4.79) is substituted into (4.77). The result follows since

$$E\left[\frac{S}{(k+2)}||g(x, y')||^2_{Q^*} \frac{F(t)}{f(t)} + 2\frac{S}{k+2} div_x(g(x, y')) \frac{F(t)}{f(t)}\right] \leq 0$$

as $\frac{SF(t)}{(k+2)f(t)} \geq 0$ and $||g(x, y')||^2_{Q^*} + 2 div_x(g(x, y')) \leq 0$ for all $(x,y')$ by assumption.

4.6 Numerical Comparisons

In this section we develop the risk of UMVUE of $\theta$ when $E[Y] = E[X] = \theta$ for a multivariate normal distribution and then compare the risk of the UMVUE estimator to that of estimators developed in Section 2.1 under squared error loss. More specifically we assume

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}(\begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix})$$

with $\Sigma$ known and $p \geq 3$ and set $\eta = 0$ to develop the UMVUE estimator. For the numerical comparison we set covariance structure of $\Sigma$ to be:
\[
\Sigma = \begin{pmatrix}
\sigma^2 I_p & \rho \sigma \tau I_p \\
\rho \sigma \tau I_p & \tau^2 I_p
\end{pmatrix}
\]

and then compare the risk of the UMVUE estimator developed when \(\eta = 0\), to that of estimators developed in Section 2.1 under square error loss for various values of \(\|\eta\|^2\), \(\sigma\), \(\tau\), and \(\rho\). In that way we can get a sense as to how well the estimators developed in Section 2.1 compare to the best unbiased linear combination when \(\eta = 0\), and how much the estimators in Section 2.1 can save when the investigator is mistaken about the bias of estimator \(Y\).

When \(\eta = 0\) the density of the pair of random variables \(\begin{pmatrix} X \\ Y \end{pmatrix}\) is

\[
f(x, y) = |2\pi \Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2} \left( \begin{pmatrix} x - \theta \\ y - \theta \end{pmatrix} \right) \Sigma^{-1} \left( \begin{pmatrix} x - \theta \\ y - \theta \end{pmatrix} \right)'}.
\] (4.80)

Setting

\[
X_{2,1} = Y - \Sigma_{21} \Sigma_{11}^{-1} X
\]

so that \(X\) is independent of \(X_{2,1}\), and

\[
Y^* = (I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} X_{2,1}
\] (4.81)

so that,

\[
E[Y^*] = (I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} (I - \Sigma_{21} \Sigma_{11}^{-1}) \theta = \theta,
\] (4.82)

\(Y^*\) is an unbiased estimator of \(\theta\) with variance

\[
\text{var}(Y^*) = \Sigma_{Y^*} = (I - \Sigma_{21} \Sigma_{11}^{-1})^{-1} [\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}] (I - \Sigma_{21} \Sigma_{11}^{-1})^{-t}.
\] (4.83)
The joint density of \( \begin{pmatrix} X \\ Y^* \end{pmatrix} \) is

\[
f(x, y^*) = k e^{-\frac{1}{2} \left( \begin{pmatrix} x - \theta \\ y^* - \theta \end{pmatrix}' \begin{pmatrix} 0 & 
\Sigma^{-1} \\
0 & 0 \end{pmatrix} \begin{pmatrix} x - \theta \\ y^* - \theta \end{pmatrix} \right)}
\]

where \( k \) is the normalizing constant in (4.84). By expanding the quadratic form in the exponential of density (4.84) we can find a complete sufficient statistic for the parameter \( \theta \) that will be unbiased and thus will be the UMVUE estimator. Expanding the quadratic form in (4.84) yields

\[
x' \Sigma_{11}^{-1} x + \theta' \Sigma_{11}^{-1} \theta - 2 \theta' \Sigma_{11}^{-1} x + y^* \Sigma_{Y^*}^{-1} y^* + \theta' \Sigma_{Y^*}^{-1} \theta - 2 \theta' \Sigma_{Y^*}^{-1} y^*
\]

which implies the density in (4.84) is an exponential family of the form

\[
h(x, y^*) g(\theta) e^{\theta' T(X, Y^*)}
\]

with complete sufficient statistic

\[
T(X, Y^*) = \Sigma_{11}^{-1} X + \Sigma_{Y^*}^{-1} Y^*.
\]

Therefore the UMVUE estimator of \( \theta \), \( \delta_{\text{combo}} \), is

\[
\delta_{\text{combo}}(X, Y) = (\Sigma_{11}^{-1} + \Sigma_{Y^*}^{-1})^{-1} (\Sigma_{11}^{-1} X + \Sigma_{Y^*}^{-1} Y^*)
\]

with variance

\[
\Sigma_{\delta_{\text{combo}}} = Cov(\delta_{\text{combo}}, \delta_{\text{combo}}) = (\Sigma_{11}^{-1} + \Sigma_{Y^*}^{-1})^{-1}.
\]
When the loss is of the form \( L_Q(d, \theta) = (d - \theta)'Q(d - \theta) \), the risk of \( \delta_{\text{combo}} \) when \( Y^* \) is unbiased for \( \theta \) is

\[
E[(\delta_{\text{combo}} - \theta)'Q(\delta_{\text{combo}} - \theta)] = tr(Q\Sigma_{\delta_{\text{combo}}}) = tr(Q(\Sigma_{11}^{-1} + \Sigma_{12}^{-1})) = \\
tr(Q[\Sigma_{11}^{-1} + (I - \Sigma_{21}\Sigma_{11}^{-1})'\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}]^{-1}(I - \Sigma_{21}\Sigma_{11}^{-1})^{-1}).
\]

When the researcher is mistaken and \( Y^* \) is a biased for \( \theta (\eta \neq 0) \), with bias

\[
E[\delta_{\text{combo}} - \theta] = E[(\Sigma_{11}^{-1} + \Sigma_{12}^{-1})^{-1}(\Sigma_{11}^{-1}\theta + \Sigma_{12}^{-1}(\theta + \eta) - \theta)] = \\
(\Sigma_{11}^{-1} + \Sigma_{12}^{-1})^{-1}\Sigma_{12}^{-1}\eta,
\]

the risk of using \( \delta_{\text{combo}} \) as an estimator of \( \theta \) will be

\[
E[\|\delta_{\text{combo}}(X, Y) - \theta\|^2_Q] = tr(Q\Sigma_{\delta_{\text{combo}}}) + \eta\Sigma_{Y^*}^{-1}(\Sigma_{11}^{-1}\theta + \Sigma_{12}^{-1}(\theta + \eta) - \theta) = \\
(\Sigma_{11}^{-1} + \Sigma_{12}^{-1})^{-1}\Sigma_{12}^{-1}\eta,
\]

which is unbounded in \( \eta \) unlike the estimators develop in Section 2 which have bounded risk.

