Asymptotic Consistency of the James-Stein Shrinkage Estimator

Alex Samuel Mungo, Victor Mooto Nawa

School of Natural Sciences, University of Zambia, Lusaka, Zambia
Email: alexsmungo8@yahoo.com, mnawa@unza.zm

Abstract

The study explores the asymptotic consistency of the James-Stein shrinkage estimator obtained by shrinking a maximum likelihood estimator. We use Hansen's approach to show that the James-Stein shrinkage estimator converges asymptotically to some multivariate normal distribution with shrinkage effect values. We establish that the rate of convergence is of order $\frac{1}{k\sqrt{n}}$ and rate $k\sqrt{n}$, hence the James-Stein shrinkage estimator is $k\sqrt{n}$-consistent. Then visualise its consistency by studying the asymptotic behaviour using simulating plots in R for the mean squared error of the maximum likelihood estimator and the shrinkage estimator. The latter graphically shows lower mean squared error as compared to that of the maximum likelihood estimator.

1. Introduction

A shrinkage estimator is an estimator that, either explicitly or implicitly, incorporates the effects of shrinkage. In loose terms this means that a naive or raw estimate is improved by combining it with other information. The term relates to the notion that the improved estimate is made closer to the "true value" by the supplied information than the raw estimate. Shrinkage estimation is a technique used in inferential statistics to reduce the mean squared error (MSE) of a given estimator. The idea of shrinking an estimator came in 1956 when Stein [1] established that we can reduce the MSE of an estimator if we give up a little on bias.

This means that given an estimator we can shrink it to obtain another estimator with lower MSE and the efficiency of the new estimator is desirable in the way it estimates the "true" parameter value. This works well when the number of parameters is more than two ($p \geq 3$) called the "James-Stein classical condition". When we shrink a maximum likelihood estimator (MLE) under the "James-Stein condition", we obtain a new shrinkage estimator which is closer to the assumed true value compared to the MLE. The magnitude of the improvement depends on the distance between the "true" parameter value and the parametric restriction which yields a shrinkage target denoted by $\tilde{\theta}_n^o$. With all these modifications and restrictions to achieve this estimator, we ask ourselves if this desirable shrinkage estimator is asymptotically consistent and efficient.

Literature on shrinkage estimators is a lot but we just mention a few of the most relevant contributions to our study. James and Stein [2] used shrinking techniques to come up with an estimator called the James-Stein shrinkage estimator (JSSE) which has lower squared risk loss compared to the MLE. Baranchik [3] showed that the positive part James-Stein shrinkage estimator has a lower risk than an ordinary JSSE. Berger [4] gives a discussion on selecting a minimax estimator of a multivariate normal mean by considering different types of James-Stein type estimators. Stein [5] used shrinking techniques to estimate the mean for a multivariate normal distribution. Carter and Ullah [6] constructed the sampling distribution and F-ratios for a James-Stein shrinkage estimator obtained by shrinking an ordinary least square (OLS) in regression models. George [7] proposed a new minimax multiple shrinkage estimator that allows multiple specifications for selection of a set of targets to shrink a given estimator. Geyer [8] looked at the asymptotics of constrained M-estimators which also fall in the class of shrinkage estimators. Then Hansen [9] constructed a generalised James-Stein shrinkage estimator obtained by shrinking an MLE, and Hansen [10] derived its asymptotic distribution and showed that we can shrink towards a sub-parameter space.

2. Preliminaries

The theory of shrinkage techniques plays an important role in developing efficient statistical estimators which play a key role in statistical decision theory. Therefore, a clear understanding of the asymptotic behaviour of the James-Stein shrinkage estimator $\hat{\beta}_n^*$ provides knowledge on the stability and efficiency of the estimator when the sample size value n grows without bound.

This paper will investigate the asymptotic consistency and efficiency of the James-Stein shrinkage estimator (JSSE) obtained by shrinking a maximum likelihood estimator (MLE) when we have observed variables $\mathbf{X} \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. Though the shrinkage estimator we are interested in is biased its study is important because there is a realisation that efficiency (lower risk) dominates all other properties in estimation. Efron [11] discusses how bias dominates unbiasedness in estimation.

We proceed by considering the asymptotic distributions of all the three estimators important to this study by considering results in Hansen 2016. Using the asymptotic distribution derived by Hansen [10], we employ Taylor's theorem and some limit theorems to show that $\hat{\boldsymbol{\beta}}_o^* \to_p \boldsymbol{\theta}_o$ the "true" parameter value as $n \to \infty$. Then we evaluate the asymptotic distributional bias (ADB) for the estimators $\hat{\boldsymbol{\beta}}_n$, $\tilde{\boldsymbol{\theta}}_n^o$ and $\hat{\boldsymbol{\beta}}_o^*$ and show that the variance of the latter achieves the Cramér-Rao bound (CRB) as $n \to \infty$. The analysis is done along the sequences $\boldsymbol{\theta}_n$ as $n \to \infty$. Simulation plots are produced in a statistical package R to compare the JSSE and MLE in terms of mean squared error (MSE), consistency and convergence.

The paper is organised as follows. Section 2.1 presents the parametric set up. Section 2.2 gives the form of the JSSE considered in the study while Section 2.3 discusses the asymptotic distributions of the estimators. In Section 3 we present the main results by first presenting a lemma on the convergence in probability of the shrinking factor as Section 3.1. We then show the consistency of the shrinkage estimator in Theorem 1. In Section 3.2 we evaluate the ADBs of the estimators in play. Then we show that the James-Stein shrinkage estimator is asymptotically efficient in Section 3.3 and also establish the rate of convergence in Section 3.4. In Section 3.5 we present MSE plots comparing the JSSE and MLE and in Section 4 we give a discussion and analysis of the whole study. Then conclude our study by stating the main results.

The following definitions are used to establish the asymptotic consistency and efficiency of the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$.

Definition 1

An estimator $\mathbf{T}_n = h(X_1, \cdots, X_n)$ is said to be consistent for $\boldsymbol{\theta}_n$ if it converges in probability to $\boldsymbol{\theta}_n$. That is, if for all $\varepsilon > 0$

$$\lim_{n \to \infty} Pr\left(\left|\mathbf{T}_n - \boldsymbol{\theta}_n\right| < \varepsilon\right) = 1$$

or

$$\lim_{n \to \infty} Pr\left(\left|\mathbf{T}_n - \boldsymbol{\theta}_n\right| > \varepsilon\right) = 0$$

where n is the sample size value.

Definition 2

Let $X_1, \cdots, X_n$ be independent and identically distributed (iid) according to a probability density $f_{\boldsymbol{\theta}}(X)$ satisfying suitable regularity conditions. Suppose that $\mathbf{T}_n = h(X_1, \cdots, X_n)$ is asymptotically normal say that $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}_n) \to_p N_p(0, \boldsymbol{\Sigma}_n)$ for a positive definite matrix $\boldsymbol{\Sigma}_n$ where $\mathbf{T}_n$ is estimating $\boldsymbol{\theta}_n$. Then a sequence of estimators $\{\mathbf{T}_n\} = \{h(X_1, \cdots, X_n)\}$ satisfying

$$\lim_{n \to \infty}\left[nVar(\mathbf{T}_n)\right] = \mathbf{J}_n(\boldsymbol{\theta})^{-1}$$

for the fisher information $\mathbf{J}_n(\boldsymbol{\theta})$ is said to be **asymptotically efficient**.

We now consider a statistical model. We describe the set up of the parameter of interest and the shrinking strategy used in the study.

