A theoretical study of Stein's covariance estimator

Bala Rajaratnam, Dario Vincenzi

A theoretical study of Stein's covariance estimator

Bala Rajaratnam[a], Dario Vincenzi[b]

[a]*Department of Statistics, 390 Serra Mall - Sequoia Hall, Stanford University, Stanford, CA 94305, USA*
[b]*Université Nice Sophia Antipolis, CNRS, Laboratoire Jean Alexandre Dieudonné, UMR 7351, 06100 Nice, France*

Abstract

The traditional covariance estimator, the sample covariance matrix (which is also the MLE), is known to be a poor estimator, unless the sample size is much larger than the dimension of the covariance matrix. Stein's estimator has often been regarded as a much better alternative to the MLE in small sample sizes and is traditionally used with an isotonizing algorithm, the purpose of which is to retain positivity and the original order of the sample eigenvalues. Despite the superior performance of Stein's isotonized estimator in numerical investigations, its theoretical properties have not been explored in detail, and important questions still remain unanswered. One particular question of interest is to identify the regimes under which Stein's estimator is guaranteed to perform well. A second goal is to determine the extent to which the performance of Stein's estimator depends on the isotonizing algorithm. The presence of the ad hoc isotonizing algorithm, however, renders a theoretical analysis rather difficult, and consequently risk functions are not easily quantifiable for comparison purposes. Hence formal decision theoretical results are difficult to obtain and have been elusive ever since the estimator was introduced. Despite these hurdles, in this paper we show that an analysis of Stein's covariance estimator within the unbiased estimator of risk (UBEOR) framework can nevertheless lead to important theoretical and methodological insights that are relevant for applications. Our analysis demonstrates that Stein's estimator may give only modest risk reductions when it is not isotonized, and when it is isotonized, the risk reductions can be significant. In particular, three broad regimes are identified regarding the behavior of Stein's UBEOR. The theoretical insights are then affirmed at the level of risk functions via numerical simulations.

Keywords: Covariance estimation, Unbiased estimator of risk, Eigenvalues, Shrinkage, Steinian estimation

1. Introduction

The covariance matrix estimation problem arises in various applications ranging from genomics and environmental sciences to geophysics and finance [3–5, 7, 11, 16]. The covariance matrix is a critical ingredient in many statistical procedures such as principal components analysis, discriminant analysis, etc. and serves to provide estimates of the spectrum and other statistical quantities. The sample covariance matrix S, which is also the maximum likelihood estimator (MLE), is a poor estimator unless the sample size n is much larger than the dimension p. Many modern applications are not always endowed with large sample sizes, and thus standard procedures can lead to highly non-optimal estimates of the population covariance matrix, with possibly undesirable consequences. Obtaining good estimates of the covariance matrix is a challenging problem for many reasons, not excluding the fact that (a) the number of entries in the population covariance matrix is of $O(p^2)$, and (b) covariance matrices lie in the open convex cone of positive definite matrices. Research in many scientific fields have contributed to the body of work on this important topic and various estimators have been proposed in the literature (see Refs. [5, 8, 9] for a comprehensive summary). Several of these are in the traditional Steinian type framework (see [1, 2, 4, 5, 16] to name just a few) and yield estimators which "shrink" sample quantities like the sample covariance matrix to a desired target covariance matrix.

A method, proposed by Stein himself [13–15], takes an interesting approach that lies within the shrinkage framework. Stein notes that the sample spectrum is severely distorted unless $n \gg p$. In particular, Stein observes that the larger sample eigenvalues tend to significantly overestimate their population counterparts, whereas the smaller sample eigenvalues tend to significantly underestimate their population counterparts. This leads to a larger spread in the sample spectrum as compared to the population spectrum. He notes that one possible way to address this problem

is to "shrink" the sample eigenvalues close together. This modification could potentially serve to overcome the distortion of the spectrum. Stein therefore considers the class of orthogonally invariant estimators: such estimators only modify the sample eigenvalues and leave the sample eigenvectors untouched. To this end, Stein derives the so-called unbiased estimator of risk (UBEOR) for the covariance estimation problem. This approach allows Stein to choose an optimal way to shrink the sample eigenvalues under the entropy loss function. The resulting estimator, often dubbed *Stein's covariance estimator*, has been considered a gold standard in the literature [1, 5, 6]. Though Stein's estimator requires $n > p$ in order to be well defined, it has one important distinct advantage over many contemporary methods, especially those inspired in the lasso framework. Stein's estimator does not impose sparsity: a possibly restrictive assumption which can be rather unrealistic in many applications. Numerical studies have demonstrated the superior risk properties of Stein's estimator. The modification of the sample eigenvalues in Stein's estimator can however lead to negative eigenvalue estimates or a different ordering from the original sample spectra. To mitigate this, Stein proposes an isotonizing algorithm that pools sample eigenvalues together so that the original spectral ordering and positivity are retained.

A numerical study of Stein's estimator has been previously performed by Lin and Perlman [6]. However, to our knowledge, the theoretical properties of Stein's estimator have not been formally studied in detail because of the presence of the isotonizing algorithm. This algorithm, which has no formal statistical basis, to a large extent prevents a theoretical analysis of the estimator. In this paper we systematically study the theoretical properties of Stein's estimator by formally investigating the use of the unbiased estimator of risk (UBEOR) approach that Stein proposes. The merits of the UBEOR approach can be seen from the following line of reasoning. According to Stein's theory, the expectation of the UBEOR is equal to the overall risk. The form of the population covariance matrix determines which part of the domain of the UBEOR contributes mostly to this risk. By combining the knowledge of the values that the population covariance matrix and UBEOR take, it is thus possible to better understand the behavior of the risk in different parts of the parameter space.

The paper is divided into two main components to facilitate the ultimate goal of understanding the properties of Stein's isotonized estimator. The first component analyzes Stein's "raw" estimator (that is without the isotonizing algorithm), whereas the second component analyzes Stein's estimator coupled with the isotonizing algorithm. We first derive some basic properties of Stein's raw estimator, formally characterize the regimes in which it does not yield negative eigenvalue estimates, and compare it with the MLE within the UBEOR approach. More specifically, the UBEOR of Stein's raw estimator is lower in value than that of the MLE only when the eigenvalues are far apart. Our theoretical analysis also reveals that the region within the set of ordered eigenvalues where the UBEOR of Stein's raw estimator is lower than the MLE is rather small. Thus, the UBEOR approach indicates that Stein's raw estimator may not yield significant risk reductions over the MLE. A formal analysis when p is large indicates that this phenomenon is more pronounced in higher dimensions. This question then naturally leads to investigating Stein's estimator with the isotonizing algorithm, since reported results systematically use this order preserving procedure. First, a careful analysis in small dimensions reveals that the isotonizing algorithm produces a significant drop in the UBEOR. Substantial reductions are mainly observed when the eigenvalues are close to each other or when the sample size is small. In these cases, Stein's estimator makes extensive use of the isotonizing correction because a large number of order and sign violations would be obtained otherwise. However, there exists an intermediate regime in which the sample eigenvalues are only moderately separated and simultaneously n/p is sufficiently large. In this regime, the UBEOR of Stein's raw estimator exceeds that of the MLE and Stein's covariance estimator makes only moderate use of the isotonizing algorithm; as a result, the performance of Stein's isotonized estimator is poor.

Hence, the theoretical investigation in this paper allows us to explain how the performance of Stein's isotonized estimator depends on the eigenstructure of the theoretical covariance matrix and to explore the role of the ad hoc isotonizing algorithm in risk reductions. These insights are then validated at the level of risk functions via Monte Carlo simulations. The understanding that we establish is also useful in practical applications, since there are many instances where broad prior knowledge of covariance regimes is available. Thus, our analysis lets a practitioner understand when Stein's estimator is expected to perform better than the MLE and what is driving the risk reductions.

The approach in this paper is not designed to lead to a traditional optimality or decision-theoretic result. Existing numerical results already demonstrate that there are some settings in which Stein's estimator gives slightly higher risk than the MLE. Our primary objective rather is to demonstrate that the UBEOR approach yields a deeper theoretical understanding of the behavior of Stein's isotonized estimator.

The remainder of the paper is structured as follows. Sections 2 and 3 introduce Stein's estimator and the UBEOR

framework. Section 4 undertakes a detailed study of the risk properties of Stein's raw estimator (that is without the isotonizing algorithm). Section 5 undertakes a formal analysis of Stein's estimator when used together with the isotonizing algorithm. Section 6 validates the effectiveness of the UBEOR approach via risk calculations. Section 7 concludes by summarizing the work in the paper. A supplemental section with appendices, which is not a part of the paper, is also provided and serves to give more detail on some of the results in the paper.

2. Preliminaries

In this section, we briefly recall the derivation of Stein's raw estimator and of its isotonized version. Consider a random sample, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$, from a p-dimensional normal distribution $\mathcal{N}_p(0, \Sigma)$ with $n \geqslant p$. The sample covariance matrix S (up to a multiplicative constant) is defined as $S = \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^t$. Note that $S \sim W_p(\Sigma, n)$, where $W_p(\Sigma, n)$ denotes the p-dimensional Wishart distribution with scale matrix Σ and n degrees of freedom. The matrix S can be written as $S = HLH^t$, where H is orthogonal, and $L = \text{diag}(l_1, l_2, ..., l_p)$ with $l_1 \geqslant l_2 \geqslant ... \geqslant l_p > 0$ being the ordered eigenvalues of S. Stein considers the class of *orthogonally invariant* estimators of Σ of the general form [13–15]:

$$\widehat{\Sigma} = H\Phi(\boldsymbol{l})H^t, \tag{1}$$

where $\Phi(\boldsymbol{l}) = \text{diag}(\varphi_1(\boldsymbol{l}), \varphi_2(\boldsymbol{l}), ..., \varphi_p(\boldsymbol{l}))$ and $\varphi_j = \varphi_j(\boldsymbol{l})$ estimates the jth largest eigenvalue of Σ, $j = 1, 2, ..., p$.

The maximum likelihood estimator S/n corresponds to the choice:

$$\widehat{\varphi}_j^{\text{ml}}(\boldsymbol{l}) := l_j/n. \tag{2}$$

The estimator $\widehat{\varphi}_j^{\text{ml}}(\boldsymbol{l})$ is known to be biased upwards for larger eigenvalues and biased downwards for smaller eigenvalues. Hence, the sample spectrum "severely distorts" the population eigenvalues [6]. These biases can be significant when p/n is not small or when $\Sigma \approx kI$ for some $k > 0$. Stein [13–15] proposes an approach to rectify this problem by obtaining improved eigenvalue estimates and using them in estimators of the form in (1). Stein uses a fundamental Wishart identity to obtain the risk of the estimator $\widehat{\Sigma}$ in (1) under the entropy loss function:

$$L_1(\widehat{\Sigma}, \Sigma) = \text{tr}(\widehat{\Sigma}\Sigma^{-1}) - \ln\det(\widehat{\Sigma}\Sigma^{-1}) - p.$$

Stein's loss function above has several key advantages in that it is generally tractable, allows for the calculation of the UBEOR, and has a functional form that is similar to the Gaussian likelihood. Under this loss function Stein obtains the following expression for the risk:

$$R_1(\widehat{\Sigma}, \Sigma) := \mathbb{E}_\Sigma[L_1(\widehat{\Sigma}, \Sigma)] = \mathbb{E}_\Sigma[F(\boldsymbol{l})] \tag{3}$$

with

$$F(\boldsymbol{l}) = \sum_{j=1}^{p} \left[\left(n - p + 1 + 2l_j \sum_{i \neq j} \frac{1}{l_j - l_i} \right) \psi_j(\boldsymbol{l}) + 2l_j \frac{\partial \psi_j}{\partial l_j} - \ln(\psi_j(\boldsymbol{l})) \right] - c_{p,n}, \tag{4}$$

where $\psi_j(\boldsymbol{l}) := \phi_j(\boldsymbol{l})/l_j$ and $c_{p,n}$ is defined as follows:

$$c_{p,n} = \mathbb{E}\Big(\sum_{j=1}^{p} \ln \chi^2_{n-j+1} \Big) + p = \sum_{j=1}^{p} \frac{\Gamma'(\frac{1}{2}(n-j+1))}{\Gamma(\frac{1}{2}(n-j+1))} + p\ln 2 + p. \tag{5}$$

Stein observes that $F(\boldsymbol{l})$ is an unbiased estimator of the risk (UBEOR) of $\widehat{\Sigma}$ [13–15], and notes that this formulation can be used to choose $\psi_j = \psi_j(\boldsymbol{l})$ in order to minimize $F(\boldsymbol{l})$. When the derivatives $\partial\psi_j/\partial l_j$ are disregarded, minimizing the UBEOR with respect to the ψ_j for $j = 1, 2, ..., p$ yields:

$$\widehat{\psi}_j^{\text{St}}(\boldsymbol{l}) := \frac{1}{\alpha_j(\boldsymbol{l})}, \qquad \alpha_j(\boldsymbol{l}) := n - p + 1 + 2l_j \sum_{i \neq j} \frac{1}{l_j - l_i}, \qquad j = 1, 2, \ldots, p. \tag{6}$$

The modified eigenvalue estimator above (known as Stein's covariance estimator) is therefore given as

$$\widehat{\varphi}_j^{\text{St}}(\boldsymbol{l}) := l_j/\alpha_j(\boldsymbol{l}). \tag{7}$$

Note first that $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l}) \approx \widehat{\varphi}_j^{\mathrm{ml}}(\boldsymbol{l})$ for large n, i.e., in such settings the improvements that Stein's estimator can give over the MLE should be modest. In a numerical study of Stein's estimator, Lin and Perlman [6] also note that the modified estimates $\widehat{\varphi}_j^{\mathrm{St}}$ will differ most from the maximum likelihood estimates when some or all of l_j are close together and when p/n is not small. It is evident from (6) and (7) that Stein's estimator uses adjacent eigenvalues estimates to "shrink" or pull the sample eigenvalue estimates closer together. That said, there are two serious problems with Stein's estimator which do not allow for it to be used directly. First, the $\widehat{\varphi}_1^{\mathrm{St}}(\boldsymbol{l}), \widehat{\varphi}_2^{\mathrm{St}}(\boldsymbol{l}), \ldots, \widehat{\varphi}_p^{\mathrm{St}}(\boldsymbol{l})$ can violate the original ordering as given by $l_1 \geqslant l_2 \geqslant \cdots \geqslant l_p > 0$. Second, the $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l})$ can yield negative eigenvalue estimates. Stein proposes using an isotonizing algorithm that pools adjacent estimators to eliminate order and sign violation [13–15]. The "pooled estimate" using $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l}), \widehat{\varphi}_{j+1}^{\mathrm{St}}(\boldsymbol{l}), \ldots, \widehat{\varphi}_{j+s}^{\mathrm{St}}(\boldsymbol{l})$ is defined as:

$$\widehat{\varphi}_j^{\mathrm{iso}}(\boldsymbol{l}) = \widehat{\varphi}_{j+1}^{\mathrm{iso}}(\boldsymbol{l}) = \cdots = \widehat{\varphi}_{j+s}^{\mathrm{iso}}(\boldsymbol{l}) := \frac{l_j + l_{j+1} + \cdots + l_{j+s}}{\alpha_j(\boldsymbol{l}) + \alpha_{j+1}(\boldsymbol{l}) + \cdots + \alpha_{j+s}(\boldsymbol{l})}\,. \tag{8}$$

Thus, Stein's isotonized estimator is defined as follows:

$$\widehat{\varphi}_j^{\mathrm{St+iso}}(\boldsymbol{l}) := \begin{cases} \widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l}) & \text{if } \dfrac{l_1}{\alpha_1(\boldsymbol{l})} \geqslant \dfrac{l_2}{\alpha_2(\boldsymbol{l})} \geqslant \cdots \dfrac{l_p}{\alpha_p(\boldsymbol{l})} > 0 \\ \widehat{\varphi}_j^{\mathrm{iso}}(\boldsymbol{l}) & \text{otherwise.} \end{cases} \tag{9}$$

Accordingly, we also define new quantities which will be useful later as follows:

$$\widehat{\psi}_j^{\mathrm{St+iso}}(\boldsymbol{l}) := \widehat{\varphi}_j^{\mathrm{St+iso}}(\boldsymbol{l})/l_j. \tag{10}$$

The isotonizing algorithm does not have a decision-theoretic basis and is simply a technique to preserve order and positivity of the estimates. The reader is referred to [6] for more details. Despite the need for this isotonizing algorithm, Stein's estimator, when it is defined (i.e., when $n \geqslant p$), has been shown to consistently perform well in comparison with other competing estimators, and is still widely regarded as the "gold standard". In particular, newly proposed estimators are often compared to Stein's estimator as it has established itself as a benchmark [1, 5, 6]. To distinguish between Stein's original estimator and the version supplemented by the isotonizing algorithm, we shall refer to the latter as Stein's "isotonized" estimator and to the former as Stein's "raw" estimator unless the context is clear.