In order to get a direct comparison of how the risk of \( \delta_{\text{combo}} \) compares to that of the estimators developed in Section 2, we assume for simplicity that

\[
\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2p}\left( \begin{pmatrix} \theta \\ \theta + \eta \end{pmatrix}, \begin{pmatrix} \sigma^2I_p & \rho\sigma\tau I_p \\ \rho\sigma\tau I_p & \tau^2I_p \end{pmatrix} \right)
\]

and the loss is the quadratic loss, \( L_I(d, \theta) \). Then

\[
Y^* = \left( \frac{\sigma^2}{\sigma^2 - \rho\sigma\tau} \right)(Y - \frac{\rho\tau}{\sigma}X)
\]

so that

\[
\Sigma_{Y^*} = \frac{\sigma^4\tau^2(1 - \rho^2)}{(\sigma^2 - \rho\sigma\tau)^2}I_p
\]

and thus \( \delta_{\text{combo}} \) has the form

\[
\delta_{\text{combo}} = \left[ \frac{\sigma^2\tau^2(1 - \rho^2)}{\tau^2(1 - \rho^2) + (\sigma - \rho\tau)^2} \right] \frac{1}{\sigma^2}X + \left( \frac{\sigma^2 - \rho\sigma\tau}{\tau^2\sigma^2(1 - \rho^2)} \right)(Y - \frac{\rho\tau}{\sigma}X)
\]

with
\[ \Sigma_{\delta_{\text{combo}}} = \frac{\sigma^4 \tau^2 (1 - \rho^2)}{\tau^2 \sigma^2 (1 - \rho^2) + \sigma^2 (\sigma - \rho \tau)^2} I_p \]

and risk

\[ R(\delta_{\text{combo}}, \eta) = \text{tr}(I_p) \frac{\sigma^4 \tau^2 (1 - \rho^2)}{\tau^2 \sigma^2 (1 - \rho^2) + \sigma^2 (\sigma - \rho \tau)^2} + \eta' \eta \frac{\sigma^4 (\sigma - \rho \tau)^4}{\tau^2 \sigma^2 (1 - \rho^2) + \sigma^2 (\sigma - \rho \tau)^2}. \]

Consider now the shrinkage estimators developed in Section 2. By Corollary 4.1 setting \( a = p - 2 \) and \( B = (I_p - \Sigma_{21} \Sigma_{11}^{-1})^{-1} \) implies the estimator

\[ \delta_1(X, Y) = Z + (1 - \frac{\sigma^2(p - 2)}{\|X - Z\|^2}) (X - Z) \quad (4.91) \]

with

\[ Z = \left( \frac{\sigma^2}{\sigma^2 - \rho \sigma \tau} \right) (Y - \frac{\rho \tau}{\sigma} X) \quad (4.92) \]

will be minimax. The positive-part version of \( \delta_1(X, Y) \),

\[ \delta_1^+(X, Y) = Z + \max \{0, (1 - \frac{\sigma^2(p - 2)}{\|X - Z\|^2})\} (X - Z) \quad (4.93) \]

will also be minimax by Theorem 4.1. Theorem 4.2 with

\[ \Lambda = \left\{ \frac{(\sigma^2 - \rho \sigma \tau)^2 + (\tau^2 \sigma^2 - \rho^2 \sigma^2 \tau^2)}{(\sigma^2 - \rho \sigma \tau)^2 \sigma^2} \right\} I, \]

\[ \mu_{\theta \eta} = \frac{-\sigma^2}{\sigma^2 - \rho \sigma \tau} \eta, \]

\[ Q_1(\mu_{\theta \eta}) = \frac{\sigma^2 \eta' \eta}{(\sigma^2 - \rho \sigma \tau)^2 + (\tau^2 \sigma^2 - \rho^2 \sigma^2 \tau^2)}, \]

and

\[ Q_2(\mu_{\theta \eta}) = \frac{\eta' \eta}{(\sigma^2 - \rho \sigma \tau)^2} \]

gives the risk the following bounds for the risk of \( \delta_1(X, Y) \) (whose risk depends on the mean only through \( \eta \) due to the choice of \( B \))

\[ p \sigma^2 - (p - 2) \sigma^2 \min \left\{ \frac{(p + 2) A}{(p - 2)(p + 2) |A + B| + \sigma^2 |\eta'|}, \frac{A}{(p - 2)(p - 4) |A + B| + \sigma^2 |\eta'|} \right\} \leq R(\delta_1, \eta) \leq \]
\[ p\sigma^2 - (p - 2)^2 \frac{\sigma^2 A}{(p-2)((A+B)\tau^2 \eta'\eta)} \]

where

\[ A = (\sigma^2 - \varrho \sigma \tau)^2 \] (4.94)

and

\[ B = \tau^2 \sigma^2 - \varrho^2 \sigma^2 \tau^2 \] (4.95)

Figure 4.1 plots the simulated risk of \( \delta_1(X,Y) \), as a function of \( \|\eta\|^2 \), for various values of \( \sigma^2 = \tau^2 = 1 \) along with the associated bounds developed in Theorem 3 when \( p = \text{dim}(X) = 6 \). Since the risk depends on the mean only through the variable \( \eta \), the risk is plotted for various values of \( \|\eta\|^2 \). Since the eigenvalues of \( \Sigma \) are constant, for \( \eta = 0 \) the upper and lower bound coincide so that the risk will be exact when \( \eta = 0 \). Figure 1 shows that the correlation (\( \varrho \)) between the variables \( X \) and \( Y \) can have a dramatic impact on how much the shrinking procedure saves in terms of the risk. In this case the a negative correlation will improve the risk. The risk are asymptotically approaching the risk of the estimator, \( \delta_0(X) = X \) from below.

![Risk Bounds and Simulated Risk](Image)

Figure 4.1: Simulated risk of \( \delta_1(X,Y) \) (4.91) with \( \sigma^2 = \tau^2 = 1 \) for varying \( \varrho \)

Figure 4.2 compares the risk of \( \delta_1(X,Y) \), as a function of \( \|\eta\|^2 \), for varying \( \tau^2 \) when \( \sigma^2 = 1 \) for when \( \varrho = .7 \) and 0 respectively. Of note is that positive correlation between \( X, Y \) improves the risk when \( \tau^2 \) is 49, while a positive correlation increases the risk when \( \tau^2 = .5 \).
Figure 4.2: Simulated risk for $\delta_1(X,Y)$ (4.91) for $\sigma^2 = 1$ and varying $\tau^2$ and $\varrho$

Figure 4.3 compares, for fixed $\sigma^2$, $\tau^2$, the risk of the positive part version of $\delta_1(X,Y)$, $\delta_1^+$, to the risk of $\delta_1(X,Y)$, as a function of $\|\eta\|^2$, when $\text{dim}(X) = 6 = p$. $\delta_1^+(X,Y)$ dominates $\delta_1(X,Y)$, however the savings in terms of risk depends on the correlation, $\varrho$. In this case, the larger the correlation, $\varrho$, the closer the risk of $\delta_1(X,Y)$ is to $\delta_1^+(X,Y)$.

In order to compare the risk of $\delta_1(X,Y)$ to the risk of $\delta_{\text{combo}}(X,Y)$ we first compare the risk of using $\delta_{\text{combo}}$ when $\eta = 0$, and show that $\delta_{\text{combo}}$ will have uniformly smaller risk than $\delta_1$ when $\eta = 0$. We then give sufficient conditions for when risk of $\delta_1$ will dominate the risk of $\delta_{\text{combo}}$ by comparing the upper bound for the risk of $\delta_1(X,Y)$ developed in Theorem 2 to the exact risk of using $\delta_{\text{combo}}$. When $\eta'\eta = 0$, setting

$$A := \sigma^2\tau^2(1 - \varrho^2)$$
\( B := (\sigma^2 - \varrho \sigma \tau)^2, \)

\[
\begin{align*}
p\sigma^2 - \frac{(p-2)\sigma^2 B}{A+B} & \leq R(\delta_1(X, Y), \theta, \eta) \leq p\sigma^2 - \frac{(p-2)\sigma^2 B}{A+B}.
\end{align*}
\]

by Theorem 4.2, and

\[
R(\delta_{\text{combo}}(X, Y), \theta, \eta) = \frac{p\sigma^2 A + B}{A+B} = \frac{p\sigma^2 (\sigma^2 \tau^2 (1 - \varrho^2))}{\sigma^2 \tau^2 (1 - \varrho^2) + (\sigma^2 - \varrho \sigma \tau)^2}
\]

so that

\[
R(\delta_1(X, Y), \theta, 0) - R(\delta_{\text{combo}}(X, Y), \theta, 0) = p\sigma^2 - \frac{(p-2)\sigma^2 B}{A+B} - \frac{p\sigma^2 A}{A+B} =
\]

\[
p\sigma^2 \frac{B}{A+B} - (p-2)\sigma^2 \frac{B}{A+B} = \frac{2p^2 B}{A+B} \geq 0
\]

implying that \( R(\delta_{\text{combo}}(X, Y), \theta, 0) < R(\delta_1(X, Y), \theta, 0). \)