2.1. Parametric Structure

Consider an unbiased estimator $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta} \in \Omega$ such that $\hat{\boldsymbol{\theta}}_n \sim N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}_n)$ is a p-multivariate normal, where elements of Ω are p-dimensional parameter vectors. Let $\tilde{\boldsymbol{\theta}}_n^o$ (shrinkage target) be a restricted maximum likelihood estimator (RMLE) for $\boldsymbol{\theta} \in \Omega_o$ a sub-parameter space partitioned from the whole parameter space Ω by a parametric restriction $\Omega_o = \{\boldsymbol{\theta} \in \Omega : \mathbf{a}(\boldsymbol{\theta}) = 0\}$ where $\mathbf{a}(\boldsymbol{\theta}) : \mathbb{R}^p \to \mathbb{R}^m$. The sub-parameter space Ω_o provides a simple model of interest to shrink to. If $m = p$ then $\Omega_o = \{0\}$ ($\boldsymbol{\theta}$ is the kernel of $\mathbb{R}^p$) a singleton zero vector and if $m < p$, we create a sub-model of particular interest. Let $\mathbf{A}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{a}'$ where $\mathbf{A}$ is a shrinkage matrix of dimension $p \times m$. We introduce another matrix $\mathbf{G}$ which harmonises the dimension of the RMLE from m to p. Hence we have a mapping $\mathbf{g}(\boldsymbol{\theta}) : \mathbb{R}^m \to \mathbb{R}^p$ such that $\mathbf{G}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbf{g}'$. The matrix $\mathbf{G}$ is an $m \times p$ matrix when we consider a sub-parameter space Ω_o and $\mathbf{G} = \mathbf{I}_p$ is a p-dimensional identity matrix when we have the whole parameter space Ω. We note that the matrix $\mathbf{G}$ is used to increase the dimension of the RMLE since it will be m-dimensional. Therefore we have a plug-in restricted maximum likelihood estimator $\mathbf{g}(\tilde{\boldsymbol{\theta}}_n^o) = \tilde{\boldsymbol{\beta}}_n^o$. The matrix $\mathbf{G}$ is a matrix harmonising the dimension due to shrinkage and the actual dimension of the parameters of interest p. The plug-in unrestricted MLE $\hat{\boldsymbol{\theta}}_n$ in the shrinkage sense is denoted by $\hat{\boldsymbol{\beta}}_n$. With all parameters set, we present the generalised James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ in the next section.

2.2. Positive Part James-Stein Shrinkage Estimator

Let $\hat{\boldsymbol{\theta}}_n$ be the MLE for $\boldsymbol{\theta} \in \Omega$ and $\tilde{\boldsymbol{\theta}}_n^o$ be a restricted maximum likelihood estimator for $\boldsymbol{\theta} \in \Omega_o$ a sub-parameter space of the whole parameter space Ω such that the elements of the parameter space are in $\mathbb{R}^p$ as described before. Let $g(\tilde{\boldsymbol{\theta}}_n^o) = \hat{\boldsymbol{\beta}}_n^o$ be the plug-in estimator of the RMLE of p-dimension. Then the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ obtained by shrinking the MLE towards the target $\tilde{\boldsymbol{\theta}}_n^o$ is given by

$$\hat{\boldsymbol{\beta}}_n^* = \hat{\boldsymbol{\beta}}_n - \left(\frac{p-2}{n(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o)} \right)_+ (\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o) \tag{1}$$

where $(a)_+$ is positive trimming function and $p \geq 3$. The shrinkage estimator in (1) can be expressed as a weighted average by letting

$$D_n = n(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o) \tag{2}$$

a distance statistic which is the same as the loss function $n\ell(\hat{\boldsymbol{\beta}}_n, \tilde{\boldsymbol{\beta}}_n^o)$ where $\boldsymbol{\Sigma}$ is a covariance matrix and

$$\hat{w} = \left(1 - \frac{\tau}{D_n} \right)_+ \tag{3}$$

for $\tau = p - 2$. Then (1) becomes

$$\hat{\boldsymbol{\beta}}_n^* = \hat{w}\hat{\boldsymbol{\beta}}_n + \left(1 - \hat{w}\right)\tilde{\boldsymbol{\beta}}_n^o \tag{4}$$

which is a weighted average of $\hat{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_n^o$. The James-Stein shrinkage estimator presented above has lower risk compared to the MLE as shown by Hansen [10] and James and Stein [2]. To check whether the shrinkage estimator is asymptotically consistent we need its asymptotic distribution. We therefore present the asymptotic distributions for the MLE, RMLE and JSSE in the next section.

2.3. Asymptotic Distribution

We assume that the maximum likelihood estimator satisfies the assumptions for regularity conditions given in Hansen [10] and Newey [12]. With these assumptions in mind, the asymptotic distributions of $\tilde{\boldsymbol{\beta}}_n^o$ and $\hat{\boldsymbol{\beta}}_n^*$ are analysed along the sequences $\boldsymbol{\theta}_n = \boldsymbol{\theta}_o + n^{-\frac{1}{2}}\mathbf{h}$ where $\boldsymbol{\theta}_o$ is the assumed true parameter value and $\mathbf{h} \in \mathbb{R}^p$ is a constant providing a neighbourhood for the true parameter value $\boldsymbol{\theta}_o$. From the normality of the MLE we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\right) \to_d Z \sim N_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \tag{5}$$

as $n \to \infty$. Using (5), Hansen [10] obtained the asymptotic distribution of the restricted maximum likelihood estimator as

$$\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n^o - \boldsymbol{\theta}_n\right) \to_p Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\left(Z + \mathbf{h}\right) \tag{6}$$

which has some shrinkage value effect $\mathbf{k} = \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}$. As a consequence of the convergence in (5) and (6) we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o\right) \to_d \mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\left(Z + \mathbf{h}\right) \tag{7}$$

which is an asymptotic distribution the MLE converges to when it is estimating the RMLE where $\boldsymbol{\beta}_n = g\left(\boldsymbol{\theta}_n\right)$. The distance statistic D_n in Equation (2) converges to some distribution given by a non central chi-squared distribution as described by Hansen [10]

$$D_n = n\ell\left(\hat{\boldsymbol{\beta}}_n, \tilde{\boldsymbol{\beta}}_n^o\right) \to_d \left(Z + \mathbf{h}\right)^\mathrm{T}\mathbf{B}\left(Z + \mathbf{h}\right) = \xi \sim \chi_p^2\left(\mathbf{h}^\mathrm{T}\mathbf{B}\mathbf{h}\right) \tag{8}$$

where matrix $\mathbf{B} = \mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{G}\boldsymbol{\Sigma}^{-1}\mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}$. Using (2) Hansen [10] showed that

$$\hat{w} \to_d w(Z) = \left(1 - \frac{p-2}{\xi}\right)_+ \tag{9}$$

which is a positive asymptotic distribution of an inverse of a chi-squared distribution with constant $\tau = p - 2$ for $p \geq 3$. Therefore using (9) as $n \to \infty$, Hansen [10] showed that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d w(Z)\mathbf{G}^\mathrm{T}Z + \left(1 - w(Z)\right)\left(\mathbf{G}^\mathrm{T}Z - \mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\left(Z + \mathbf{h}\right)\right) \tag{10}$$

which is normally distributed with some shrinkage value effect. With the asymptotic distribution of the shrinkage estimator in place we now present the main results.

3. Main Results

In this section we present our main results. We show that the James-Stein shrinkage estimator $\hat{\beta}_n^*$ is asymptotically consistent. Secondly, we evaluate the asymptotic distribution bias of the three estimators in play. We then show that the shrinkage estimator is asymptotically efficient by showing that its variance achieves the Cramér-Rao bound. Further, we explore the convergence rate of the James-Stein shrinkage estimator and present the simulation plots for the MSE done in R.

3.1. Consistency of the James-Stein Shrinkage Estimator

We present Lemma 1 which shows the convergence in probability of the weight (shrinkage factor) $\hat{w}$. The result is used when establishing the consistency of the James-Stein shrinkage estimator $\hat{\beta}_n^*$.

Lemma 1

From Equation (8) we have

$$\hat{w} \to_d w(Z) = \left(1 - \frac{\tau}{\xi}\right)_+ \tag{11}$$

where $\xi = (Z + \mathbf{h})^{\mathrm{T}} \mathbf{B}(Z + \mathbf{h}) \sim \chi_p^2(\mathbf{h}^{\mathrm{T}} \mathbf{B} \mathbf{h})$ is a non central Chi-squared distribution with non centrality parameter $\mathbf{h}^{\mathrm{T}} \mathbf{B} \mathbf{h}$, $\tau = p - 2$ and $p \geq 3$. Along the sequences θ_n, if $\mathbf{h} \to \infty$ then

$$w(Z) \to_p 1 \tag{12}$$

and if $\mathbf{h}$ is fixed then

$$w(Z) \to_p 0 \quad \text{otherwise} \quad w(Z) \to_p \mathbf{r} \tag{13}$$

where $\mathbf{r}$ is a constant such that $0 < \mathbf{r} < 1$ and $(a)_+$ in (11) is a positive trimming function which keeps what is in the brackets greater than or equal to zero.