3. The unbiased estimator of risk (UBEOR): basic properties

Recall that the function F defined in (4) is an unbiased estimator of the risk of $\widehat{\Sigma}$. In this section, we study some of the basic properties of F when $\varphi_j(\boldsymbol{l})$ corresponds to 1) the MLE, 2) Stein's raw estimator, and 3) Stein's isotonized estimator. Let us define the domain of the function F as:

$$\mathscr{D}_p := \{\boldsymbol{l} \in \mathbb{R}^p : l_1 \geqslant l_2 \geqslant \cdots \geqslant l_p > 0\}. \tag{11}$$

The UBEOR of the MLE, denoted by $F^{\mathrm{ml}}(\boldsymbol{l})$, can be obtained by replacing $\psi_j(\boldsymbol{l})$ in (4) with $\widehat{\varphi}_j^{\mathrm{ml}}(\boldsymbol{l})/l_j$ as in (2):

$$F^{\mathrm{ml}}(\boldsymbol{l}) = K_{p,n}^{\mathrm{ml}}, \qquad K_{p,n}^{\mathrm{ml}} := p(1 + \ln n) - c_{p,n}. \tag{12}$$

Thus, the UBEOR for the MLE is a constant function. To obtain (12) we have made use of the following identity:

$$\sum_{j=1}^{p} \alpha_j(\boldsymbol{l}) = np \qquad \forall\, \boldsymbol{l} \in \mathscr{D}_p. \tag{13}$$

Likewise, the UBEOR of Stein's raw estimator, denoted by $F^{\mathrm{St}}(\boldsymbol{l})$, is obtained by substituting $\widehat{\psi}_j^{\mathrm{St}}(\boldsymbol{l})$ into (4):

$$F^{\mathrm{St}}(\boldsymbol{l}) = \sum_{j=1}^{p} \left[1 + \frac{4l_j}{\alpha_j^2(\boldsymbol{l})} \sum_{i \neq j} \frac{l_i}{(l_j - l_i)^2} + \ln(\alpha_j(\boldsymbol{l}))\right] - c_{p,n}. \tag{14}$$

Note that unlike $F^{\mathrm{ml}}(\boldsymbol{l})$ the domain of F^{St} is not the whole set $\mathscr{D}_p$, as F^{St} is defined only on the following subset:

$$\mathscr{E}_{p,n} := \{\boldsymbol{l} \in \mathscr{D}_p : l_i \neq l_j \ \forall\, i, j = 1, \ldots, p \text{ and } \alpha_j(\boldsymbol{l}) > 0 \ \forall\, j = 1, \ldots, p\} = \bigcap_{j=2}^{p} \{\boldsymbol{l} \in \mathscr{D}_p : \alpha_j(\boldsymbol{l}) > 0\},$$

where the equality follows from the fact that if $\tilde{\boldsymbol{l}} \in \mathscr{D}_p$ is such that $\tilde{l}_k = \tilde{l}_j$ for some $1 \leqslant k < j \leqslant p$, then some of the $\alpha_j(\boldsymbol{l})$ must be negative in a neighborhood of $\tilde{\boldsymbol{l}}$ (the condition $\alpha_1(\boldsymbol{l}) > 0$ is not imposed because it is satisfied for all $\boldsymbol{l} \in \mathscr{D}_p$). The fact that the domain of F^{St} is not the whole set $\mathscr{D}_p$ implies that the theoretical risk of Stein's raw estimator is not defined, i.e., the expected value of the UBEOR is either infinity or takes complex values. It is also worth remarking that $\mathscr{E}_{p,n}$ depends not only on the dimension of the covariance matrix but also on the sample size. Note also that the function α_j diverges as two eigenvalues of S approach each other (see (6)), hence $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l})$ may yield a zero estimate. Thus, Stein's raw estimator does not give a meaningful value on the boundary of $\mathscr{D}_p$. The potential negativity of the $\alpha_j(\boldsymbol{l})$ however (and hence of the $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l})$) is a somewhat more serious concern with Stein's raw estimator, since the $\alpha_j(\boldsymbol{l})$ take negative values on an open subset of $\mathscr{D}_p$ (vs. problems just on the boundary). Consequently, any improvement to Stein's raw estimator should focus on the violation of the positivity of the $\alpha_j(\boldsymbol{l})$.

To study the behavior of Stein's estimator, it is useful to observe that F^{St} depends on the l_j only through their ratios l_j/l_i, $j \neq i$. This property can be checked by rewriting $F^{\mathrm{St}}(\boldsymbol{l})$ from (14) in the following form:

$$F^{\mathrm{St}}(\boldsymbol{l}) = \sum_{j=1}^{p} \left[1 + \frac{4 \sum_{i=1}^{j-1} \frac{l_j/l_i}{(l_j/l_i - 1)^2} + 4 \sum_{i=j+1}^{p} \frac{l_i/l_j}{(1 - l_i/l_j)^2}}{\left(n - p + 1 + 2 \sum_{i=1}^{j-1} \frac{l_j/l_i}{l_j/l_i - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - l_i/l_j} \right)^2} + \ln\left(n - p + 1 + 2 \sum_{i=1}^{j-1} \frac{l_j/l_i}{l_j/l_i - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - l_i/l_j} \right) \right] - c_{p,n}, \tag{15}$$

where the sums have been split in such a way that all the ratios of the type l_k/l_q are less than 1. Moreover, if $i < j$, l_j/l_i can be expressed in terms of the ratios of the adjacent eigenvalues through the following telescoping product:

$$\frac{l_j}{l_i} = \frac{l_j}{l_{j-1}} \frac{l_{j-1}}{l_{j-2}} \cdots \frac{l_{i+2}}{l_{i+1}} \frac{l_{i+1}}{l_i} = \prod_{k=i}^{j-1} \frac{l_{k+1}}{l_k}, \qquad 1 \leqslant i < j \leqslant p.$$

Thus, F^{St} can be regarded as a function of the $p - 1$ ratios of the adjacent eigenvalues of S, i.e.,

$$F^{\mathrm{St}}(\boldsymbol{l}) = f^{\mathrm{St}}(\boldsymbol{x}) := \sum_{j=1}^{p} \left\{ 1 + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[1 - \pi_j^i(\boldsymbol{x})]^2} + \ln(a_j(\boldsymbol{x})) \right\} - c_{p,n}, \tag{16}$$

where $\boldsymbol{x} = (x_1, \ldots, x_{p-1})$ with $x_j = l_{j+1}/l_j$ and

$$a_j(\boldsymbol{x}) := n - p + 1 + 2 \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{\pi_i^j(\boldsymbol{x}) - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - \pi_j^i(\boldsymbol{x})}, \quad j = 1, \ldots, p \quad \text{and} \quad \pi_i^j(\boldsymbol{x}) := \prod_{k=i}^{j-1} x_k, \quad 1 \leqslant i < j \leqslant p. \tag{17}$$

Note that by construction $\pi_i^j(\boldsymbol{x}) \leqslant 1$ for all $1 \leqslant i < j \leqslant p$. The function f^{St} is defined on the set:

$$\Omega_{p,n} := \bigcap_{j=2}^{p} \{\boldsymbol{x} \in (0,1)^{p-1} \ : \ a_j(\boldsymbol{x}) > 0\}.$$

The study of F^{St} can therefore be reduced to the study of the function f^{St}, where the latter explicitly represents the dependence of the UBEOR on the ratios of the l_j. In terms of the new variables $\boldsymbol{x}$, the requirement that the order of the l_j is preserved by Stein's raw estimator, i.e., $\widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l}) \geqslant \widehat{\varphi}_{j+1}^{\mathrm{St}}(\boldsymbol{l})$, $j = 1, 2, \ldots, p - 1$, can be rewritten in the form:

$$x_j \leqslant a_{j+1}(\boldsymbol{x})/a_j(\boldsymbol{x}), \qquad j = 1, \ldots, p - 1. \tag{18}$$

It is easily verified that the UBEOR of Stein's isotonized estimator, $F^{\mathrm{St+iso}}$, is also a function of the ratios l_{j+1}/l_j. Hence the representation given by $f^{\mathrm{St+iso}}$ (vs. $F^{\mathrm{St+iso}}$) is more convenient to study, since the non-compact domain involving ordered eigenvalues is now reduced to the unit cube.

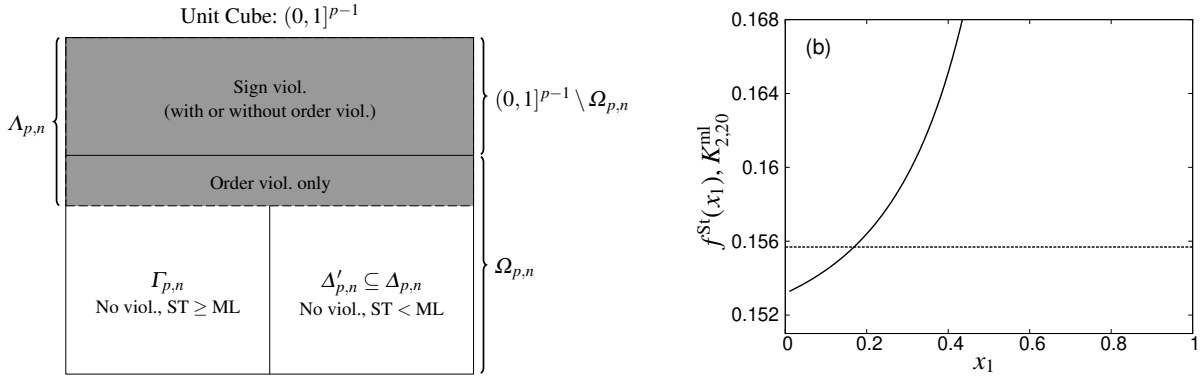


Figure 1: (a) Schematic representation of the sets $\Omega_{p,n}$, $\Delta_{p,n}$, $\Delta'_{p,n}$, $\Gamma_{p,n}$, and $\Lambda_{p,n}$; (b) Graph of the UBEOR for Stein's raw estimator (solid line) and for the MLE (dashed line) for $p = 2$ and $n = 20$;

In the rest of the paper, we shall also consider the following subsets of $(0,1]^{p-1}$:

$$\Delta_{p,n} := \{\boldsymbol{x} \in \Omega_{p,n} : f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{p,n}\}, \tag{19}$$

$$\Delta'_{p,n} := \{\boldsymbol{x} \in \Omega_{p,n} : f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{p,n} \text{ and } x_j \leqslant a_{j+1}(\boldsymbol{x})/a_j(\boldsymbol{x}) \;\forall\, j = 1,\ldots,p-1\}, \tag{20}$$

$$\Gamma_{p,n} := \{\boldsymbol{x} \in \Omega_{p,n} : x_j \leqslant a_{j+1}(\boldsymbol{x})/a_j(\boldsymbol{x}) \;\forall\, j = 1,\ldots,p-1, f^{\mathrm{St}}(\boldsymbol{x}) \geqslant K^{\mathrm{ml}}_{p,n}\}, \tag{21}$$

$$\Lambda_{p,n} := (0,1]^{p-1} \setminus \left(\Delta'_{p,n} \cup \Gamma_{p,n}\right). \tag{22}$$

$\Delta_{p,n}$ is the set where the UBEOR of Stein's raw estimator, $f^{\mathrm{St}}(\boldsymbol{x})$, is lower than that of the MLE ($K^{\mathrm{ml}}_{p,n}$ is defined in (12)). $\Delta'_{p,n}$ is the set where $f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{p,n}$ and, in addition, the order of the sample eigenvalues is preserved by Stein's raw estimator (see (18)); by construction $\Delta'_{p,n} \subseteq \Delta_{p,n} \subseteq \Omega_{p,n} \subseteq (0,1]^{p-1}$. In fact, we observe $\Delta'_{p,n} = \Delta_{p,n}$, i.e., when $f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{p,n}$ the order of the eigenvalues is preserved. On $\Gamma_{p,n}$, both the positivity and the order of the sample eigenvalues are preserved by Stein's raw estimator (and hence the isotonizing algorithm is not required), but the UBEOR of Stein's raw estimator is greater than that of the MLE. Finally, $\Lambda_{p,n}$ is the set where the positivity or the order of the sample eigenvalues are violated by Stein's raw estimator, and hence the isotonizing algorithm is required. Note that $\Omega_{p,n} \cap \Lambda_{p,n} \neq \emptyset$. A schematic representation of the above sets is shown in Fig. 1(a).

4. A study of the UBEOR of Stein's raw estimator

We first present a detailed study of the cases $p = 2, 3, 4$ to understand the UBEOR of Stein's raw estimator (i.e., without the isotonizing correction) in small dimensions, so as to obtain intuition for the general case. The results will then be extended to arbitrary dimensions. The study will focus on the characterization of the domain $\Omega_{p,n}$ and of the set $\Delta_{p,n}$ where f^{St} is less than the UBEOR of the MLE. Recall that the UBEOR comparison framework amounts to contrasting two functions inside the expectation operator: the difference signals the quality of the type of risk reductions that Stein's estimator can yield.

4.1. The case $p = 2$

For $p = 2$, f^{St} is a function of the variable x_1 representing the ratio l_2/l_1:

$$f^{\mathrm{St}}(x_1) = 2 + 4x_1\left\{\frac{1}{[n+1-(n-1)x_1]^2} + \frac{1}{[(n+1)x_1-(n-1)]^2}\right\} + \ln\left(n + \frac{1+x_1}{1-x_1}\right) + \ln\left(n - \frac{1+x_1}{1-x_1}\right) - c_{2,n}. \tag{23}$$

The domain of f^{St} is the interval $\Omega_{2,n} = (0, \tilde{x}_1)$ with $\tilde{x}_1 = (n-1)/(n+1)$; hence $\mathrm{vol}(\Omega_{2,n}) = \tilde{x}_1$. The graph of f^{St} is given in Fig. 1(b) for $n = 20$. The behavior of f^{St} for an arbitrary n can be deduced from the following properties[1]:

[1] Although $\Omega_{p,n}$ does not in principle contain the point $\boldsymbol{x} = \boldsymbol{0}$, the function f^{St} can be defined at the origin. Therefore, to simplify the notation in the remainder of the paper, we shall write $f^{\mathrm{St}}(\boldsymbol{0})$ instead of $\lim_{\|\boldsymbol{x}\|\to 0} f^{\mathrm{St}}(\boldsymbol{x})$.

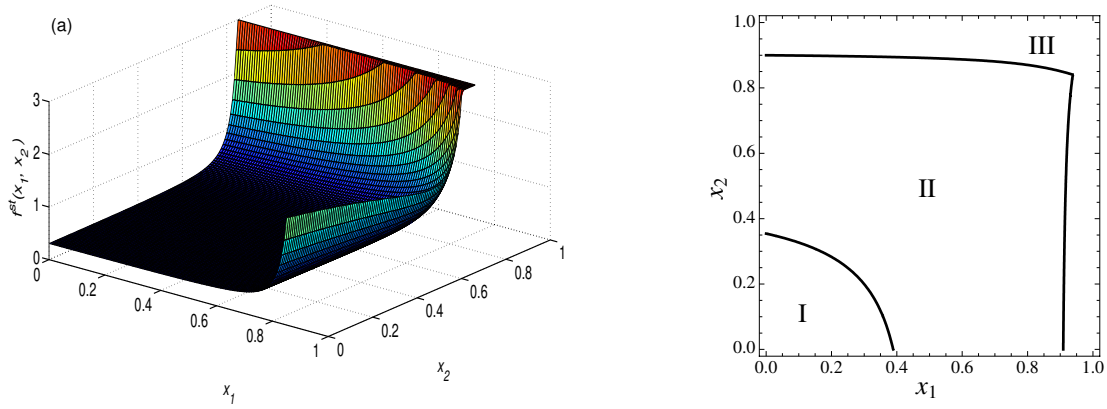


Figure 2: (a) Graph of f^{St} as a function of (x_1, x_2) for $p = 3$ and $n = 20$; (b) Set I: $f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{2,n}$, set II: $f^{\mathrm{St}}(\boldsymbol{x}) > K^{\mathrm{ml}}_{2,n}$. f^{St} is not defined on set III;

1. $f^{\mathrm{St}}(0) = 2 + \ln(n^2 - 1) - c_{2,n} < 2 + \ln n^2 - c_{2,n} = K^{\mathrm{ml}}_{2,n}$;
2. f^{St} is monotonically increasing as a function of x_1 (see Sect. A in the Supplemental Material);
3. As $x_1 \to \tilde{x}_1^-$, $f^{\mathrm{St}}(x_1) \sim c_n(x_1 - \tilde{x}_1)^{-2}$ with $c_n = 4(n-1)/(n+1)^3$. Thus, f^{St} grows without bound.

As a consequence of the above properties, there exists one and only one $x_1^\star \in \Omega_{2,n}$ such that $f^{\mathrm{St}}(x_1^\star) = K^{\mathrm{ml}}_{2,n}$, and the set in which $f^{\mathrm{St}}(x_1)$ is less than $K^{\mathrm{ml}}_{2,n}$ is $\Delta_{2,n} = (0, x_1^\star)$. For a given n, $x_1^\star$ can be calculated by numerically solving the equation $f^{\mathrm{St}}(x_1^\star) = K^{\mathrm{ml}}_{2,n}$ from (23). Such a calculation shows that, as $n \to \infty$, $x_1^\star$ rapidly converges to an asymptotic value (Fig. 3(a)), which can be formally determined by considering the large-n expansion of $f^{\mathrm{St}}(x_1)$:

$$f^{\mathrm{St}}(x_1) - K^{\mathrm{ml}}_{2,n} = -\frac{x_1^2 - 6x_1 + 1}{(x_1 - 1)^2}\left(\frac{1}{n^2}\right) + O\left(\frac{1}{n^4}\right), \qquad x_1 \in \Omega_{2,n}. \tag{24}$$

Solving the equation $x_1^2 - 6x_1 + 1 = 0$ and taking the solution less than 1 yield:

$$\lim_{n\to\infty} x_1^\star = 3 - 2\sqrt{2} \simeq 0.172. \tag{25}$$

As $n \to \infty$, the domain $\Omega_{2,n}$ tends to the whole interval $(0, 1)$. In contrast, $\mathrm{vol}(\Delta_{2,n}) = x_1^\star$ tends to a value much less than 1, as can be seen from (25). Therefore the region on which $f^{\mathrm{St}}(x_1) < K^{\mathrm{ml}}_{2,n}$ is relatively small. Furthermore, (24) shows that, as $n \to \infty$, the convergence of $f^{\mathrm{St}}(x_1)$ to $K^{\mathrm{ml}}_{2,n}$ is not uniform in x_1, since $\lim_{x_1 \to \tilde{x}_1^-}\left[f^{\mathrm{St}}(x_1) - K^{\mathrm{ml}}_{2,n}\right] = \infty$ $\forall\, n \geqslant 2$. Taking limits in the opposite direction (i.e., n approaching $p = 2$), the length of $\Delta_{2,n}$ is much less than 1 (Fig. 3(b)). Furthermore, at $x_1 = 0$, the percentage reduction in the UBEOR of Stein's raw estimator relative to the MLE is a decreasing function of n and attains its maximum when $n = p = 2$ (Fig. 3(c)), though the reduction is not very large.