When the upper bound for the risk of \( \delta_1 \) given by Theorem 4.2 is less than the risk of \( \delta_{\text{combo}} \)

\[
p\sigma^2 - \frac{(p-2)\sigma^2 B}{(p-2)[(B+A) + \sigma^2 \eta'] \eta} < \frac{p\sigma^2 A + B}{A+B} + \frac{\eta' \eta B^2}{[A+B]^2} \iff
\]

\[
\frac{p\sigma^2 B}{A+B} - \frac{(p-2)\sigma^2 B}{(p-2)[(B+A) + \sigma^2 \eta'] \eta} < \frac{\eta' \eta B^2}{[A+B]^2} \iff
\]

\[
p\sigma^2 B(A+B)[(p-2)(B+A) + \sigma^2 \eta'] - (p-2)\sigma^2 B [A+B]^2 < \eta' \eta B^2 [(p-2)(B+A) + \sigma^2 \eta'].
\]

(4.96)

Setting \( x = \eta' \eta \) where \( x \in [0, \infty) \) and

\[
Q(x) = ax^2 + bx + c
\]

with

\[
a = \sigma^2 B^2,
\]

\[
c = -(p-2)\sigma^2 B(A+B)^2[p-1],
\]
and
\[ b = B(B + A)[(p - 2)B - p\sigma^4], \]
expression (4.96) will be satisfied once \( Q(x) > 0 \). Since \( a > 0 \) and \( c < 0 \), \( Q(x) \) will have 2 real roots denoted by \( r_1 \) and \( r_2 \) respectively, where \( r_1 < 0 \) and \( r_2 > 0 \). Since \( a > 0 \), \( x > r_2 \) implies \( Q(x) > 0 \) and hence for large values of the \( \eta'\eta \) the shrinkage estimator dominates the combo estimator. In terms of the original parameters once
\begin{align*}
\eta'\eta &> -\frac{(\sigma^2 - \rho\sigma\tau)^2((\sigma^2 - \rho\sigma\tau)^2 + \sigma^2\tau^2(1 - \rho^2))((p - 2)(\sigma^2 - \rho\sigma\tau)^2 - p\sigma^4)}{2\sigma^4((\sigma^2 - \rho\sigma\tau)^2 + \sigma^2\tau^2(1 - \rho^2))^2} + \\
&\frac{\sqrt{((\sigma^2 - \rho\sigma\tau)^2 + \sigma^2\tau^2(1 - \rho^2))^2((p - 2)(\sigma^2 - \rho\sigma\tau)^2 - p\sigma^4)^2 + 4p^2(\rho^2 + 1)(\sigma^2 - \rho\sigma\tau)^2}}{2\sigma^2((\sigma^2 - \rho\sigma\tau)^2 + \sigma^2\tau^2(1 - \rho^2))^2},
\end{align*}
\( R(\delta_1(X, Y), \theta, \eta) < R(\delta^{\text{combo}}(X, Y), \theta, \eta) \).

Simulated values of the the risk of \( \delta_1(X, Y) \) were plotted against the exact values of the risk for \( \delta^{\text{combo}} \) as a function of \( \|\eta\|^2 \) in Fig 4.4, and 4.5 for varying values of \( \tau^2 \) and \( \rho \), with \( \sigma^2 \) held fix at the value 1 when \( \dim(X) = 6 = p \).

Figure 4.4 plots, as a function of \( \|\eta\|^2 \), the risk of the combo estimator and the simulated risk of the shrinkage estimator when \( \sigma^2 = \tau^2 = 1 \), for varying \( \rho \). Although the risk of \( \delta^{\text{combo}} \) dominates the risk of \( \delta_1(X, Y) \) when \( \eta = 0 \), the risk is comparable while for large values \( \|\eta\|^2 \) the risk of \( \delta_1(X, Y) \) stays bounded and is minimax while the risk of \( \delta^{\text{combo}} \) increases linearly in \( \|\eta\|^2 \). In this case a negative correlation helped both \( \delta^{\text{combo}} \) and \( \delta_1(X, Y) \), however the slope of the risk of \( \delta^{\text{combo}} \) was much larger when \( \rho = -.5 \) than when \( \rho = .5 \).

![Figure 4.4: Simulated risk of \( \delta_1(X, Y) \) (4.91) and risk of \( \delta^{\text{combo}}(X, Y) \) (4.90) for \( \sigma^2 = \tau^2 = 1 \) and varying \( \rho \).]
Figure 4.5 plots, as a function of $\|\eta\|^2$, the risk of the combo estimator and the simulated risk of the shrinkage estimator for varying $\tau$ when $\sigma^2 = 1$ and $\varrho = -.5$. Just as in Figure 4.4 the risk between the two estimators is comparable when $\eta = 0$, while for large values of $\|\eta\|^2$ the risk of the shrinkage estimator is bounded and dominates the combo estimator. In this case a larger value for $\tau^2$ resulted in an increased the risk at the origin for both the combo estimator and shrinkage estimator. The rate of increase in risk when $\tau^2 = .05$ is larger for both the combination estimator, $\delta^{combo}$, and the shrinkage estimator, $\delta_1$, than when $\tau^2 = 49$ however.

Figure 4.5: Simulated risk of $\delta_1(X,Y)$ (4.91) vs combo estimator (4.90) for $\sigma^2 = 1$, $\varrho = -.5$ and varying $\tau^2$

Figure 4.6 shows that although the positive-part version of the shrinkage estimator dominates the original shrinkage estimator, at the origin the combo estimator still has smaller risk than both the positive-part version of the shrinkage estimator, and the shrinkage estimator.
Figure 4.6: Simulated risk of $\delta_1(X,Y)$ (4.91) and $\delta_1^+(X,Y)$ (4.93) versus risk of combo estimator (4.90) for $\sigma^2 = 1, \tau^2 = .5$ and $\varrho = -.5$

4.7 Summary and Conclusions

The shrinkage estimator developed in Section 2 seems to be a logical candidate to estimate the mean when the auxiliary information is presented to the statistician. Under conditions given in Section 2 the shrinkage estimator dominates the MLE estimator $\delta_0(X) = X$ under general quadratic loss. In section 6 the proposed shrinkage estimator was compared to the UMVUE when no bias is present, $\delta^{combo}$. It compared favorably for small values of bias ($\|\eta\|^2$) while dominating $\delta^{combo}$ for larger values of bias ($\|\eta\|^2$) when the parameter $B$ is chosen so that the risk of the shrinkage estimator depends on the mean only through the bias $\eta$. Furthermore extensions were developed for normal distributions with residual vectors whose covaraince is known up to a scale, and broad classes of elliptically symmetric distributions. Further research is needed to see what modifications to the proposed estimator are needed when the covaraince structure of the estimators is completely unknown, and if other linear combinations of $Z$ in (4.3) yield better results if $\eta$ is assumed to be in some affine space of $\mathbb{R}^p$. 
Chapter 5

Appendix

5.1 Appendix A: Proof of Lemma 3.2 and Theorem 3.4

Proof of Lemma 3.2.

