Shrinkage efficiency bounds: An extension

Giuseppe De Luca
University of Palermo, Palermo, Italy

Jan R. Magnus
Vrije Universiteit Amsterdam and Tinbergen Institute,
Amsterdam, The Netherlands

June 27, 2022

Abstract: Hansen (2005) obtained the efficiency bound in the p-dimensional normal location model when $p \geq 3$, generalizing an earlier result of Magnus (2002) for the one-dimensional case ($p = 1$). The classes of estimators considered are, however, different in the two cases. We provide an alternative bound to Hansen's which is a more natural generalization of the one-dimensional case, and we compare the classes and the bounds.

1 Introduction

Consider p independent normally distributed observations $x_1, x_2, \ldots, x_p$ where x_i has an unknown mean θ_i and variance 1, in other words $x \sim \mathrm{N}(\theta, I_p)$, where $x = (x_1, \ldots, x_p)'$, $\theta = (\theta_1, \ldots, \theta_p)'$, and I_p denotes the identity matrix of order p. This is the so-called (multivariate) normal location model.

Suppose we wish to estimate θ by an estimator of the form

$$t(x) = \lambda(w_p)x, \tag{1}$$

where λ depends on x only through $w_p = x'x$. The mean squared error (MSE) of $t(x)$ is the positive semidefinite $p \times p$ matrix

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}[(\hat{\theta} - \theta)(\hat{\theta} - \theta)'] \tag{2}$$

and its trace is called the *risk* of the estimator, which we denote by $R(\theta, \lambda)$.

If we replace x by $y = Sx$ where S is an orthogonal matrix, then $y'y = x'x$ so that λ remains the same. Hence λ is orthogonally invariant, and the class of estimators defined by (1) is called the class of *orthogonally invariant estimators*. Within this class we have

$$\sup_{\theta} R(\theta, \lambda) \geq p \tag{3}$$

for all λ (Van der Vaart, 1998, Proposition 8.6). When equality occurs in (3) then $t(x)$ is called *minimax*. For $p = 1$ and $p = 2$ there is a unique minimax estimator, namely the maximum-likelihood estimator $t(x) = x$, sometimes called the 'usual' estimator. But for $p \geq 3$ there are many minimax estimators.

The following two questions now arise. First, how do we choose between these minimax estimators? And second, should we also allow estimators that are not minimax but satisfy some other optimality criterion?

In the case $p = 1$ there is only one minimax estimator, namely $t(x) = x$ with risk 1, but there are many shrinkage estimators of the form $t(x) = \lambda(x)x$ whose risk is sometimes larger than 1 and sometimes smaller than 1. These estimators are not minimax, but have other desirable (arguably more desirable) properties. One possibility to distinguish between these estimators is through the minimax regret criterium. To use this criterium we need a lower bound for the class of estimators we are interested in. Magnus (2002) defined this class (for $p = 1$) by the following four restrictions:

(a) $0 \leq \lambda(x) \leq 1$ for all x,

(b) $\lambda(-x) = \lambda(x)$ for all x,

(c) λ is non-decreasing on $[0, \infty)$, and

(d) λ is continuous except possibly on a set of measure zero,

and proved that the lower bound is equal to $\theta^2/(1 + \theta^2)$ (Magnus, 2002, Appendix, Theorem A.7). The resulting minimax regret solution is obviously not minimax.

Hansen (2015) takes a different route. He considers only minimax estimators and obtains minimax regret estimators *within* this class. In order to derive these estimators he needs the efficiency bound (the lowest achievable risk) for the class of minimax estimators of the form (1). When $p = 1$ or $p = 2$ the efficiency bound is equal to p, because there is only one λ which leads to a minimax estimator, namely $\lambda \equiv 1$. But when $p \geq 3$, then Hansen's (2015, Theorem 2) efficiency bound is a rather intricate formula which we denote by $R_H(p, \alpha)$, where $\alpha = \theta'\theta/p$ is the average noncentrality.