Proof.

We begin by considering the first case when $\mathbf{h}$ diverges to infinity. Suppose that $\mathbf{h} \to \infty$ then

$$(Z + \mathbf{h}) \to \infty \quad \text{as} \quad n \to \infty. \tag{14}$$

Therefore from (14) we have

$$(Z + \mathbf{h})^{\mathrm{T}} \mathbf{B}(Z + \mathbf{h}) = \xi \to \infty \quad \text{as} \quad \mathbf{h} \to \infty \quad \text{and} \quad n \to \infty. \tag{15}$$

Now considering that

$$\hat{w} \to_d w(Z) = \left(1 - \frac{\tau}{\xi}\right)_+$$

as $n \to \infty$. Using (15) we have

$$w(Z) \to_p 1 \quad \text{as} \quad n \to \infty. \tag{16}$$

Hence we have established (12).

Secondly suppose that **h** is fixed, then we have the sequence

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}_o + n^{-\frac{1}{2}} \mathbf{h}$$

which becomes $\boldsymbol{\theta}_n = \boldsymbol{\theta}_o$ as $n \to \infty$ implying that $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_o$ as $n \to \infty$. Suppose

$$\xi = \chi_p^2\left(\mathbf{h}^{\mathrm{T}} \mathbf{B} \mathbf{h}\right) \to_p D \quad \text{as} \quad n \to \infty \tag{17}$$

where D is a constant, **h** is fixed and **B** is not affected by an increase in n, then

$$w(Z) \to_p \left(1 - \frac{\tau}{D}\right)_+$$

where $p \geq 3$. If $\dfrac{\tau}{D} = 1$ as $n \to \infty$, then

$$w(Z) \to_p 0.$$

If $\dfrac{\tau}{D} > 1$ then $\left(1 - \dfrac{\tau}{D}\right)$ will be negative and by definition of the positive streaming function we end up with zero. This will vary as p changes but still considering $\xi \sim \chi_p^2\left(\mathbf{h}^{\mathrm{T}} \mathbf{B} \mathbf{h}\right)$ the probability of ξ depends on the degrees of freedom p and will vary according to the chi-squared distribution, implying that the ratio $\dfrac{\tau}{D} = M \geq 1$ as $n \to \infty$. Therefore we have

$$w(Z) \to_p \left(1 - \frac{\tau}{D}\right)_+ \quad \text{as} \quad n \to \infty,$$

$$w(Z) \to_p \left(1 - M\right)_+ \quad \text{as} \quad n \to \infty$$

for a constant $M > 1$. Proceeding in the same way we have

$$w(Z) \to_p \left(F\right)_+ \quad \text{as} \quad n \to \infty, \text{ for } 1 - M = F < 0,$$

$$w(Z) \to_p 0 \quad \text{as} \quad n \to \infty$$

by definition of $\left(x\right)_+$. Thus

$$w(Z) \to_p 0 \quad \text{as} \quad n \to \infty. \tag{18}$$

Otherwise, if the ratio $\dfrac{\tau}{D}$ is such that $0 < \dfrac{\tau}{D} < 1$ as $n \to \infty$, we have

$$w(Z) \to_p \mathbf{r}$$

where $\mathbf{r} \in (0,1)$.

Lemma 1 above establishes convergence of the weight $\hat{w}$ which determines shrinkage. In this case we realise that the same weight determines the convergence in distribution and probability of the shrinkage estimator. From the regularity conditions we know that the MLE $\hat{\boldsymbol{\theta}}_n$ is asymptotically consistent and

this consistency extends to the RMLE $\bar{\theta}_n^o$. With this fact in mind, we now present the main result which shows the consistency of the shrinkage estimator $\hat{\beta}_n^*$.

Theorem 1

Let $\theta \in \Omega$, where Ω is a parameter space with elements in $\mathbb{R}^p$. Suppose we have a James-Stein shrinkage estimator $\hat{\beta}_n^*$ which is obtained by shrinking the maximum likelihood estimator $\hat{\theta}_n$ of $\theta \in \Omega$ where the shrinkage target $\tilde{\theta}_n^o$ is the restricted maximum likelihood estimator of $\theta \in \Omega_o$ a partitioned sub-parameter space from Ω by the restriction described in Section 2.1. Then the JSSE is given by

$$\hat{\beta}_n^* = \hat{\beta}_n - \left(\frac{p-2}{n\left(\hat{\beta}_n - \tilde{\beta}_n^o\right)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \left(\hat{\beta}_n - \tilde{\beta}_n^o\right)} \right)_+ \left(\hat{\beta}_n - \tilde{\beta}_n^o\right)$$

where $\hat{\beta}_n = \mathbf{g}(\hat{\theta}_n)$, $\tilde{\beta}_n^o = \mathbf{g}(\tilde{\theta}_n^o)$, $p \geq 3$ and $(x)_+$ is a positive trimming function. If $\hat{\theta}_n$ is consistent for θ_n as $n \to \infty$ then the James-Stein shrinkage estimator $\hat{\beta}_n^*$ is also consistent for θ_n as $n \to \infty$, where the sequence θ_n is as defined in Section 2.3.

Proof.

Let $\beta_n = \theta_n$. To show that $\hat{\beta}_n^*$ is consistent for θ_n as $n \to \infty$ we consider the value of $\mathbf{h}$ which determines the neighbourhood of the sequence θ_n, when it diverges to infinity and when it is just fixed. Suppose that $\mathbf{h}$ diverges to infinity, to evaluate

$$\sqrt{n}\left(\hat{\beta}_n^* - \beta_n\right) \to_d w(Z)\mathbf{G}^{\mathrm{T}}Z + \left(1 - w(Z)\right)\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\left(Z + \mathbf{h}\right)\right) \quad (19)$$

as $n \to \infty$, we first consider $w(Z)$ from (3). By Lemma 1

$$w(Z) \to_p 1 \quad \text{as} \quad n \to \infty. \quad (20)$$

Hence from (20) and substituting $w(Z)$ by 1 in (19) we have

$$\sqrt{n}\left(\hat{\beta}_n^* - \beta_n\right) \to_d \mathbf{G}^{\mathrm{T}}Z \sim \mathbf{G}^{\mathrm{T}}N_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$$

which gives

$$\sqrt{n}\left(\hat{\beta}_n^* - \beta_n\right) \to_d N_p\left(\mathbf{0}, \boldsymbol{\Sigma}_\beta\right) \quad (21)$$

as $n \to \infty$ where $\boldsymbol{\Sigma}_\beta = \mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{G}$. Thus we have

$$\hat{\beta}_n^* \to_p \beta_n$$

if

$$\lim_{n \to \infty} P\left(\left|\hat{\beta}_n^* - \beta_n\right| > \varepsilon\right) = 0$$

for any $\varepsilon > 0$. Hence $\hat{\beta}_n^*$ is consistent for $\beta_n = \theta_n$.

Secondly, suppose $\mathbf{h}$ is fixed within an imaginable value, then the sequence

$$\theta_n = \theta_o + n^{-\frac{1}{2}}\mathbf{h}$$

becomes $\boldsymbol{\theta}_n = \boldsymbol{\theta}_o$ as $n \to \infty$. From this equality we have $\hat{\boldsymbol{\theta}}_n = \hat{\boldsymbol{\theta}}_o$ and two conditions arise. The first one is that the sequence $\boldsymbol{\theta}_n$ will be within the restricted parameter space Ω_o with $\boldsymbol{\theta}_o \in \Omega_o$. From the restriction $\{\boldsymbol{\theta} \in \Omega : \mathbf{a}(\boldsymbol{\theta}) = 0\}$ of Ω_o this means that the shrinkage target is exactly at the true value and our consideration will be just on one parameter space. Therefore, we have $\hat{\boldsymbol{\theta}}_n = \tilde{\boldsymbol{\theta}}_n^o$, but from (6)

$$\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n^o - \boldsymbol{\theta}_n\right) \to_d \zeta = Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})$$

which will be the same as the asymptotic distribution of $\hat{\boldsymbol{\theta}}_n$ since we only consider the sub-parameter space Ω_o, and the shrinkage value $\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})$ affects it. Thus

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n\right) \to_d \zeta = Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h}) \quad \text{as} \quad n \to \infty \tag{22}$$

because we are estimating $\boldsymbol{\theta} \in \Omega_o$ and from Section 2.1 there will be no difference between the MLE and RMLE. Due to this equality of the two maximum likelihood estimators, from (4)

$$\hat{\boldsymbol{\beta}}_n^* = \hat{w}\hat{\boldsymbol{\beta}}_n + (1 - \hat{w})\tilde{\boldsymbol{\beta}}_n^o$$

and substituting (22) in (19) we have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d w(Z)\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right)$$
$$+ (1 - w(Z))\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right)$$

as $n \to \infty$, which becomes

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d \left[w(Z)\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right)\right]$$
$$- \left[w(Z)\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right)\right]$$
$$+ \mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right)$$

as $n \to \infty$, and then simplifies to

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d \mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right) \tag{23}$$

as $n \to \infty$, which is the same as the asymptotic distribution of $\tilde{\boldsymbol{\beta}}_n^o = \mathbf{g}\left(\tilde{\boldsymbol{\theta}}_n^o\right)$. Therefore using the consistency of the RMLE and (23), the consistency of the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ follows from the consistency of $\tilde{\boldsymbol{\theta}}_n^o$.