Proof. Let $Z = \frac{X}{\sigma}$, $\eta = \frac{\theta}{\sigma}$, and $S = \frac{\|X\|^2}{\sigma^2} = \|Z\|^2$. Then

$$-b\|\delta_r(X) - \theta\|^2 = -b\sigma^2\|\tau(S)Z - \eta\|^2 =$$

$$-b\sigma^2[S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\eta^T(1 - \frac{\tau(S)}{S})].$$  \(5.1\)

Let $Y = CZ$ where $C$ is an orthogonal matrix whose first row is given by $(\frac{\theta_1}{\|\theta\|}, \ldots, \frac{\theta_p}{\|\theta\|}) = (\frac{m_1}{\|m\|}, \ldots, \frac{m_p}{\|m\|})$. Since $C$ is orthogonal, $Y^TY = Z'CC'Z = S$ so that (5.1) becomes

$$-b\sigma^2[S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\eta^T(1 - \frac{\tau(S)}{S})] =$$

$$-b\sigma^2[S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\|\eta\|y_1 + 2\|\eta\|y_1 \frac{\tau(S)}{S}].$$

Since $C$ is an orthogonal Matrix, $Y$ has a Multivariate Normal distribution where $Y_1 \sim N(\|\eta\|, 1), Y_i \sim N(0, 1)$ for $i = 2, \ldots, p$ and $\{Y_i\}_{i=1}^p$ is independent.

$$E_{\theta}[e^{-b\|\delta_r(X) - \theta\|^2}] =$$  \(5.2\)

$$E_{(\|\eta\|, 0, \ldots, 0)}[e^{-b\sigma^2[S + \frac{\tau^2(S)}{S} - 2\tau(S) + \|\eta\|^2 - 2\|\eta\|y_1 + 2\|\eta\|y_1 \frac{\tau(S)}{S}]}]$$  \(5.3\)
where expectation (5.2) is taken with respect to the distribution of $X$, while expectation (5.3) is taken with respect to the distribution of $Y$. Making the change of variables $\sum_{i=2}^{n} Y_i = U$, where $U \sim \chi^2_{p-1}$, expression (5.3) becomes

$$
\int_{0}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(y_1-\|\eta\|)^2} \frac{dy_1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2} u \frac{p-1}{2}} \frac{du}{2^{\frac{p-1}{2}} \Gamma(\frac{p-1}{2})} =
$$

$$
\int_{0}^{\infty} \int_{-\infty}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1-\|\eta\|)^2} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} dy_1
$$

The inner integral of (5.4)

$$
\int_{-\infty}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1-\|\eta\|)^2} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} dy_1 =
$$

$$
\int_{-\infty}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1-\|\eta\|)^2} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} dy_1 =
$$

$$
\int_{-\infty}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1^2 + \|\eta\|^2)} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} dy_1 =
$$

Making the transformation $y \to -y$ in (5.5),

$$
\int_{-\infty}^{0} e^{-(b \sigma + \frac{1}{2})^2(y_1^2 + \|\eta\|^2)} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} dy_1 =
$$

so that the inner integral in (5.4) becomes:

$$
\int_{0}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1^2 + \|\eta\|^2)} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) * dy_1
$$

$$
\int_{0}^{\infty} e^{2(b \sigma + \frac{1}{2})^2\|\eta\|y_1 - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} + e^{-(b \sigma + \frac{1}{2})^2\|\eta\|y_1 - 2b \sigma^2 \|\eta\|y_1 \frac{\tau(y_1^2 + u)}{y_1^2 + u} )} dy_1 =
$$

$$
\int_{0}^{\infty} e^{-(b \sigma + \frac{1}{2})^2(y_1^2 + \|\eta\|^2)} e^{-\frac{b \sigma^2 \tau(y_1^2 + u)}{y_1^2 + u}} + 2b \sigma \tau(y_1^2 + u) * dy_1
$$
\[ [(\sum_{r=0}^{\infty} \frac{((2\|\eta\|y_1 \{ ((\beta^2 + \frac{1}{2}) - \beta \tau(y_1^2 + u) \})^r}{r!} + \sum_{r=0}^{\infty} (-1)^r \frac{((2\|\eta\|y_1 \{ ((\beta^2 + \frac{1}{2}) - \beta \tau(y_1^2 + u) \})^r}{r!} ]dy_1 = \]

\[ 2 \int_0^{\infty} e^{-((\beta^2 + \frac{1}{2})\|\eta\|^2) - \frac{\beta \tau^2(y_1^2 + u)}{\|\eta\|^2} + 2\beta \tau y_1^2 + u} \sum_{r=0}^{\infty} \left( \frac{(2\|\eta\|y_1)^{2r}}{(2r)!} \left( \frac{(\beta^2 + \frac{1}{2}) - \beta \tau y_1^2 + u}{y_1^2 + u} \right)^{2r} \right) dy_1. \]

\[ \text{(5.6)} \]

Making the change of variables \( v = y_1^2 \) in (5.6), results in (5.6)

\[ \int_0^{\infty} e^{-((\beta^2 + \frac{1}{2})\|v\|^2) - \frac{\beta \tau^2(v+u)}{\|v\|^2} + 2\beta \tau v + u} \sum_{r=0}^{\infty} \left( \frac{(2\|v\|v)^{2r}}{(2r)!} \left( \frac{(\beta^2 + \frac{1}{2}) - \beta \tau(v+u)}{v+u} \right)^{2r} \right) dv. \]

Hence

\[ E_\theta[e^{-b\|\theta(X) - \theta\|}]. \]

\[ \int_0^{\infty} \left[ \int_0^{\infty} e^{-((\beta^2 + \frac{1}{2})\|v\|^2) - \frac{\beta \tau^2(v+u)}{\|v\|^2} + 2\beta \tau v + u} \sum_{r=0}^{\infty} \left( \frac{(2\|v\|v)^{2r}}{(2r)!} \left( \frac{(\beta^2 + \frac{1}{2}) - \beta \tau(v+u)}{v+u} \right)^{2r} \right) dv \right] \frac{1}{\pi^\frac{1}{2} 2^\frac{r}{2} \Gamma(\frac{r-1}{2})} \left. du \right. \]

\[ \]. \]

\[ \text{(5.8)} \]

Using the equality \( \Gamma(2r + 1) = \Gamma(r + \frac{1}{2}) r! 2^{2r} \pi^{-\frac{1}{2}} \) and making the change of variables

\[ w_1 = u + v \]

\[ w + 2 = \frac{v}{u+v} \]

do that \( v = w_1 w_2, u = w_1 - w_1 w_2 \) with inverse Jacobin \( w_1, \) (5.8) is equivalent to

\[ \int_0^{\infty} \int_0^{\infty} e^{-((\beta^2 + \frac{1}{2})(w_1 w_2 + \|\eta\|^2) - \frac{\beta \tau^2(w_1 w_2)}{w_1} + 2\beta \tau w_1 w_2) \} w_1^{-\frac{1}{2}} w_2^{-\frac{1}{2}} \}

\[ \sum_{r=0}^{\infty} \left( \frac{\|\eta\|^2 (\beta^2 + \frac{1}{2})}{r! \Gamma(r + \frac{1}{2})} \right)^r \left( \frac{(\beta^2 + \frac{1}{2})}{w_1 (\beta^2 + \frac{1}{2})} \right)^{2r} \right) \] \[ e^{-(\beta^2 + \frac{1}{2})(w_1 - w_1 w_2)} w_1^{-\frac{1}{2} - 1} (1 - w_2) \frac{1}{2} - \frac{1}{2} - 1 w_1 dw_1 \frac{dw_2}{2^r \Gamma(\frac{r-1}{2})} = \]

\[ e^{-(\beta^2 + \frac{1}{2})\|\eta\|^2} \sum_{r=0}^{\infty} \left( \frac{\|\eta\|^2 (\beta^2 + \frac{1}{2})}{r! \Gamma(r + \frac{1}{2})} \right)^r \]
\[ \int_0^\infty (1 - \frac{b\sigma^2(w_1)}{w_1(b\sigma^2 + \frac{1}{2})})^{2r}(b\sigma^2 + \frac{1}{2})^r w_1^{r+\frac{p}{2}-1} e^{-(b\sigma^2 + \frac{1}{2})w_1 - \frac{b\sigma^2w_1}{w_1} + 2b\sigma^2\tau(w_1)} dw_1 \]