The above analysis suggests two competing phenomena when comparing f^{St} and f^{ml} as n increases: 1) the difference $f^{\mathrm{St}} - K^{\mathrm{ml}}_{2,n}$ at 0 decreases and 2) the size of the region $\Delta_{2,n}$ on which $f^{\mathrm{St}} < K^{\mathrm{ml}}_{2,n}$ increases. We shall see from our numerical work (see Section 6) that relative risk reductions compared to the MLE are higher for smaller n — indicating that the former phenomenon outweighs the latter. This point also demonstrates that neither the function values nor the volume of $\Delta_{p,n}$ should be analyzed in isolation when studying Stein's estimator in the UBEOR framework.

4.2. The cases $p = 3$ and $p = 4$

The graph of the UBEOR of Stein's raw estimator when $p = 3$ and $n = 20$ is given in Fig. 2(a). The domain $\Omega_{3,n}$ is described by the inequalities $a_2(\boldsymbol{x}) > 0$ and $a_3(\boldsymbol{x}) > 0$ and corresponds to

$$n(1 - x_1)(1 - x_2) - 2(1 - x_2) + 2(1 - x_1) > 0 \quad \text{and} \quad (n + 2)(1 - x_1x_2)(1 - x_2) - 2(1 - x_2) - 2(1 - x_1x_2) > 0. \tag{26}$$

These inequalities correspond to the union of the sets denoted by I and II in Fig. 2(b). In Fig. 2(b), the set $\Delta_{3,n}$, where $f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{3,n}$, is denoted by I: it is a connected set and contains a neighborhood of the origin in the first quadrant. The

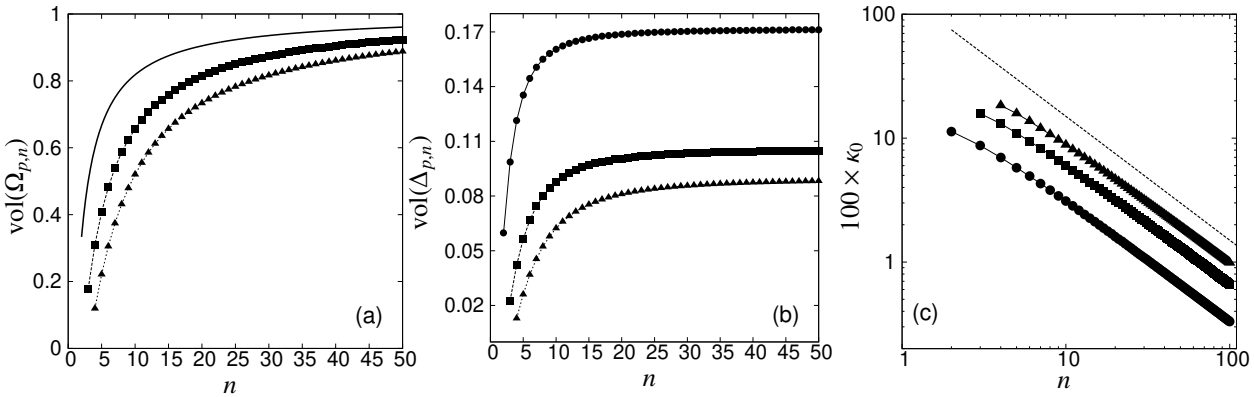


Figure 3: (a) Volume of $\Omega_{p,n}$ as a function of n for $p = 2$ (solid line), $p = 3$ (■), $p = 4$ (▲); (b) Volume of $\Delta_{p,n}$ as a function of n for $p = 2$ (●), $p = 3$ (■), $p = 4$ (▲); (c) Percentage reduction in the UBEOR at $\boldsymbol{x} = \mathbf{0}$, $100 \times \kappa_0 := 100 \times [K^{\mathrm{ml}}_{p,n} - f^{\mathrm{St}}(\mathbf{0})]/K^{\mathrm{ml}}_{p,n}$, as a function of n for $p = 2$ (●), $p = 3$ (■), $p = 4$ (▲). The dashed straight line is proportional to $1/n$. Both the vertical and horizontal axis are scaled logarithmically.

function f^{St} is not defined on region III. Figures 3(a) and (b) report the volumes of $\Omega_{3,n}$ and $\Delta_{3,n}$ as a function of n; the behavior is similar to the one observed for $p = 2$.

The function f^{St} diverges positively when $a_2(\boldsymbol{x})$ or $a_3(\boldsymbol{x})$ goes to zero, but, unlike in the $p = 2$ case, f^{St} is not monotonic in x_2 for all fixed x_1. Thus a simple characterization of $\Omega_{3,n}$ and $\Delta_{3,n}$ for the $p = 3$ case is not available. Note also that f^{St} is neither convex nor concave (see Sect. B in the Supplemental Material). At $\boldsymbol{x} = 0$, the percentage reduction in the UBEOR of Stein's raw estimator relative to the MLE decreases approximately linearly with increasing n, as in the $p = 2$ case (Fig. 3(c)). Moreover, for all $\boldsymbol{x} \in \Omega_{p,n}$, the difference $f^{\mathrm{St}}(\boldsymbol{x}) - K^{\mathrm{ml}}_{p,n}$ is $O(1/n^2)$ as $n \to \infty$. This property will be derived directly for arbitrary p in the next section. The $p = 4$ case yields results which are similar to the $p = 3$ case. The reader is referred to Sect. C in the Supplemental Material for details.

4.3. Summary of the cases $p = 2, 3, 4$

The comparison of the UBEOR of Stein's raw estimator vs. that of the MLE for $p = 2, 3, 4$ is summarized below:

1. The UBEOR of Stein's raw estimator is not defined for all values of the l_j, since it contains logarithmic terms whose argument can be negative. This occurs in an open subset of the domain of the UBEOR. Furthermore, the UBEOR of Stein's raw estimator diverges positively as the argument of one of the logarithmic terms approaches zero. The implications of the above are threefold: (a) the risk of Stein's raw estimator is either infinite or complex-valued; (b) it provides compelling evidence that Stein's raw estimator needs to be rectified or modified in such cases, and thus motivates the isotonizing algorithm (the UBEOR of which is studied in the second part of this paper); (c) it motivates the need to characterize the region over which the UBEOR diverges in order to assess the severity of the problem.
2. When the UBEOR of Stein's raw estimator is rewritten in terms of the ratios of the adjacent eigenvalues of S, the domain of definition of the estimator, $\Omega_{p,n}$, is a strict subset of $(0, 1)^{p-1}$. For a given p, the volume of $\Omega_{p,n}$ increases with increasing n and tends to 1 as $n \to \infty$. The volume of $\Omega_{p,n}$ decreases with increasing p when n is held constant (Fig. 3(a)). Hence the problem that the UBEOR of Stein's raw estimator may not even be defined is alleviated as n increases, but is exacerbated as p increases.
3. The UBEOR of Stein's raw estimator takes lower values than that of the MLE only when the adjacent eigenvalues of S are sufficiently separated, i.e., when the ratios x_i are close to zero. The set $\Delta_{p,n}$, where Stein's raw estimator is less in value than the MLE, is connected and contains a neighborhood of the origin. The volume of $\Delta_{p,n}$ increases with n and tends to an asymptotic value less than 1. For a given n, the volume of $\Delta_{p,n}$ decreases as the dimension p increases (Fig. 3(b)).
4. Although $f^{\mathrm{St}}(\boldsymbol{x}) < K^{\mathrm{ml}}_{p,n}$ in the neighborhood of $\boldsymbol{x} = \mathbf{0}$, the percentage reduction at the origin in the UBEOR of Stein's raw estimator relative to the MLE is modest and decreases linearly with increasing n (Fig. 3(c)). Hence Stein's estimator is not expected to yield substantive gains over the MLE when either the population eigenvalues are well separated and/or when the sample size is very high. The latter is consistent with the fact

that Stein's estimator tends to the MLE as $n \to \infty$. We also note that despite diminishing reductions in UBEOR as n increases, there is a dimension effect: the modest reduction in UBEOR at $\boldsymbol{x} = \mathbf{0}$ for fixed n is higher for larger p (see Fig. 3(c)).

5. We note here for completeness that the comparison between $f^{\mathrm{St}}(\boldsymbol{x})$ and $K^{\mathrm{ml}}_{p,n}$ in the space of the sample eigenvalues yields similar results.
6. For $n \to \infty$, the UBEOR of Stein's raw estimator tends to that of the MLE as follows: $f^{\mathrm{St}}(\boldsymbol{x}) - K^{\mathrm{ml}}_{p,n} = O(1/n^2)$ for all $\boldsymbol{x} \in \Omega_{p,n}$. Moreover, the convergence is not uniform in $\boldsymbol{x}$. We shall see in the Supplemental Material (see (D.11)) that the leading order of the UBEOR of the MLE is $O(1/n)$. This point indicates that in an asymptotic sense the difference in the UBEOR of the two estimators is relatively small, which points to possibly only modest risk reductions when using Stein's estimator instead of the MLE. We shall see later in the paper that this will no longer always be the case when the isotonizing algorithm is invoked (see Subsect. 5.3).
7. The function $f^{\mathrm{St}}(\boldsymbol{x})$ is not convex in general, nor is it monotonic in each of the variables when the other variables are held constant. The exception to the above behavior is the $p = 2$ case (see Subsect. 4.1). The lack of convexity/monotonicity in the $p > 2$ cases implies that the region in which the UBEOR of Stein's raw estimator takes lower values than that of the MLE is not easily characterizable.

4.4. The arbitrary p case

Several observations reported in Subsect. 4.3 for small p can be proved for any general $p > 1$. We shall maintain the formulation of Stein's UBEOR in terms of the ratios $x_j = l_{j+1}/l_j$. The results proved for $f^{\mathrm{St}}(\boldsymbol{x})$ can be easily translated into analogous properties of $F^{\mathrm{St}}(\boldsymbol{l})$. The proofs of the lemmas and propositions in this subsection are provided in Sect. D in the Supplemental Material.

We have already noted that f^{St} may have singularities at those points at which the argument $a_j(\boldsymbol{x})$ of one of the logarithmic terms vanishes. For $j \geqslant 2$, $a_j(\boldsymbol{x})$ is the sum of two terms of opposite sign (see (17)):

$$a_j(\boldsymbol{x}) = \underbrace{n - p + 1 + 2\sum_{i=j+1}^{p} \frac{1}{1-\pi_j^i(\boldsymbol{x})}}_{>0} + \underbrace{2\sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{\pi_i^j(\boldsymbol{x})-1}}_{<0}, \tag{27}$$

Thus, if one of the products $\pi_i^j(\boldsymbol{x})$ for $i < j$ is sufficiently close to 1 (i.e. l_j is sufficiently close to l_i for some $i < j$), then $a_j(\boldsymbol{x})$ may vanish or become negative. Indeed, the following Lemma shows that there exists some $\boldsymbol{x} \in (0,1)^{p-1}$ such that at least one of the $a_j(\boldsymbol{x})$ is equal to zero.

Lemma 1. *For all $n \geqslant p > 1$, there exists $\tilde{\boldsymbol{x}} \in (0,1)^{p-1}$ and $M \subset \{2,\ldots,p\}$, $|M| > 1$, such that $a_j(\tilde{\boldsymbol{x}}) = 0$ for all $j \in M$ and $a_j(\tilde{\boldsymbol{x}}) > 0$ for $j \notin M$.*

Remark 1. The points $\boldsymbol{x} \in (0,1)^{p-1}$ such that $a_j(\boldsymbol{x}) = 0$ for all $j \in M$ for some $M \subseteq \{2,3,\ldots,p\}$ and $a_j(\boldsymbol{x}) > 0$ for all $j \notin M$ belong to the boundary of $\Omega_{p,n}$. Owing to the continuity of the a_j, such points are accumulation points of $\Omega_{p,n}$.

The following proposition asserts tha f^{St} becomes unbounded as $\boldsymbol{x}$ approaches one of the points where at least one of the a_j vanishes.

Proposition 1. *Let $n \geqslant p > 1$ and $M \subseteq \{2,\ldots,p\}$. If $\tilde{\boldsymbol{x}} \in (0,1)^{p-1}$ is such that $a_j(\tilde{\boldsymbol{x}}) = 0$ for all $j \in M$ and $a_j(\tilde{\boldsymbol{x}}) > 0$ for all $j \notin M$, then $\lim_{\boldsymbol{x}\to\tilde{\boldsymbol{x}}} f^{\mathrm{St}}(\boldsymbol{x}) = +\infty$.*

Recall that in order to obtain the estimator $\widehat{\psi}_j^{\mathrm{St}}$, Stein disregards the derivative terms $2l_j\partial\psi_j/\partial l_j$ in the UBEOR. The proof of Proposition 1 (see Sect. D in the Supplemental Material) demonstrates that these derivative terms determine the behavior of the UBEOR near the boundary of its domain and are therefore not negligible.

The large-n behavior of the UBEOR of Stein's raw estimator is described by the following proposition.

Proposition 2. *For any $p > 1$ and $\boldsymbol{x} \in \Omega_{p,n}$,*

$$f^{\mathrm{St}}(\boldsymbol{x}) = K_{p,n}^{\mathrm{ml}} + \left(\frac{1}{n^2}\right) \sum_{j=1}^{p} \left\{ 4 \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} + 4 \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[1 - \pi_j^i(\boldsymbol{x})]^2} - \frac{1}{2} \left[1 - p + 2 \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{\pi_i^j(\boldsymbol{x}) - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - \pi_j^i(\boldsymbol{x})} \right]^2 \right\} + O\left(\frac{1}{n^3}\right) \quad (28)$$

as $n \to \infty$. Furthermore, the convergence is not uniform in $\boldsymbol{x}$.

The analysis in the cases $p = 2, 3, 4$ highlighted an important property of f^{St}: when the ratios l_{j+1}/l_j are sufficiently small (i.e., $\|\boldsymbol{x}\|$ is close to 0), $f^{\mathrm{St}} < K_{p,n}^{\mathrm{ml}}$. The next proposition proves this property for arbitrary dimension.

Proposition 3. *For any $n \geqslant p > 1$, there exists an open ball $B_\delta(\mathbf{0})$ with radius $\delta > 0$ and center $\boldsymbol{x} = \mathbf{0}$ such that, for all $\boldsymbol{x} \in B_\delta(\mathbf{0}) \cap \Omega_{p,n}$, $f^{\mathrm{St}}(\boldsymbol{x}) < K_{p,n}^{\mathrm{ml}}$ and the order of the sample eigenvalues is preserved by Stein's raw estimator. Moreover, for a given n, $K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0})$ is a monotonically increasing function of p for $1 < p \leqslant n$.*

Finally, the following Lemma describes the large-n behavior of the percentage reduction in the UBEOR of Stein's raw estimator relative to the MLE at $\boldsymbol{x} = \mathbf{0}$ and asserts that the reduction of the UBEOR given by Stein's raw estimator is most pronounced for small sample sizes.

Lemma 2. *For $p > 1$,*

$$\kappa_0 := \frac{K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0})}{K_{p,n}^{\mathrm{ml}}} \sim \frac{p-1}{3n} \qquad (n \to \infty). \quad (29)$$

We observe from the analysis conducted in the previous sections that the domain on which Stein's raw estimator is defined diminishes as p increases (see, e.g., Fig. 3(a)). However, Proposition 3 and Lemma 3 demonstrate that, for large n, the local behavior of Stein's raw estimator at $\boldsymbol{x} = \mathbf{0}$ improves with increasing p. The above comparisons thus have two implications: 1) studying only the local behavior of the UBEOR of Stein's estimator can lead to a misinformed understanding of the potential overall risk gains of Stein's estimator, i.e., Stein's estimator does not perform uniformly well in all parts of the parameter space. Hence the risk gains of Stein's estimator seen in numerical work in previous papers should be interpreted in this context, i.e., global statements have been inferred regarding the performance of the estimator based on risk gains for specific parameter values. This is perhaps an unintended pitfall of using only numerical risk experiments; 2) If prior information is known regarding how well separated the eigenvalues are, Stein's raw estimator can in fact be used effectively to yield risk gains.

A useful and perhaps equally interesting question is to understand the relative risk reduction in Stein's UBEOR at the origin when the dimension p is also allowed to grow. Further asymptotic calculations (not given here for the sake of brevity), but this time as p is allowed to grow and $n = p$, show that the relative risk reduction at $\boldsymbol{x} = \mathbf{0}$ is in fact bounded, i.e., $O(1)$ as $p \to \infty$. Hence in this regime the relative risk reduction at $\boldsymbol{x} = \mathbf{0}$ does not increase with p.