Lastly, we consider the case when we have two well defined parameter spaces, Ω_o and $\Omega - \Omega_o$. Then we have $\hat{\boldsymbol{\theta}}_n \neq \tilde{\boldsymbol{\theta}}_n^o$. Analysing (19) further, we consider the shrinkage effect value $\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$ which is not affected by the sample size value n but it is a value affected by an increase or decrease in the number of parameters p. Since

$$Z \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$$

then

$$Z + \mathbf{h} \sim N_p(\mathbf{h}, \Sigma) \tag{24}$$

by linearity property of the normal distribution. Also implying that

$$\eta(Z + \mathbf{h}) \sim N_p(\boldsymbol{\eta}\mathbf{h}, \boldsymbol{\eta}^{\mathrm{T}}\Sigma\boldsymbol{\eta}) \tag{25}$$

for some matrix $\boldsymbol{\eta}$ of dimension $p \times p$. From (13) of Lemma 1 we have

$$w(Z) \to_p 0 \quad \text{if} \quad \frac{\tau}{D_n} \geq 1 \tag{26}$$

as $n \to \infty$. Therefore (19) becomes

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d w(Z)\mathbf{G}^{\mathrm{T}}Z + (1 - w(Z))\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\eta}(Z + \mathbf{h})\right)$$

for some shrinkage value effect matrix $\boldsymbol{\eta} = \Sigma\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\Sigma\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$. Evaluating this asymptotic distribution as $n \to \infty$ we have

$$w(Z)\mathbf{G}^{\mathrm{T}}Z + (1 - w(Z))\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\eta}(Z + \mathbf{h})\right) \to_d \mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\eta}(Z + \mathbf{h})\right)$$

as $n \to \infty$ since $w(Z) \to_p 0$. Thus

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d \mathbf{G}^{\mathrm{T}}\left(Z - \Sigma\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\Sigma\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}(Z + \mathbf{h})\right) \quad \text{as} \quad n \to \infty.$$

Hence the consistency of $\hat{\boldsymbol{\beta}}_n^*$ follows from the consistency of $\tilde{\boldsymbol{\beta}}_n^o = \mathbf{g}(\tilde{\boldsymbol{\theta}}_n^o)$ which is consistent since $\tilde{\boldsymbol{\theta}}_n^o$ is consistent. Similarly if $w(Z) \to \mathbf{r} \in (0,1)$, the consistency of the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ follows from the consistency of the restricted maximum likelihood estimator and also the fact that $\hat{\boldsymbol{\theta}}_n$ is consistent for $\boldsymbol{\theta}_n$. Thus the shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ is asymptotically consistent for $\boldsymbol{\theta}_n$.

In Theorem 1, we first consider the case when the sequence $\boldsymbol{\theta}_n$ has a neighbourhood which is not restricted by letting $\mathbf{h}$ diverge to infinity. When this is the case, the entire parameter space becomes of interest and for $\mathbf{h} \to \infty$ we obtain $\xi \to_p \infty$ and $\hat{w} \to_p 1$ as $n \to \infty$. Hence there is no difference on how the parameters in Ω_o and Ω are asymptotically distributed. As a result the asymptotic distribution of the James-Stein shrinkage estimator is the same as that of the initial maximum likelihood estimator under this condition. Therefore the consistency of the James-Stein shrinkage estimator follows from the consistency of the maximum likelihood estimator.

In the second case we take $\mathbf{h}$ as a fixed imaginable value. In this case the two parameter spaces are well defined and distinctive in terms of where the parameters of interest are located. When $n \to \infty$ then $\boldsymbol{\theta}_n = \boldsymbol{\theta}_o$ since $n^{-\frac{1}{2}}\mathbf{h} \to 0$. Thus when we are within the restricted sub-parameter space Ω_o the maximum likelihood estimator and the restricted likelihood estimator are asymptotically distributed the same. The consequence of having the two maximum likelihood estimators (MLE and RMLE) distributed the same results in the James-Stein shrinkage estimator having the same asymptotic distribution as the MLE and RMLE. Furthermore, we have $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \tilde{\boldsymbol{\beta}}_n^o\right) \to_p 0$ as $n \to \infty$. Stone [13] obtained similar results though under invariant estimators. In this case the two

maximum likelihood estimators do not have to be necessarily invariant. Therefore the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ is asymptotically consistent for $\boldsymbol{\theta}_n$.

In the next section we investigate the asymptotic distribution bias of $\hat{\boldsymbol{\beta}}_n^*$. The results in this section are used in showing the asymptotic efficiency of the shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$.

3.2. Asymptotic Distributional Bias

We study the asymptotic distributional bias (ADB) for the three estimators by analysing the asymptotic bias values. The ADB of an estimator $\mathbf{T}_n$ is given by

$$\mathbf{ADB}(\mathbf{T}_n) = \lim_{n \to \infty} \mathbf{E}\left[\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}_n)\right] \tag{27}$$

where the estimator $\mathbf{T}_n$ is estimating $\boldsymbol{\theta}_n$. We present the asymptotic distributional bias for $\hat{\boldsymbol{\theta}}_n$, $\tilde{\boldsymbol{\theta}}_n^o$ and $\hat{\boldsymbol{\beta}}_n$ in the theorem below.

Theorem 2

Suppose that regularity assumptions for the MLE and RMLE hold. Then under $\{P_n\}$ a sequence of parameter dimension with the sample size value n and $p \geq 3$, the ADBs of the estimators $\hat{\boldsymbol{\theta}}_n$, $\tilde{\boldsymbol{\theta}}_n^o$ and $\hat{\boldsymbol{\beta}}_n^*$ are respectively

1. $\mathbf{ADB}(\hat{\boldsymbol{\theta}}_n) = 0$

2. $\mathbf{ADB}(\tilde{\boldsymbol{\theta}}_n^o) = -\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}$

3. $\mathbf{ADB}(\hat{\boldsymbol{\beta}}_n^*) = -\vartheta\mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}$

where $\vartheta = \mathbf{E}_\theta\left[\dfrac{p-2}{\xi}\right]$.

Proof.

1.

$$\begin{aligned}
\mathbf{ADB}(\hat{\boldsymbol{\theta}}_n) &= \lim_{n \to \infty} \mathbf{E}_\theta\left(\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n)\right) \\
&= \lim_{n \to \infty} 0 \\
&= 0 \\
\therefore \mathbf{ADB}(\hat{\boldsymbol{\theta}}_n) &= 0
\end{aligned} \tag{28}$$

since $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) \to_d Z \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$ as $n \to \infty$.

2.

$$\begin{aligned}
\mathbf{ADB}(\tilde{\mathbf{q}}_n^o) &= \lim_{n \to \infty} \mathbf{E}_\theta\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n^o - \boldsymbol{\theta}_n)\right) \\
&= \lim_{n \to \infty}\left(-\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}\right) \\
&= -\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{h} \\
\therefore \mathbf{ADB}(\tilde{\mathbf{q}}_n^o) &= -\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}
\end{aligned} \tag{29}$$

from Equation (6).