Hansen considers his bound $R_H(p, \alpha)$ to be a generalization of the efficiency bound derived in Magnus (2002) for the case $p = 1$. But the class of estimators in Magnus (2002) is quite different to the class of minimax estimators considered by Hansen. Magnus purposely considers estimators that are *not* minimax and asks whether any of them have better properties than $t(x) = x$, the minimax estimator. Following the latter strategy, we shall prove the following result.

Proposition (efficiency bound): *Under the assumption that λ is nondecreasing, the efficiency bound for the risk in the class of estimators defined by (1) is given by*

$$R_M(p, \alpha) = \frac{\alpha p}{1 + \alpha}. \tag{4}$$

The proposition contains the bound obtained in Magnus (2002) as a special case ($p = 1$). Note that Hansen (2015) does not require that λ is nondecreasing. Hence, there are estimators in Hansen's class that are not in our class and there are estimators in our class (in particular, $\lambda(w) = \lambda$ for some constant $0 \leq \lambda < 1$) that are not in his class.

2 Why should λ be nondecreasing

Before we prove the proposition, we argue that it does not make much sense to consider λ functions that are not nondecreasing. To see this we consider the simple regression model where y is generated by

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u, \qquad u \sim \mathrm{N}(0, \sigma^2 I_n). \tag{5}$$

where the regressor matrices X_1 and X_2 have dimensions $n \times k_1$ and $n \times k_2$, respectively.

We are interested in estimating β_1 and consider two models: the unrestricted model (5) and the restricted model where $\beta_2 = 0$. We denote the least-squares (LS) estimators of β_1 and β_2 from the unrestricted model by $\hat\beta_{1u}$ and $\hat\beta_{2u}$, and the LS estimator of β_1 from the restricted model by $\hat\beta_{1r}$.

Let $M_1 = I_n - X_1(X_1'X_1)^{-1}X_1'$ and define the parameter vector θ and its estimator $\hat\theta$ by

$$\theta = (X_2'M_1X_2)^{1/2}\beta_2, \qquad \hat\theta = (X_2'M_1X_2)^{1/2}\hat\beta_{2u}. \tag{6}$$

Then $\hat\theta \sim \mathrm{N}(\theta, \sigma^2 I_{k_2})$. The estimation of β_1 and the estimation of θ are intimately connected, as we shall see shortly.

Let s^2 be the LS estimator of σ^2 in the unrestricted model. Then $F = \hat\theta'\hat\theta/(k_2 s^2)$ is the F-statistic for testing $\beta_2 = 0$, and the typical pretest estimator of β_1 takes the form

$$\hat\beta_{1p} = \begin{cases} \hat\beta_{1r} & \text{if } F \le c, \\ \hat\beta_{1u} & \text{if } F > c \end{cases} \tag{7}$$

for some $c > 0$. The continuous version of the pretest estimator is the model-averaging estimator

$$\hat\beta_{1m} = \lambda\hat\beta_{1u} + (1-\lambda)\hat\beta_{1r}, \qquad \lambda = \lambda(\hat\theta, s^2). \tag{8}$$

If we take λ to be a function of F, then the larger is F the more weight we should put on $\hat\beta_{1u}$, so that λ should be nondecreasing in F.

The "equivalence theorem" of Magnus and Durbin (1999) shows that

$$\mathrm{MSE}(\hat\beta_{1m}) = \sigma^2(X_1'X_1)^{-1} + Q\,\mathrm{MSE}(\lambda\hat\theta)Q', \tag{9}$$

where

$$Q = (X_1'X_1)^{-1}X_1'X_2(X_2'M_1X_2)^{-1/2}. \tag{10}$$

A good estimator (in the MSE sense) $\lambda\hat\theta$ of θ thus automatically implies a good estimator $\hat\beta_{1m}$ of β_1, and hence, if λ is a nondecreasing function of F in the regression setup, then it should also be a nondecreasing function of F in the setup with $\hat\theta \sim \mathrm{N}(\theta, \sigma^2 I_{k_2})$.

3 Proof of the proposition

The proof is in two steps. We first consider the simplest case, namely the case where λ is a constant. Then we consider the general case.