5. A study of Stein's isotonized estimator in the UBEOR framework

In practice, Stein's "raw" estimator has been shown to perform very well when coupled with the isotonizing algorithm. Recall that this algorithm pools adjacent eigenvalues together when positivity or order is violated (see (8)). Therefore, a natural question that arises concerns the precise role played by the isotonizing algorithm in the risk reductions enjoyed by Stein's estimator. Note that the isotonizing algorithm is ad hoc and renders a decision theoretic analysis of Stein's estimator rather intractacble. Recall however that our goal is not to get a decision theoretic result concerning Stein's isotonized estimator, but rather to obtain a theoretical understanding of its behavior by exploiting its UBEOR. Note also that when no order or sign violations are present Stein's raw and isotonized estimators coincide.

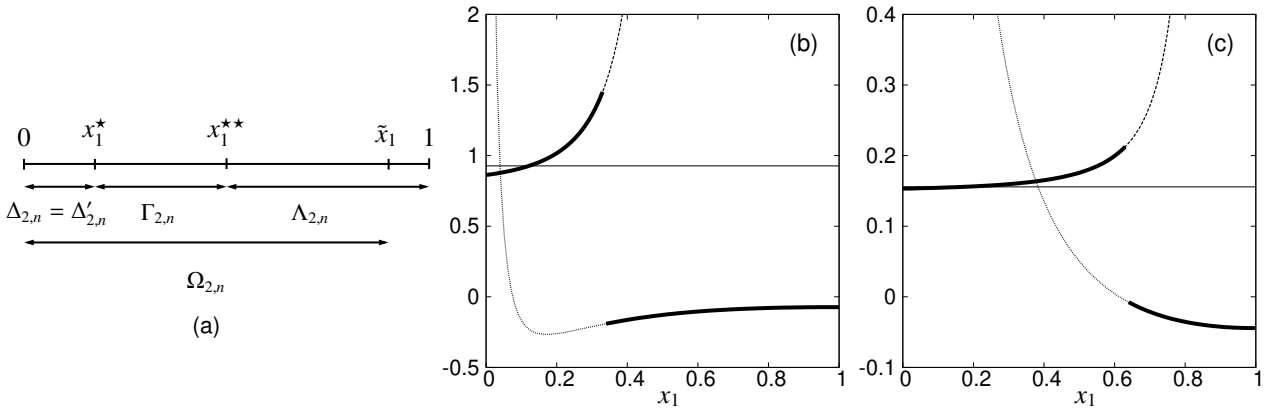


Figure 4: (a) Schematic representation of the $p = 2$ case (we refer the reader to Sect. 3 for the definitions of the sets $\Omega_{2,n}$, $\Delta_{2,n}$, $\Gamma_{2,n}$, and $\Lambda_{2,n}$). The function f^{St} is defined on $\Omega_{2,n} = (0, \tilde{x}_1)$ only. The positivity of the sample eigenvalues is preserved by Stein's raw estimator on $\Omega_{2,n}$ and is violated on $(0, 1] \setminus \Omega_{2,n}$. The order of the sample eigenvalues is preserved on $\Delta'_{2,n} \cup \Gamma_{2,n} = (0, x_1^{\star\star}]$ and is violated on $\Lambda_{2,n} = (x_1^{\star\star}, 1]$. The function $f^{\mathrm{St}}(x_1)$ is less than $K_{2,n}^{\mathrm{ml}}$ on $\Delta_{2,n} = (0, x_1^\star)$ and greater than $K_{2,n}^{\mathrm{ml}}$ on $(x_1^\star, \tilde{x}_1)$. (b) and (c) Graph of the functions f^{St} (dashed line), f^{iso} (dotted line), $f^{\mathrm{St+iso}}$ (thick solid line), and $K_{2,n}^{\mathrm{ml}}$ (thin solid flat line) for: $n = 4$ (b) and $n = 20$ (c). The thick solid line is superimposed on the graphs of f^{St} and f^{iso} when they are the same as the graph of $f^{\mathrm{St+iso}}$.

5.1. The case $p = 2$

For $p = 2$, the set $\Lambda_{2,n}$ on which Stein's raw estimator yields either sign or order violations thus invoking the isotonizing algorithm (see (22) for the definition) consists of the points $x_1 \in (0, 1]$ such that

$$(n-1)x_1^2 - 2(n+1)x_1 + (n-1) < 0 \quad \text{(order violation)} \qquad \text{or} \qquad x_1 \geqslant \tilde{x}_1 \quad \text{(violation of positivity)}, \tag{30}$$

where $\tilde{x}_1 = (n-1)/(n+1)$ is the highest value of x_1 for which $f^{\mathrm{St}}(x_1)$ is defined (see Subsect. 4.1). It is easily checked from (30) that order violation occurs for $x_1 > x_1^{\star\star}$, where $x_1^{\star\star} := (n + 1 - 2\sqrt{n})/(n-1)$. Hence $\Lambda_{2,n} = (x_1^{\star\star}, 1]$.

Furthermore, recall that the value $x_1^\star$ is defined such that $f^{\mathrm{St}}(x_1) < K_{2,n}^{\mathrm{ml}}$ for $x_1 < x_1^\star$ and $f^{\mathrm{St}}(x_1) > K_{2,n}^{\mathrm{ml}}$ for $x_1 > x_1^\star$ (see Subsect. 4.1) and note that $x_1^\star < x_1^{\star\star} < \tilde{x}_1 \ \forall\, n \geqslant 2$. Therefore, the set on which the UBEOR of Stein's raw estimator is greater than that of the MLE, but where the isotonizing correction does not apply, is $\Gamma_{2,n} = [x_1^\star, x_1^{\star\star}]$ and is non-empty (Fig. 6(a)). A schematic description of the $p = 2$ case is given in Fig. 4(a).[2]

The isotonized estimator thus takes the form $\widehat{\varphi}_j^{\mathrm{St+iso}}(\boldsymbol{I}) = l_j \widehat{\psi}_j^{\mathrm{St+iso}}(l_2/l_1)$ with

$$\widehat{\psi}_1^{\mathrm{St+iso}}(x_1) = \begin{cases} \dfrac{1-x_1}{n+1-(n-1)x_1} & 0 < x_1 \leqslant x_1^{\star\star} \\ \dfrac{1+x_1}{2n} & x_1^{\star\star} < x_1 \leqslant 1 \end{cases} \qquad \text{and} \qquad \widehat{\psi}_2^{\mathrm{St+iso}}(x_1) = \begin{cases} \dfrac{1-x_1}{n-1-(n+1)x_1} & 0 < x_1 \leqslant x_1^{\star\star} \\ \dfrac{1+x_1}{2nx_1} & x_1^{\star\star} < x_1 \leqslant 1. \end{cases} \tag{31}$$

Correspondingly, the UBEOR of Stein's isotonized estimator is given by:

$$f^{\mathrm{St+iso}}(x_1) = \begin{cases} f^{\mathrm{St}}(x_1) & 0 < x_1 \leqslant x_1^{\star\star} \\ f^{\mathrm{iso}}(x_1) & x_1^{\star\star} < x_1 \leqslant 1, \end{cases}$$

where the explicit expression of $f^{\mathrm{St}}(x_1)$ is given by (23) and where $f^{\mathrm{iso}}(x_1)$ is the UBEOR when the isotonizing algorithm is used:

$$f^{\mathrm{iso}}(x_1) = 1 - \frac{1}{n} + \frac{x_1}{2} - \frac{3x_1}{2n} + \frac{n-3}{2nx_1} - \ln\left(\frac{1+x_1}{2n}\right) - \ln\left(\frac{1+x_1}{2nx_1}\right) - c_{2,n}.$$

[2] In particular, the inequality $x_1^\star < x_1^{\star\star} < \tilde{x}_1 \ \forall\, n \geqslant 2$ also implies that $\Delta'_{2,n} = \Delta_{2,n}$ for all $n \geqslant 2$ (see (20) for the definition of $\Delta'_{2,n}$), i.e., the original order of the sample eigenvalues is never violated when the UBEOR of Stein's raw estimator is less than that of the MLE.

Naturally, f^{iso} is a well-defined function on the entire domain $(0, 1]$ and can be regarded as an independent function regardless of its use within Stein's estimator. A study of the function $f^{\text{St+iso}}$ shows that $f^{\text{iso}}(1) = 2\left(1 - \frac{2}{n} + \ln n\right) - c_{2,n} < K^{\text{ml}}_{2,n}$. Furthermore, for $n = 2, 3$, $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ for all $x_1 \in (0, 1]$, while, for $n \geqslant 4$, there exists one and only one $\bar{x}_1 \in (0, 1)$ such that $f^{\text{iso}}(x_1) > K^{\text{ml}}_{2,n}$ for $0 < x_1 < \bar{x}_1$ and $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ for $\bar{x}_1 < x_1 \leqslant 1$. The graph of $f^{\text{St+iso}}$ is reported in Figs. 4(b) and (c) for $n = 4, 20$.

The comparison of the UBEOR of Stein's isotonized estimator vs. that of the MLE in the two-dimensional case therefore leads to the following insights:

1. There exists an interval on which the UBEOR of Stein's isotonized estimator is greater than that of the MLE;
2. The isotonizing algorithm is invoked on a considerable portion of the domain containing the region where the ratio of the sample eigenvalues is near 1. In the set in which the isotonizing algorithm applies, the reduction in the UBEOR compared to the MLE is considerable and significantly greater than that observed in those sets where isotonizing is not used (Figs. 4(b) and (c)). This observation can be made more precise by considering the quantities
$$\kappa_1 = \frac{K^{\text{ml}}_{p,n} - f^{\text{St+iso}}(1, \ldots, 1)}{K^{\text{ml}}_{p,n} - f^{\text{St}}(0, \ldots, 0)} \qquad \text{and} \qquad \kappa_2 = \frac{K^{\text{ml}}_{p,n} - f^{\text{St+iso}}(1, \ldots, 1)}{K^{\text{ml}}_{p,n}}. \tag{32}$$
The first quantity compares the difference between $K^{\text{ml}}_{2,n}$ and $f^{\text{St}}(\boldsymbol{x})$ at the origin with the same difference at $\boldsymbol{x} = (1, \ldots, 1)$. At $\boldsymbol{x} = (0, \ldots, 0)$ the reduction in the UBEOR comes from Stein's raw estimator, whereas at $\boldsymbol{x} = (1, \ldots, 1)$ all the eigenvalues are pooled together and the reduction is a result of isotonization. The ratio κ_1, therefore, is one simple way to quantify the relative efficacy of the isotonizing algorithm in reducing the UBEOR in comparison with Stein's raw estimator. The second quantity is the relative reduction in the UBEOR at $\boldsymbol{x} = (1, \ldots, 1)$, where the reduction is entirely due to the isotonizing algorithm, and is analogous to κ_0 (see (29)). For $p = 2$ (see Figs. 6(b) and (c)):
$$\kappa_1 = 4\left[n \ln\left(\frac{n^2}{n^2 - 1}\right)\right]^{-1} \sim 4n \quad \text{and} \quad \kappa_2 = \frac{4}{n[2(1 + \ln n) - c_{2,n}]} \sim \frac{4}{3} - \frac{26}{27n} \qquad (n \to \infty). \tag{33}$$
The expressions above affirm the fact that the relative reduction in UBEOR afforded by the isotonizing algorithm is non-negligible and has a complex relationship with the sample size n.
3. There exist points $x_1 < x_1^{\star\star}$ such that $f^{\text{iso}}(x_1) < K^{\text{ml}}_{2,n}$ but $f^{\text{St+iso}}(x_1) = f^{\text{St}}(x_1) > K^{\text{ml}}_{2,n}$. In particular, there exists a subset of the domain in which: (a) the function $f^{\text{iso}}(x_1)$ is considerably less than the UBEOR of the MLE and simultaneously (b) Stein's estimator does not require the isotonizing correction, though the UBEOR of the raw estimator is greater than that of the MLE. For these points, although an isotonizing correction is not required, using it will nevertheless reduce the UBEOR. In other words, the "shrinkage" given by the isotonizing correction becomes useful even way before its originally intended purpose of order preservation. These insights can potentially be useful for constructing new estimators which can combine the strengths of Stein's raw estimator and the isotonizing algorithm (see [10]).
4. From Figs. 4(b) and (c) we observe that the behavior of $f^{\text{St+iso}}$ in comparison to $K^{\text{ml}}_{p,n}$ starts to exhibit well defined trends. In fact three distinct regimes start to emerge: 1) The first is near the origin where the eigenvalues are well separated and the UBEOR of Stein's isotonized estimator (which in this case is identical to the UBEOR of Stein's raw estimator) is less than $K^{\text{ml}}_{p,n}$ (though the reduction is small), 2) The second regime where the eigenvalues are only moderately separated and the UBEOR of Stein's estimator takes values higher than $K^{\text{ml}}_{p,n}$, and 3) The third regime where the eigenvalues are close to each other, thus invoking the isotonizing algorithm due to sign/order violations. In this third regime, the UBEOR of Stein's isotonized estimator is significantly lower than that of the MLE.

5.2. The cases $p = 3$ and $p = 4$

The exact functional form of the isotonized estimator for the $p = 3$ case (analogous to (31) in the $p = 2$ case) is described in detail in Sect. E in the Supplemental Material. Only the main properties of the UBEOR are presented here. The UBEOR of the isotonized estimator, when rewritten in terms of the ratios $x_1 = l_2/l_1$ and $x_2 = l_3/l_2$, is a

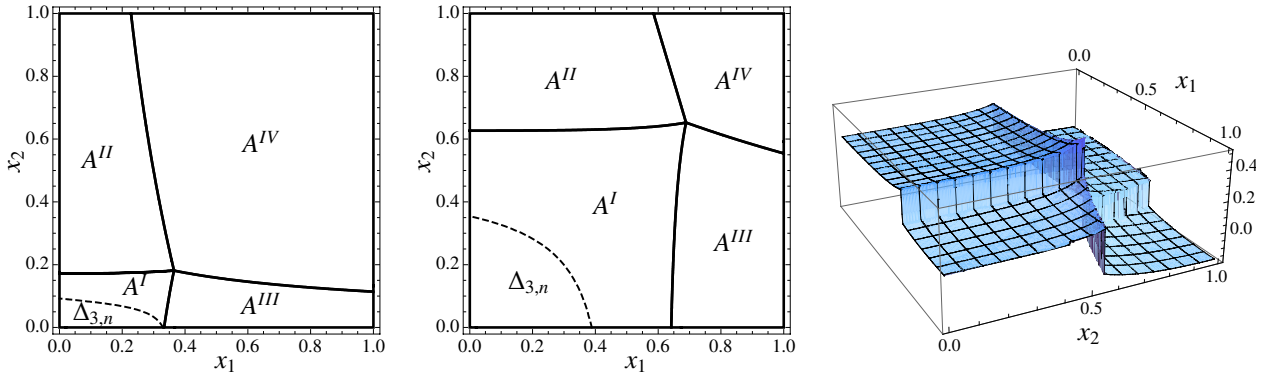


Figure 5: Left and center panels: Subsets A^I, A^{II}, A^{III}, and A^{IV} for $n = 3$ (left) and $n = 20$ (center). The set $\Delta_{3,n}$ is contained in A^I. The set $\Gamma_{3,n}$ is equal to $A^I \setminus \Delta_{3,n}$, while $\Lambda_{3,n} = A^{II} \cup A^{III} \cup A^{IV}$. The sets $\Delta_{3,n}$ and $\Delta'_{3,n}$ are numerically indistinguishable. Right panel: Graph of $f^{\text{St+iso}}$ for $p = 3$ and $n = 20$. Note that $K^{\text{ml}}_{3,20} = 0.318$.

function on $(0, 1] \times (0, 1]$. The domain $(0, 1] \times (0, 1]$ is divided into four subsets, A^I, A^{II}, A^{III}, and A^{IV}, on each of which the estimator of the eigenvalues of Σ, and hence the UBEOR, takes a different form (see Fig. 5):[3]

1. in A^I, all the estimated eigenvalues are positive, their order is preserved, and consequently the isotonized estimator and Stein's raw estimator coincide;
2. in A^{II}, $\widehat{\varphi}^{\text{St}}_2(\boldsymbol{l})$ and $\widehat{\varphi}^{\text{St}}_3(\boldsymbol{l})$ are pooled together because of either order reversal or negative values, while the estimator of the first eigenvalue is $\widehat{\varphi}^{\text{St}}_1(\boldsymbol{l})$;
3. in A^{III}, $\widehat{\varphi}^{\text{St}}_1(\boldsymbol{l})$ and $\widehat{\varphi}^{\text{St}}_2(\boldsymbol{l})$ are pooled together, while the estimator of the third eigenvalue is $\widehat{\varphi}^{\text{St}}_3(\boldsymbol{l})$;
4. in A^{IV}, $\widehat{\varphi}^{\text{St}}_1(\boldsymbol{l})$, $\widehat{\varphi}^{\text{St}}_2(\boldsymbol{l})$, and $\widehat{\varphi}^{\text{St}}_3(\boldsymbol{l})$ are pooled together.

The set $\Lambda_{3,n}$, where the isotonizing algorithm is required (see (22)), is thus $\Lambda_{3,n} = A^{II} \cup A^{III} \cup A^{IV}$. We shall denote by $\psi^I_j(\boldsymbol{x})$ Stein's isotonized estimator on the set A^I (the functions $\psi^{II}_j(\boldsymbol{x})$, $\psi^{III}_j(\boldsymbol{x})$, and $\psi^{IV}_j(\boldsymbol{x})$ are defined analogously—see Sect. E in the Supplemental Material for details).