3. $\mathbf{ADB}(\hat{\boldsymbol{\beta}}_n^*) = \lim_{n \to \infty} \mathbf{E}_\theta\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n)\right)$. From Equation (19) of Theorem 1 we

have

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}_n\right) \to_d w(Z)\mathbf{G}^{\mathrm{T}}Z + \left(1 - w(Z)\right)\mathbf{G}^{\mathrm{T}}\left(Z - \boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\left(Z + \mathbf{h}\right)\right)$$

where

$$w(Z) = \left(1 - \frac{p-2}{\xi}\right)_+ .$$

Therefore,

$$\mathbf{E}_\theta\left[w(Z)\right] = \left[1 - \left(\frac{p-2}{\xi}\right)\right]$$

$$= 1 - \mathbf{E}_\theta\left[\frac{p-2}{\xi}\right]$$

$$= 1 - \vartheta$$

where $\vartheta = \mathbf{E}_\theta\left[\frac{p-2}{\xi}\right]$, $p \geq 3$ and $\mathbf{E}_\theta(Z) = 0$ as $n \to \infty$ since $Z \sim N_p\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$.

Then

$$\mathbf{ADB}\left(\hat{\boldsymbol{\beta}}_n^*\right) = \lim_{n \to \infty}\left[\left(1 - \vartheta\right)0 + \left(1 - 1 + \vartheta\right)\left(-\mathbf{G}^{\mathrm{T}}\left(\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h}\right)\right)\right]$$

$$= \lim_{n \to \infty}\left[-\vartheta\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h}\right] \qquad (30)$$

$$= -\vartheta\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h}$$

$$\therefore \mathbf{ADB}\left(\hat{\boldsymbol{\beta}}_n^*\right) = -\vartheta\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h}$$

where $\vartheta = \mathbf{E}_\theta\left[\frac{p-2}{\xi}\right]$ for $p \geq 3$.

Remark 1

When the fixed constant $\mathbf{h} = \mathbf{0}$, the asymptotic distributional bias values of the three estimators are zero. Therefore $\mathbf{h} \neq \mathbf{0}$.

From Equation (28) of Theorem 2, the maximum likelihood estimator is asymptotically unbiased. Equations (29) and (30) show that the restricted maximum likelihood estimator and the James-Stein shrinkage estimator are both asymptotically biased. This means that both shrinking and partitioning of a parameter space brings bias to estimators.

Using the bias of the shrinkage estimator obtained above, we now analyse whether the James-Stein estimator $\hat{\boldsymbol{\beta}}_n^*$ is asymptotically efficient.

3.3. Asymptotic Efficient

To check whether the shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ is asymptotically efficient, we use the Cramér-Rao bound for biased estimators. In the theorem below we show that the variance of the JSSE achieves this bound as $n \to \infty$. We use concepts of the study by Hogdes and Lehman [14].

Theorem 3

Let $\hat{\boldsymbol{\beta}}_n^*$ be a James-Stein shrinkage estimator obtained by shrinking a maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ where the two estimators are as defined in Section 2.2. Given the asymptotic bias $b_\theta\left(\hat{\boldsymbol{\beta}}_n^*\right)$ of the JSSE $\hat{\boldsymbol{\beta}}_n^*$, the Cramér-Rao bound for $\hat{\beta}_{n_j}^*$ is given by

$$\mathbf{CRB} = \frac{\left[1 + \mathbf{b}_{\theta_j}'\left(\hat{\beta}_n^*\right)\right]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} \quad \text{for} \quad j = 1, 2, \cdots, p \tag{31}$$

where $\mathbf{J}(\boldsymbol{\theta})$ is the fisher information and $\mathbf{b}_{\theta_j}'$ is the derivative of the jth element of the bias vector. Then

$$\frac{\mathbf{CRB}}{\Sigma_{jj}\left(\hat{\beta}_n^*\right)} = 1 \quad \text{as} \quad n \to \infty,$$

and thus the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ is **asymptotically efficient** for all $j = 1, 2, \cdots, p$.

Proof. We analyse asymptotic efficiency by evaluating the Cramér-Rao bound as $n \to \infty$. Consider the **bias** *of the estimator* $\hat{\boldsymbol{\beta}}_n^*$ *from part 3 of Theorem 2*

$$\mathbf{b}_\theta\left(\hat{\beta}_n^*\right) = -\vartheta \mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\mathbf{h} \tag{32}$$

where $\vartheta = \mathbf{E}\left[\dfrac{p-2}{\xi}\right]$ for $p \geq 3$. The expectation $\mathbf{E}\left[\dfrac{p-2}{\xi}\right]$ of the fraction $\dfrac{p-2}{\xi}$ which follows a distribution determined by the distribution

$\xi \sim \chi_p^2(h'\mathbf{B}h)$ has a value (constant) free of the parameter $\boldsymbol{\theta}$. Therefore we regard it as a constant. Let $\alpha = -\vartheta$ then (32) becomes

$$\mathbf{b}_\theta\left(\hat{\beta}_n^*\right) = \alpha \mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\mathbf{h} \tag{33}$$

and $\mathbf{b}_\theta'\left(\hat{\beta}_n^*\right)$ will be

$$\begin{aligned}
\mathbf{b}_\theta'\left(\hat{\beta}_n^*\right) &= \frac{\partial}{\partial \boldsymbol{\theta}} b_\theta(\hat{\mathbf{b}}_n^*) \\
&= \alpha \frac{\partial}{\partial \boldsymbol{\theta}}\left(\mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}\right)
\end{aligned} \tag{34}$$

a matrix of dimension p. Using the definition of the CRB and then combining (32) and (34) we obtain

$$\frac{\left[1 + \mathbf{b}_{\theta_j}'\left(\hat{\beta}_n^*\right)\right]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} = \frac{\left[1 + \alpha\dfrac{\partial}{\partial \boldsymbol{\theta}_j}\left(\mathbf{G}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}\mathbf{A}\right)^{-1}\mathbf{A}^\mathrm{T}\mathbf{h}\right)\right]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} \tag{35}$$

for $j = 1, 2, \cdots, p$ where $\dfrac{\partial}{\partial \boldsymbol{\theta}_j}$ is the partial derivative of the jth element,

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(-\sum_{i=1}^n \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\log f_{\boldsymbol{\theta}}(X_i)\right)^{-1}, \quad \mathbf{A} = \mathbf{A}(\boldsymbol{\theta}_o) = \frac{\partial}{\partial \boldsymbol{\theta}}\mathbf{a}(\boldsymbol{\theta})' \quad \text{and}$$

$$\mathbf{J} = \mathbf{J}(\boldsymbol{\theta}) = \mathbb{E}\left[\sum_{i=1}^{n} -\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log f_{\boldsymbol{\theta}}(X_i)\right].$$

We begin our analysis of the bound by considering the terms involved. Thus

$$\frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{A} = \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathbf{a}(\boldsymbol{\theta})'$$

remains the same as $n \to \infty$. We have

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(-\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log f_{\boldsymbol{\theta}}(X_i)\right)^{-1} \to \boldsymbol{\Sigma}_p \tag{36}$$

as $n \to \infty$ where the elements of $\boldsymbol{\Sigma}_p$ are zeros apart from the diagonal elements $\boldsymbol{\Sigma}_{jj}(\boldsymbol{\theta})$ which are ones for $j = 1, 2, \cdots, p$ since the observations are iid and follow a p-multivariate standard normal distribution. Thus from (36) we have

$$\frac{\partial}{\partial\boldsymbol{\theta}_j}\boldsymbol{\Sigma}_{jj}(\boldsymbol{\theta}) \to 0 \quad \text{and} \quad \frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma} \to \mathbf{0} \tag{37}$$

as $n \to \infty$ for $j = 1, 2, \cdots, p$. This implies that

$$\frac{\partial}{\partial\boldsymbol{\theta}_j}\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h} \to 0 \quad \text{and} \quad \frac{\partial}{\partial\boldsymbol{\theta}}\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h} \to \mathbf{0} \tag{38}$$

for $j = 1, 2, \cdots, p$ as $n \to \infty$. Then from (38) we have

$$\mathbf{b}'_{\theta_j}(\hat{\beta}_n^*) \to 0 \quad \text{and} \quad \mathbf{b}'_\theta(\hat{\beta}_n^*) \to \mathbf{0} \tag{39}$$

for $j = 1, 2, \cdots, p$ as $n \to \infty$. Therefore from (38) and (39), then using (35) we have