\[ \int_0^1 w_2^{r+\frac{1}{2}-1}(1 - w_2)^{\frac{p-1}{2}-1} \frac{dw_2}{2^{\frac{p-1}{2}}\Gamma(\frac{p-1}{2})\Gamma(r+\frac{1}{2})} = \]

\[ e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{(||\eta||^2(b\sigma^2 + \frac{1}{2}))^r}{r!} \]

\[ \int_0^\infty (1 - \frac{b\sigma^2\tau(w_1)}{(b\sigma^2 + \frac{1}{2})w_1})^{2r}(b\sigma^2 + \frac{1}{2})^r w_1^{r+\frac{p}{2}-1} e^{-(b\sigma^2 + \frac{1}{2})w_1 - \frac{b\sigma^2w_1}{w_1} + 2b\sigma^2\tau(w_1)} \frac{dw_1}{2^{\frac{r}{2}}\Gamma(r + \frac{p}{2})} \]

Making the change of variables \( t = (b\sigma^2 + \frac{1}{2})w_1 \) in (5.9) establishes the result since (5.9) is equivalent to

\[ e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{(||\eta||^2(b\sigma^2 + \frac{1}{2}))^r}{r!} \]

after the change of variables.

\[ \Box \]

Proof of Theorem 3.4

**Proof.** Suppose \( \forall r \geq 0, I_b(r) \geq 1 \). Then the minimaxity of the estimator \( \delta_r(X) \) follows from the following inequalities:

\[ E_\theta[1 - e^{-b||\delta_r(X) - \theta||^2}] = \]

\[ 1 - \left[\left(\frac{1}{2b\sigma^2+1}\right)^{\frac{p}{2}} e^{-\phi} \sum_{r=0}^\infty \frac{\phi^r}{r!} I_b(r)\right] \leq \]

\[ 1 - \left[\left(\frac{1}{2b\sigma^2}\right)^{\frac{p}{2}} e^{-\phi} \sum_{r=0}^\infty \frac{\phi^r}{r!}\right] = \]

\[ 1 - \left(\frac{1}{2b\sigma^2+1}\right)^{\frac{p}{2}} E_\theta[1 - e^{-b||X - \theta||^2}] \]
where $\phi = (b + \frac{1}{2\sigma^2})\|\theta\|^2$. To show that $I_b(r) \geq 1$, let:

$$\tau_0(t) = \tau\left(\frac{-2t}{2b\sigma^2+1}\right)$$

$$q(t) = t\left(1 - \frac{b\sigma^2\tau_0(t)}{t}\right)^2$$

then

$$-q(t) = -t\left(1 - \frac{b\sigma^2\tau_0(t)}{t}\right)^2 = -t\left(1 + \frac{b^2\sigma^4\tau_0^2(t)}{t^2} - 2b\sigma^2\tau_0(t)\right) =$$

$$-t - \frac{b^2\sigma^4\tau_0^2(t)}{t} + 2b\sigma^2\tau_0(t)$$

so that the term:

$$e^{-t - \frac{b\sigma^2\tau_0(t)}{t} + 2b\sigma^2\tau_0(t)}$$

in $I_b(r)$ can be re-expressed as

$$e^{-t - \frac{b^2\sigma^4\tau_0^2(t)}{t} + 2b\sigma^2\tau_0(t)}$$

Furthermore the term

$$(1 - \frac{b\sigma^2\tau_0(t)}{t})^{2r}r^{r+\frac{p}{2}-1} = [(1 - \frac{b\sigma^2\tau_0(t)}{t})^{2}]^r * \left[(1 - \frac{b\sigma^2\tau_0(t)}{t})^{2}\right]^\frac{p}{2} - 1$$

in $I_b(r)$ can be re-expressed as

$$[(1 - \frac{b\sigma^2\tau_0(t)}{t})^{2}]^r + \frac{p}{2} - 1 * (1 - \frac{b\sigma^2\tau_0(t)}{t})^{-p-2} =$$

$$q(t)^{r+\frac{p}{2}} - 1 * (1 - \frac{b\sigma^2\tau_0(t)}{t})^{-p-2}.$$  

Substitution of (5.11) and (5.13) into $I_b(r)$ implies

$$I_b(r) = \int_0^\infty e^{-q(t)}(q(t))^{r+\frac{p}{2}} - 1 * e^{-\frac{b\sigma^2\tau_0^2(t)}{2t}} \{1 - \frac{b\sigma^2\tau_0(t)}{t}\}^{-p-2} \frac{dt}{\Gamma(r+\frac{p}{2})}.$$  

Let
\[ k = \sup \{ t > 0 | \frac{\tau(t)}{2t + 1} \geq \frac{1}{b\sigma^2} \}. \]

Since \( \tau \) is continuous, positive, and bounded by assumption, the \( \lim_{u \to 0} \frac{\tau}{2u + 1} = \infty \) and 
\( \lim_{u \to \infty} \frac{\tau}{2u + 1} = 0. \) By the continuity of \( \frac{\tau}{2u + 1} \) for \( u > 0, \) there exist a \( k > 0 \) such that 
\( \frac{\tau}{2k} = \frac{1}{\sigma^2} \) and \( \forall t > k, \frac{\tau(t)}{2t + 1} \leq \frac{1}{b\sigma^2}. \)

For \( c > 0 \) and \( 0 \leq z < 1 \) the inequality:
\[ (\frac{1}{1-z})^c \geq e^{cz} \]

implies for \( t \geq k \)
\[ e^{-\frac{b\sigma^2 \tau^2(t)}{2t}} \{ 1 - \frac{b\sigma^2 \tau_0(t)}{t} \}^{-(p-2)} \geq e^{-\frac{b\sigma^2 \tau^2(t)}{2t}} e^{(p-2)\frac{b\sigma^2 \tau_0(t)}{t}} = e^{-\frac{b\sigma^2 \tau_0(t)(\tau_0(t)-2(p-2))}{t}} \geq 1 \]

since \( \tau_0(t) \leq 2(p-2) \) by assumption i). For \( t \geq k \) the derivative with respect to \( t \) of \( q(t) \) satisfies
\[ q'(t) = 1 - 2b\sigma^2 \tau'_0(t) + b^2 \sigma^4 \left[ \frac{2\tau_0(t)}{t^2} \right] \]

as \( b^2 \sigma^4 \frac{\tau_0(t)}{t^2} \) is non-negative and \( 2b\sigma^2 \tau'_0(t)(1 - \frac{b\sigma^2 \tau_0(t)}{t}) \) is non-negative as \( \tau'_0(t) \) is non-negative by assumption ii) and \( 1 - \frac{b\sigma^2 \tau_0(t)}{t} \geq 0, \) and
\[ e^{-q(t)(q(t))^{r+\frac{p}{2}-1} e^{-\frac{b\sigma^2 \tau^2(t)}{2t}} \{ 1 - \frac{b\sigma^2 \tau_0(t)}{t} \}^{-(p-2)} \geq 0. \]