The four subsets A^I, A^{II}, A^{III}, A^{IV} are represented in Fig. 5 for $n = p = 3$ and for $n = 20$. As seen in Section 4, the portion of the domain where the isotonizing correction applies is substantial, especially when n is close to p, i.e., the isotonizing algorithm "kicks in" when the sample size is relatively small. The set A^I contains $\Delta_{3,n}$, on which $f^{\text{St}}(\boldsymbol{x}) < K^{\text{ml}}_{3,n}$. The set $\Delta'_{3,n}$, where $f^{\text{St}}(\boldsymbol{x}) < K^{\text{ml}}_{p,n}$ and there are no sign/order violations, was also computed numerically. We note, similarly to the $p = 2$ case, that $\Delta'_{3,n}$ and $\Delta_{3,n}$ are indistinguishable, i.e., on the domain on which $f^{\text{St}}(\boldsymbol{x}) < K^{\text{ml}}_{3,n}$ Stein's estimator retains positivity and the original order of the sample eigenvalues. The difference $A^I \setminus \Delta_{3,n}$ is the set $\Gamma_{3,n}$, in which Stein's estimator is not isotonized, but its UBEOR is greater than that of the MLE (see (21) for the definition of $\Gamma_{p,n}$). Figure 5 shows that, for $n = 3$ and $n = 20$, the set $\Gamma_{3,n}$ is non-empty. The same results hold for other values of n, as demonstrated by the computation of the volume of $\Gamma_{3,n}$ (Fig. 6(a)). Therefore, when the sample eigenvalues are moderately separated, the UBEOR of Stein's isotonized estimator is greater than that of the MLE. In fact, more than 40% of the total volume is covered by $\Gamma_{3,n}$.

Given that the isotonized estimator takes different forms on different subsets of the domain, its UBEOR $f^{\text{St+iso}}$ is a function defined piecewise that is continuous within each part (see Fig. 5, right panel, for $n = 20$). Clearly, in subsets A^{II}, A^{III}, A^{IV} the isotonizing correction produces a significant reduction in the UBEOR. A similar behavior is observed for other values of n, but is not reported here. Once again, as in the $p = 2$ case, we observe that the behavior of the UBEOR of Stein's estimator in comparison to that of the MLE is characterized by three distinct regimes. These regimes correspond to the sets $\Delta_{p,n}$, $\Gamma_{p,n}$, and $\Lambda_{p,n}$.

Even in the $p = 3$ case, similarly to the $p = 2$ case, there exists a subset of the domain in which: (a) $f^{\text{iso}}(\boldsymbol{x})$ is considerably less than the UBEOR of the MLE and simultaneously (b) Stein's estimator does not require the isotonizing correction, though the UBEOR of the raw estimator is greater than that of the MLE (see Sect. E in the Supplemental Material).

[3] For the precise mathematical definitions of A^I, A^{II}, A^{III}, A^{IV}, the reader is referred to Sect. E in the Supplemental Material.

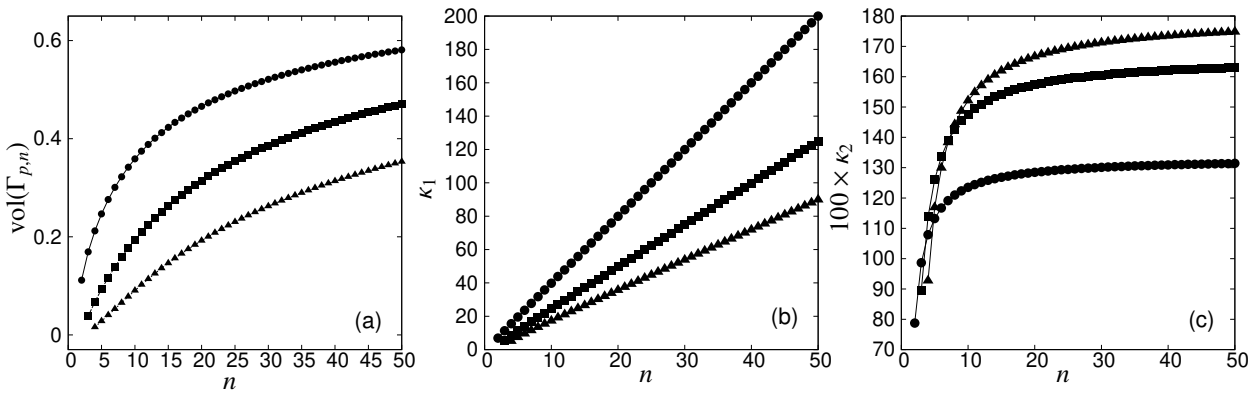


Figure 6: (a) Volume of $\Gamma_{p,n}$ as a function of n for $p = 2$ (●), $p = 3$ (■), $p = 4$ (▲); (b) κ_1 as a function of n for $p = 2$ (●), $p = 3$ (■), $p = 4$ (▲) and $p \leqslant n \leqslant 50$; (c) $100 \times \kappa_2$ as a function of n for $p = 2$ (●), $p = 3$ (■), $p = 4$ (▲) and $p \leqslant n \leqslant 50$.

As in the $p = 2$ case, the magnitude of the effect of the isotonizing algorithm on the UBEOR can be quantified by considering the quantities κ_1 and κ_2 (see (32) for the definition); they are plotted in Figs. 6(b) and (c) as a function of n for fixed p. Both quantities increase with increasing n; the former one increases approximately linearly, while the latter one tends to a finite value. The plots indicate that the reduction in UBEOR due to the isotonizing algorithm compared to that ofStein's raw estimator increases with sample size (see Fig. 6(b)). These increases are more pronounced for lower dimensions. On the contrary, the reduction in UBEOR due to the isotonizing algorithm is quite robust and actually seems to improve as p gets larger (see Fig. 6(c)). The asymptotic behaviors shown in Figs. 6(b) and (c) are proved for a general p in the next subsection. Analysis of the $p = 4$ case continues the trends observed for the $p = 3$ case, as illustrated by Figs. 6(a), (b) and (c)).

5.3. The arbitrary p case

We now evaluate the impact of the isotonizing correction on the UBEOR using κ_1 and κ_2 as before (see Sect. D in the Supplemental Material for the proof of the lemma):

Lemma 3. *For $p > 1$,*

$$\kappa_1 \sim \frac{6[p(p+1)-2]n}{p(p^2-1)} \quad (n \to \infty) \qquad \textit{and} \qquad \kappa_2 = 2\left[1 - \frac{2}{p(p+1)}\right]\left\{1 - \frac{2[p(2p+3)-1]}{12(p+1)n}\right\} + O\left(\frac{1}{n^2}\right) \quad (n \to \infty).$$

The behavior of κ_2 demonstrates that when all the eigenvalues are pooled together the use of the isotonizing algorithm can lead to a considerable reduction in the UBEOR relative to $K^{\mathrm{ml}}_{p,n}$. Moreover, the horizontal asymptotes observed empirically when studying the small p case (see Fig. 6(c)) can now be theoretically explained. A surprising phenomena emerges: the relative reduction in the UBEOR compared to the MLE persists even as the sample size increases. In addition, the dimension effect seen in Fig. 6(c) is also apparent from the expression of κ_2.

The behavior of κ_1 shows that the reduction in the UBEOR of Stein's estimator is orders of magnitude greater when the sample eigenvalues are close to each other as compared to when they are well separated (i.e., the first regime vs. the the third regime outlined at the end of the $p = 2$ section). In other words, the isotonizing algorithm reduces the UBEOR much more than Stein's raw estimator does. This suggests that some of the substantial risk reductions seen in Stein's estimator may actually be attributable to the isotonizing algorithm. It is also worth remarking that the convergence of the UBEOR of Stein's isotonized estimator to that of the MLE is not faster than $O(1/n)$ (see (D.15) in Sect. D in the Supplemental Material). This behavior differs from the analogous rate of $O(1/n^2)$ of Stein's raw estimator (see Proposition 2); the isotonizing correction has an effect of shifting Stein's estimator away from the MLE by an entire order of magnitude. The ratio of the two rates yields the linear function (of n) that was empirically observed in the analysis of the small p case (see Fig. 6(b)).

Lemma 3 quantify the maximal reduction in the UBEOR that are potentially achievable because of applying the isotonozing algorithm. Analyzing the UBEOR for the sample covariance structure which corresponds to the full-multiplicity case ($\boldsymbol{x} = (1, 1, \ldots, 1)$) is convenient in this regard as the isotonozing algorithm is being used in this

setting. Such an analysis can be regarded as corresponding to the case when the population structure is a multiple of the identity, i.e., when all the population eigenvalues are equal. Considering this particular settings yields analytic expressions that can be theoretically examined (without resorting to numerical calculations). The effect of the isotonizing algorithm can however also be examined under more general covariance structures. The study of the cases $p = 2, 3, 4$ addressed this issue broadly. The numerical study undertaken in the next section also provides further insights into the role of the isotonizing algorithm in reducing the risk of Stein's covariance estimator. This numerical analysis affirms the insights from the UBEOR framework and quantifies the actual risks of Stein's raw and isotonized estimators.

We also note that a theoretical analysis similar to the one above can be undertaken as both $p(n), n \to \infty$, and gives qualitatively similar results. Such analysis has been omitted for the sake of brevity.

6. Validation of the UBEOR approach via risk calculations

Recall that the UBEOR approach entails comparing two or more functions in order to qualitatively understand their respective expectations. The UBEOR approach asserts that $\mathbb{E}_\Sigma[F(\boldsymbol{l})] = \mathbb{E}_\Sigma[L_1(\widehat{\Sigma}, \Sigma)]$ (see (3)). The part of the parameter space in which Σ lies determines which part of the domain of $F^{\text{St+iso}}$ mainly contributes to the expectation $\mathbb{E}_\Sigma[F(\boldsymbol{l})]$. It is thus possible to understand the behavior of $\mathbb{E}_\Sigma[F^{\text{St+iso}}(\boldsymbol{l})]$ by combining the knowledge of $F^{\text{St+iso}}$ and Σ. It thus allows a better understanding of the regimes in which Stein's estimator and its isotonized version perform well.

The purpose of the numerical examples presented in this section is to demonstrate that the UBEOR approach does indeed yield a useful theoretical tool for understanding the performance of Stein's estimator globally over the parameter space. The case $p = 3$ is examined first and has the advantage that the graph of the UBEOR can be visualized. Thereafter a reference numerical study [6] on Stein's estimator is revisited in order to assess if insights obtained in the UBEOR framework also hold up in higher dimensions.

6.1. Validation of the UBEOR approach for the $p = 3$ case

In Sects. 4 and 5, the behavior of Stein's raw and isotonized estimators is explored via ratios of the type l_2/l_1 and l_3/l_2. From Fig. 5 (left and center panels), it is evident that considering the case where $l_2/l_1 = l_3/l_2 = r$ for $r \in (0, 1]$ corresponds to the straight line connecting $(0, 0)$ and $(1, 1)$. This parametrization will cover all three regions $\Delta'_{3,n}$ (or equivalently $\Delta_{p,n}$), $\Gamma_{3,n}$, and $\Lambda_{3,n}$ that correspond to the three different regimes relevant to the behavior of Stein's estimator in comparison to the MLE. Thus, we consider covariance matrices of the form: $\Sigma(r) = \text{diag}(1, r, r^2)$, $r \in (0, 1]$. In particular, by varying r from 0 to 1, $\Sigma(r)$ moves from the well separated case to the full multiplicity case. For a given sample from the multivariate normal distribution $\mathcal{N}_3(0, \Sigma(r))$, the ratios of the adjacent sample eigenvalues will of course not be exactly equal to r, but will be distributed around the point $(l_2/l_1 = r, l_3/l_2 = r)$. The goal of the study is to assess if the behavior of the UBEOR mirrors what is happening at the level of its expectation, the risk itself. In the numerical simulations that follow, the risk is computed for both the MLE and for Stein's isotonized estimator for covariance matrices of the above form. The percentage reduction in risk with respect to the MLE is given by

$$\gamma_{L_1} = \frac{R_1(\widehat{\Sigma}^{\text{ml}}, \Sigma) - R_1(\widehat{\Sigma}^{\text{St+iso}}, \Sigma)}{R_1(\widehat{\Sigma}^{\text{ml}}, \Sigma)} \times 100, \tag{34}$$

where $\widehat{\Sigma}^{\text{ml}}$ and $\widehat{\Sigma}^{\text{St+iso}}$ denote the MLE and Stein's isotonized estimator of Σ respectively. The risk γ_{L_1} is reported in Fig. 7(a) as a function of r for $n = 3, 20, 100$.

First consider the case when n is small. In this setting, the isotonizing correction is needed on a large portion of $(0, 1] \times (0, 1]$, and thus it produces a significant reduction in the UBEOR (Fig. 5, left panel). The insight given by the UBEOR approach is perfectly mirrored in actual risk calculations. In particular, for n close to p, Stein's isotonized estimator outperforms the MLE for all r (see Fig. 7(a) for $n = p = 3$). The percentage reduction in risk increases with increasing r. Indeed, as r increases, the point (r, r) moves from $(0, 0)$ towards $(1, 1)$ and out of the region $\Delta'_{3,n}$ where Stein's estimator does not require the isotonizing correction. It quickly moves into the region $\Lambda_{3,n}$ where at least two of the eigenvalues are pooled together (Fig. 5, center panel).

For greater n, the area of the set $\Delta'_{3,n}$, where $K^{\text{ml}}_{3,n} > f^{\text{St}}$, increases and converges rather rapidly to its asymptotic value (Figs. 3(b) and 5, left and center panels). If $(r, r) \in \Delta'_{3,n}$, Stein's estimator is expected to yield better results than

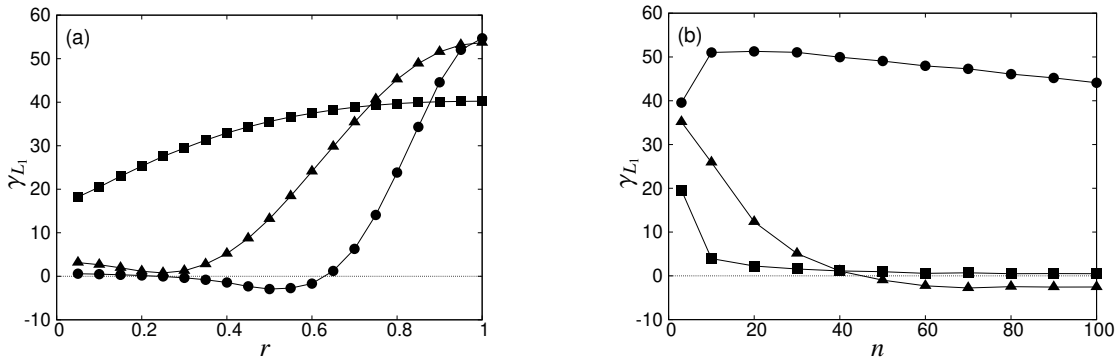


Figure 7: Percentage reduction in average loss for $p = 3$: (a) $n = p = 3$ (■), $n = 20$ (▲), $n = 100$ (●) and $r \in (0, 1]$; (b) $r = 0.1$ (■), $r = 0.5$ (▲), $r = 0.9$ (●) and $p \leqslant n \leqslant 100$.

the MLE, but given that the UBEOR is not modified by the isotonizing correction in $\Delta'_{3,n}$, the improvement is modest and decreases with increasing n (see Lemma 3, Fig. 3(c), and Fig. 7(a) for $n = 20, 100$).

As n increases, the region $\Gamma_{3,n}$ also increases in size (although at a lower rate as compared to $\Delta'_{3,n}$) and includes increasingly large values of r (Figs. 5 and 6(a)). We remind the reader that in $\Gamma_{3,n}$ Stein's estimator does not require the isotonizing correction despite the fact that $f^{\text{St}} > K^{\text{ml}}_{p,n}$. It is thus for $(r, r) \in \Gamma_{3,n}$ that the performance of Stein's estimator is expected to be the least satisfactory, and this effect is more pronounced for large n (Fig. 7(a) for $n = 100$).

While $\Delta'_{3,n}$ and $\Gamma_{3,n}$ expand as n increases, the region where the isotonizing correction applies shrinks towards values of r that are close to 1 (Fig. 5). Consequently, the larger the n, the closer r must be to 1 to be able to observe the benefit arising from the isotonizing algorithm (Fig. 7(a)). Once r is sufficiently close to 1 the improvement with respect to the MLE is much greater than the one observed for small r (see the behavior of κ_1 in Lemma 3 and Fig. 6(b) in the $p = 3$ case, as well as Fig 7(a)). Finally, for $r = 1$ and for large n, the value of γ_{L_1} does not vary appreciably with n. This behavior is in accordance with Lemma 3 (see also Fig. 6(c)).

The above constitutes an analysis in which n is fixed and r is changing. We now proceed to a fixed r and changing n analysis. For a fixed r, the dependence of the relative risk reduction γ_{L_1} upon n can once again be explained by considering the properties of the UBEOR. For small and intermediate r (i.e., when the eigenvalues are well separated or moderately separated), γ_{L_1} decreases with increasing n (see Fig. 7(b) for $r = 0.1$ and $r = 0.5$). This behavior is due to the contraction of the set $\Lambda_{3,n}$ (Fig. 5) and the simultaneous decrease of the percentage reduction in the UBEOR at $\boldsymbol{x} = (0, 0)$ (Lemma 2 and Fig. 3(c)). However, if (r, r) is well inside $\Delta'_{3,n}$, the relative risk reduction γ_{L_1} remains positive since $f^{\text{St}}(\boldsymbol{x}) < K^{\text{ml}}_{3,n}$. On the contrary, for intermediate values of r, the point (r, r) belongs to $\Gamma_{3,n}$, where $f^{\text{St+iso}}(\boldsymbol{x}) = f^{\text{St}}(\boldsymbol{x}) > K^{\text{ml}}_{3,n}$, and hence γ_{L_1} is negative (see Fig. 7(b) for $r = 0.5$). For r close to 1, γ_{L_1} is much less sensitive to the value of n, in the sense that the point (r, r) lies in the region where the isotonizing algorithm applies extensively (Fig. 5). The initial increase of the curve shown in Fig. 7(b) for $r = 0.9$ can be explained as the combination of two competing phenomena: a) the percentage reduction in the UBEOR at $\boldsymbol{x} = (1, 1)$ rapidly increases at small values of n (see Fig. 6(c) and Lemma 3), and b) although the set $\Lambda_{3,n}$ is contracting (Fig. 5), it is not contracting fast enough to counter the preceding effect. The subsequent decrease of γ_{L_1} can be understood in the context of the tapering off of the reduction in the UBEOR (see once more Fig. 6(c)). This tapering off as n increases can also be seen theoretically from the behavior of κ_2 in Lemma 3 and from the contraction of the set $\Lambda_{3,n}$ where the isotonizing correction is used (Fig. 5).