$$\begin{aligned}
\frac{\left[1 + \mathbf{b}'_{\theta_j}(\hat{\beta}_n^*)\right]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} &= \frac{\left[1 + \alpha\frac{\partial}{\partial\boldsymbol{\theta}_j}\left(\mathbf{G}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A}(\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{h}\right)\right]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} \\
&= \frac{[1+0]^2}{\mathbf{J}_{jj}(\boldsymbol{\theta})} \\
&= \mathbf{J}_{jj}(\boldsymbol{\theta})^{-1}
\end{aligned} \tag{40}$$

as $n \to \infty$ for $j = 1, 2, \cdots, p$. Since for all $j = 1, 2, \cdots, p$ we have $\boldsymbol{\Sigma}_{jj}(\boldsymbol{\theta}) = \mathbf{J}_{jj}(\boldsymbol{\theta})^{-1}$ then $\boldsymbol{\Sigma} = \mathbf{J}^{-1}$ as $n \to \infty$. Hence from (40) we have the variance for the James-Stein shrinkage estimator $\hat{\beta}_{n_j}^*$, $\boldsymbol{\Sigma}_{jj}(\hat{\beta}_n^*)$ converges to the **CRB** as $n \to \infty$ for all $j = 1, 2, \cdots, p$. This means that

$$\frac{\mathbf{CRB}}{\boldsymbol{\Sigma}_{jj}(\hat{\beta}_n^*)} \to \frac{\boldsymbol{\Sigma}_{jj}(\hat{\beta}_n^*)}{\boldsymbol{\Sigma}_{jj}(\hat{\beta}_n^*)} = 1 \quad \text{as} \quad n \to \infty$$

for all $j = 1, 2, \cdots, p$. Thus the James-Stein shrinkage estimator $\hat{\beta}_n^*$ is asymptotically efficient.

Theorem 3 above shows that the James-Stein shrinkage estimator obtained

from shrinking the MLE achieves the CRB asymptotically. This means that the shrinkage estimator is asymptotically efficient and hence stable when we have large sample size values. Since the initial estimator (MLE) is known to be asymptotically efficient then we see that the shrinking process has no effect on the asymptotic efficiency of an estimator we are shrinking.

3.4. Rate of Convergence

We now investigate the rate of convergence of the shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ (JSSE) by using concepts applied on the MLE discussed in Hoeffding [15]. To proceed we consider the shrinkage estimator of the form in (1) using plug-in maximum likelihood estimators $\hat{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_n^o$. Let $\hat{\boldsymbol{\beta}}_n^*$ be a James-Stein shrinkage estimator obtained when we shrink the MLE $\hat{\boldsymbol{\beta}}_n = \mathbf{g}\left(\hat{\boldsymbol{\theta}}_n\right)$ defined earlier before for $p \geq 3$. We proceed to find the rate of convergence of this estimator by using its relationship with the MLE. Since the shrinkage target value may have no effect on the convergence rate, for easier transformation of our sequence $\boldsymbol{\theta}_n$ we set $\tilde{\boldsymbol{\theta}}_n^o = \mathbf{0}$ implying $\tilde{\boldsymbol{\beta}}_n^o = \mathbf{0}$. Thus we have

$$\hat{\boldsymbol{\beta}}_n^* = \hat{\boldsymbol{\beta}}_n - \left(\frac{p-2}{n\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{V}^{-1}\hat{\boldsymbol{\beta}}_n\right)}\right)_+ \hat{\boldsymbol{\beta}}_n$$

which becomes

$$\hat{\boldsymbol{\beta}}_n^* = \left(1 - \frac{p-2}{\hat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{V}^{-1}\hat{\boldsymbol{\beta}}_n}\right)_+ \hat{\boldsymbol{\beta}}_n$$

when we factor out $\hat{\boldsymbol{\beta}}_n$ and drop out the n in the denominator to have a form with a lower MSE according to the James-Stein shrinkage strategy. Let

$$k = \left(1 - \frac{p-2}{\hat{\boldsymbol{\beta}}_n^{\mathrm{T}}\mathbf{V}^{-1}\hat{\boldsymbol{\beta}}_n}\right)_+, \tag{41}$$

then

$$\hat{\boldsymbol{\beta}}_n^* = k\hat{\boldsymbol{\beta}}_n. \tag{42}$$

Now consider the sequence

$$\hat{\boldsymbol{\beta}}_{n_j} = \boldsymbol{\beta}_{o_j} + O_p\left(\frac{1}{\sqrt{n}}\right) \tag{43}$$

for $j = 1, 2, \cdots, p$ where $\boldsymbol{\beta}_{o_j}$ is the "true" jth parameter value. From the equality in (42) we have

$$\hat{\boldsymbol{\beta}}_n = \frac{1}{k}\hat{\boldsymbol{\beta}}_n^* \tag{44}$$

for the shrinkage value k. Therefore substituting the right hand side of (44) in (43) the sequence $\hat{\boldsymbol{\beta}}_{n_j}$ becomes

$$\frac{1}{k}\hat{\boldsymbol{\beta}}_{n_j}^* = \boldsymbol{\beta}_{o_j} + O_p\left(\frac{1}{\sqrt{n}}\right),$$

hence we have the sequence

$$\hat{\boldsymbol{\beta}}^*_{n_j} = k\boldsymbol{\beta}_{o_j} + O_p\left(\frac{1}{\sqrt{n}}\right)k \tag{45}$$

which is in terms of the shrinkage estimator with the shrinking effect value k such that $0 < k \le 1$. Analysing this sequence further shows that it satisfies the smoothness regularity conditions for the MLE, therefore we can proceed.

Let $\boldsymbol{\beta}^*_{o_j} = k\boldsymbol{\beta}_{o_j}$ be the true value in the shrinkage sense which is obtained when we scale the true value $\boldsymbol{\beta}_{o_j}$ with the shrinkage factor k. Then the sequence (45) becomes

$$\hat{\boldsymbol{\beta}}^*_{n_j} = \boldsymbol{\beta}^*_{o_j} + O_p\left(\frac{1}{\sqrt{n}}\right)k \tag{46}$$

implying that

$$\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j} = O_p\left(\frac{1}{\sqrt{n}}\right)k \tag{47}$$

for all $j = 1, 2, \cdots, p$. This means that $\left(\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j}\right)$ is still within the neighbourhood of $\frac{1}{\sqrt{n}}$ since $0 < k \le 1$. Therefore, using the second order Taylor's theorem we have

$$\ln\frac{\prod_{i=1}^{n} f_{\hat{\boldsymbol{\beta}}^*_{n_j}(x_i)}}{\prod_{i=1}^{n} f_{\boldsymbol{\beta}^*_{o_j}(x_i)}} = \left(\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j}\right)\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}^*_o\right)}Z_j - \frac{n}{2}\left(\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j}\right)^2 \mathbf{I}_{jj}\left(\boldsymbol{\beta}^*_o\right) + O_p(1) \tag{48}$$

for $j = 1, 2, \cdots, p$. Since

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ln\mathbf{L}\left(\hat{\boldsymbol{\beta}}_n\right) = \mathbf{0}$$

for the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_n$, then also

$$\frac{\partial}{\partial\boldsymbol{\beta}}\ln\mathbf{L}\left(\hat{\boldsymbol{\beta}}^*_n\right) = \mathbf{0}$$

implying that

$$\frac{\partial}{\partial\boldsymbol{\beta}_j}\ln\mathbf{L}\left(\hat{\boldsymbol{\beta}}^*_{n_j}\right) = 0 \tag{49}$$

for all $j = 1, 2, \cdots, p$. Assuming that the log-likelihood function is differentiable, from (48) and (49) we have

$$\frac{\partial}{\partial\hat{\boldsymbol{\beta}}^*_{n_j}}\ln\left(\frac{\prod_{i=1}^{n} f_{\hat{\boldsymbol{\beta}}^*_{n_j}(x_i)}}{\prod_{i=1}^{n} f_{\boldsymbol{\beta}^*_{o_j}(x_i)}}\right) \tag{50}$$

$$= \left(\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j}\right)^{1-1}\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}^*_o\right)}Z_j - \frac{2n}{2}\left(\hat{\boldsymbol{\beta}}^*_{n_j} - \boldsymbol{\beta}^*_{o_j}\right)^{2-1}\mathbf{I}_{jj}\left(\boldsymbol{\beta}^*_o\right) + O_p(1)$$

and simplifying (50) becomes

$$\frac{\partial}{\partial \hat{\boldsymbol{\beta}}_{n_j}^*} \ln\left(\frac{\prod_{i=1}^{n} f_{\hat{\boldsymbol{\beta}}_{n_j}^*(x_i)}}{\prod_{i=1}^{n} f_{\boldsymbol{\beta}_{o_j}^*(x_i)}}\right) = \sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}Z_j - n\left(\hat{\boldsymbol{\beta}}_{n_j}^* - \boldsymbol{\beta}_{o_j}^*\right)\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right) + O_p(1) = 0$$

implying

$$\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}Z_j - n\left(\hat{\boldsymbol{\beta}}_{n_j}^* - \boldsymbol{\beta}_{o_j}^*\right)\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right) + O_p(1) = 0. \qquad (51)$$