Therefore for \( t \geq k \)
\[ I_b(r) \geq \int_k^\infty e^{-q(t)(q(t))^{r+\frac{p}{2}-1} e^{-\frac{b\sigma^2 \tau^2(t)}{2t}} \{ 1 - \frac{b\sigma^2 \tau_0(t)}{t} \}^{-(p-2)} \frac{dt}{\Gamma(r+\frac{p}{2})} \geq \int_k^\infty e^{-q(t)(q(t))^{r+\frac{p}{2}-1} e^{-\frac{b\sigma^2 \tau_0(t)}{2t}} (\tau_0(t)-2(p-2)) \frac{dt}{\Gamma(r+\frac{p}{2})} \geq \int_k^\infty e^{-q(t)(q(t))^{r+\frac{p}{2}-1} q'(t) \frac{dt}{\Gamma(r+\frac{p}{2})} = 1 \] (5.14)

after the substitution \( u = q(t) \) in the (5.14).
5.2 Appendix B: Proof of Lemma 3.4

Proof of Lemma 3.4

Proof.

\[ E_\theta \left[ e^{-b\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \right] = e^{-b\sigma^2\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \]  \quad (5.15)

\[ E_\theta \left[ e^{-b\sigma^2\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \right] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^p} e^{-b\sigma^2\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \|x\|^2 f(x|\sigma^2)dx f(u|\sigma^2)du. \]  \quad (5.16)

where the expectation in (5.16) is taken with respect to the density of \( \begin{pmatrix} X \\ U \end{pmatrix} \). Since

\[ f((x, u)|\sigma^2) = (2\pi\sigma^2)^{-\frac{p+m}{2}} e^{-\frac{\|x-u\|^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{p}{2}} e^{-\frac{\|x\|^2}{2\sigma^2}} (2\pi\sigma^2)^{-\frac{m}{2}} e^{-\frac{\|u\|^2}{2\sigma^2}} = f_1(x|\sigma^2)f_2(u|\sigma^2), \]

the distribution of \( X \) is independent of the distribution of \( U \) so that

\[ E_\theta \left[ e^{-b\sigma^2\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \right] = \int_{\mathbb{R}^m} \int_{\mathbb{R}^p} e^{-b\sigma^2\|X - \theta\|^2} \left( \frac{\|U\|^2 r}{\|X\|^2} \right) X - \theta \|x\|^2 f(x|\sigma^2)dx f(u|\sigma^2)du. \]  \quad (5.17)

Let \( Z = \frac{X}{\sigma}, U^* = \frac{\|U\|^2}{\sigma^2(m+2)}, \) and \( \eta = \frac{\theta}{\sigma} \) in (5.17) so that (5.17) is equivalent to

\[ \int_0^\infty \left[ \int_{\mathbb{R}^p} e^{-b\sigma^2\left(1-\frac{u^* r}{\|x\|^2}\right)\|z-\eta\|^2} \left( 2\pi \right)^{-\frac{p}{2}} e^{-\frac{\|z\|^2}{2\sigma^2}} dz \right] e^{\frac{u^* (m+2)}{2}} \Gamma\left(\frac{m}{2}\right) \frac{1}{\Gamma\left(\frac{m+2}{2}\right)} du^* = \]

\[ E[E_\theta \left[ e^{-b\sigma^2\left(1-\frac{u^* r}{\|x\|^2}\right)\|z-\eta\|^2} \right]] \]  \quad (5.18)

where the inner expectation (5.18) is taken with respect to the \( N_p(\eta, I_p) \) distribution, and the outer expectation in (5.18) taken with respect to a \( Gamma\left(\frac{m}{2}, \frac{2}{m+2}\right) \) distribution.

Letting \( W = \frac{U^* (m+2)}{2}, \) (5.18) becomes

\[ \int_0^\infty \left[ \int_{\mathbb{R}^p} e^{-b\sigma^2\left(1-\frac{w^* r}{\|x\|^2}\right)\|z-\eta\|^2} \left( 2\pi \right)^{-\frac{p}{2}} e^{-\frac{\|z\|^2}{2\sigma^2}} dz \right] e^{-w^* \frac{m}{2} - \frac{1}{2}} \Gamma\left(\frac{m}{2}\right) dw \]  \quad (5.19)
where \( \tau_0(\frac{t}{w}) = \frac{2\tau(\frac{t+2\tau}{w})}{m+2} \) in (5.19). The inner integral in (5.19),
\[
\int_{\mathbb{R}^p} e^{-\sigma^2\|z-\eta\|^2} (2\pi)^{-\frac{p}{2}} e^{-\frac{\|z-\eta\|^2}{2}} dz =
\]
\[
E_\eta \left[ e^{-\sigma^2\|Z\|^2 + \frac{w^2(\|\eta\|^2 + \|Z\|^2)}{2}} - 2\sigma^2\|\eta\|^2 \right]
\]
(5.20)
\[
E_\eta \left[ e^{-\sigma^2\|Z\|^2 + \frac{w^2(\|\eta\|^2 + \|Z\|^2)}{2}} - 2\sigma^2\|\eta\|^2 \right] =
\]
(5.21)

where the expectation in (5.20) and (5.21) are with respect to a \( N_p(\eta, I_p) \) distribution.

To simplify (5.21), let \( Y = CZ \), where \( C \) is an orthogonal transformation with first row equal to \((\theta_{11}, ..., \theta_{p1})'\). Each \( Y_i \) are mutually independent as the co-variance of \( Y \) is the identity matrix and \( Y_1 \sim N(\|\eta\|, 1) \), and \( Y_i \sim N(0, 1) \) for \( i = 2, ..., n \). Then \( Z = \sum_{i=1}^n Y_i^2 \) has a \( \chi^2_{(p-1)} \) distribution so that (5.21) becomes
\[
\int_0^\infty \int_{\mathbb{R}} e^{-\sigma^2 y^2 + \frac{w^2(\eta^2 + \|Z\|^2)}{2}} - 2\sigma^2\|\eta\|^2 \frac{2\eta w \tau_0}{y^2} dy_1 dz =
\]
(5.22)
\[
e^{-\eta^2 \sum y^2} \frac{u+1}{\sqrt{2\pi}} \frac{dz}{G(\frac{u+1}{2})}
\]

Looking at the inner integral of (5.22),
\[
\int_{\mathbb{R}} e^{-\sigma^2 \frac{y}{2} + \frac{1}{2}} (y_1 - \|\eta\||y_1 - \|\eta\|) - \sigma^2 w^2 \frac{y^2 \tau_0}{y_1^2} + 2\sigma^2 w \tau_0 \frac{y_1 y^2}{y_1^2} dy_1 =
\]
\[
\int_{\mathbb{R}} e^{-\sigma^2 \frac{y}{2} + \frac{1}{2}} (y_1 - \|\eta\||y_1 - \|\eta\|) - \sigma^2 w^2 \frac{y^2 \tau_0}{y_1^2} + 2\sigma^2 w \tau_0 \frac{y_1 y^2}{y_1^2} dy_1 =
\]
(5.23)

Using a taylor series expansion about the point 0, expression (5.23)
\[ e^{2(b\sigma^2 + \frac{1}{2})\|\eta\|^2 y_1 - 2b\sigma^2\|\eta\|^2 y_1 w_{\alpha}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})} + e^{-2(b\sigma^2 + \frac{1}{2})\|\eta\|^2 y_1 + 2b\sigma^2\|\eta\|^2 y_1 w_{\alpha}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})} = \]

\[ 2 \sum_{r=0}^{\infty} \frac{(2\|\eta\|y_1)^{2r}}{(2r)!} [(b\sigma^2 + \frac{1}{2}) - \frac{b\sigma^2 w_{\eta}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})}{y_{1}^{2+z}}]^{2r} = \]

\[ 2 \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r (y_{1}^{2})^r (b\sigma^2 + \frac{1}{2})^r [1 - \frac{b\sigma^2 w_{\eta}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})}{(b\sigma^2 + \frac{1}{2}) (y_{1}^{2+z})}]^{2r}}{(2r)!} = \]

\[ 2 \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r (y_{1}^{2})^r (b\sigma^2 + \frac{1}{2})^r [1 - \frac{b\sigma^2 w_{\eta}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})}{(b\sigma^2 + \frac{1}{2}) (y_{1}^{2+z})}]^{2r}}{r! \Gamma(r + \frac{1}{2}) \pi^{-\frac{1}{2}}} \]