To demonstrate the broad generality and relevance of the insights from our UBEOR approach for higher values of p, an analogous investigation was performed for the $p = 6$ case with $\Sigma(r) = \text{diag}(1, r, r^2, r^3, r^4, r^5)$, $r \in (0, 1]$. The results shown in the Supplemental Material (Sect. F) for the $p = 6$ case confirm the explanation given in the $p = 3$ case.

In conclusion, Stein's estimator yields significantly better results than the MLE when some of the eigenvalues are close to each other, or when n is close to p, or when both these conditions are met (Regime III). It gives a modest amount of risk reduction when eigenvalues are sufficiently separated (Regime I). If $n \gg p$ and the adjacent eigenvalues are sufficiently separated, Stein's isotonized estimator does not yield a significant improvement over the MLE, and

can even yield slightly worse results (Regime II).

An analysis of the UBEOR of Stein's raw and isotonized estimators is able to qualitatively explain the complex risk reductions that are seen without explicitly computing the risk functions.

6.2. Validation of the UBEOR approach in the context of Lin and Perlman's [6] examples

An important numerical study of the performance of Stein's isotonized estimator is conducted by Lin and Perlman [6]. They study a variety of covariance matrices with structures that are important in both theoretical and applied statistics. All the covariance matrices have dimension $p = 6$, and at a first glance seem small, though it is important to note that there are 21 parameters in total. Moreover, the sample sizes that are considered, namely, $n = 6, 15, 30, 60, 100$ represent different finite-sample regimes, ranging all the way from when the estimator is barely defined to the moderate sample size regime. Moreover, we also confirm the analysis of the $p = 6$ case carries over to higher dimensions such as $p = 100, 200$ but is omitted here for brevity. In Ref. [6], Stein's isotonized estimator is found to perform significantly better than the MLE when the eigenvalues of Σ are close or can be divided into clusters of approximately equal eigenvalues. The improvement, however, is modest when the eigenvalues of Σ are more dispersed. We show below that the UBEOR analysis is able to give theoretical insight into Lin and Perlman's results.

For the purpose of comparison, the same examples as in Lin and Perlman [6] are repeated here. Define the p-dimensional vector $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_p)$ and the $p(p-1)/2$-dimensional vector $\boldsymbol{\rho} = (\rho_{21}; \rho_{31}, \rho_{32}; \ldots; \rho_{p1} \ldots, \rho_{p(p-1)})$. Let R denote the symmetric matrix with off-diagonal elements ρ_{ij}. The covariance matrices considered are of the form: $\Sigma = \mathrm{diag}(\boldsymbol{\sigma})R(\boldsymbol{\rho})\mathrm{diag}(\boldsymbol{\sigma})$. The following five 6×6 test matrices Σ_α $(1 \leqslant \alpha \leqslant 5)$ have been examined in Ref. [6]:

$\boldsymbol{\sigma}_1 = (1, 1, 1, 1, 1, 1)$, $\boldsymbol{\rho}_1 = (0; 0, 0; \ldots; 0, \ldots, 0)$;

$\boldsymbol{\sigma}_2 = (1, 1, 1, 1, 1, 1)$, $\boldsymbol{\rho}_2 = (0.9; 0.9, 0.9; \ldots; 0.9, \ldots, 0.9)$;

$\boldsymbol{\sigma}_3 = (3.08, 2.66, 3.00, 2.55, 4.73, 2.93)$,
$\boldsymbol{\rho}_3 = (0.60; -0.38, -0.45; 0.61, 0.43, -0.61; 0.09, 0.34, -0.51, 0.63; -0.36, 0.08, 0.36, -0.21, 0.20)$;

$\boldsymbol{\sigma}_4 = (1, 1, 1, 1, 1, 1)$, $\boldsymbol{\rho}_4 = (0.58; 0.61, 0.58; 0.60, 0.53, 0.94; 0.57, 0.53, 0.87, 0.88; 0.60, 0.55, 0.88, 0.88, 0.92)$;

$\boldsymbol{\sigma}_5 = (1, 2, 3, 4, 5, 6)$, $\boldsymbol{\rho}_5 = \boldsymbol{\rho}_4$.

The ratios of the adjacent eigenvalues of the matrices Σ_α are given in Table 1. As mentioned earlier, the covariance matrices considered above arise naturally in practice. The first matrix Σ_1 corresponds to the well known case of "sphericity" and is ubiquitous in hypothesis tests of covariance structure. The second matrix Σ_2 describes collections of random variables which are all highly positively correlated with each other. The corresponding population eigenvalue ensemble confirms that much of the variation in the random variables is explained by just one leading principal component, which is indicative of the intrinsic "low dimensional" nature of the covariance matrix. The third matrix Σ_3 describes the setting where there are both positive and negative correlations and is typical of various applications in genomics, environmental sciences, and finance. The population eigenvalue ensemble varies in such a way that the eigenvalues increase by powers of two. The fourth covariance matrix Σ_4 describes collections of random variables with equal variances that are all moderately positively correlated with each other and have just one leading principal component. Though Σ_4 is similar to Σ_2 in terms of correlations, the case where the correlations are only moderately large is more typical in applications. The fifth covariance matrix Σ_5 is similar to Σ_4 except that the variances (i.e., the diagonal terms) are increasing. This added flexibility in Σ_5 (as compared to the homoscedastic assumption of Σ_4) is also more realistic and has the effect of increasing the variation explained by the largest principal component.

The risk of Stein's isotonized estimator was computed for the matrices Σ_α, and the outcomes were compared with those of the MLE. The percentage reduction in risk (see (34)) is reported in Table 1 for the five test matrices Σ_α and for $n = 6, 15, 30, 60, 100$. Note that the very minor deviations from Lin and Perlman's numerical results [6] are essentially due to random variation (in addition, the number of simulations used here is one hundred times larger than in Ref. [6]). We are now in a position to interpret the above numerical results in the UBEOR framework.

The adjacent eigenvalues of the matrices Σ_1 and Σ_2 have ratios equal or close to 1. In these cases, as suggested by the UBEOR approach, Stein's estimator makes extensive use of the isotonizing algorithm and yields excellent results. As n increases, γ_{L_1} initially increases and eventually tapers off. This behavior is comparable to the one described in Subsect. 6.1 for fixed r close to 1 and changing n (see, e.g., Fig. F.1(b) for $r = 0.9$). The significant risk reductions

Σ_α	Ratios of eigenvalues	$n = 6$	$n = 15$	$n = 30$	$n = 60$	$n = 100$
Σ_1	$(1, 1, 1, 1, 1)$	53.3	71.3	74.5	75.7	76.1
Σ_2	$(0.02, 1, 1, 1, 1)$	47.0	54.2	53.4	52.5	52.0
Σ_3	$(0.5, 0.5, 0.5, 0.5, 0.5)$	38.7	20.5	7.1	0.2	-0.9
Σ_4	$(0.16, 0.60, 0.40, 0.47, 0.75)$	37.8	21.4	10.3	3.7	1.4
Σ_5	$(0.04, 0.86, 0.82, 0.29, 0.78)$	39.5	26.5	19.4	15.3	12.2

Table 1: Percentage reduction in average risk, γ_{L_1}, for $\widehat{\varphi}_j^{\text{St+iso}}$. Note the small differences in the risks from those given in Ref. [6].

for cases like Σ_1 and Σ_2 have given much credit to Stein's estimator, though it is clear from our investigation here that such risk reductions are primarily attributable to the isotonizing algorithm.

For Σ_3, the ratios of the adjacent eigenvalues are all equal to 0.5. Hence, for large n, they fall into the region where Stein's estimator does not require the isotonizing correction even though its UBEOR is greater than that of the MLE. Accordingly, γ_{L_1} decreases rapidly with increasing n and becomes negative for $n \gg p$ (see also Fig. F.1(b) for $r = 0.5$). A comparable decrease is observed for Σ_4 and Σ_5 owing to the contraction of the set where the isotonizing algorithm applies; however, the effect is less dramatic since some of the ratios between adjacent eigenvalues are very small or close to 1. The risk reduction in the cases Σ_3 and Σ_4 monotonically decreases as a function of n and is comparable to the risk of the MLE for $n = 60, 100$. This suggests that when n increases the need to isotonize decreases. We conclude by noting once more that the risk reductions observed mirror the UBEOR analysis conducted in earlier sections.

7. Summary and concluding remarks

In this paper a comprehensive theoretical investigation of Stein's covariance estimator in the UBEOR framework was undertaken, and three different regimes were identified regarding the behavior of Stein's UBEOR $F^{\text{St}}(\boldsymbol{l})$. These findings provide a theoretical means to understand how the risk of Stein's estimator depends on various covariance structures Σ. In the first regime, when all the eigenvalues of Σ are sufficiently separated, the sample eigenvalues are mainly distributed in the part of the domain in which the UBEOR of Stein's raw estimator is smaller than that of the MLE, with the latter taking a constant value $K_{p,n}^{\text{ml}}$. Thus, $\mathbb{E}_\Sigma[F^{\text{St+iso}}(\boldsymbol{l})]$ is less than $K_{p,n}^{\text{ml}}$, and hence the risk of Stein's estimator is less than that of the MLE. In this case, the isotonizing correction is not required and the reduction of the UBEOR (and hence of the risk) is to be attributed only to Stein's raw estimator. The reduction, however, is not very large and decreases as the sample size n increases. In the second regime, a) some of the eigenvalues of Σ are close together or b) are moderately separated but n is close to p. In this regime, Stein's estimator requires extensive use of the isotonizing correction, and has the effect of considerably reducing the UBEOR. Here the risk of Stein's isotonized estimator is considerably less than that of the MLE. Our analysis shows that the size of the set in which the isotonizing algorithm applies increases as the sample size n decreases and attains its maximum at $n = p$. Lastly, in the third regime, that is when the eigenvalues of Σ are only moderately separated and the sample size n is sufficiently greater than the dimension p, the UBEOR of Stein's raw estimator is greater than that of the MLE. Here the isotonizing algorithm does not apply, although were it to be applied, it would lead to a reduction in the UBEOR. As a consequence, the risk of Stein's estimator is comparable to that of the MLE (or even greater than it). These theoretical insights have also been corroborated at the level of risk functions via Monte Carlo simulations. We note that the UBEOR approach does not lead to a decision theoretic or optimality result: it is not meant to be, since it is already known from numerical work that Stein's estimator does not render the MLE inadmissible. The UBEOR approach is nevertheless able to obtain important insights into the theoretical workings of Stein's isotonized estimator. Such results are valuable in the sense that a theoretical understanding of Stein's estimator has been elusive since its introduction about 40 years ago.

8. Acknowledgments

The authors would like to thank Charles Stein for his encouragement for the work and Brett Naul for helping to draw figure 1 and for reading a previous version of the paper when it was completed. The authors acknowledge

funding from the France–Stanford Center for Interdisciplinary Studies which facilitated their collaborations. B.R. was supported in part by NSF Grant Nos. DMS 0906392, DMS-CMG 1025465, AGS-1003823, DMS-1106642 and grants NSA H98230-11-1-0194, DARPA-YFA N66001-11-1-4131, and SUWIEVP10-SUFSC10-SMSCVISG0906.

References

[1] Daniels, M.J., Kass, R.E.: Shrinkage estimators for covariance matrices. Biometrics **57**, 1173–1184 (2001)
[2] Haff, L.R.: The variational form of certain Bayes estimator. Ann. Statist. **19**, 1163–1190 (1991)
[3] Hamimeche, S., Lewis, A.: Properties and use of CMB power spectrum likelihoods. Phys. Rev. D **79**, 083012 (2009)
[4] Khare, K., Rajaratnam, B.: Wishart distributions for decomposable covariance graph models. Ann. Statist. **39**, 514–555 (2011)
[5] Ledoit, O., Wolf, M.: A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. **88**, 365–411 (2004)
[6] Lin, S., Perlman, M.: A Monte-Carlo comparison of four estimators of a covariance matrix. In *Multivariate Analisis VI* (P.R. Krishnaiah, ed.) 411–429. North Holland, Amsterdam (1985)
[7] Pope, A.C., Szapudi, I.: Shrinkage estimation of the power spectrum covariance matrix. Mon. Not. R. Astron. Soc. **389**, 766–774 (2008)
[8] Pourahmadi, M., Covariance Estimation: The GLM and Regularization Perspectives. Statistical Science **26**, 369–387 (2011)
[9] Rajaratnam, B., Massam, H., Carvalho, C.M.: Flexible covariance estimation in graphical Gaussian models. Ann. Statist. **36**, 2818–2849 (2008)
[10] Rajaratnam, B., Vincenzi, D.: A note on covariance estimation in the unbiased estimator of risk framework, submitted (2015)
[11] Schäfer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statist. Appl. Genet. Mol. Biol. **4**, 32 (2005).
[12] Stein, C.: Some problems in multivariate analysis. Technical report No. 6, Stanford University (1956)
[13] Stein, C.: Estimation of a covariance matrix. In *Riesz Lecture. 39th Annual Meeting*, IMS, Atlanta, GA (1975)
[14] Stein, C.: Lectures on the theory of estimation of many parameters (in Russian). In *Studies in the Statistical Theory of Estimation, Part I* (I.A. Ibragimov and M.S. Nikulin, eds.), *Proceedings of Scientific Seminars of the Steklov Institute, Leningrad Division* **74**, 4–65 (1977)
[15] Stein, C: Lectures on the theory of estimation of many parameters. J. Math. Sci. **34**, 1373–1403 (1986)
[16] Won, J., Kim, S.J., Lim, J., Rajaratnam, B.: Condition Number-Regularized Covariance Estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75:427–450 (2013)

Supplemental Material

A. Monotonicity of the UBEOR for $p = 2$

The derivative of f^{St} is given by:

$$\frac{df^{\mathrm{St}}}{dx_1} = \frac{4[n+1+(n-1)x_1]}{[n+1-(n-1)x_1]^3} + \frac{4[(n+1)x_1+n-1]}{[n-1-(n+1)x_1]^3} + \frac{4(x_1+1)}{(1-x_1)[n+1-(n-1)x_1][(n+1)x_1-(n-1)]}.$$

We want to show that $df^{\mathrm{St}}/dx_1 > 0$ for all $x_1 \in \Omega_{2,n} = (0, \tilde{x}_1)$. The first and the second term on the right-hand-side are strictly positive, whereas the third one is negative for all $x_1 \in \Omega_{2,n}$. We focus on the sum of the second and the third term:

$$\frac{4}{(n+1)x_1-(n-1)}\left\{-\frac{(n+1)x_1+n-1}{[(n+1)x_1-(n-1)]^2} + \frac{x_1+1}{(1-x_1)[n+1-(n-1)x_1]}\right\}.$$

In the above expression, the prefactor is strictly negative for all $x_1 \in \Omega_{2,n}$. We thus need to consider only the terms inside the curly bracket and show that their sum is negative. We note that since $x_1 > 0$,

$$|(n+1)x_1-(n-1)| < n+1-(n-1)x_1,$$

and therefore

$$\frac{1}{|(n+1)x_1-(n-1)|} > \frac{1}{n+1-(n-1)x_1}.$$

To prove that $df^{\mathrm{St}}/dx_1 > 0$, it is then sufficient to show that

$$\frac{(n+1)x_1+n-1}{|(n+1)x_1-(n-1)|} > \frac{1+x_1}{1-x_1}.$$

This inequality is satisfied for all $x_1 \in \Omega_{2,n}$ since on $\Omega_{2,n}$ it reduces to $x_1 > 0$.

B. The UBEOR of Stein's raw estimator for $p = 3$

Figure B.1 shows that for $p = 3$ the function f^{St} is neither convex nor concave.

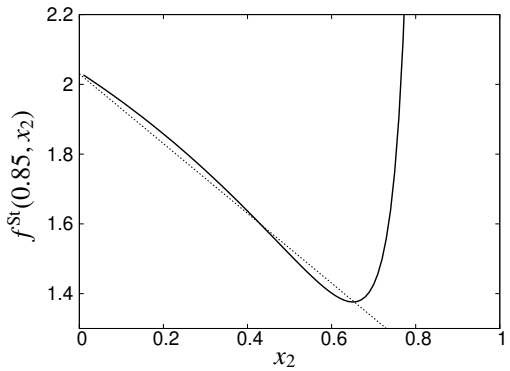


Figure B.1: Behavior of $f^{\mathrm{St}}(x_1, x_2)$ as a function of x_2 for $x_1 = 0.85$, $p = 4$, $n = 20$ (solid line). The dashed line has been drawn to show that f^{St} is neither convex nor concave.