Rearranging (51) we have

$$n\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)\left(\hat{\boldsymbol{\beta}}_{n_j}^* - \boldsymbol{\beta}_{o_j}^*\right) + O_p(1) = \sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}Z_j \qquad (52)$$

for $j = 1, 2, \cdots, p$ where $Z_j \sim \left(0, \mathbf{V}_{\boldsymbol{\beta}_j}\right)$ where $\mathbf{V}_{\boldsymbol{\beta}_j}$ is the variance of the jth element of the covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}$ of the distribution $= \mathbf{G}^{\mathrm{T}} N_p\left(\mathbf{0}, \mathbf{V}\right)$ and thus Z_j is the standard normal distribution for the jth element of $\hat{\boldsymbol{\beta}}_n$. Now dividing the left and right hand side of (52) by $\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}$ we obtain

$$\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}\left(\hat{\boldsymbol{\beta}}_{n_j}^* - \boldsymbol{\beta}_{o_j}^*\right) + O_p(1) = Z_j \qquad (53)$$

where Z_j is the distribution of the jth element of $\hat{\boldsymbol{\beta}}_n$ and $Z \sim N_p\left(\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}}\right)$. Using sequence (45), Equation (53) becomes

$$k\sqrt{n}\sqrt{\mathbf{I}_{jj}\left(\boldsymbol{\beta}_o^*\right)}\left(\hat{\boldsymbol{\beta}}_{n_j} - \boldsymbol{\beta}_{o_j}\right) + O_p(1) = Z_j \qquad (54)$$

for some $\boldsymbol{\beta}_{o_j}^* \to \boldsymbol{\beta}_{o_j}$ and $j = 1, 2, \cdots, p$ where $Z \sim N_p\left(\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}}\right)$ and $\mathbf{V}_{\boldsymbol{\beta}} = \mathbf{G}^{\mathrm{T}}\mathbf{V}\mathbf{G}$. The distribution Z_j is a normal distribution for each jth element of the plug-in estimator $\hat{\boldsymbol{\beta}}_n$ as described before in the analysis above.

Thus Equation (54) establishes the condition which implies local asymptotic normality (LAN) and differentiability in quadratic mean (DQM) for the estimator $\hat{\boldsymbol{\beta}}_n^*$ which implies that the rate of convergence is of order $\frac{1}{k\sqrt{n}}$ and rate $k\sqrt{n}$. This can also be achieved if we use the fact that the risk bound of the James-Stein shrinkage estimator is bounded by that of the MLE and the latter converges at the rate $\sqrt{n}$. Hence the James-Stein shrinkage estimator is $k\sqrt{n}$-consistent.

3.5. Simulation Plots

In this section the behaviour of the mean squared error (MSE) of the maximum likelihood estimator $\hat{\theta}_n$ is compared to that of the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ as the sample size n increases. The statistical package R is used to simulate plots of the MSE for different sample size values of n using the R package library (MASS) to stimulate data which follow a multivariate normal distribution. The data is generated using a 3×3 correlation matrix (ρ) to get the covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.3 & 0.1 \\ 0.3 & 1 & 0.2 \\ 0.1 & 0.2 & 1 \end{bmatrix}$$

which is symmetric and the variance in the major diagonal is 1 representing a standard normal variance. Thus we take $p = 3$ and this meets the James-Stein classical condition of $p \geq 3$. Since $X \sim N_3(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, we have

$$\hat{\boldsymbol{\theta}}_n = \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_{in} \quad \text{for} \quad p = 3 \tag{55}$$

making the MLE $\hat{\boldsymbol{\theta}}_n$ a 3×1 matrix which implies that the dimension for the shrinkage estimator is also 3×1. Now knowing that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is unbiased and the James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ is biased, the following expressions were used to calculate the mean squared error (MSE) of the two estimators. Using (55), we have

$$\mathbf{MSE}_\theta(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\Sigma}_n(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\Sigma}_n(\overline{X}_n) = Var(\overline{X}_n) \tag{56}$$

as the mean squared error of $\hat{\boldsymbol{\theta}}_n$ for $p = 3$ since $\mathbf{b}_\theta(\hat{\boldsymbol{\theta}}_n) = 0$. Similarly we have

$$\begin{aligned} \mathbf{MSE}_\theta(\hat{\boldsymbol{\beta}}_n^*) &= \boldsymbol{\Sigma}_n(\hat{\boldsymbol{\beta}}_n^*) + \left[\mathbf{b}_\theta(\hat{\boldsymbol{\beta}}_n^*)\right]^2 \\ &= Var(\hat{\boldsymbol{\beta}}_n^*) + \left[\mathbf{b}_\theta(\hat{\boldsymbol{\beta}}_n^*)\right]^2 \end{aligned} \tag{57}$$

for the James-Stein shrinkage estimator which becomes

$$\begin{aligned} \mathbf{MSE}_\theta(\hat{\boldsymbol{\beta}}_n^*) &= \boldsymbol{\Sigma}_n(k\hat{\boldsymbol{\theta}}_n) + \left[\mathbf{b}_\theta(k\hat{\boldsymbol{\theta}}_n)\right]^2 \\ &= Var(k\hat{\boldsymbol{\theta}}_n) + \left[\mathbf{b}_\theta(k\hat{\boldsymbol{\theta}}_n)\right]^2 \end{aligned} \tag{58}$$

where k is a shrinkage value which shrinks the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ to a James-Stein shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ for $p = 3$. Thus the mean squared error of the shrinkage estimator $\hat{\boldsymbol{\beta}}_n^*$ in (58) is obtained by using (56). The shrinkage value k is evaluated using the expression

$$k = \left(1 - \frac{1}{\overline{X}_n^{\mathrm{T}} \boldsymbol{\Sigma}_n(\overline{X}_n) \overline{X}_n}\right) = 1 - \frac{1}{\overline{X}_n^{\mathrm{T}} Var(\overline{X}_n) \overline{X}_n} \tag{59}$$

for $p = 3$. The commands for all expressions and plots produced in R are provided in the appendix.

We present MSE plots obtained by simulating the mean squared error using the sample size values of $n = 30$, 2000 and 100,000. The MSE plots for both the James-Stein shrinkage estimator and maximum likelihood estimator for each sample size value considered are plotted on the same graph for easy comparison of the MSE trends. We begin by considering a small sample size value of $n = 30$ to compare the the way the MSE line plots change from one particular point to the other. Since we are interested in asymptotic behaviour, we increase the sample size value to 2000 and then 100,000 to analyse the MSE trends and the rate at

which the line plots become smooth.

MSE Plots for JSSE-MLE: n=30

Figure 1. MSE plots for the MLE and JSSE for $n = 30$.