In the special case where λ is a constant, we write

$$\lambda x - \theta = \lambda(x - \theta) - \theta(1-\lambda), \tag{11}$$

and obtain the risk

$$R(\theta, \lambda) = \mathrm{E}(\lambda x - \theta)'(\lambda x - \theta) = p\lambda^2 + \theta'\theta(1-\lambda)^2. \tag{12}$$

For given θ, the risk is minimized when

$$0 = \frac{\partial R(\theta, \lambda)}{\partial \lambda} = 2p\lambda - 2\theta'\theta(1-\lambda), \tag{13}$$

which leads to $\lambda^* = \alpha/(1+\alpha)$, and thus shows that

$$R(\theta, \lambda) \geq R(\theta, \lambda^*) = \frac{\alpha p}{1+\alpha} = R_M(p, \alpha) \tag{14}$$

for all (constant) λ with equality if and only if $\lambda = \alpha/(1+\alpha)$.

The previous argument shows that, in the class of constant λ-functions, the bound in (14) is not just a bound but in fact the efficiency bound of the risk $R(\theta, \lambda)$. Hence, if we can show that $R(\theta, \lambda) \geq \alpha p/(1+\alpha)$ in a larger class of λ-functions, then this will also be the efficiency bound (and not just a bound) in this larger class. This is what we shall prove next.

In the general case we have $t(x) = \lambda(w_p)\,x$, where $w_p = x'x$ follows a noncentral χ^2 distribution with p degrees of freedom and noncentrality $\theta'\theta$. Following De Luca and Magnus (2021, Proposition 2), we write

$$\begin{aligned} R(\theta, \lambda) &= \mathrm{E}[(t(x)-\theta)'(t(x)-\theta)] \\ &= \mathrm{E}[\lambda^2(w_p)x'x] - 2\theta' \,\mathrm{E}[\lambda(w_p)x] + \theta'\theta \\ &= p\,\mathrm{E}\,\lambda^2(w_{p+2}) + \theta'\theta\,\mathrm{E}\,\lambda^2(w_{p+4}) - 2\theta'\theta\,\mathrm{E}\,\lambda(w_{p+2}) + \theta'\theta, \end{aligned} \tag{15}$$

where the last equality follows from Bock (1975, Theorems A and B), who showed that

$$\mathrm{E}[\lambda(w_p)x_i] = \theta_i\,\mathrm{E}[\lambda(w_{p+2})] \tag{16}$$

and

$$\mathrm{E}[\lambda^2(w_p)x_i^2] = \mathrm{E}[\lambda^2(w_{p+2})] + \theta_i^2\,\mathrm{E}[\lambda^2(w_{p+4})] \tag{17}$$

for any function $\lambda : [0, \infty) \to (-\infty, \infty)$. Note that w_p, w_{p+2}, and w_{p+4} denote noncentral χ^2 random variables with *the same* noncentrality $\theta'\theta$ and degrees of freedom p, $p+2$, and $p+4$, respectively.

Let $r = \alpha/(1+\alpha)$. Then we can write (15) as

$$\begin{aligned} R(\theta, \lambda)/p &= \mathrm{E}\,\lambda^2(w_{p+2}) + \alpha\,\mathrm{E}\,\lambda^2(w_{p+4}) - 2\alpha\,\mathrm{E}\,\lambda(w_{p+2}) + \alpha \\ &= (1+\alpha)\,\mathrm{E}\,\lambda^2(w_{p+2}) - 2\alpha\,\mathrm{E}\,\lambda(w_{p+2}) + \alpha + \alpha[\mathrm{E}\,\lambda^2(w_{p+4}) - \mathrm{E}\,\lambda^2(w_{p+2})] \\ &= r + (1+\alpha)\,\mathrm{E}[\lambda(w_{p+2}) - r]^2 + \alpha[\mathrm{E}\,\lambda^2(w_{p+4}) - \mathrm{E}\,\lambda^2(w_{p+2})]. \end{aligned} \tag{18}$$