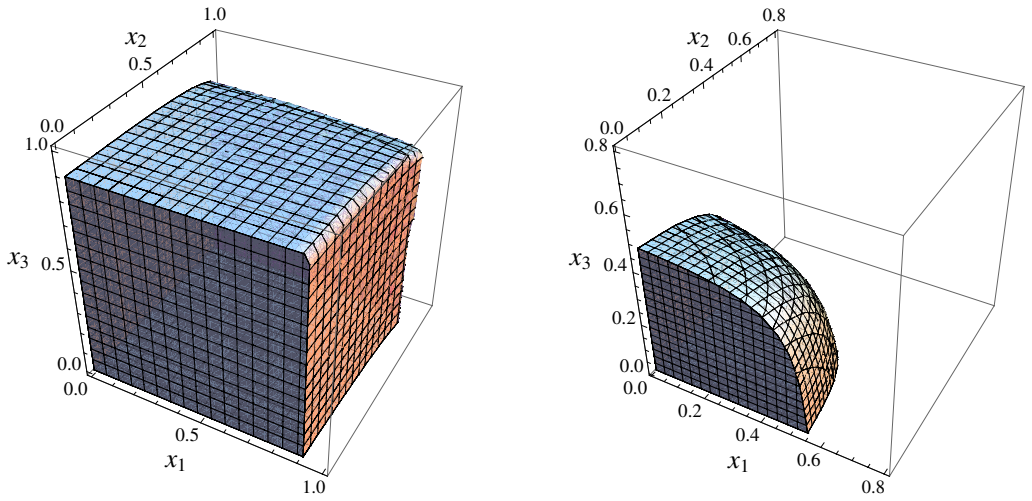
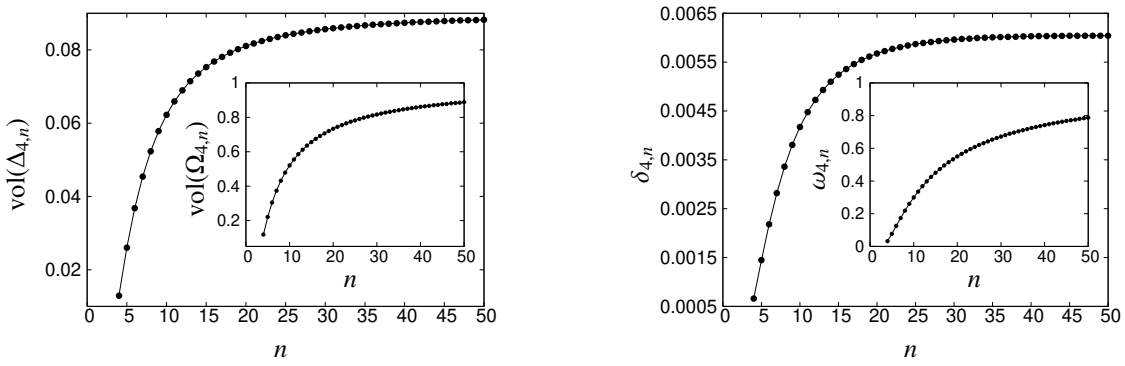


Figure C.1: Left: the set $\Omega_{4,20}$. Right: the set $\Delta_{4,20}$. The sets $\Delta_{4,20}$ and $\Delta'_{4,20}$ are numerically indistiguishable.



(a) (b)

Figure C.2: (a) Volumes of the sets $\Delta_{4,n}$ and $\Omega_{4,n}$ as a function of n; (b) Volume ratios occupied by the preimages of the sets $\Delta_{4,n}$ and $\Omega_{4,n}$ under $\boldsymbol{\nu}$.

C. The UBEOR of Stein's raw estimator for $p = 4$

For $p = 4$, the graph of f^{St} cannot be visualized, but the sets $\Omega_{4,n}$ and $\Delta_{4,n}$ can be computed numerically (see Fig. C.1). Once more, we observe that f^{St} is less than $K^{\mathrm{ml}}_{4,n}$ when $\|\boldsymbol{x}\|$ is sufficiently small, and that $\Delta_{4,n}$ is a connected set. The volumes of the two sets are reported in Fig. 2(a) as a function of n. When the comparison between Stein's UBEOR and that of the MLE is undertaken in the original $\boldsymbol{l}$-space, the volume of the set where $F^{\mathrm{St}}(\boldsymbol{l})$ is less than $K^{\mathrm{ml}}_{4,n}$ turns out to be very small as compared to the volume of $\mathscr{D}_p$ (see Fig. 2(b)).

The function f^{St} is not monotonic in each separate variable x_i, nor is it convex (Fig. C.3).

At $\boldsymbol{x} = 0$, the percentage reduction in the UBEOR of Stein's raw estimator relative to the MLE (see (29)) decreases approximately linearly with increasing n (Fig. 3(c)).

As mentioned in Subsect. 4.2, the above properties are similar to those of the $p = 3$ case.

D. Proofs of Lemmas and Propositions

Proof of Lemma 1. We prove that for any $n \geqslant p > 1$ and for any $q \geqslant 2$, there exists $\tilde{\boldsymbol{x}} \in (0,1)^{p-1}$ such that $a_q(\tilde{\boldsymbol{x}}) = 0$ and $a_j(\tilde{\boldsymbol{x}}) > 0$ for all $j = 1, \ldots, p$, $j \neq q$. The Lemma follows from a straightforward generalization of the following proof.

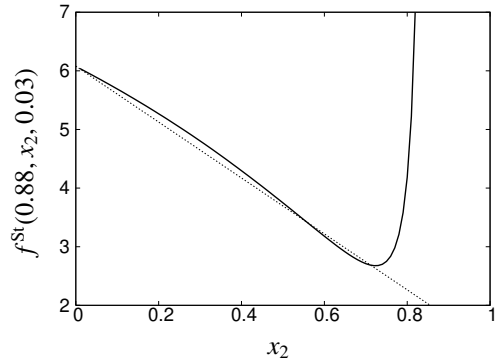


Figure C.3: Behavior of $f^{\mathrm{St}}(x_1, x_2, x_3)$ as a function of x_2 for $x_1 = 0.88$, $x_3 = 0.03$, $p = 4$, $n = 20$ (solid line). The dashed line has been drawn to show that f^{St} is neither convex nor concave.

Let $q \geqslant 2$, $\epsilon \in (0, 1)$, and $\eta \in (0, 1)$. Consider $\tilde{\boldsymbol{x}} \in (0, 1)^{p-1}$ such that

$$\begin{aligned} \tilde{x}_k &= \epsilon \qquad \forall\, k = 1, \ldots, p, \quad k \neq q-1, \\ \tilde{x}_{q-1} &= 1 - \eta. \end{aligned}$$

Then, for $i < j$,

$$\pi_i^j(\tilde{\boldsymbol{x}}) = \begin{cases} \epsilon^{j-i} & \text{if } 1 \leqslant j < q \\ (1-\eta)\epsilon^{q-i-1} & \text{if } j = q \text{ and } 1 \leqslant i < q-1 \\ (1-\eta) & \text{if } j = q \text{ and } i = q-1 \\ (1-\eta)\epsilon^{j-i-1} & \text{if } q < j \leqslant p \text{ and } 1 \leqslant i < q \\ \epsilon^{j-i} & \text{if } q < j \leqslant p \text{ and } q \leqslant i < j, \end{cases}$$

whence the negative contribution to $a_j(\tilde{\boldsymbol{x}})$ (see (27)) behaves for $\epsilon, \eta \to 0$ as follows:

$$s_j(\tilde{\boldsymbol{x}}) := \sum_{i=1}^{j-1} \frac{\pi_i^j(\tilde{\boldsymbol{x}})}{\pi_i^j(\tilde{\boldsymbol{x}}) - 1} = \begin{cases} -\epsilon + O\left(\epsilon^2\right) & \text{if } 2 \leqslant j < q, \\ -\dfrac{1}{\eta} + 1 - \epsilon + O(\epsilon^2) + O(\epsilon)O(\eta) & \text{if } j = q, \\ -2\epsilon + O\left(\epsilon^2\right) + O(\epsilon)O(\eta) & \text{if } j = q+1, \\ -\epsilon + O\left(\epsilon^2\right) + O(\epsilon)O(\eta) & \text{if } q+1 < j \leqslant p. \end{cases}$$

For $j \neq q$, $|s_j(\tilde{\boldsymbol{x}})|$ can be made arbitrarily small by reducing ϵ, and hence there exists $\epsilon \in (0, 1)$ such that $a_j(\tilde{\boldsymbol{x}})$ is positive for all $j \neq q$. Simultaneously, η can be adjusted to increase $|s_q(\tilde{\boldsymbol{x}})|$ in such a way that $a_q(\tilde{\boldsymbol{x}})$ vanishes.

□

Proof of Proposition 1. Recall the definition of f^{St} given in (16). It can easily be proved that if $\boldsymbol{x}$ belongs to a neighborhood of $\tilde{\boldsymbol{x}}$, then $x_k < 1$ for all $k = 1, \ldots, p$ and hence $\pi_i^j(\boldsymbol{x}) < 1$ for all i, j. Thus, in a neighborhood of $\tilde{\boldsymbol{x}}$ the sums

$$\sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} \qquad \text{and} \qquad \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[\pi_j^i(\boldsymbol{x}) - 1]^2}$$

are bounded for all $j = 1, \ldots, p$. The same holds for all the $a_j(\boldsymbol{x})$, $j = 1, \ldots, p$. As a result, for all $j \notin M$, the j-th summand in (16) satisfies:

$$\left| 1 + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[1 - \pi_j^i(\boldsymbol{x})]^2} + \ln(a_j(\boldsymbol{x})) \right| < +\infty \tag{D.1}$$

for all $\boldsymbol{x}$ belonging to a neighborhood of $\tilde{\boldsymbol{x}}$. By contrast, for $j \in M$, we have

$$\lim_{\boldsymbol{x}\to\tilde{\boldsymbol{x}}} \left\{ 1 + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} + \frac{4}{a_j^2(\boldsymbol{x})} \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[1 - \pi_j^i(\boldsymbol{x})]^2} + \ln(a_j(\boldsymbol{x})) \right\} = +\infty \tag{D.2}$$

given that $a_j(\boldsymbol{x}) \to 0$ as $\boldsymbol{x} \to \tilde{\boldsymbol{x}}$ and $0 < \pi_i^j(\boldsymbol{x}) < 1$ in a neighborhood of $\tilde{\boldsymbol{x}}$. Equations (D.1) and (D.2) imply that f^{St} diverges positively as $\boldsymbol{x} \to \tilde{\boldsymbol{x}}$.

□

Proof of Proposition 2. By using (17) and (16), we can rewrite f^{St} in the form:

$$f^{\mathrm{St}}(\boldsymbol{x}) = \sum_{j=1}^{p} \left\{ 1 + \frac{g_j(\boldsymbol{x})}{[n + h_j(\boldsymbol{x})]^2} + \ln(n + h_j(\boldsymbol{x})) \right\}, \tag{D.3}$$

where

$$g_j(\boldsymbol{x}) = 4 \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{[\pi_i^j(\boldsymbol{x}) - 1]^2} + 4 \sum_{i=j+1}^{p} \frac{\pi_j^i(\boldsymbol{x})}{[1 - \pi_j^i(\boldsymbol{x})]^2}$$

and

$$h_j(\boldsymbol{x}) = a_j(\boldsymbol{x}) - n = -p + 1 + 2 \sum_{i=1}^{j-1} \frac{\pi_i^j(\boldsymbol{x})}{\pi_i^j(\boldsymbol{x}) - 1} + 2 \sum_{i=j+1}^{p} \frac{1}{1 - \pi_j^i(\boldsymbol{x})}.$$

For all $n \geqslant p > 1$, (13) yields

$$\sum_{j=1}^{p} a_j(\boldsymbol{x}) = np \qquad \forall\, \boldsymbol{x} \in \Omega_{p,n}, \tag{D.4}$$

whence

$$\sum_{j=1}^{p} h_j(\boldsymbol{x}) = 0. \tag{D.5}$$

We have

$$\frac{g_j(\boldsymbol{x})}{[n + h_j(\boldsymbol{x})]^2} = \frac{g_j(\boldsymbol{x})}{n^2} + O\left(\frac{1}{n^3}\right) \qquad (n \to \infty) \tag{D.6}$$

and

$$\ln(n + h_j(\boldsymbol{x})) = \ln n + \ln\left(1 + \frac{h_j(\boldsymbol{x})}{n}\right) = \ln n + \frac{h_j(\boldsymbol{x})}{n} - \frac{h_j^2(\boldsymbol{x})}{2n^2} + O\left(\frac{1}{n^3}\right) \tag{D.7}$$

as $n \to \infty$. Equation (28) can be obtained by substituting (D.6) and (D.7) into (D.3) and by using (D.5).

Furthermore, Proposition 1 implies that the convergence is not uniform.

□

Proof of Proposition 3. The value of f^{St} at $\boldsymbol{x} = \boldsymbol{0}$ is

$$f^{\mathrm{St}}(\boldsymbol{0}) = p + \sum_{j=1}^{p} \ln(n + p + 1 - 2j) - c_{p,n}.$$

The difference between the two unbiased estimators of risk can be written as

$$K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\boldsymbol{0}) = \sum_{j=1}^{p} [\ln(n) - \ln(n + p + 1 - 2j)] = \ln\left(\prod_{j=1}^{p} \frac{n}{n + p + 1 - 2j} \right). \tag{D.8}$$

Thus, proving that $K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0})$ is an increasing function of p reduces to showing that

$$\prod_{j=1}^{p+1} \frac{n}{n+p+2-2j} > \prod_{j=1}^{p} \frac{n}{n+p+1-2j}$$

or, equivalently,

$$\frac{n}{n-p} \prod_{j=1}^{p} \frac{n+p+1-2j}{n+p+2-2j} > 1. \tag{D.9}$$

The above inequality can be proved by induction. For $p = 2$ and $p = 3$, inequality (D.9) is trivially satisfied for all $n > 1$. We now assume that inequality (D.9) holds for a given p (inductive hypothesis). For $p + 2$, the left-hand-side of (D.9) is written

$$\begin{aligned}\frac{n}{n-p+2} \prod_{j=1}^{p+2} \frac{n+p+3-2j}{n+p+4-2j} &= \frac{n}{n-p+2} \prod_{j=1}^{p+2} \frac{n+p+1-2(j-1)}{n+p+2-2(j-1)} \\ &= \frac{n}{n-p+2} \prod_{k=0}^{p+1} \frac{n+p+1-2k}{n+p+2-2k} = \frac{n^2-(p+1)^2}{n^2-(p+2)^2} \frac{n}{n-p} \prod_{k=1}^{p} \frac{n+p+1-2k}{n+p+2-2k},\end{aligned}$$

whence

$$\frac{n}{n-p+2} \prod_{j=1}^{p+2} \frac{n+p+3-2j}{n+p+4-2j} > \frac{n^2-(p+1)^2}{n^2-(p+2)^2} > 1.$$

This argument holds $\forall\ 1 < p \leqslant n$. The difference $K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0})$ therefore increases with increasing p. Since $K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(0) = 0$ for $p = 1$, $K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0})$ is strictly positive for all $p > 1$.

As for the order of the estimated eigenvalues, the restriction $l_j/\alpha_j(\boldsymbol{l}) \geqslant l_{j+1}/\alpha_{j+1}(\boldsymbol{l})$ is equivalent to (see (18))

$$x_j \leqslant \frac{a_{j+1}(\boldsymbol{x})}{a_j(\boldsymbol{x})},$$

where $\boldsymbol{x}$ is such that $x_j = l_{j+1}/l_j$. For $\boldsymbol{x} \to \mathbf{0}$, the above inequality is written:

$$0 \leqslant \frac{n+p-1-2j}{n+p+1-2j},$$

and is satisfied for all $j = 1, \ldots, p-1$.

The Proposition then follows from the continuity of f^{St} and of the a_j.