(5.24)

where equality $\Gamma(2r + 1) = \Gamma(r + \frac{1}{2}) \Gamma(r + 1) 2^{2r} \pi^{-\frac{1}{2}}$ is used in (5.24). Therefore the inner integral of (5.22) becomes

\[ 2 \sum_{r=0}^{\infty} \frac{(2\|\eta\|y_1)^{2r}}{(2r)!} [(b\sigma^2 + \frac{1}{2}) - \frac{b\sigma^2 w_{\eta}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})}{y_{1}^{2+z}}]^{2r} \]

\[ e^{-(b\sigma^2 + \frac{1}{2})\|\eta\|^2 y_1 - 2b\sigma^2\|\eta\|^2 y_1 w_{\alpha}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})} \int_0^{\infty} (y_{1}^{2})^r (b\sigma^2 + \frac{1}{2})^r [1 - \frac{b\sigma^2 w_{\eta}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})}{(b\sigma^2 + \frac{1}{2}) (y_{1}^{2+z})}]^{2r} \]

(5.25)

To simplify expression (5.25) let $u = y_1^2$ so that (5.25) is expressible as

\[ e^{-(b\sigma^2 + \frac{1}{2})\|\eta\|^2 u - 2b\sigma^2\|\eta\|^2 u w_{\alpha}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})} \int_0^{\infty} u^{r-\frac{1}{2}} (b\sigma^2 + \frac{1}{2})^r [1 - \frac{b\sigma^2 w_{\eta}(\frac{u+z}{u})}{(b\sigma^2 + \frac{1}{2}) (u+z)}]^{2r} \]

(5.26)

Substituting expression (5.26) into (5.22) yields

\[ \int_0^{\infty} \int_0^{\infty} e^{-(b\sigma^2 + \frac{1}{2})\|\eta\|^2 u - 2b\sigma^2\|\eta\|^2 u w_{\alpha}(\frac{y_{1}^{2+z}}{y_{1}^{2+z}})} \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} \]

\[ u^{r-\frac{1}{2}} (b\sigma^2 + \frac{1}{2})^r [1 - \frac{b\sigma^2 w_{\eta}(\frac{u+z}{u})}{(b\sigma^2 + \frac{1}{2}) (u+z)}]^{2r} e^{-(b\sigma^2 + \frac{1}{2})u - 2b\sigma^2 u w_{\alpha}(\frac{u+z}{u})} \]

\[ e^{-(b\sigma^2 + \frac{1}{2})z^2 \frac{u-z}{2}} \frac{du}{\Gamma(r + \frac{1}{2}) 2^{2r} \Gamma(\frac{r-1}{2})} \]

(5.27)

To evaluate the double integral let,

\[ v_1 = u + z \quad \text{and} \quad v_2 = \frac{u}{u+z} \quad \text{so that} \quad du dv = |v_1| dv_1 dv_2 \]
so that (5.27) becomes:

\[
\int_0^1 \int_0^\infty e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{((b\sigma^2 + \frac{1}{2})||\eta||^2)^r}{r!} v_1^{r-\frac{1}{2}} v_2^{r-\frac{1}{2}} (b\sigma^2 + \frac{1}{2}) v_1^2 [1 - \frac{b\sigma^2 w_\tau v_1}{(b\sigma^2 + \frac{1}{2}) v_1}] 2r e^{-(b\sigma^2 + \frac{1}{2}) v_1 v_2 - b\sigma^2 w^2 \frac{2b\sigma^2 w v_1}{v_1^2} + 2b\sigma^2 w_\tau v_1} \ast \\
e^{-(b\sigma^2 + \frac{1}{2})(v_1 - v_2)} v_1^{\frac{p-3}{4}} (1 - v_2)^{\frac{p-1}{4}} v_1 \frac{dv_1}{\Gamma(r + \frac{1}{2})} \frac{dv_2}{2^\frac{p-1}{2} \Gamma(\frac{p-1}{2})} 
\]

(5.28)

Integration of \( v_2 \) in expression (5.28)

\[
\int_0^\infty v_1^{r+\frac{p-1}{2}} (b\sigma^2 + \frac{1}{2}) v_1^2 [1 - \frac{b\sigma^2 w_\tau v_1}{(b\sigma^2 + \frac{1}{2}) v_1}] 2r e^{(b\sigma^2 + \frac{1}{2}) v_1 - b\sigma^2 w^2 \frac{2b\sigma^2 w v_1}{v_1^2} + 2b\sigma^2 w_\tau v_1} dv_1 = \\
\int_0^\infty v_1^{r+\frac{p-1}{2}} (b\sigma^2 + \frac{1}{2}) v_1^2 [1 - \frac{b\sigma^2 w_\tau v_1}{(b\sigma^2 + \frac{1}{2}) v_1}] 2r e^{(b\sigma^2 + \frac{1}{2}) v_1 - b\sigma^2 w^2 \frac{2b\sigma^2 w v_1}{v_1^2} + 2b\sigma^2 w_\tau v_1} dv_1 \\
= e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{((b\sigma^2 + \frac{1}{2})||\eta||^2)^r}{r!} I_{b\sigma^2}(r) 
\]

(5.29)

Setting \( t = (b\sigma^2 + \frac{1}{2}) v_1 \), expression (5.29) is equivalent to

\[
\frac{1}{2b\sigma^2 + 1} \frac{1}{2} e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{((b\sigma^2 + \frac{1}{2})||\eta||^2)^r}{r!} I_{b\sigma^2}(r) 
\]

(5.30)

where

\[
I_{b\sigma^2}(r) = \int_0^\infty t^{r+\frac{p-1}{2}} [1 - \frac{b\sigma^2 w}{t} w_\tau v_1 (\frac{t}{(b\sigma^2 + \frac{1}{2}) v_1})] e^{-\frac{t}{2} - \frac{b\sigma^2 w}{t} w_\tau v_1 (\frac{t}{(b\sigma^2 + \frac{1}{2}) v_1})} \frac{dt}{\Gamma(r + \frac{p-1}{2})} 
\]

(5.31)

Thus expression using (5.30), expression (5.19),

\[
\int_0^\infty \int_{\mathbb{R}^2} e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} (1 - \frac{u^v}{||\eta||^2}) z - \eta^2 (2\pi)^{-\frac{p}{2}} e^{\frac{-||z - \eta||^2}{2}} \frac{dz}{1(||z||^2)} \frac{e^{-w \frac{w^2}{2} - 1}}{\Gamma(\frac{p}{2})} \\
= \frac{1}{2b\sigma^2 + 1} \frac{1}{2} e^{-(b\sigma^2 + \frac{1}{2})||\eta||^2} \sum_{r=0}^\infty \frac{((b\sigma^2 + \frac{1}{2})||\eta||^2)^r}{r!} I_{b\sigma^2}(r) e^{-w \frac{w^2}{2} - 1} \frac{dw}{\Gamma(\frac{p}{2})} 
\]
Defining \( I^{*}_{b\sigma^2}(r) = \int_{0}^{\infty} I_{b\sigma^2}(r) e^{-u^2} w^m \frac{du}{\Gamma}\), if \( I^{*}_{b\sigma^2}(r) > 1 \) for every \( r \) then

\[
E_\theta[e^{-b\|Y\|^2}] = \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
E_\theta[e^{-b\|X\|^2}] = \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

so that \( R(X, \theta) > R(\delta_r, \theta) \) establishing the result.