Hence, if we can show that $\mathrm{E}\,\lambda^2(w_{p+4}) \geq \mathrm{E}\,\lambda^2(w_{p+2})$, then it will follow that $R(\theta, \lambda)/p \geq r$, and the proof is complete. To prove this, let $f_p(w; \theta'\theta)$ denote the noncentral χ^2 distribution with p degrees of freedom and noncentrality $\theta'\theta$. Then,

$$\begin{aligned} \mathrm{E}\lambda^2(w_{p+4}) &= \int \lambda^2(w) f_{p+4}(w; \theta'\theta)\,dw \\ &= \iint \lambda^2(w+z) f_{p+2}(w; \theta'\theta) f_2(z; 0)\,dw\,dz \\ &\geq \iint \lambda^2(w) f_{p+2}(w; \theta'\theta) f_2(z; 0)\,dw\,dz \\ &= \int \lambda^2(w) f_{p+2}(w; \theta'\theta)\,dw = \mathrm{E}\,\lambda^2(w_{p+2}), \end{aligned} \tag{19}$$

since λ^2 is nondecreasing on $(0, \infty)$.

4 Comparison with Hansen's bound

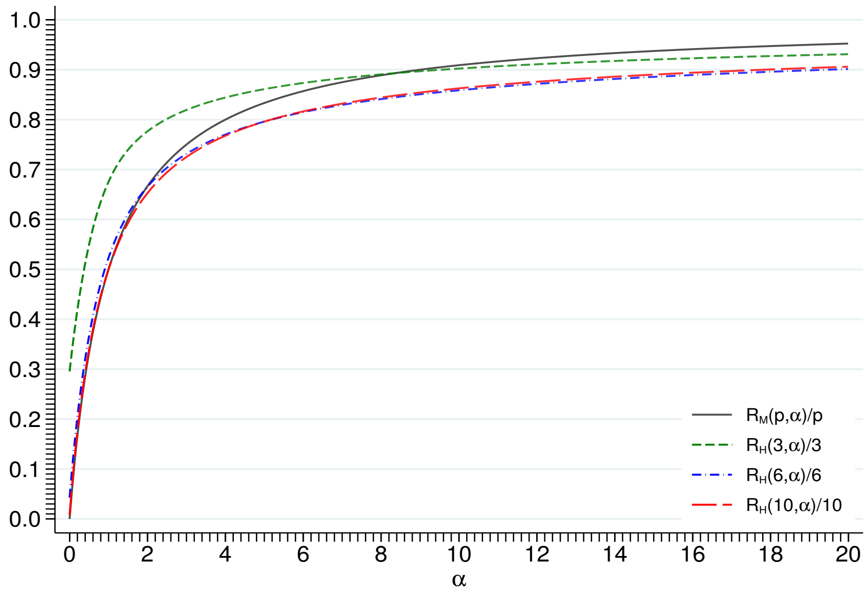

Figure 1: Efficiency bounds $R_H(p,\alpha)/p$ and $R_M(p,\alpha)/p$

Let us compare Hansen's efficiency bound with ours. In Figure 1 we draw our bound $R_M(p,\alpha)/p$ which is equal to $\alpha/(1+\alpha)$ by Proposition 1, and thus does not depend on p. In contrast, Hansen's $R_H(p,\alpha)/p$ does depend on p and we have drawn it here for three values of p: 3, 6, and 10. We see that the bounds are similar but not the same, and that R_H is sometimes larger than R_M (for small values of α) and sometimes smaller (for large values of α). The larger is p, the smaller is the value of α where $R_H = R_M$. The "silly" estimator $t(x) = 0$ defined by $\lambda = 0$ has risk αp and hence it has zero risk at $\alpha = 0$. This estimator is included in our class but not in Hansen's class, which is why our efficiency bound is zero at $\alpha = 0$.

References

Bock, M. E. (1975). Minimax estimators of the mean of a multivariate normal distribution. *The Annals of Statistics*, 3, 209–218.

De Luca, G. and J.R. Magnus (2021). Weak versus strong dominance of shrinkage estimators. *Journal of Quantitative Economics*, 19, 239–266.

Hansen, B. E. (2015). Shrinkage efficiency bounds. *Econometric Theory*, 31, 860–879.

Magnus, J. R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal*, 5, 225–236.

Magnus, J. R. and J. Durbin (1999). Estimation of regression coefficients of interest when other regression coefficients are on no interest. *Econometrica*, 67, 639–643.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.