MSE plots for JSSE-MLE: n=2000
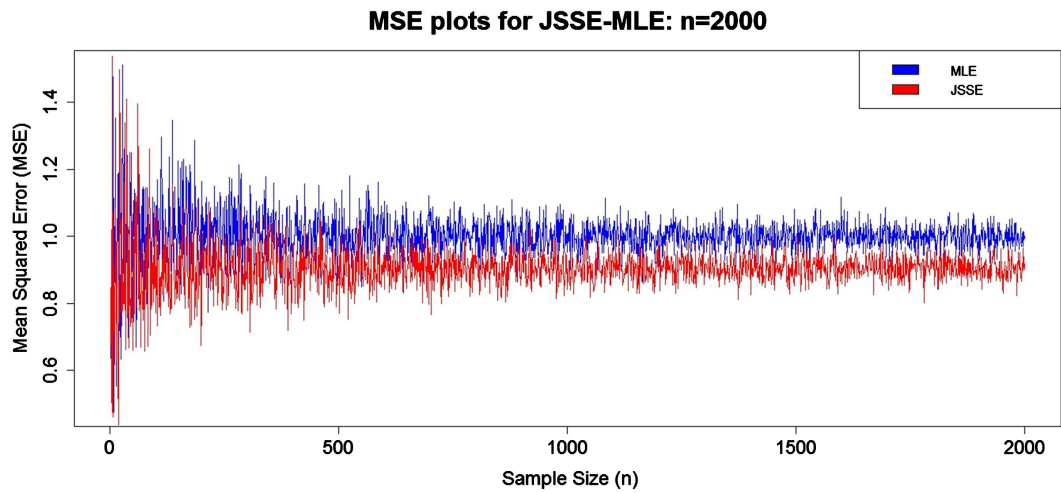
Figure 2. MSE for the JSSE and MLE for $n = 2000$.

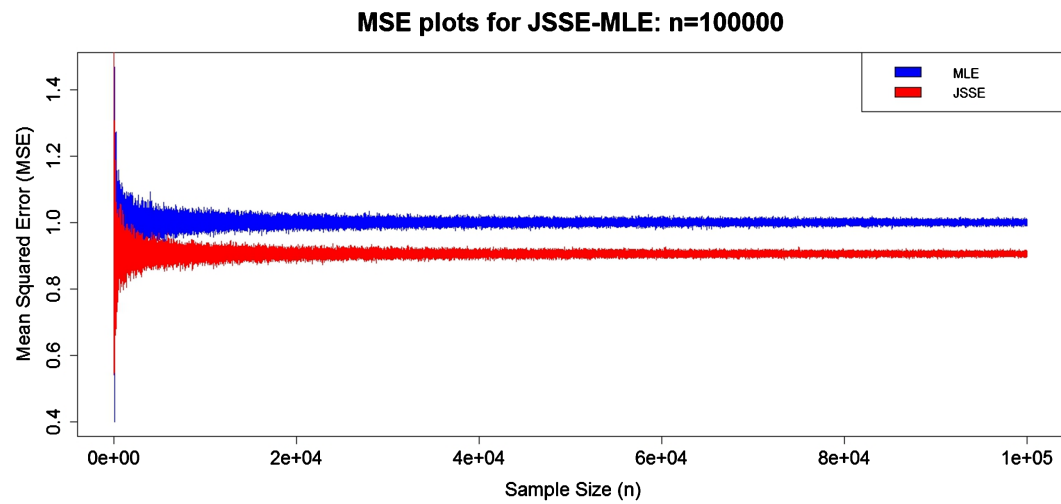**MSE plots for JSSE-MLE: n=100000**

Figure 3. MSE plots for JSSE and MLE for $n = 10,000$.

Collectively the scaled plots show that there is some reduction in the mean squared error of the James-Stein shrinkage estimator compared to that of the initial estimator (MLE). The trend in mean squared error for both the maximum likelihood estimator and the James-Stein shrinkage estimator shows that as the sample size value n increases, the MSE values converge to some value. The MSE plots suggest that the James-Stein shrinkage estimator converges to a lower MSE value 0.9 compared to the maximum likelihood estimator which converges to a MSE value 1.0. They also show that the James-Stein shrinkage estimator converges faster compared to the MLE though the difference is minimal.

4. Conclusions and Suggestions

In the study, we explored the asymptotic properties of the James-Stein shrinkage estimator $\hat{\beta}_n^*$ which is obtained by shrinking a MLE $\hat{\theta}_n$. Asymptotic consistency and efficiency of the shrinkage estimator $\hat{\beta}_n^*$ were investigated. From the regularity conditions, the MLE is known to be unbiased, consistent and efficient as the sample size value $n \to \infty$. Therefore, the study analysed these asymptotic properties by checking whether the new estimator (shrinkage) obtained after shrinking possesses them. Thus the study examined whether the shrinking process has an effect on these properties. The results show that the James-Stein shrinkage estimator $\hat{\beta}_n^*$ is asymptotically consistent and efficient. The study also showed that the shrinkage estimator (JSSE) is asymptotically biased, a property it possesses even for small values of the sample size value n. The bias is brought by the shrinking factor k given in Equation (59). We therefore see that the shrinking process introduces bias to the estimators obtained but it preserves asymptotic consistency and efficiency, and more importantly, it reduces the MSE.

Thus the James-Stein shrinkage estimator obtained by shrinking techniques proves to be useful though it is biased. This estimator is more effective than the maximum likelihood estimator as shown in this study and by Hansen [10]. The study also showed that the JSSE is stable for large values of the sample size value n, making it suitable in practical applications since large sample size values are normally considered for effective estimation. Since error is always there in estimation, we justify shrinking (minimising error) as a very important technique for yielding effective estimators.

The study has investigated the asymptotic behaviour of the James-Stein shrinkage estimator. Asymptotic properties analysed in the study include rate of convergence, consistency and efficiency. The results show that the James-Stein shrinkage estimator has a lower mean squared error compared to the maximum likelihood estimator though it is biased. The results further show that the JSSE is asymptotically consistent and efficient.

Acknowledgements

managing editor Alline Xiao for a rare attitude of high quality.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Stein, C. (1956) Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In: Neyman, J., ed., *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume I, Statistical Laboratory of the University of California, Berkeley, 197-206. https://doi.org/10.1525/9780520313880-018

[2] James, W. and Stein, C. (1961) Estimation with Quadratic Loss. In: Neyman, J., ed., *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Statistical Laboratory of the University of California, Berkeley.

[3] Baranchik, A.J. (1964) Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution. Technical Report No. 51. Department of Statistics, Stanford University, Stanford, CA.

[4] Berger, J.O. (1976) Minimax Estimation of Multivariate Normal Mean with Arbitrary Quadratic Loss. *Journal of Multivariate Analysis*, **6**, 256-264. https://doi.org/10.1016/0047-259X(76)90035-X

[5] Stein, C. (1981) Estimation of the Mean of a Multivariate Normal Distribution. *Annals of Statistics*, **9**, 1135-1151. https://doi.org/10.1214/aos/1176345632

[6] Carter, R.L. and Ullah, A. (1984) The Sampling Distribution of Estimators and Their F-Ratios in Regression Model. *Journal of Econometrics*, **25**, 109-122. https://doi.org/10.1016/0304-4076(84)90040-X

[7] George, E. I. (1986) Minimax Multiple Shrinkage Estimation. *Annals of Statistics*, **14**, 188-205. https://doi.org/10.1214/aos/1176349849

[8] Geyer, C.J. (1994) On the Asymptotic of Constrained M-Estimation. *Annals of Statistics*, **22**, 1993-2010. https://doi.org/10.1214/aos/1176325768

[9] Hansen, E.B. (2008) Generalized Shrinkage Estimators. https://web-docs.stern.nyu.edu/old_web/emplibrary/shrink3.pdf

[10] Hansen, E.B. (2016) Efficient Shrinkage in Parametric Models. *Journals of Econometrics*, **190**, 188-205. https://doi.org/10.1016/j.jeconom.2015.09.003

[11] Efron, B. (1975) Biased versus Unbiased Estimation. In *Advances in Mathematics*, Academic Press, New York. https://doi.org/10.1016/0001-8708(75)90114-0

[12] Newey, W.K. and Mcfadden, D.L. (1994) Large Sample Estimation and Hypothesis Testing. University Press, Holland, 2113-2245. https://doi.org/10.1016/S1573-4412(05)80005-4

[13] Stone, C.J. (1974) Asymptotic Properties of Estimators of a Location Parameter. *The Annals of Statistics*, **6**, 1127-1137. https://doi.org/10.1214/aos/1176342869

[14] Hogdes, J.L. and Lehman, E.L. (1975) Some Applications of the Cramér-Rao Inequality. In *Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, 13-22.

[15] Hoeffding, W. (1948) A Class of Statistics with Asymptotically Normal Distribution. *Annals of Mathematical Statistics*, **19**, 293-325. https://doi.org/10.1214/aoms/1177730196