To that end let \( t = wu \) in \( I^{*}_{b\sigma}(r) \) so that

\[
I^{*}_{b\sigma}(r) = \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
e^{-t - \frac{b\sigma^2(t)}{2} w^2} \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
e^{-t} \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
\int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
e^{-t} \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} \int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
\int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

\[
\int_{0}^{\infty} I \sum_{r=0}^{\infty} \frac{((b\sigma^2 + \frac{1}{2})\|\eta\|^2)^r}{r!} I^{*}_{b\sigma^2}(r) > 0
\]

Defining \( u_\sigma = \sup\{u > 0 | \frac{\tau(u)}{u} \geq \frac{1}{b\sigma^2}\} = \sup\{u > 0 | \frac{\tau(u)}{u} \geq \frac{1}{b\sigma^2}\} > 0 \) follows form \( \lim_{u \to \infty} \frac{\tau(u)}{u} = 0 \) and \( \lim_{u \to \infty} \frac{\tau(u)}{u} = \infty \) and the continuity of the function \( \frac{\tau(u)}{u} \) so that exist a \( u \), such that \( \frac{\tau(u)}{u} \) is less than \( 1 + \frac{1}{2b\sigma^2} \) and so for all other values greater than this value the value of \( \frac{\tau(u)}{u} \) will be less than \( \frac{1}{b\sigma^2} \). Therefore \( I^{*}_{b\sigma^2}(r) \) satisfies

\[
I^{*}_{b\sigma^2}(r) > \int_{u_\sigma}^{\infty} (u + 1 + \frac{b\sigma^2(u+\frac{1}{2})}{u}) \tau(u) \frac{du}{B(r+\frac{1}{2}, \frac{m}{2})}
\]
For ease of notation, let \( \tau_{0,\sigma}(u) = \tau_0\left(\frac{u}{(ba^2+\frac{1}{2})}\right) \) so that

\[
I_{ba^2}^*(r) > \int_{u_\sigma}^\infty \frac{\left[u(1 - \frac{ba^2\tau_0^2(u)}{u})\right]^2}{1 + \frac{ba^2\tau_0^2(u)}{2u}} \, du
\]

(5.31)

\[
[1 - \frac{ba^2\tau_0^2(u)}{u}]^{-(p+2)}[1 + \frac{ba^2\tau_0^2(u)}{2u}]^{-(\frac{m}{2}+1)} \geq \frac{e^{(p-2)\frac{ba^2\tau_0^2(u)}{u} - \left(\frac{m}{2}+1\right)\frac{ba^2\tau_0^2(u)}{2u}}}{e^{\frac{ba^2\tau_0^2(u)}{u}[(p-2)\frac{ba^2\tau_0^2(u)}{2u}] - \tau_0^2(u)} > 1
\]

if \( 0 < \tau_{0,\sigma}(u) < \frac{4(p-2)}{m+2} \) which is satisfied by assumption since \( 0 < \tau(t) < 2(p-2) \) for all \( t > 0 \).

Define \( q(u) = \frac{u(1 - \frac{ba^2\tau_0^2(u)}{u})}{1 + \frac{ba^2\tau_0^2(u)}{2u}} \) and supposing that \( q'(u) \leq 1 \) for all \( u > u_\sigma \),

\[
I_{ba^2}^*(r) > \int_{u_\sigma}^\infty \frac{\left[u(1 - \frac{ba^2\tau_0^2(u)}{u})\right]^2}{1 + \frac{ba^2\tau_0^2(u)}{2u}} \, du
\]

(5.32)
\[
\int_{u_\sigma}^{\infty} \frac{(q(u))^{r+\frac{q}{2}-1}}{(1+q(u))^{r+\frac{q}{2}+\frac{m}{2}}} q'(u) \frac{du}{B(r+\frac{q}{2}, \frac{m}{2})} = 1. \tag{5.33}
\]

by the change of variables \( w = q(u) \) in (5.33).

To show \( q'(u) \leq 1 \) for all \( u > u_\sigma \),

\[
q(u) = \frac{u(1-\frac{\sigma^2 \tau(u)}{2u})^2}{1+\frac{\sigma^2 \tau(u)}{2u}} = \frac{2(u-\beta \sigma^2 \tau_0(u))}{2u+\beta \sigma^2 \tau_0(u)}
\]

so that:

\[
q'(u) = \frac{(2u+\beta \sigma^2 \tau_0(u))4(u-\beta \sigma^2 \tau_0(u))4(1-\beta \sigma^2 \tau_0(u))-2(u-\beta \sigma^2 \tau_0(u))2(2+2\beta \sigma^2 \tau_0(u))\tau_0(u)}{(2u+\beta \sigma^2 \tau_0(u))^2} = \frac{2(u-\beta \sigma^2 \tau_0(u))[4(1-\beta \sigma^2 \tau_0(u))-2(u-\beta \sigma^2 \tau_0(u))2(2+2\beta \sigma^2 \tau_0(u))\tau_0(u)]}{(2u+\beta \sigma^2 \tau_0(u))^2} \tag{5.34}
\]

For each of notation let \( h(u) = \tau_0(u) \). The numerator in (5.34) becomes after expansion

\[
4u^2 + 4ub\sigma^2h(u) - 8u^2b\sigma^2h'(u) - 4u^2b\sigma^2h(u)h'(u) - 4b^2\sigma^4h^2(u) - 4b^2\sigma^4h^3(u) + 8ub^2\sigma^4h(u)h'(u) + 4ub^2\sigma^4h^2(u)h'(u)
\]

Since \( q'(u) \leq 1 \) iff

\[
4u^2 + 4ub\sigma^2h(u) - 8u^2b\sigma^2h'(u) - 4u^2b\sigma^2h(u)h'(u) - 4b^2\sigma^4h^2(u) - 4b^2\sigma^4h^3(u) + 8ub^2\sigma^4h(u)h'(u) + 4ub^2\sigma^4h^2(u)h'(u) \leq 4u^2 + b^2\sigma^4h^4(u) + 4b\sigma^2h^2(u)
\]

or equivalently:

\[
b^2\sigma^4h^4(u) + 4b^2\sigma^4h^2(u) + 4b^2\sigma^4h^3(u) + 8u^2b\sigma^2h'(u) + 4u^2b\sigma^2h(u)h'(u) - 8ub^2\sigma^4h'(u)h(u) - 4ub^2\sigma^4h'(u)h^2(u) =
\]

\[
b^2\sigma^4h^2(u)[h^2(u) + 4 + 4h(u)] + 4ub\sigma^2h'(u)[2u + uh(u) - 2b\sigma^2h(u) + b\sigma^2h^2(u)] =
\]
\[ b^2 \sigma^4 h^2(u)[h(u) + 2]^2 + 4ub\sigma^2 h'(u)[u(2 + h(u)) - b\sigma^2 h(u)(2 + h(u))] = \]

\[ b^2 \sigma^4 h^2(u)[h(u) + 2]^2 + 4ub\sigma^2 h'(u)[(2 + h(u))(u - b\sigma^2 h(u))] \geq 0 \quad (5.35) \]

Since the (5.35) holds for \( u > u_\sigma \) as \( \frac{\tau_0 \left( \frac{u}{b\sigma^2 + \frac{1}{2}} \right)}{u} \leq \frac{1}{b\sigma^2} \iff u - b\sigma^2 \tau_0 \left( \frac{u}{b\sigma^2 + \frac{1}{2}} \right) \geq 0 \) and since \( h'(u) \geq 0 \) follows from the assumption \( \tau' \geq 0 \Rightarrow \tau'_0 \left( \frac{u}{b\sigma^2 + \frac{1}{2}} \right) \frac{1}{b\sigma^2 + \frac{1}{2}} = h'(u) \geq 0 \). Therefore \( q'(u) \leq 1 \) for \( u \geq u_\sigma \) so that \( I^*_{b\sigma^2}(r) > 1 \) establishing the result.
Bibliography