□

Proof of Lemma 2. Recall the asymptotic expansion of the psi function (Abramowitz, M., Stegun, I.A.: Handbook of Mathematical Functions. Dover Publications, 1964; formula 6.3.18):

$$\psi^{(0)}(z) := \frac{\Gamma'(z)}{\Gamma(z)} = \ln z - \frac{1}{2z} - \frac{1}{12z^2} + O\left(\frac{1}{z^4}\right) \qquad (z \to \infty). \tag{D.10}$$

The asymptotic expansion of the constant $c_{p,n}$, whose definition is given in (5), follows from those of $\ln z$ and of $\psi^{(0)}(z)$ above:

$$c_{p,n} = p(1 + \ln n) - \frac{p(p+1)}{2n} - \frac{p[p(2p+3)-1]}{12n^2} + O\left(\frac{1}{n^3}\right) \qquad (n \to \infty).$$

Hence

$$K_{p,n}^{\mathrm{ml}} = \frac{p(p+1)}{2n} + \frac{p[p(2p+3)-1]}{12n^2} + O\left(\frac{1}{n^3}\right) \qquad (n \to \infty). \tag{D.11}$$

Furthermore, (D.8) yields:

$$K_{p,n}^{\mathrm{ml}} - f^{\mathrm{St}}(\mathbf{0}) = -\ln \prod_{j=1}^{p} \left(1 + \frac{p+1-2j}{n}\right) = -\sum_{j=1}^{p} \ln\left(1 + \frac{p+1-2j}{n}\right) \sim \sum_{j=1}^{p} \frac{(p+1-2j)^2}{2n^2} = \frac{p(p^2-1)}{6n^2} \qquad (n \to \infty). \tag{D.12}$$

The Lemma then follows from (D.11) and (D.12). □

Proof of Lemma 3. The Lemma is more easily proved by using the original variables $\boldsymbol{l}$ rather than their ratios. When all the eigenvalues are pooled together, $\widehat{\psi}_j^{\mathrm{St+iso}}(\boldsymbol{l})$ is written:

$$\widehat{\psi}_j^{\mathrm{St+iso}}(\boldsymbol{l}) = \frac{1}{l_j} \frac{\sum_{k=1}^{p} l_k}{\sum_{k=1}^{p} \alpha_k(\boldsymbol{l})} = \frac{1}{l_j} \frac{\sum_{k=1}^{p} l_k}{np} \qquad j = 1, \ldots, p, \tag{D.13}$$

where the last equality follows from (13). Moreover,

$$l_j \frac{\partial}{\partial l_j} \widehat{\psi}_j^{\mathrm{St+iso}} \bigg|_{\boldsymbol{l}} = -\frac{1}{l_j} \frac{\sum_{k=1}^{p} l_k}{np} + \frac{1}{np} \qquad j = 1, \ldots, p. \tag{D.14}$$

Substituting (D.13) and (D.14) in (4) yields:

$$F^{\mathrm{St+iso}}(\boldsymbol{l}) = \sum_{j=1}^{p} \left[\frac{(n-p-1)}{np} \sum_{k=1}^{p} \frac{l_k}{l_j} + \frac{2}{np} - \ln\left(\frac{1}{np} \sum_{k=1}^{p} \frac{l_k}{l_j} \right) \right] - c_{p,n}$$

with the last equality following from:

$$\sum_{j=1}^{p} \sum_{i \neq j} \frac{1}{l_j - l_i} = 0.$$

If $\boldsymbol{l}^\star$ is such that $l_1^\star = l_2^\star = \cdots = l_p^\star$, then

$$F^{\mathrm{St+iso}}(\boldsymbol{l}^\star) = p(1 + \ln n) - \frac{p(p+1)-2}{n} - c_{p,n} = K_{p,n}^{\mathrm{ml}} - \frac{p(p+1)-2}{n},$$

and hence

$$f^{\mathrm{St+iso}}(1, 1, \ldots, 1) = K_{p,n}^{\mathrm{ml}} - \frac{p(p+1)-2}{n}. \tag{D.15}$$

The Lemma can then be proved by using (D.12) and (D.15).

The expansion of κ_2 is a consequence of (D.11) and (D.15). □

E. Stein's isotonized estimator for $p = 3$

For $p = 3$ Stein's isotonized estimator takes the form reported below. Different cases should be distinguished depending on the point $\boldsymbol{l} \in \mathscr{D}_p$:[4]

1. if $\alpha_2(\boldsymbol{l}), \alpha_3(\boldsymbol{l}) > 0$ and

   a) $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} \geqslant \dfrac{l_2}{\alpha_2(\boldsymbol{l})} \geqslant \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_j^{\mathrm{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_j^{\mathrm{St}}(\boldsymbol{l}) = \frac{l_j}{\alpha_j(\boldsymbol{l})}, \qquad j = 1, 2, 3;$$

[4]Note that $\alpha_1(\boldsymbol{l}) > 0$ for all $\boldsymbol{l} \in \mathscr{D}_p$ and hence the analysis of the isotonized algorithm needs to consider the sign of $\alpha_2(\boldsymbol{l})$ and $\alpha_3(\boldsymbol{l})$ and whether the order is violated by the estimates $(l_1/\alpha_1(\boldsymbol{l}), l_2/\alpha_2(\boldsymbol{l}), l_3/\alpha_3(\boldsymbol{l}))$.

b) $\dfrac{l_2}{\alpha_2(\boldsymbol{l})} < \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$ and $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} \geqslant \dfrac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_1^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1}{\alpha_1(\boldsymbol{l})} \quad \text{and} \quad \widehat{\varphi}_2^{\text{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_3^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})};$$

c) $\dfrac{l_2}{\alpha_2(\boldsymbol{l})} < \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$ and $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} < \dfrac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2+l_3}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})} = \frac{l_1+l_2+l_3}{3n}, \qquad j = 1,2,3,$$

with the last equality following from (13);

d) $\dfrac{l_2}{\alpha_2(\boldsymbol{l})} \geqslant \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} < \dfrac{l_2}{\alpha_2(\boldsymbol{l})}$, and $\dfrac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} \geqslant \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_1^{\text{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_2^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} \quad \text{and} \quad \widehat{\varphi}_3^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_3}{\alpha_3(\boldsymbol{l})};$$

e) $\dfrac{l_2}{\alpha_2(\boldsymbol{l})} \geqslant \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} < \dfrac{l_2}{\alpha_2(\boldsymbol{l})}$, and $\dfrac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} < \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2+l_3}{3n}, \qquad j = 1,2,3;$$

2. if $\alpha_2(\boldsymbol{l}) > 0$, $\alpha_3(\boldsymbol{l}) \leqslant 0$, and

a) $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} \geqslant \dfrac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})} > 0$, then

$$\widehat{\varphi}_1^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1}{\alpha_1(\boldsymbol{l})} \quad \text{and} \quad \widehat{\varphi}_2^{\text{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_3^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})};$$

b) $\dfrac{l_1}{\alpha_1(\boldsymbol{l})} < \dfrac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})}$ or $\dfrac{l_2+l_3}{\alpha_2(\boldsymbol{l})+\alpha_3(\boldsymbol{l})} \leqslant 0$, then

$$\widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2+l_3}{3n}, \qquad j = 1,2,3;$$

3. if $\alpha_2(\boldsymbol{l}) \leqslant 0$, $\alpha_3(\boldsymbol{l}) > 0$, and

a) $\dfrac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} \geqslant \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_1^{\text{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_2^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} \quad \text{and} \quad \widehat{\varphi}_3^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_3}{\alpha_3(\boldsymbol{l})};$$

b) $\dfrac{l_1+l_2}{\alpha_1(\boldsymbol{l})+\alpha_2(\boldsymbol{l})} < \dfrac{l_3}{\alpha_3(\boldsymbol{l})}$, then

$$\widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2+l_3}{3n}, \qquad j = 1,2,3;$$

4. if $\alpha_2(\boldsymbol{l}) \leqslant 0$, $\alpha_3(\boldsymbol{l}) \leqslant 0$, then

$$\widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l}) = \frac{l_1+l_2+l_3}{3n}, \qquad j = 1,2,3.$$

The UBEOR of the isotonized estimator is obtained by setting $\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{l}) = \widehat{\varphi}_j^{\text{St+iso}}(\boldsymbol{l})/l_j$, $j = 1, 2, 3$, and by substituting the above expressions in (4). The result is thus a function defined piecewise.

As was noted in Section 5, $\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{l})$ can then be rewritten in terms of the ratios $x_j = l_{j+1}/l_j$. Define the sets:

$$A^I = \left\{\boldsymbol{x} \in (0,1]^2 : a_{j+1}(\boldsymbol{x}) > 0 \text{ and } x_j \leqslant \frac{a_{j+1}(\boldsymbol{x})}{a_j(\boldsymbol{x})} \quad \forall\, j = 1, 2\right\},$$

$$B_1^{II} = \left\{\boldsymbol{x} \in (0,1]^2 : a_2(\boldsymbol{x}) > 0, a_3(\boldsymbol{x}) > 0, x_2 > \frac{a_3(\boldsymbol{x})}{a_2(\boldsymbol{x})}, \text{ and} x_1(1+x_2) \leqslant \frac{a_2(\boldsymbol{x}) + a_3(\boldsymbol{x})}{a_1(\boldsymbol{x})}\right\},$$

$$B_2^{II} = \left\{\boldsymbol{x} \in (0,1]^2 : a_2(\boldsymbol{x}) > 0, a_3(\boldsymbol{x}) \leqslant 0, \text{ and } \frac{1}{a_1(\boldsymbol{x})} \geqslant \frac{x_1(1+x_2)}{a_2(\boldsymbol{x}) + a_3(\boldsymbol{x})} > 0\right\},$$

$$B_1^{III} = \left\{\boldsymbol{x} \in (0,1]^2 : a_2(\boldsymbol{x}) > 0, a_3(\boldsymbol{x}) > 0, x_2 \leqslant \frac{a_3(\boldsymbol{x})}{a_2(\boldsymbol{x})}, x_1 > \frac{a_2(\boldsymbol{x})}{a_1(\boldsymbol{x})}, \text{ and } \frac{1+x_1}{a_1(\boldsymbol{x}) + a_2(\boldsymbol{x})} \geqslant \frac{x_1 x_2}{a_3(\boldsymbol{x})}\right\},$$

$$B_2^{III} = \left\{\boldsymbol{x} \in (0,1]^2 : a_2(\boldsymbol{x}) \leqslant 0, a_3(\boldsymbol{x}) > 0, \text{ and } \frac{1+x_1}{a_1(\boldsymbol{x}) + a_2(\boldsymbol{x})} \geqslant \frac{x_1 x_2}{a_3(\boldsymbol{x})}\right\}.$$

Depending on the value of $\boldsymbol{x} \in (0,1)$, $\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{x})$ takes one of the following four forms:[5]

$$\widehat{\psi}_j^I(\boldsymbol{x}) = \frac{1}{a_j(\boldsymbol{x})}, \quad j = 1, 2, 3, \tag{E.1}$$

if $\boldsymbol{x} \in A^I$;

$$\widehat{\psi}_1^{II}(\boldsymbol{x}) = \frac{1}{a_1(\boldsymbol{x})}, \quad \widehat{\psi}_2^{II}(\boldsymbol{x}) = \frac{1+x_2}{a_2(\boldsymbol{x}) + a_3(\boldsymbol{x})}, \quad \widehat{\psi}_3^{II}(\boldsymbol{x}) = \frac{1+x_2^{-1}}{a_2(\boldsymbol{x}) + a_3(\boldsymbol{x})}, \tag{E.2}$$

if $\boldsymbol{x} \in A^{II} = B_1^{II} \cup B_2^{II}$;

$$\widehat{\psi}_1^{III}(\boldsymbol{x}) = \frac{1+x_1}{a_1(\boldsymbol{x}) + a_2(\boldsymbol{x})}, \quad \widehat{\psi}_2^{III}(\boldsymbol{x}) = \frac{1+x_1^{-1}}{a_1(\boldsymbol{x}) + a_2(\boldsymbol{x})}, \quad \widehat{\psi}_3^{III}(\boldsymbol{x}) = \frac{1}{a_3(\boldsymbol{x})} \tag{E.3}$$

if $\boldsymbol{x} \in A^{III} = B_1^{III} \cup B_2^{III}$;

$$\widehat{\psi}_1^{IV}(\boldsymbol{x}) = \frac{1 + x_1 + x_1 x_2}{3n}, \quad \widehat{\psi}_2^{IV}(\boldsymbol{x}) = \frac{1 + x_1^{-1} + x_2}{3n}, \quad \widehat{\psi}_3^{IV}(\boldsymbol{x}) = \frac{1 + x_2^{-1} + x_1^{-1} x_2^{-1}}{3n}, \tag{E.4}$$

if $\boldsymbol{x} \in A^{IV} = ((0,1] \times (0,1]) \setminus \left(A^I \cup A^{II} \cup A^{III}\right)$. Thus, for $j = 1, 2, 3$, $\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{x})$ is written as:

$$\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{x}) = \begin{cases} \widehat{\psi}_j^I(\boldsymbol{x}) & \boldsymbol{x} \in A^I \\ \widehat{\psi}_j^{II}(\boldsymbol{x}) & \boldsymbol{x} \in A^{II} \\ \widehat{\psi}_j^{III}(\boldsymbol{x}) & \boldsymbol{x} \in A^{III} \\ \widehat{\psi}_j^{IV}(\boldsymbol{x}) & \boldsymbol{x} \in A^{IV}. \end{cases} \tag{E.5}$$

Finally, the UBEOR of the isotonized estimator can be rewritten in terms of the ratios $x_j = l_{j+1}/l_j$ by using the above expressions for $\widehat{\psi}_j^{\text{St+iso}}(\boldsymbol{x})$.

Denote by $f^{II}(\boldsymbol{x})$ the UBEOR obtained by using the estimator $\psi_j^{II}(\boldsymbol{x})$ (see (E.2)) everywhere on $(0,1] \times (0,1]$. If $f_{|A^{II}}^{\text{St+iso}}$ denotes the function $f^{\text{St+iso}}$ restricted to the subset A^{II}, then f^{II} clearly represents the extension of $f_{|A^{II}}^{\text{St+iso}}$ to

[5] With a slight abuse of notation, we continue to write $\widehat{\psi}_j^{\text{St+iso}}$ to denote the estimator expressed in terms of the ratios l_{j+1}/l_j.

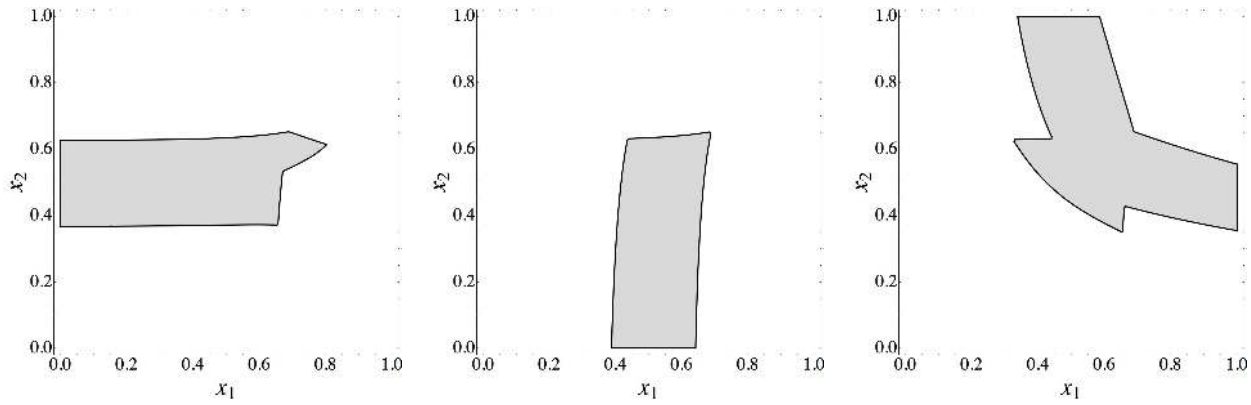


Figure E.1: Left: the set where $f^{II} < f^{\text{St+iso}}$ for $n = 20$. Center: the set where $f^{III} < f^{\text{St+iso}}$ for $n = 20$. Right: the set where $f^{IV} < f^{\text{St+iso}}$ for $n = 20$.

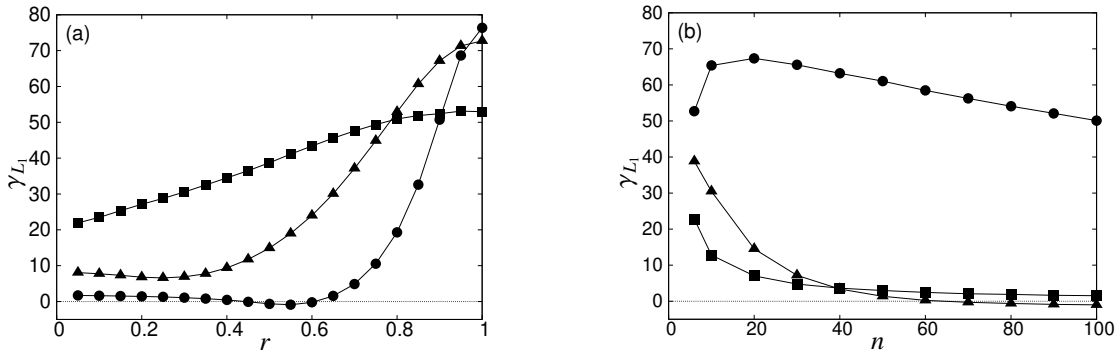


Figure F.1: Percentage reduction in average loss for $p = 6$: (a) $n = 6$ (■), $n = 20$ (▲), $n = 100$ (●) and $r \in (0, 1]$; (b) $r = 0.1$ (■), $r = 0.5$ (▲), $r = 0.9$ (●) and $p \leqslant n \leqslant 100$.

the entire domain $(0, 1] \times (0, 1]$. Likewise, denote by $f^{III}(\boldsymbol{x})$ and $f^{IV}(\boldsymbol{x})$ the unbiased estimators of risk corresponding to the use of $\psi_j^{III}(\boldsymbol{x})$ and $\psi_j^{IV}(\boldsymbol{x})$ (see (E.3) and (E.4)) everywhere in $(0, 1] \times (0, 1]$. As the functions f^{II}, f^{III}, and f^{IV} are continuous on $(0, 1] \times (0, 1]$, it is possible to infer from Fig. 5 (right panel) that there are subsets of the domain on which f^{II}, f^{III}, and f^{IV} are less than $f^{\text{St+iso}}$. A more precise illustration of this fact is reported in Fig. E.1 for $n = 20$. We also note that the size of the region where at least one of the functions f^{II}, f^{III}, and f^{IV} is less than $f^{\text{St+iso}}$ actually increases with decreasing n. Thus, for small sample sizes, these functions suggest an approach to substantially reduce the UBEOR. An analogous behavior is observed for other values of n. In conclusion, as in the $p = 2$ case, there are parts of the domain on which the isotonizing correction does not apply, even though its use over these regions would give a lower UBEOR.

F. Validation of the UBEOR approach for the $p = 6$ case

The percentage reduction in risk with respect to the MLE for $p = 6$ and $\Sigma(r) = \text{diag}(1, r, r^2, r^3, r^4, r^5)$ is shown in Fig. F.1(a) for fixed n as a function of r and in Fig. F.1(b) for fixed r as a function of n.